Supplementary information
for
Can co-location be used as a proxy for face-to-face contacts?

Mathieu Génois∗
Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France and
GESIS, Leibniz Institute for the Social Sciences,
Unter Sachsenhausen 6-8, 50667 Köln, Germany

Alain Barrat
Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France and
Data Science Laboratory, ISI Foundation, Torino, Italy
(Dated: December 18, 2017)

∗mathieu.genois@gesis.org
<table>
<thead>
<tr>
<th></th>
<th>InVS13</th>
<th>InVS15</th>
<th>LH10</th>
<th>LyonSchool</th>
<th>SFHH</th>
<th>Thiers13</th>
</tr>
</thead>
<tbody>
<tr>
<td>\langle w \rangle</td>
<td>8.17</td>
<td>6.89</td>
<td>20.3</td>
<td>9.14</td>
<td>5.77</td>
<td>20.2</td>
</tr>
<tr>
<td>\langle w^2 \rangle</td>
<td>1.53×10^6</td>
<td>6.93×10^5</td>
<td>1.02×10^6</td>
<td>834</td>
<td>545</td>
<td>1.05×10^6</td>
</tr>
<tr>
<td>\langle n \rangle</td>
<td>4.02</td>
<td>3.63</td>
<td>8.87</td>
<td>6.01</td>
<td>2.77</td>
<td>7.71</td>
</tr>
<tr>
<td>\langle n^2 \rangle</td>
<td>1.72</td>
<td>62.4</td>
<td>478</td>
<td>207</td>
<td>65.5</td>
<td>67.7</td>
</tr>
<tr>
<td>\langle \tau_c \rangle</td>
<td>2.03</td>
<td>1.90</td>
<td>2.29</td>
<td>1.52</td>
<td>2.68</td>
<td>2.61</td>
</tr>
<tr>
<td>\langle \tau_c^2 \rangle</td>
<td>84.1</td>
<td>22.5</td>
<td>15.6</td>
<td>6.02</td>
<td>16.2</td>
<td>48.4</td>
</tr>
<tr>
<td>\langle \tau_c \rangle</td>
<td>4.14</td>
<td>11.9</td>
<td>6.81</td>
<td>3.96</td>
<td>7.79</td>
<td>18.5</td>
</tr>
<tr>
<td>\langle \tau_c^2 \rangle</td>
<td>24.3</td>
<td>14.8</td>
<td>24.9</td>
<td>15.8</td>
<td>8.09</td>
<td>47.1</td>
</tr>
<tr>
<td>\langle \tau_c \rangle</td>
<td>1.07</td>
<td>1.23</td>
<td>1.65</td>
<td>1.09</td>
<td>1.14</td>
<td>1.05</td>
</tr>
<tr>
<td>\langle \tau_c^2 \rangle</td>
<td>130</td>
<td>74.1</td>
<td>93.8</td>
<td>41.5</td>
<td>52.7</td>
<td>171</td>
</tr>
<tr>
<td>\langle \tau_c \rangle</td>
<td>25.2</td>
<td>206</td>
<td>24.8</td>
<td>425</td>
<td>37.0</td>
<td>926</td>
</tr>
</tbody>
</table>

**TABLE I.** Comparison of the moments of the temporal distributions between the contact and the co-presence events. For each dataset we compute the first and second moment of the distributions of link weights, event frequencies, event durations and inter-event durations, along with the ratio of the two.
FIG. 1. Temporal distributions for co-presence and contact events — InVS13. We show for both the contact and co-presence of the same data set the distributions of event and inter-event duration, link weights (as total contact duration) and number of contacts per link.

FIG. 2. Temporal distributions for co-presence and contact events — LH10. We show for both the contact and co-presence of the same data set the distributions of event and inter-event duration, link weights (as total contact duration) and number of contacts per link.

FIG. 3. Temporal distributions for co-presence and contact events — LyonSchool. We show for both the contact and co-presence of the same data set the distributions of event and inter-event duration, link weights (as total contact duration) and number of contacts per link.

FIG. 4. Temporal distributions for co-presence and contact events — SFHH. We show for both the contact and co-presence of the same data set the distributions of event and inter-event duration, link weights (as total contact duration) and number of contacts per link.
FIG. 5. Temporal distributions for co-presence and contact events — Thiers13. We show for both the contact and co-presence of the same data set the distributions of event and inter-event duration, link weights (as total contact duration) and number of contacts per link.
FIG. 6. Contact and co-presence matrices — InVS13. Comparison between the average matrices of link density for the contacts and the co-presence daily aggregated networks. Values are averaged over all days of the data collection. Both plots have the same colour scale.

FIG. 7. Contact and co-presence matrices — LH10. Comparison between the average matrices of link density for the contacts and the co-presence daily aggregated networks. Values are averaged over all days of the data collection. Both plots have the same colour scale.
FIG. 8. Contact and co-presence matrices — Lyon-School. Comparison between the average matrices of link density for the contacts and the co-presence daily aggregated networks. Values are averaged over all days of the data collection. Both plots have the same colour scale.

FIG. 9. Contact and co-presence matrices — Thiers13. Comparison between the average matrices of link density for the contacts and the co-presence daily aggregated networks. Values are averaged over all days of the data collection. Both plots have the same colour scale.
FIG. 10. Number of contacts as a function of the number of individuals present — InVS13. We plot the number of contacts $c_a(t)$ occurring at time $t$ in a certain area $a$ as a function of the number of individuals $n_a(t)$ present at the same time in $a$. The red line shows the median of the scatter plot, with error bars defined by the 25% and 75% percentiles.

FIG. 11. Number of contacts as a function of the number of individuals present — LH10. We plot the number of contacts $c_a(t)$ occurring at time $t$ in a certain area $a$ as a function of the number of individuals $n_a(t)$ present at the same time in $a$. The red line shows the median of the scatter plot, with error bars defined by the 25% and 75% percentiles.

FIG. 12. Number of contacts as a function of the number of individuals present — LyonSchool. We plot the number of contacts $c_a(t)$ occurring at time $t$ in a certain area $a$ as a function of the number of individuals $n_a(t)$ present at the same time in $a$. The red line shows the median of the scatter plot, with error bars defined by the 25% and 75% percentiles.

FIG. 13. Number of contacts as a function of the number of individuals present — SFHH. We plot the number of contacts $c_a(t)$ occurring at time $t$ in a certain area $a$ as a function of the number of individuals $n_a(t)$ present at the same time in $a$. The red line shows the median of the scatter plot, with error bars defined by the 25% and 75% percentiles.
FIG. 14. Number of contacts as a function of the number of individuals present — Thiers13. We plot the number of contacts $c_a(t)$ occurring at time $t$ in a certain area $a$ as a function of the number of individuals $n_a(t)$ present at the same time in $a$. The red line shows the median of the scatter plot, with error bars defined by the 25% and 75% percentiles.

FIG. 15. Number of contacts as a function of the number of individuals present (loglog) — InVS15.

FIG. 16. Number of contacts as a function of the number of individuals present (loglog) — InVS13.
FIG. 17. Number of contacts as a function of the number of individuals present (loglog) — LH10.

FIG. 18. Number of contacts as a function of the number of individuals present (loglog) — LyonSchool.

FIG. 19. Number of contacts as a function of the number of individuals present (loglog) — SFHH.

FIG. 20. Number of contacts as a function of the number of individuals present (loglog) — Thiers13.
FIG. 21. Number of contacts as a function of the number of individuals present (loglog) — All data sets.
FIG. 22. **Properties of the sampled co-presence networks — InVS13.** We compare several properties of the contact network from the original data set with the surrogate contacts obtained by sampling of the co-presence data: overall timeline of contact activity, distributions of degree, weight $w$ and number of contacts per link $n$ in the network aggregated over the whole data collection period, and distributions of the contact duration $\tau_c$ and inter-contact duration $\tau_i$. 
FIG. 23. Properties of the sampled co-presence networks — LH10. We compare several properties of the contact network from the original data set with the surrogate contacts obtained by sampling of the co-presence data: overall timeline of contact activity, distributions of degree, weight $w$ and number of contacts per link $n$ in the network aggregated over the whole data collection period, and distributions of the contact duration $\tau_c$ and inter-contact duration $\tau_i$. 
FIG. 24. Properties of the sampled co-presence networks — LyonSchool. We compare several properties of the contact network from the original data set with the surrogate contacts obtained by sampling of the co-presence data: overall timeline of contact activity, distributions of degree, weight $w$ and number of contacts per link $n$ in the network aggregated over the whole data collection period, and distributions of the contact duration $\tau_c$ and inter-contact duration $\tau_i$. 
FIG. 25. Properties of the sampled co-presence networks — SFHH. We compare several properties of the contact network from the original data set with the surrogate contacts obtained by sampling of the co-presence data: overall timeline of contact activity, distributions of degree, weight $w$ and number of contacts per link $n$ in the network aggregated over the whole data collection period, and distributions of the contact duration $\tau_c$ and inter-contact duration $\tau_i$. 
FIG. 26. Properties of the sampled co-presence networks — Thiers13. We compare several properties of the contact network from the original data set with the surrogate contacts obtained by sampling of the co-presence data: overall timeline of contact activity, distributions of degree, weight $w$ and number of contacts per link $n$ in the network aggregated over the whole data collection period, and distributions of the contact duration $\tau_c$ and inter-contact duration $\tau_i$. 
FIG. 27. Evolution of the mean aggregated degree and strength. We compare how the average degree $\langle k(t) \rangle$ and the average strength $\langle s(t) \rangle$ grow as we aggregate the network along time, for the real contact data and each sampling method. InVS13 dataset.

FIG. 28. Evolution of the mean aggregated degree and strength. We compare how the average degree $\langle k(t) \rangle$ and the average strength $\langle s(t) \rangle$ grow as we aggregate the network along time, for the real contact data and each sampling method. LH10 dataset.

FIG. 29. Evolution of the mean aggregated degree and strength. We compare how the average degree $\langle k(t) \rangle$ and the average strength $\langle s(t) \rangle$ grow as we aggregate the network along time, for the real contact data and each sampling method. LyonSchool dataset.

FIG. 30. Evolution of the mean aggregated degree and strength. We compare how the average degree $\langle k(t) \rangle$ and the average strength $\langle s(t) \rangle$ grow as we aggregate the network along time, for the real contact data and each sampling method. SFHH dataset.
FIG. 31. **Evolution of the mean aggregated degree and strength.** We compare how the average degree $\langle k(t) \rangle$ and the average strength $\langle s(t) \rangle$ grow as we aggregate the network along time, for the real contact data and each sampling method. Thiers13 dataset.

FIG. 32. **Node ranking similarity.** We plot for each model the Jaccard similarity between the top $N\%$ nodes, ranked according to their degree $k$, their strength $s$, their betweenness centrality $b$, as function of $N$. The plot shows the median similarity and the 90% confidence interval. InVS13 dataset.

FIG. 33. **Node ranking similarity.** We plot for each model the Jaccard similarity between the top $N\%$ nodes, ranked according to their degree $k$, their strength $s$, their betweenness centrality $b$, as function of $N$. The plot shows the median similarity and the 90% confidence interval. LH10 dataset.
FIG. 34. **Node ranking similarity.** We plot for each model the Jaccard similarity between the top $N\%$ nodes, ranked according to their degree $k$, their strength $s$, their betweenness centrality $b$, as function of $N$. The plot shows the median similarity and the 90% confidence interval. LyonSchool dataset.

FIG. 35. **Node ranking similarity.** We plot for each model the Jaccard similarity between the top $N\%$ nodes, ranked according to their degree $k$, their strength $s$, their betweenness centrality $b$, as function of $N$. The plot shows the median similarity and the 90% confidence interval. SFHH dataset.

FIG. 36. **Node ranking similarity.** We plot for each model the Jaccard similarity between the top $N\%$ nodes, ranked according to their degree $k$, their strength $s$, their betweenness centrality $b$, as function of $N$. The plot shows the median similarity and the 90% confidence interval. Thiers13 dataset.
FIG. 37. Epidemic prevalence. We plot the fraction of the total number of outbreaks that reach at least 20% of the population (crosses) and the distribution of the sizes of these outbreaks (boxplots) as functions of the reproductive number $R$, for the original data and the sampled co-presence networks — InVS13 case.

FIG. 38. Epidemic prevalence. We plot the fraction of the total number of outbreaks that reach at least 20% of the population (crosses) and the distribution of the sizes of these outbreaks (boxplots) as functions of the reproductive number $R$, for the original data and the sampled co-presence networks — LH10 case.

FIG. 39. Epidemic prevalence. We plot the fraction of the total number of outbreaks that reach at least 20% of the population (crosses) and the distribution of the sizes of these outbreaks (boxplots) as functions of the reproductive number $R$, for the original data and the sampled co-presence networks — LyonSchool case.

FIG. 40. Epidemic prevalence. We plot the fraction of the total number of outbreaks that reach at least 20% of the population (crosses) and the distribution of the sizes of these outbreaks (boxplots) as functions of the reproductive number $R$, for the original data and the sampled co-presence networks — SFHH case.
FIG. 41. **Epidemic prevalence.** We plot the fraction of the total number of outbreaks that reach at least 20% of the population (crosses) and the distribution of the sizes of these outbreaks (boxplots) as functions of the reproductive number $R$, for the original data and the sampled co-presence networks — Thiers13 case.
FIG. 42. **Vaccination strategies.** We plot the ratio between the vaccination and no vaccination cases of the fraction of the total number of outbreaks that reach at least 20% of the population (top), and of the median size of these outbreaks (bottom) for different vaccination strategies, for the original data and the reconstructed networks. The **group**<sub>n</sub> strategies consist in vaccinating one or several groups entirely; the **group_rand** strategy vaccinates \(n_g\) random nodes, where \(n_g\) is the average group size; the **rand**<sub>n</sub> strategies randomly vaccinates a specified fraction of nodes; the \(b_n, k_n, s_n\) strategies vaccinate the top \(n\%\) nodes according to, respectively, betweenness centrality, degree and strength ranking. InVS13 case.

FIG. 43. **Vaccination strategies.** We plot the ratio between the vaccination and no vaccination cases of the fraction of the total number of outbreaks that reach at least 20% of the population (top), and of the median size of these outbreaks (bottom) for different vaccination strategies, for the original data and the reconstructed networks. The **group**<sub>n</sub> strategies consist in vaccinating one or several groups entirely; the **group_rand** strategy vaccinates \(n_g\) random nodes, where \(n_g\) is the average group size; the **rand**<sub>n</sub> strategies randomly vaccinates a specified fraction of nodes; the \(b_n, k_n, s_n\) strategies vaccinate the top \(n\%\) nodes according to, respectively, betweenness centrality, degree and strength ranking. LH10 case.
FIG. 44. **Vaccination strategies.** We plot the ratio between the vaccination and no vaccination cases of the fraction of the total number of outbreaks that reach at least 20% of the population (top), and of the median size of these outbreaks (bottom) for different vaccination strategies, for the original data and the reconstructed networks. The \textit{group-n} strategies consist in vaccinating one or several groups entirely; the \textit{group_rand} strategy vaccinates \( n_g \) random nodes, where \( n_g \) is the average group size; the \textit{rand-n} strategies randomly vaccinates a specified fraction of nodes; the \( b_n, k_n, s_n \) strategies vaccinate the top \( n \% \) nodes according to, respectively, betweenness centrality, degree and strength ranking. LyonSchool case.

FIG. 45. **Vaccination strategies.** We plot the ratio between the vaccination and no vaccination cases of the fraction of the total number of outbreaks that reach at least 20% of the population (top), and of the median size of these outbreaks (bottom) for different vaccination strategies, for the original data and the reconstructed networks. The \textit{group-n} strategies consist in vaccinating one or several groups entirely; the \textit{group_rand} strategy vaccinates \( n_g \) random nodes, where \( n_g \) is the average group size; the \textit{rand-n} strategies randomly vaccinates a specified fraction of nodes; the \( b_n, k_n, s_n \) strategies vaccinate the top \( n \% \) nodes according to, respectively, betweenness centrality, degree and strength ranking. SFHH case.
FIG. 46. Vaccination strategies. We plot the ratio between the vaccination and no vaccination cases of the fraction of the total number of outbreaks that reach at least 20% of the population (top), and of the median size of these outbreaks (bottom) for different vaccination strategies, for the original data and the reconstructed networks. The $\text{group}_n$ strategies consist in vaccinating one or several groups entirely; the $\text{group}_\text{rand}$ strategy vaccinates $n_g$ random nodes, where $n_g$ is the average group size; the $\text{rand}_n$ strategies randomly vaccinates a specified fraction of nodes; the $b_n$, $k_n$, $s_n$ strategies vaccinate the top $n\%$ nodes according to, respectively, betweenness centrality, degree and strength ranking. Thiers13 case.