Department of Mathematics and Computer Science
Department of Education and Psychology
Freie Universität Berlin

Social Network Analysis
Dr. Robert Krause
Wintersemester 2021/2022

Networks of Health - Networks of Disease
Understanding hospital contact patterns using actor-oriented network modelling

Midterm report
by
Raphael Leuner
5094927

Contents

1 Motivation

Ever since humans have started to systematically and scientifically study diseases, one of the most important questions has been about how diseases are transmitted between individuals. While the SARS-CoV2 pandemic has brought this topic to the forefront of political and scientific discussions in recent years, it has been an intensively researched question for centuries. Not just the SARS-CoV2 pandemic has shown that transmissions in places with high attendance and small distances over a prolonged period of time can be especially consequential. In hospitals, this is exacerbated by the fact that here potential pathogens from infected patients can meet people with a weakened immune system, for example during chemotherapy. According to the Federal Ministry of Health (BMG), between 400,000 and 600,000 patients get infected with hospital diseases in Germany every year with 10,000 - 20,000 of them dying as a result of one of these infections [1].

Hospitals are also often a breeding ground for Antibiotic Resistant Germs (ARG) because of a high exposure to patients being treated with antibiotics. Therefore, reducing the number of patients falling ill from hospital infections can also reduce the number of ARG infections drastically.

This project looks at modelling the contacts in hospitals based on a complete network approach in order to understand the effects that influence the patterns of contact formation in the health sector. For modelling contacts, two different approaches are chosen here. The first one (Stochastic Actor-Oriented Models (SAOMs)) is based on discrete time steps and the aggregated development of the network in between those moments [4], the second approach (Dynamic Network Actor Models (DyNAMs)) is using continuous, time stamped data [5].

2 Data and Preprocessing

The data on which both modelling approaches are built was collected in a short stay geriatric unit with 19 beds in Lyon, France [8]. It comprises of 97 hours of continuous contact data between Monday 06/12/2010, 01:00 PM and Friday 10/12/2010, 02:00 PM. The data was collected by RFID chips with sender/receiver units placed on the chest of the participants. The contacts were recorded when people were facing each other within a 1-1.5 meter distance for at least 20 seconds. The data is provided in a continuous format in 20 second intervals. There are a total of 32,424 datapoints that result in 14,037 recorded contacts. The data is undirected, so no inferences can be made as to who initiated the contact. Out of the 50 staff members of the hospital, 46 (27 nurses, 11 doctors and 8 administrative staff) participated in the data collection as well as 29 of the 31 patients admitted. In the dataset, each participant is denoted with a unique ID and the occupation of the participant is provided. Unsurprisingly, the cumulative distribution of contacts presented in figure 1 shows a very uneven distribution of contact patterns across the day. This resembles the workday of health care workers (HCWs) as well as administrative staff. The number of contacts during nights is minimal and there are some nights where there are no recorded contacts between patients and medical staff for hours.

Figure 2 shows a heatmap plot of the contact patterns for all 75 participants of the study. It indicates that HCWs generally have an increased number of contacts during their worktime, while the contacts of patients are more sporadic, but also more spread out across the day. It also shows that the ward has been accepting new patients during the course of the study as well as discharging patients at the end of their stationary treatment. This information was used as a basis for data preparation and preprocessing, which was performed in Python.

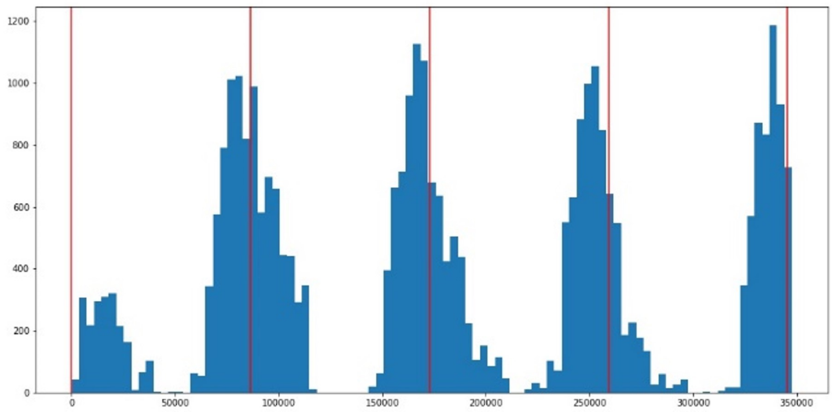


Figure 1: Contact frequency distribution for 97 hours, red lines indicate division of data during processing

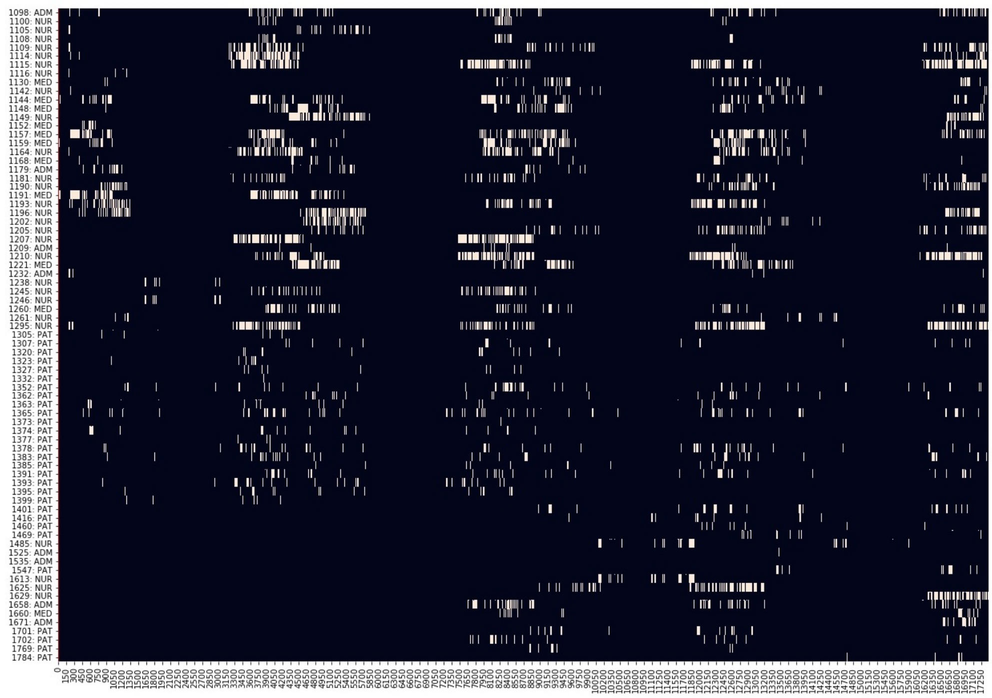


Figure 2: Heatmap visualization of all contact events for all study participats

2.1 Data Aggregation

Analysing the data using SAOMs requires the aggregating contacts over a set period of time. Therefore, the time stamped contact data was aggregated into 4 discrete contact matrices, each of them covering a 24 hour interval. The aggregation windows are visualized with the red lines in figure 1. If in one interval a participant did not record any contacts, the participant was assumed to be absent, therefore all possible contacts were marked with a 10. Tables 1 and 2 are showing summary statistics for the dataset after the aggregation of the data had been performed.

Periods	0 => 0	0 => 1	1 => 0	1 => 1	Distance	Jaccard	Missing
1 ==> 2	2056	287	227	205	504	0.285	0 (0%)
2 ==> 3	2036	247	288	204	516	0.276	0 (0%)
3 ==> 4	2060	264	262	189	492	0.264	0 (0%)

Table 1: Edge changes between subsequent observations:

Observation time	1	2	3	4
density	0.156	0.177	0.163	0.163
average degree	11.520	13.120	12.027	12.080
number of ties	432	492	451	453
missing fraction	0.000	0.000	0.000	0.000

Table 2: Network density indicators:

2.2 Continuous Data Preparation

Since the original data source already provides time stamped data, no further preprocessing apart from adaptation of the data format was required. However, marking missing data in this continuous format proved to be more challenging than for the aggregated data, as the original dataset does not provide any information about admission and discharge dates for patients or shift hours for medical staff. Therefore it was assumed that patients would be present the whole time from their first to their last contact at the hospital. The estimation of shift hours for HCWs followed the same pattern, but assuming that if there was a pause between contacts that was longer than 3 hours, the next contact would be considered to be part of a new shift. The choice of a 3 hour interval was made as it provided the most reasonable separation of shifts (see figure 3 for visualization of shifts and admission periods) and did not result in the creation of two separate shifts within one day for any of the HCWs.

Figure 3: Heatmap visualization of the derived shifts/presence of all participants

3 Background and Hypotheses

There has been extensive research on the transmission of diseases in environments with many people working or living together in close proximity. A special focus has been put on transmissions in hospitals. Here, according to the World Health Organization (WHO), a distinction needs to be made between different routes of transmission [9]. These can be broken down into two paths, for which social networks can be an adequate model:

The first transmission type requires contact between the infected person and the susceptible person. That can either be a direct, person-to-person contact, or, more likely according to the WHO, an indirect contact transmission via a contaminated surface or instrument.

The other main contact dependent infection route is one that does not require direct physical contact and in which disease vectors are transmitted across distances. This includes both airborne diseases in which the microbes survive in the air for a prolonged duration as well as droplet transmissions, where microbes are present in small droplets in the room.

Other transmission routes include the transmission by external vectors, such as insects or rats as well as the transmission through infected food water or pharmaceuticals. Since these transmission routes are not related to person-to-person contact networks in hospitals, they will not be regarded further in this work.

For a long time the predominant theory in infectious disease epidemiology was that the number of transmissions by any one individual infected patient would be more or less equal across the population. However, this assumption has been challenged in the last years and been replaced by the 20/80 rule. This empirical rule states that only about 20% of the host population for a disease is responsible for at least 80% of the total number of transmissions. This pattern has been described in many different disease transmission scenarios [10]. Similar transmission patterns have also been observed in many past disease outbreaks and pandemics and has been coined "Superspreading" in the aftermath of the SARS pandemic $2002 - 2003$ [6]. The term gained huge public attention during the ongoing Covid19-pandemic and many researchers have been suggesting, that the 20/80 rule holds for the SARS-CoV2 virus aswell. In one study tracing an outbreak in Hunan, China, researchers have shown that 80% of infections can be traced back to only 15% of primary infections [7].

There are numerous theories and explanation attempts on why people might become superspreaders and it is still an ongoing discussion whether this is mostly driven by biological factors (such as immunodeficiency) or by behavioral factors (for example personal hygiene). But it seems to be clear that the number and type of contacts that people have is an important factor for whether someone becomes a superspreader or not and that controlling the contacts of potential superspreaders can be a vital contribution to controlling the spread of a disease [3]. Furthermore, it has been shown that superspreaders also play an important role in the emergence of new disease variants [2], further increasing the importance of understanding disease transmissions at places with a high numbers of vulnerable people.

Average daily contacts	Nurses	Doctors	Patients	Administration
Day 1	19.8	16.8	12.5	17.2
Day 2	27.0	20.5	11.5	24.75
Day 3	22.8	21.3	10.5	18.5
Day 4	22.3	16.5	11.2	14.9

Table 3: Average daily contacts for all four participant groups

Table 3 shows the average number of different contacts that members of the four participant groups had at any given day. It shows, that the hospital staff has significantly more contacts compared to patients. This could suggest that medical personnel, especially nurses are at a higher risk of becoming superspreaders at hospitals. Based on this basic assumption the following 5 hypothesis are formed:

Hypothesis 1:
Medical staff, both nurses and doctors, see a faster rate and more opportunities of change in their contacts. Patients will usually encounter the same doctors and the same nurses at each days, while new patients will be admitted into the hospital and previous patients discharged. Since administrative staff does not have direct contact to patients on a regular basis, no significant rate effect is expected for this group. More opportunities to change contact patterns create a higher risk for nurses and doctors to become superspreaders.
This is modelled using the RateX effect with individual interactions for administrative staff, nurses and medical staff.

Hypothesis 2:
Medical staff, both nurses and doctors, have more contacts compared to patients. As medical staff frequently visits different patients in their rooms, it can be expected that they have an increased number of contacts compared to patients. The same effect is not expected with administrative staff, since they do not visit patients and therefore do not move around the hospital that much.
This is modelled using the egoPlusAltX effect for all three personnel at the hospital (admin, nurses, medical) as interaction.

Hypothesis 3:
All hospital employees (nurses, doctors and administrative staff) have a stronger tendency to form contacts with other members of their own employee group. This assumption is based on structural properties of a hospital, such as strict hierarchies between doctors and nurses, a spatial division between offices of administrative staff and the hospital ward as well as separate socializing and break rooms for different groups of employees. Since the data is derived from a short term hospital ward, patients will most likely not know each other and do not find many opportunities to get to know each other.
This is modelled using the interaction between the egoPlusAltX effect and the sameX effect. This is again modelled for all three personnel at the hospital (admin, nurses, medical) as interaction for egoPlusAltX and numeric encoding of the jobs for the sameX effect.

Hypothesis 4:
For potential pathogens to be passed around in an hospital setting over a prolongued period of time, potential superspreaders (or simply participants with many contacts and therefore many opportunities to get infected) need to be in contact with each other. This suggests that one of the risk factors for superspreader-like transmissions in a hospital setting is the existence of more contacts between nodes who themselves have many contacts.
This is modelled using the degPlus effect, an effect that combines the degree popularity and degree activity of two nodes.

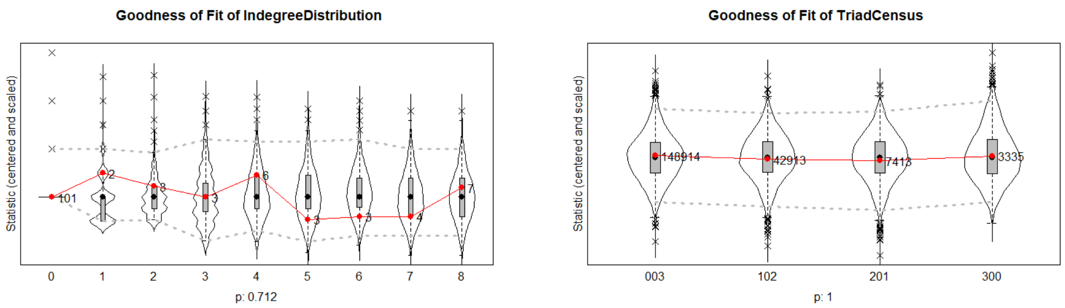
Hypothesis 5:
Since the data is based on physical contact in a limited space, a contact between A and B and

a contact between B and C makes a contact between A and C more likely. Therefore a positive triadic effect can be expected in the network aswell.
This is modelled using the transTriads effect.

4 Results

4.1 Modelling with RSiena

The results of the model shown in table 4 were created using the Siena implementation in R. The model shows convergence as all of the convergence t-ratios are below 0.1 in absolute values and the overall maximum convergence ratio is at 0.1725.
In figure 4 the Goodness of Fit (GOF) for both the in/outdegree distribution as well as the triad distribution is shown. The model achieves sufficient GOF and therefore the effects will be interpreted further.



(a) Indegree/Outdegree distribution

(b) Triad distribution

Figure 4: Goodness of Fit estimation for the presented model

Hypothesis 1 (effects 4-6):
Nurses and doctors have more opportunities to change their contacts (different for administrative staff)

The rate effects for nurses and medical staff are positive, while the rate effect for administrative staff are negative. However, only the effect for nurses achieves statistical significance with a p-value below 0.05. This suggests, that the hypothesis holds for nurses and they do have more opportunities to form contacts, while the same can not be said for doctors and administrative staff.

Hypothesis 2 (effects 10-12):
Nurses and doctors have more contacts compared to patients (different for administrative staff)

Administrative staff and nurses have a slight positive value, indicating a tendency to form more contacts. However, both of these effects are non-significant. For doctors, this effect is negative and significant. This suggests that doctors have less contacts in general compared to other employees in the hospitals.

Hypothesis 3 (effects 13-16):
All hospital employees (nurses, doctors and administrative staff) tend to form more contacts within their own group

The sameX parameter for the role in the hospital overall is negative, however it does not reach statistical significance. All three interaction effects between the sameX effect and the egoPlusAltX effect reach statistical significance and are positive, with doctors showing the strongest effect. This suggests that contact patterns are to some extend segregated between the employment groups in the hospital. The strong effect for doctors indicates that most of their increased activity to form contacts happens within the same employment group.

Hypothesis 4 (effect 9):
An increased number of contacts between people with many contacts is expected

The degree activity and popularity is slightly positive and significant. This suggests that while there is a certain tendency for contacts between people with many contacts themselves, this effect is not that strong in the context of a hospital.

Hypothesis 5 (effect 8):
A positive effect for triad closure is expected, since contacts resemble proximity

The transitive triads parameter is positive, which indicates that, as expected due to the demands and constraints of physical contact patterns, there is a tendency for triad connections to be closed.

Effect 7: Degree (density):
This is a standard parameter in any model and is strongly negative in this case, suggesting that contacts in the hospital are rare in general.

Effect Number	Parameter	Estimate	Standard Error	Normal Variate	Convergence t-Ratio	P-Value
1	"constant contacts rate (period 1)"	17.816	3.264	5.46	-0.0353	0
2	"constant contacts rate (period 2)"	24.248	4.621	5.25	-0.0334	0
3	"constant contacts rate (period 3)"	15.897	2.963	5.36	0.0010	0
4	"effect adm on rate"	-0.115	0.535	-0.21	-0.0191	0.8302
5	"effect nur on rate"	1.131	0.383	2.95	-0.0779	0.0031
6	"effect med on rate"	0.253	0.484	0.52	0.0290	0.6015
7	"degree (density)"	-1.586	0.165	-9.58	0.0598	0
8	"transitive triads"	0.064	0.016	4.08	0.0229	0
9	"degree act+pop"	0.013	0.004	2.86	0.0403	0.0043
10	"adm ego and alt"	0.019	0.094	0.2	-0.0329	0.8424
11	"nur ego and alt"	0.017	0.103	0.17	0.0824	0.8686
12	"med ego and alt"	-0.281	0.08	-3.49	0.0403	0.0005
13	"same ro"	-0.161	0.11	-1.46	0.0677	0.1434
14	"int. same ro x nur ego and alt"	0.701	0.15	4.67	0.0857	0
15	"int. same ro x med ego and alt"	1.279	0.156	8.18	0.0160	0
16	"int. same ro x adm ego and alt"	0.681	0.241	2.82	-0.0672	0.0048

Table 4: Model parameter estimates using RSiena's siena07 estimation algorithm

4.2 Modelling with goldfish

Modelling of the continuous data was done using the goldfish package, which implements the DyNAM model [5]. The two models presented here are rather simple and are meant to show that an application of DyNAM models on this dataset is feasible.

In the DyNAM approach, two separate models are required for rate and choice modelling. Hypothesis 2 is modelled using the ego effects in the rate model. In accordance with the SOAM model, this is done for all three groups of staff. Table 5 shows the estimates for this model. The effects are positive for administrative staff as well as nurses and doctors and these effects are significant for all three estimates. This shows, that in general all three groups have an increased contact activity. This does not contradict the results from the SOAM modelling approach (where the only significant effect for the egoPlusAltX effect was for doctors and negative), as the separate effects included in the SOAM modelled an increased number of contacts within their own group (see Hypothesis 3). These additional effects were not included in the DyNAM model.

In the choice model, presented in table 6, hypothesis 3 and hypothesis 5 are being tested. For hypothesis 3, the same effect was used modelling the choice of contacts within the same group for all three types of staff. These effects were positive and significant, supporting the findings from the SOAM model presented previously. For hypothesis 5, similarly, the transitivity was modelled and here again the results from before could be confirmed with a slightly positive and significant effect.

Effect	Estimate	Standard Error	z-Value	Pr(>|z|)
Intercept	-9.84578435	0.03158437	-311.730	< 2.2e-16 ***
indeg	0.04728795	0.00037156	127.270	< 2.2e-16 ***
ego admin	1.53151432	0.03837312	39.911	< 2.2e-16 ***
ego nurse	2.35524815	0.03246525	72.547	< 2.2e-16 ***
ego med	2.48477475	0.03325152	74.727	< 2.2e-16 ***

Table 5: Rate Model, Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Effect	Estimate	Standard Error	z-Value	Pr(>|z|)
inertia	3.68045491	0.03216582	114.421	< 2.2e-16 ***
trans	0.02700342	0.00079618	33.916	< 2.2e-16 ***
same admin	0.46270367	0.02385684	19.395	< 2.2e-16 ***
same nurse	0.36423232	0.01409350	25.844	< 2.2e-16 ***
same med	1.32826355	0.02013705	65.961	< 2.2e-16 ***

Table 6: Choice Model, Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5 Discussion

One of the main goals of this project was to show the feasibility of analyzing automatically recorded, physical contact pattern data with two different social network modelling approaches. The results show, that both models can create use- and meaningful results. That is especially promising, since the Covid19 pandemic and widespread rollout of contact tracing apps has made the creation of this sort of data much easier. Even though there are certain technological differences, especially the use of Bluetooth instead of RFID, similar data sets could be created using this technology.

The central hypothesis about the increased contact activity of HCWs has been confirmed in the analysis. The results also show, that this is mostly explained by within group activity. This was especially prominent for medical staff, who even showed a decreased contact activity outside their own group. This can be relevant information when a hospital creates policies meant to suppress for example an acute disease outbreak.

The results also suggest an increased presence of events with a high risk for superspreading activity, contacts between two people with many contacts themselves. While the effect is not very strong, it might still be warranted to further investigate these events in hospital contact analysis.

The DyNAM model results confirm that modelling hospital contacts based on continuous data is not just a promising approach for further analysis, but due to a better and deeper usage of the present dataset could potentially be the superior approach for further analysis of this and similar datasets. The initial analysis presented here can be improved upon by also taking the duration of contacts into account and analyzing the data in separate windows to understand if contact patterns differ at different times of the day.

While further analysis of this dataset overall seems very promising, some constraints of both the analysis as well as the data and its collection should be kept in mind. The contact information is in a binary format with a fixed distance threshold. While this, at 1-1.5 meters, is consistent with the generally accepted distance at which people are at a severe risk for infection by droplets, airborne infections can generally bridge much wider distances. Similarly, indirect contact transmission by handling material at different points in time without personal contact are also not adequately represented in the data. Therefore, modelling contact patterns based on this dataset can shine light on one aspect of transmissions of infectious diseases in hospitals, but will most likely not explain the entirety of contact based transmissions between health care workers and patients.

6 Code and Data

The data used in this project is available here [8]
The data preprocessing script as well as the analysis scripts for both models are available on Github:
`https://github.com/rlnrbio/hospital_contacts`

7 Supplemental - Effects

Covariate effect on Rate (RateX):
The effect of actor covariates (RateX) with values v_{hi} can be represented by the factor:

$$\lambda_{i2}^{net} = exp(\sum_h \alpha_h v_{hi})$$

Same covariate/covariate-related identity (SameX):
Defined by the number of ties of i to all other actors j who have exactly the same value on the covariate

$$s_{i88}^{net}(x) = \sum_j x_{ij} I\{v_i = v_j\}$$

where the indicator function $I\{v_i = v_j\}$ is 1 if the condition $\{v_i = v_j\}$ is satisfied, and 0 if it is not

Covariate Effect (egoPlusAltX):
This is defined as the sum of the covariate-ego and covariate-alter effects; equivalently, this effect operates as the covariate-ego combined with the covariate-alter effect under the assumption that the parameters for both are the same:

$$s_{i3}^{net}(x) = \sum_j x_{ij}(v_i + v_j)$$

Transitive triads effect (transTriads):
defined by the number of transitive patterns in i's relations, which for non-directed networks can be written as:

$$s_{i5}^{net}(x) = \sum_{j<h} x_{ij} * x_{ih} * x_{hj}$$

Degree activity plus popularity effect (degPlus):
This is defined as the sum of the in=outdegree popularity and in=outdegree activity effects; equivalently, this effect operates as the degree popularity combined with the degree activity effect under the assumption that the parameters for both are the same:

$$s_{i1}^{net}(x) = \sum_j x_{ij}(x_{j+} + x_{i+})$$

References

[1] Bundesministerium für Gesundheit. *Krankenhaushygiene*. https://www.bundesgesundheitsministerium.de/krankenhaushygiene.html. Accessed: 2022-02-02.

[2] Gabriel E. Leventhal et al. "Evolution and emergence of infectious diseases in theoretical and real-world networks". In: *Nature Communications* 6.1 (Jan. 2015). DOI: 10.1038/ncomms7101. URL: https://doi.org/10.1038/ncomms7101.

[3] Dyani Lewis. "Superspreading drives the COVID pandemic — and could help to tame it". In: *Nature* 590.7847 (Feb. 2021), pp. 544–546. DOI: 10.1038/d41586-021-00460-x. URL: https://doi.org/10.1038/d41586-021-00460-x.

[4] Tom A.B. Snijders, Gerhard G. van de Bunt, and Christian E.G. Steglich. "Introduction to stochastic actor-based models for network dynamics". In: *Social Networks* 32.1 (2010). Dynamics of Social Networks, pp. 44–60. ISSN: 0378-8733. DOI: https://doi.org/10.1016/j.socnet.2009.02.004. URL: https://www.sciencedirect.com/science/article/pii/S0378873309000069.

[5] Christoph Stadtfeld, James Hollway, and Per Block. "Dynamic Network Actor Models: Investigating Coordination Ties through Time". In: *Sociological Methodology* 47.1 (2017), pp. 1–40. DOI: 10.1177/0081175017709295. eprint: https://doi.org/10.1177/0081175017709295. URL: https://doi.org/10.1177/0081175017709295.

[6] Richard A. Stein. "Super-spreaders in infectious diseases". In: *International Journal of Infectious Diseases* 15.8 (2011), e510–e513. ISSN: 1201-9712. DOI: https://doi.org/10.1016/j.ijid.2010.06.020. URL: https://www.sciencedirect.com/science/article/pii/S1201971211000245.

[7] Kaiyuan Sun et al. "Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2". In: *Science* 371.6526 (2021), eabe2424. DOI: 10.1126/science.abe2424. eprint: https://www.science.org/doi/pdf/10.1126/science.abe2424. URL: https://www.science.org/doi/abs/10.1126/science.abe2424.

[8] Philippe Vanhems et al. "Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors". In: *PLOS ONE* 8.9 (Sept. 2013), pp. 1–9. DOI: 10.1371/journal.pone.0073970. URL: https://doi.org/10.1371/journal.pone.0073970.

[9] WHO. "Prevention of hospital-acquired infections - A practical guide". In: *Department of Communicable Disease, Surveillance and Response* (2002). URL: https://www.who.int/csr/resources/publications/whocdscsreph200212.pdf.

[10] M. E. J. Woolhouse et al. "Heterogeneities in the transmission of infectious agents: Implications for the design of control programs". In: *Proceedings of the National Academy of Sciences* 94.1 (1997), pp. 338–342. ISSN: 0027-8424. DOI: 10.1073/pnas.94.1.338. eprint: https://www.pnas.org/content/94/1/338.full.pdf. URL: https://www.pnas.org/content/94/1/338.