Graph-based risk modeling for infectious diseases in social systems

Maximilian Stäbler - Darmstadt University of Applied Sciences, Faculties Computer Science & Mathematics and Sciences

Introduction

Motivation:

- In December 2020, the World Health Organization (WHO) describes lower respiratory infections as the deadliest communicable disease in the world and the fourth leading cause of death overall.
- On December 31, 2019, the WHO China Country Office was informed of cases of pneumonia of unknown etiology detected in Wuhan, a metropolis of one million people in Hubei province. The cumulative number of confirmed SARS-CoV-2 infections is more than 114.1 million worldwide by March 1, 2021. The number of coronavirus-related deaths rose to more than 2.5 million by that date[1].
- In addition to the health, environmental, and social challenges facing humanity, the coronavirus outbreak is disrupting the global economy. The lockdown measures and distance regulations imposed have interfered with industrial processes to such an extent that companies in various industries have had to close down for extended periods of time.
- Previous models often neglect the social structure of the system under consideration and do not allow decision makers to adopt individual intervention strategies.

Research questions:

1. How can a protocol of social interactions within a complex social (sub)system be used to calculate the risk of infection for an infectious disease that is transmissible through social interactions within that system?
2. Do characteristics of contacts, such as duration of contacts between two individuals, provide information about the likelihood of infection and do topological properties of the resulting network graph allow inferences about infection dynamics?

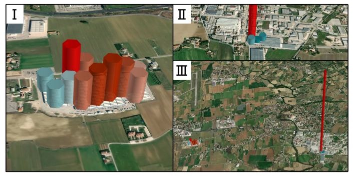
Dataset

Ⅰ

id	deviceid	devicetime	beaconid	day	gateway
138404553	ISA_4_13	2020-05-24 03:03:51.037	ISA_4_43	145	bz2139
138404560	ISA_4_27	2020-05-24 03:03:52.533	ISA_4_49	145	bz2130
138404562	ISA_4_133	2020-05-24 03:03:52.550	ISA_4_492	145	bz2138
138404567	ISA_4_90	2020-05-24 03:03:54.045	ISA_4_30	145	bz2129
138404581	ISA_4_243	2020-05-24 03:03:58.594	ISA_4_55	145	bz2131

TAB. 3.1: Extract of the social interaction dataframe containing all logged proximity alerts within give time-range. *Deviceid* (primary) and *beaconid* (secondary) describe the tokens involved in the interaction.

Ⅱ



Ⅲ

Each employee carried a button device (Ⅰ) during working hours, which reports a distance alarm when the distance between two button devices (ergo two workers) is less than 1.5 meters for more than 15 seconds. Recorded near-contact alarms were sent to a backend server via gateways and stored (Ⅱ). These gateways were installed at various locations within the plant site. The exact physical location of the gateways is known (Ⅲ). The raw data extracted from the safefactory backend contains a total of 279445 near contact alarms between 621 workers. The time period considered is from May 24, 2020, 03:03:51 to June 22, 2020 22:44:19.

Infection Model

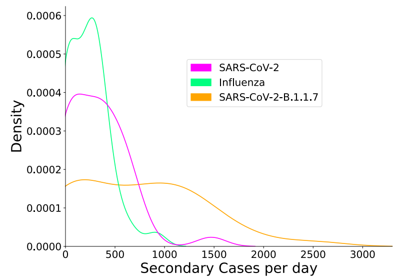
The infectious disease transmission model takes as input the contact records for a workplace with n individuals over a period of m days. The $n * m$ matrix Cijd therefore describes the number of close contact alerts between the pair of individuals (i,j) on day d for i ∈ [1,...,n] and d ∈ [1,...,m]. Assuming that primary case i becomes infected on day d and remains infected for d + T , the transmission probability from i to j can be calculated with

$$P_{i\to j,d}(p,C,T) = 1-(1-p)^{\sum_{k=d}^{d+T} c_{ijk}}$$

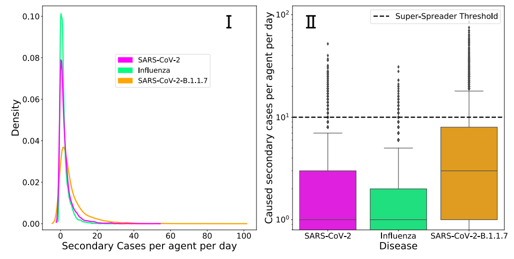
where i, j ∈ [1, . . . ,n] and d ∈ [1, . . . , m − T] and T as the infectious period. For the calculation of infection events, the risk of infection per social interaction ($P_{i\to j,d}$) is compared with a random variable from a discrete uniform distribution between 0 and 1. If $P_{i\to j,d}$ is larger, an infection event occurred; if it is smaller, none occurred. The distribution of the number of secondary cases (k) is therefore

$$P_k = \frac{1}{n} * \left(\sum_{i=1}^{n} \delta_k\left(\sum_{d=0}^{m} Y_{id}\right)\right), with\ \delta_{k,x} := \begin{cases} 1, if\ x = k \\ 0, otherwise \end{cases}$$

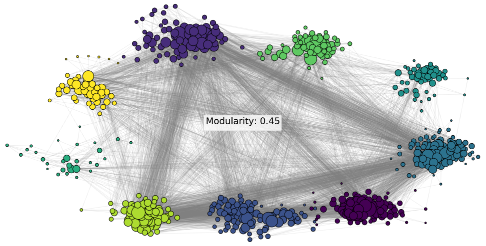
Results



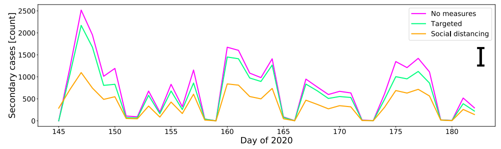
The risk of infection per social interaction was calculated for SARS-CoV-2 (0.0432), SARS-CoV-2-B-1.1.7 (0.1128), and influenza (0.1342). The distributions of secondary cases of the diseases considered have fat tails. This is consistent with reports[2] of these infectious diseases and means that a very large number of secondary cases were generated on very few days. It appears that SARS-CoV-2 has a slightly higher propensity to generate high secondary cases per day than influenza. However, the difference between the UK mutation and the other two diseases is significant. SARS-CoV-2-B.1.1.7 has a flatter distribution, resulting in a higher average number of secondary cases per day. There are also higher maximum values at the edge of the distribution of SARS-CoV-2-B.1.1.7.

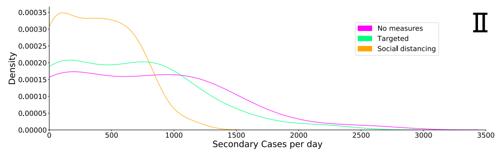


If the previously described secondary events per day exceed a threshold of 10, these social interactions are referred to as a super-spreading event (SSE). Accordingly, the total number of these SSEs per day describes the S-index. Here, the triggered secondary events per person per day are considered. Individuals infected with SARS-CoV-2 and Influenza infect a similar average number of individuals per day (5.3 II). However, the S-index of SARS-CoV-2 is 270, more than twice the S-index of Influenza (107). The broader distribution of the number of secondary cases of the UK mutation means that many of mild outliers (within 1.5×IQR) have caused SSE. Compared with this, only a few outliers infected with Influenza and SARS-CoV-2 caused SSE.



It is suspected that the gateway positions influence the number of proximity alarms recorded. Since workers usually follow certain patterns in their daily work and normally always deal with a similar group of colleagues, one would expect the formation of a community structure. For verification, the interaction graph over the complete period was divided into different communities using the Girvan-Newman method. The division with the highest calculated modularity of 0.45 devides the graph into 9 communities. The number of agents per community is evenly distributed between 29 and 115. The total degree of all agents within a community varies from 1 to 139. These results confirm the assumed community structure and show that some communities are more connected through more inter-community connections.





As a final investigation, it was evaluated whether measures that apply exclusively to the most important 50 agents in the system have significant impact on infection dynamics within the system. Only the UK SARS-CoV-2-B.1.1.7 variant is considered and social distancing as a countermeasure. It can be observed that targeted countermeasures applied to only about 15% of the agents within the social system can reduce the number of secondary cases. Figure part I shows that secondary cases per day for targeted actions are about halfway between social distancing and no actions. However, it is particularly interesting to see the right margin of the distributions in II. One can see that targeted measures for the most important agents according to Social-Network-Analysis (SNA) metrics, maximum values of secondary cases per day can be reduced by about 25%.

Discussion

Main findings:

- The presented framework allows the calculation of important infection parameters such as the S-index or the transmission probability per contact p based on social interaction data.
- Simmulations have shown that social distancing is a more efficient countermeasure than wearing masks in the considered subsystem.
- The analysis of temporary structures shows that different communities with different infection risks are formed.
- Social distancing applied only to the top 15% individuals identified by SNA metrics was able to reduce the maximum S-index values for SARS-CoV-2-B.1.1.7 by about 25%. This is an indication that when resources such as vaccine are scarce at the onset of a previously unknown infectious disease, targeted interventions can be a useful means of infection control.

Limitations:

- It could not be verified whether the calculated parameters are valid only for the social system under consideration or also for other subpopulations. The same applies to the results of the simulated countermeasures.
- Environmental parameters such as location of interaction and activity during social interaction are not considered in the model although recent studies emphasize the importance of including such parameters.
- Individual risk prevalences of each person are not considered. It is assumed that each person has the same risk of infection.
- Special metrics for identifying central nodes within the graph can again increase the efficiency of directed countermeasures.

References

[2] *M. Fukui & C. Furukawa. Power Laws in Superspreading Events: Evidence from Coronavirus Outbreaks and Implications for SIR Models. Preprint. Epidemiology, June* 2020. *DOI*: 10.1101 /2020.06.11.20128058. [1] *E. Dong; H. Du & L. Gardner. "An Interactive Web – Based Dash – board to Track COVID – 19 in Real Time". en. (May* 2020), *pp.* 533– 534. *ISSN*: 14733099. *DOI*: 10.1016/*S*1473 – 3099(20)30120 –