Mobile smartphone tracing can detect almost all SARS-CoV-2 infections

Bastian Prasse* and Piet Van Mieghem*

Delft University of Technology
June 25, 2020

Abstract

Currently, many countries are considering the introduction of tracing software on mobile smartphones with the main purpose to inform and alarm the mobile app user. Here, we demonstrate that, in addition to alarming and informing, mobile tracing can detect nearly all individuals that are infected by SARS-CoV-2, including the notorious asymptomatic infections. Besides voluntary reports of the infectious health status, our algorithm requires the location information of an overwhelming part of the population and then guarantees that almost all SARS-CoV-2 infections in that closed population can be detected. Our algorithm is based on a hidden Markov epidemic model and recursive Bayesian filtering. The potential that mobile tracing apps, in addition to medical testing and quarantining, can eradicate COVID-19 may persuade citizens to trade-off privacy against public health.

1 Introduction

The COVID-19 pandemic triggered firm lockdowns of societies and economies around the world. Lockdown measures must be released gently and, if necessary, retightened to avoid a dramatic second wave of COVID-19. To trace the pandemic, smartphone apps have recently received a lot of attention [1, 2, 3]. A particular challenge to estimating the prevalence of COVID-19 are the asymptomatic infections. Recent contact apps aim to alarm the user of a potential infection, if the user has been close to another user with a confirmed SARS-CoV-2 infection. Alarming individuals by contact apps is a particular method of social alertness [4, 5, 6, 7, 8]. If alerted, individuals are more cautious and less likely to become infected. For a comparison of the effect of social alertness and social distancing, we refer the reader to [9]. The awareness of potential infections may lead to eradication of the virus [10].

The intended use of smartphone app goes beyond alarming individuals. For instance, in the COVID Symptom Study [3], smartphone users provide their health status as a self report via an app on a daily basis. The self reports include user information, such as age and location, and potential

*Faculty of Electrical Engineering, Mathematics and Computer Science, P.O Box 5031, 2600 GA Delft, The Netherlands; email: b.prasse@tudelft.nl, p.f.a.vanmieghem@tudelft.nl
COVID-19 symptoms, such as fever or loss of smell and taste. The self-reports aid at identifying emerging geographical hotspots of SARS-CoV-2 infections.

Previous studies [11, 12, 13, 2] consider aggregated location information, in the form of mobility flow or population density. Here, we explore the full potential of location information for tracing the spread of COVID-19. More precisely, we consider that the location of the majority of the population is accessible on individual level.

2 Epidemic model

We consider the spread of SARS-CoV-2 among $N$ individuals. At any discrete time $k \in \mathbb{N}$, every individual $i$ has a viral state $X_i[k] \in \mathcal{C}$. The set of compartments equals $\mathcal{C} = \{S, E, I, I_{\text{asym}}, R, R_{\text{asym}}\}$. The state $X_i[k] = S$ denotes that individual $i$ is susceptible (healthy). The exposed state $X_i[k] = E$ denotes that individual $i$ is infected by SARS-CoV-2 but not contagious yet. After the exposed state $E$, an individual becomes either infectious symptomatic $I$ or infectious asymptomatic $I_{\text{asym}}$. Individuals in either infectious state $I$ and $I_{\text{asym}}$ are contagious to susceptible individuals in their vicinity. After some time, symptomatic infected individuals in $I$ transition to the symptomatic removed state $R$, due to recovery, quarantine, hospitalisation or death. Similarly, asymptomatic infectious individuals transition to the asymptomatic removed state $R_{\text{asym}}$. The sole difference of the compartment $R$ to $R_{\text{asym}}$ is that the respective individual is aware of a past infection. Removed individuals in $R$ or $R_{\text{asym}}$ cannot infect susceptible individuals any longer. We assume that a recovered individual is immune. Hence, multiple infections do not occur.

Throughout this work, users refer to the subset of all individuals $i = 1, ..., N$ who voluntarily report their health status via the smartphone app. We denote the fraction of individuals that are users by $c_0 \in [0, 1]$. Users of the app report a symptomatic infection or a recovery from COVID-19. Furthermore, we observe location information of the individuals $i = 1, ..., N$. There are two possibilities of incorporating location information. On the one hand, we could observe a neighbourhood $N_i[k] \subset \{1, ..., N\}$ that specifies all individuals $j \neq i$ that are sufficiently close (e.g., closer than 1.5 meters) to individual $i$ at time $k$. The neighbourhood $N_i[k]$ could be obtained from bluetooth. On the other hand, we could observe a $2 \times 1$ location vector $z_i[k] \in \mathbb{R}^2$. The vector $z_i[k]$ specifies the latitude and longitude of individual $i$ at time $k$ and can be obtained, for instance, by GPS. The neighbourhood of node $i$ is obtained from the location vector $z_i[k]$ by

$$N_i[k] = \{j = 1, ..., N, j \neq i \, | \, \|z_i[k] - z_j[k]\|_2 \leq d_{\text{inf}}\}$$

for some distance $d_{\text{inf}}$. The sole location information in our model are the neighbourhoods $N_i[k]$. We do not distinguish between neighbourhoods $N_i[k]$ that were measured directly, by bluetooth, or indirectly, by GPS coordinates.

We model the spread of COVID-19 by a hidden Markov model, which consists of two parts. First, the dynamics of the viral state $X_i[k]$. Second, the user behaviour of reporting their viral state $X_i[k]$.

2.1 Dynamics

An individual with a symptomatic infection traverses the viral states $S \rightarrow E \rightarrow I \rightarrow R$. Analogously, the course of an asymptomatic infection is $S \rightarrow E \rightarrow I_{\text{asym}} \rightarrow R_{\text{asym}}$. An infectious individual $j$,
with $X_j[k] = \mathcal{I}$ or $X_j[k] = \mathcal{I}_{asym}$, infects a susceptible individual $i$ with the infection probability $\beta$, if individual $j$ is in the neighbourhood $\mathcal{N}_i[k]$ of individual $i$. The infection probability $\beta$ depends on the contagiousness of SARS-CoV-2 and on the prevalence of facemasks and other spread reduction measures. The set

$$\mathcal{N}_{inf,i}[k] = \{ j \in \mathcal{N}_i[k] | X_j[k] = \mathcal{I} \text{ or } X_j[k] = \mathcal{I}_{asym} \}$$

consists of all infectious individuals $j$ that are close to individual $i$ at time $k$. The number of infectious neighbours of individual $i$ at time $k$ is denoted by $|\mathcal{N}_{inf,i}[k]|$. The probability of an infection of individual $i$ follows from potential infections by any individual $j$ in the set $\mathcal{N}_{inf,i}[k]$ as

$$\Pr \left[ X_i[k+1] = \mathcal{E} | X_i[k] = \mathcal{S}, \mathcal{N}_{inf,i}[k] \right] = 1 - (1 - \beta)^{|\mathcal{N}_{inf,i}[k]|} (1 - \epsilon).$$

(2)

Here, the self-infection probability $\epsilon$ accounts for infections from an individual $i$ other than $i = 1, ..., N$, whose location is inaccessible. Individuals leave the exposed state $\mathcal{E}$ with the incubation probability $\gamma$ to an infectious state,

$$\Pr \left[ X_i[k+1] = c | X_i[k] = \mathcal{E} \right] = \begin{cases} 
\gamma \alpha & \text{if } c = \mathcal{I}_{asym}, \\
\gamma (1 - \alpha) & \text{if } c = \mathcal{I}, \\
(1 - \gamma) & \text{if } c = \mathcal{E}.
\end{cases}$$

Here, $\alpha$ denotes the probability of an asymptomatic infection. Any symptomatic infected individual is removed with the removal probability $\delta$. In other words,

$$\Pr \left[ X_i[k+1] = \mathcal{R} | X_i[k] = \mathcal{I} \right] = \delta.$$  

(3)

For simplicity, we consider that the removal of an asymptomatic infection is the same as (3),

$$\Pr \left[ X_i[k+1] = \mathcal{R}_{asym} | X_i[k] = \mathcal{I}_{asym} \right] = \delta.$$  

However, the modelling framework does allow for different removal probabilities of symptomatic and asymptomatic infections. Denote the first time that individual $i$ is infected by $k_{I,i}$, $X_i[k_{I,i}] = \mathcal{I}$ and $X_i[k_{I,i} - 1] = \mathcal{E}$. Similarly, denote the first time that individual $i$ is removed by $k_{R,i}$. Since the viral state compartments are in the order $\mathcal{E} \rightarrow \mathcal{I} \rightarrow \mathcal{R}$, it holds that $k_{R,i} > k_{I,i}$. The sojourn time $k_{R,i} - k_{I,i}$ of state $\mathcal{I}$ is the number of discrete times $k$ that individual $i$ has been infected. By (3), we implicitly assume that the sojourn time follows a geometric distribution with mean $1/\delta$.

### 2.2 Observations

Users of the contact app submit a health report at every time $k$. We denote the reported viral state of user $i$ as $X_{rep,i}[k]$. At every time $k$, the reported state $X_{rep,i}[k]$ equals either healthy $\mathcal{S}$, infected $\mathcal{I}$, or removed $\mathcal{R}$. If user $i$ is infected $X_i[k] = \mathcal{I}$ or removed $X_i[k] = \mathcal{R}$, then the reported viral state equals $X_{rep,i}[k] = \mathcal{I}$ or $X_{rep,i}[k] = \mathcal{R}$, respectively. If the true viral state $X_i[k]$ equals $\mathcal{E}$, $\mathcal{I}_{asym}$ or $\mathcal{R}_{asym}$, then user $i$ has no symptoms and reports to be healthy $X_{rep,i}[k] = \mathcal{S}$. If user $i$ is healthy, $X_i[k] = \mathcal{S}$, then the reported state equals either $X_{rep,i}[k] = \mathcal{I}$, with false alarm probability $p_{fa}$, or $X_{rep,i}[k] = \mathcal{S}$, with probability $(1 - p_{fa})$. 

3
3 Who is infected?

At time $k$, we would like to know: who is infected by COVID-19? In other words, for every individual $i$, we would like to compute the symptomatic infection risk

$$\Pr [X_i[k] = I | \mathcal{M}[k]]$$

and the asymptomatic infection risk

$$\Pr [X_i[k] = I_{\text{asym}} | \mathcal{M}[k]].$$

Here, we formally define all observations, or measurements, up until time $k$ as $\mathcal{M}[k]$. More specifically, the set $\mathcal{M}[k]$ specifies the reported viral state $X_{\text{rep},i}[l]$ of every user $i$ and the neighbourhood $\mathcal{N}_{\text{inf},i}[l]$ of every individual $i$ at every time $l \leq k$. In Appendix A we propose an recursive Bayesian filtering method to (approximately) compute the infection risks $\Pr [X_i[k] = I | \mathcal{M}[k]]$ and $\Pr [X_i[k] = I_{\text{asym}} | \mathcal{M}[k]]$. As a side product, we obtain the probabilities $\Pr [X_i[k] = c | \mathcal{M}[k]]$ for the other viral states $c = S, E, R, R_{\text{asym}}$. The computation time is polynomial in the number of individuals $N$ and the number of observations $k$.

We perform simulations of the hidden Markov model (Section 2) with $N = 10,000$ individuals and vary the fraction of app users $c_0$. To generate the locations $z_i[k]$ at every time $k$, we employ a simple movement model: For every individual $i$, both entries of the initial $2 \times 1$ location vector $z_i[1]$ are set to a uniform random number in $[0, 1]$. Given the location vector $z_i[k]$ at any time $k$, we obtain the location vector at the next time $k + 1$ as follows. With a probability of 0.8, the location does not change, and hence $z_i[k + 1] = z_i[k]$. Otherwise, with a probability of 0.2, both entries of the location vector $z_i[k + 1]$ are set to a uniform random number in $[0, 1]$. To obtain the neighbourhoods $\mathcal{N}_i[k]$ from $\Pi$, we set the distance to $d_{\text{inf}} = 0.005$. The curing and infection probabilities are set to $\delta = 0.2$ and $\beta = 0.5$, respectively. The self-infection probability is set to $\epsilon = 0.001$, and the false alarm probability is set to $p_{fa} = 0.05$. Furthermore, we set the incubation probability to $\gamma = 0.5$ and the fraction of asymptomatic infections to $\alpha = 0.1$. For any individual $i$, the true initial viral state is set to $X_i[1] = I$ or $X_i[1] = S$ with a probability of 0.01 and 0.99, respectively. Furthermore, we assume that we know the true initial viral state $X_i[1]$ of 90% randomly chosen individuals. For the other 10% of individuals, we know only the prior distribution of the viral state $X_i[1]$ as $\Pr [X_i[1] = S] = 0.99$ and $\Pr [X_i[1] = I] = 0.01$.

Can we estimate the average number of infections? First, we define $I_{\text{all}}[k]$ as the true number of individuals whose viral state $X_i[k] = I$. Similarly, we define the estimated total number of infections at time $k$ as

$$\hat{I}_{\text{all}}[k] = \sum_{i=1}^{N} \Pr [X_i[k] = I | \mathcal{M}[k]].$$

For the asymptomatic infections, the quantities $I_{\text{asym,all}}[k]$ and $\hat{I}_{\text{asym,all}}[k]$ are defined analogously. Figure 1 compares the estimate $\hat{I}_{\text{all}}[k]$ to the true number symptomatic infections $I_{\text{all}}[k]$. If $c_0 = 80\%$ of the population use the contact app, then the number of symptomatic infections $\hat{I}_{\text{all}}[k]$ in the whole population is traced almost perfectly. For a user fraction of $c_0 = 40\%$, the tracing accuracy
Figure 1: **Tracing the number of symptomatic infections.** The number of symptomatic infections $I_{\text{all}}[k]$ and the estimate $\hat{I}_{\text{all}}[k]$ versus time $k$ for one realisation of the hidden Markov epidemic model.

deteriorates, but still seems acceptable. Figure 2 shows that the number of asymptomatic infections $I_{\text{asym,all}}[k]$ is traced reasonable accurate.

Can we estimate if a single individual $i$ is infected? We have computed the posterior probability $\Pr[X_i[k] = c | M[k]]$ for every compartment $c \in C$. Thus, we obtain the Bayesian estimate of the viral state $X_i[k]$ at any time $k$ as

$$\hat{X}_i[k] = \arg\max_{c \in C} \Pr[X_i[k] = c | M[k]].$$

We define $C[k]$ as the number of individuals $i$ for which the estimate at time $k$ is correct, i.e., $\hat{X}_i[k] = X_i[k]$. Figure 3 shows that the viral state $X_i[k]$ of most individuals $i$ is estimated correctly at any time $k$, also if few people use the contact app. For instance, with a fraction of $c_0 = 40\%$ users, the viral state $X_i[k]$ of more than 60\% individuals of the whole population is estimated correctly at all times $k$. Thus, provided the location $z_i[k]$ of every individual $i$ is known, the whole population significantly benefits from a fraction of individuals that use the contact app.

The spreading parameters $\alpha, \beta, \delta, \epsilon$ and $\gamma$ might not be know exactly but could be estimated from observing the viral spread \[14, 15\]. In Appendix B we show that the computation of the infection risk $\Pr[X_i[k] = I | M[k]]$ is reasonably robust to errors in the spreading parameters estimates.
Figure 2: **Tracing the number of asymptomatic infections.** The number of asymptomatic infections $I_{\text{asym,all}}[k]$ and the estimate $\hat{I}_{\text{asym,all}}[k]$ versus time $k$ for one realisation of the hidden Markov epidemic model.

Figure 3: **Tracing the viral state of single individuals.** The number of correct viral state estimates $\hat{X}_i[k]$ versus time $k$. 
4 Conclusions

This work considers the application of contact apps beyond alarming users of potential infections: the detection of SARS-CoV-2 infections. We assume that the location of individuals is accessible, but only app users report their health status. Our results indicate that, even if only a fraction of the population use the contact app, all SARS-CoV-2 infections can be detected within a reasonable accuracy, including asymptomatic infections and within a reasonable time period.

Data privacy regulations [16] prohibit the use of location information of individuals without their consent. Without consent of the respective individual, location information must be processed in aggregated or anonymised form. This work demonstrates the great value of location information for tracing COVID-19, which is an incentive to provide location data. Furthermore, our modelling framework can be used as basis for further work to unlock the full potential of location information, whilst ensuring data privacy. In particular, the lack of location information could be incorporated via the self-infection infection probability $\epsilon$.

References

[1] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, “Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing,” *Science*, vol. 368, no. 6491, 2020.

[2] N. Oliver, B. Lepri, H. Sterly, R. Lambiotte, S. Deletaille, M. De Nadai, E. Letouzé, A. A. Salah, R. Benjamins, C. Cattuto et al., “Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle,” 2020.

[3] D. A. Drew, L. H. Nguyen, C. J. Steves, C. Menni, M. Freydin, T. Varsavsky, C. H. Sudre, M. J. Cardoso, S. Ourselin, J. Wolf et al., “Rapid implementation of mobile technology for real-time epidemiology of COVID-19,” *Science*, 2020.

[4] S. Funk, M. Salathé, and V. A. Jansen, “Modelling the influence of human behaviour on the spread of infectious diseases: a review,” *Journal of the Royal Society Interface*, vol. 7, no. 50, pp. 1247–1256, 2010.

[5] I. Z. Kiss, J. Cassell, M. Recker, and P. L. Simon, “The impact of information transmission on epidemic outbreaks,” *Mathematical Biosciences*, vol. 225, no. 1, pp. 1–10, 2010.

[6] F. D. Sahneh and C. Scoglio, “Epidemic spread in human networks,” *Proc. IEEE Conf. Decision Control*, pp. 3008–3013, 2011.

[7] F. D. Sahneh, F. N. Chowdhury, and C. M. Scoglio, “On the existence of a threshold for preventive behavioral responses to suppress epidemic spreading,” *Scientific Reports*, vol. 2, p. 632, 2012.

[8] G. Theodorakopoulos, J.-Y. Le Boudec, and J. S. Baras, “Selfish response to epidemic propagation,” *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 363–376, 2012.
A Computation of the infection risk

A.1 Assumptions in the computations

We define the $N \times 1$ viral state vector as $X[k] = (X_1[k], ..., X_N[k])^T$. The reported viral state vector $X_{\text{rep}}[k]$ is defined analogously, where we formally set $X_{\text{rep},i}[k] = 0$ if individual $i$ does not use the contact app. We rely on three assumptions to compute the infection risk (4). First, we assume the conditional stochastic independence

$$
\Pr[X[k]|X_{\text{rep}}[k], M[k-1]] \approx \prod_{i=1}^{N} \Pr[X_i[k]|X_{\text{rep},i}[k], M[k-1]]. \quad (5)
$$
There are $6^N$ possible combinations of the entries of the viral state vector $X[k]$. Thus, it is practically impossible to state the full distribution of the vector $X[k]$. The assumption (5) instead implies that the distribution of the vector $X[k]$ can be decomposed into the marginal distribution of the entries $X_1[k], X_2[k], ..., X_N[k]$, which can be computed separately. Furthermore, assumption (5) might be of relevance to privacy: The full distribution $Pr [X[k]|X_{rep}[k], \mathcal{M}[k-1]]$ is sensitive data. In contrast, the single factors $Pr [X_i[k]|X_{rep,i}[k], \mathcal{M}[k-1]]$ might in parts be made accessible to some individuals.

Furthermore, we make the assumption that the viral state $X_i[k]$ does not depend on the infectious neighbourhoods $\mathcal{N}_{inf,1}[k], ..., \mathcal{N}_{inf,N}[k]$ at time $k$. More precisely,

$$Pr [X_i[k]|X_{rep,i}[k], \mathcal{N}_{inf,1}[k], ..., \mathcal{N}_{inf,N}[k], \mathcal{M}[k-1]] = Pr [X_i[k]|X_{rep,i}[k], \mathcal{M}[k-1]].$$

The viral state $X_i[k]$ does depend on the neighbourhoods $\mathcal{N}_{inf,1}[k-1]$ at the previous time step $k-1$, due to the infection probability (2). Thus, the impact of the location on the infection dynamics is delayed by one time step, and we consider assumption (5) rather technical. Third, we assume the analogue to (6) for the joint distribution of the random variables $X_1[k], ..., X_N[k]$,

$$Pr [X[k]|X_{rep}[k], \mathcal{N}_{inf,1}[k], ..., \mathcal{N}_{inf,N}[k], \mathcal{M}[k-1]] = Pr [X[k]|X_{rep}[k], \mathcal{M}[k-1]].$$ (7)

### A.2 Recursive Bayesian filtering

The infection risk (4) can be computed by iterating over time:

**Initialisation** At time $k = 1$, we assume that the probability distribution

$$Pr [X_i[1]]$$

is given for every individual $i$. Formally, we can write

$$Pr [X_i[1]] = Pr [X_i[1]|\mathcal{M}[0]],$$

since there are no observations at time $k = 0$. (Or, the set of observation $\mathcal{M}[0]$ at time $k = 0$ is empty, because we start measuring at $k = 1$.)

**Measurement update** We are given the distribution $Pr [X_i[k]|\mathcal{M}[k-1]]$ for every node $i$. (Starting with (5) at time $k = 1$.) For every user $i$, the measurement update incorporates the reported viral state $X_{rep,i}[k]$ to obtain a more accurate distribution of the viral state $X_i[k]$. For individuals $i$ who do not report their viral state $X_{rep,i}[k]$ through the app, we formally set

$$Pr [X_i[k]|X_{rep,i}[k], \mathcal{M}[k-1]] = Pr [X_i[k]|\mathcal{M}[k-1]].$$

For the other individuals $i$, whose reported state $X_{rep,i}[k]$ is available, we compute the probability $Pr [X_i[k]|X_{rep,i}[k], \mathcal{M}[k-1]]$ with Bayes’ Theorem (7) as

$$Pr [X_i[k]|X_{rep,i}[k], \mathcal{M}[k-1]] = \frac{Pr [X_{rep,i}[k]|X_i[k], \mathcal{M}[k-1]] Pr [X_i[k]|\mathcal{M}[k-1]]}{Pr [X_{rep,i}[k]|\mathcal{M}[k-1]].}$$

Given the viral state $X_i[k]$, the reported viral state $X_{rep,i}[k]$ does not depend on past measurements $\mathcal{M}[k-1]$, and hence

$$Pr [X_i[k]|X_{rep,i}[k], \mathcal{M}[k-1]] = \frac{Pr [X_{rep,i}[k]|X_i[k]] Pr [X_i[k]|\mathcal{M}[k-1]]}{Pr [X_{rep,i}[k]|\mathcal{M}[k-1]]}. \quad (9)$$
The distribution \( \Pr[X_{\text{rep},i}[k]|X_i[k]] \) is specified by the observation model in Subsection \(^{22}\). In particular, for \( X_{\text{rep},i}[k] = \mathcal{R} \), it holds that

\[
\Pr[X_{\text{rep},i}[k] = \mathcal{R}|X_i[k] = c, \mathcal{M}[k-1]] = \begin{cases} 
1 & \text{if } c = \mathcal{R}, \\
0 & \text{if } c \neq \mathcal{R}.
\end{cases}
\]

If user \( i \) reports to be healthy, \( X_{\text{rep},i}[k] = \mathcal{S} \), then we obtain that

\[
\Pr[X_{\text{rep},i}[k] = \mathcal{S}|X_i[k] = c, \mathcal{M}[k-1]] = \begin{cases} 
1 & \text{if } c \in \{\mathcal{E}, \mathcal{I}_{\text{asym}}, \mathcal{R}_{\text{asym}}\}, \\
1 - p_{\text{fa}} & \text{if } c = \mathcal{S}, \\
0 & \text{if } c \in \{\mathcal{I}, \mathcal{R}\}.
\end{cases}
\]

Similarly, if user \( i \) reports to be infected, \( X_{\text{rep},i}[k] = \mathcal{I} \), then it holds that

\[
\Pr[X_{\text{rep},i}[k] = \mathcal{I}|X_i[k] = c, \mathcal{M}[k-1]] = \begin{cases} 
1 & \text{if } c = \mathcal{I}, \\
p_{\text{fa}} & \text{if } c = \mathcal{S}, \\
0 & \text{if } c \in \{\mathcal{E}, \mathcal{R}, \mathcal{I}_{\text{asym}}, \mathcal{R}_{\text{asym}}\}.
\end{cases}
\]

The denominator in (9) follows from the law of total probability \(^{17}\) as

\[
\Pr[X_{\text{rep},i}[k]|\mathcal{M}[k-1]] = \sum_{c \in \mathcal{C}} \Pr[X_{\text{rep},i}[k]|X_i[k] = c] \Pr[X_i[k] = c|\mathcal{M}[k-1]].
\]

**Time update** The measurement update computes the distribution \( \Pr[X_i[k]|X_{\text{rep},i}[k], \mathcal{M}[k-1]] \), from which the time update obtains the distribution \( \Pr[X_i[k+1]|\mathcal{M}[k]] \). The law of total probability yields that

\[
\Pr[X_i[k+1]|\mathcal{M}[k]] = \sum_{c \in \mathcal{C}} \Pr[X_i[k+1], X_i[k] = c|\mathcal{M}[k]]
\]

\[
= \sum_{c \in \mathcal{C}} \Pr[X_i[k+1]|X_i[k] = c, \mathcal{M}[k]] \Pr[X_i[k] = c|\mathcal{M}[k]],
\]

(10)

where the last equation follows from the definition of the conditional probability. First, we consider the term \( \Pr[X_i[k] = c|\mathcal{M}[k]] \) in (10). With the definition of the set of all observations \( \mathcal{M}[k] \), it holds that

\[
\Pr[X_i[k] = c|\mathcal{M}[k]] = \Pr[X_i[k] = c|X_{\text{rep},i}[k], \mathcal{N}_{\text{inf},1}[k], ..., \mathcal{N}_{\text{inf},N}[k], \mathcal{M}[k-1]].
\]

Assumption \(^{5}\) implies that

\[
\Pr[X_i[k] = c|\mathcal{M}[k]] = \Pr[X_i[k] = c|X_{\text{rep},i}[k], \mathcal{N}_{\text{inf},1}[k], ..., \mathcal{N}_{\text{inf},N}[k], \mathcal{M}[k-1]].
\]

Then, with assumption \(^{6}\), we obtain that

\[
\Pr[X_i[k] = c|\mathcal{M}[k]] = \Pr[X_i[k] = c|X_{\text{rep},i}[k], \mathcal{M}[k-1]],
\]

(11)

which has been calculated by the previous measurement update. Second, we consider the term \( \Pr[X_i[k+1]|X_i[k] = c, \mathcal{M}[k]] \) in (10). The transition of the viral state \( X_i[k] \) from time \( k \) to \( k+1 \)
depends on the cardinality of the infectious neighbourhood \( N_{\text{inf},i}[k] \), see Subsection 2.1. However, we do not directly observe the set \( N_{\text{inf},i}[k] \) but instead the set \( N_i[k] \). Since \( N_{\text{inf},i}[k] \subset N_i[k] \), it holds that

\[ 0 \leq |N_{\text{inf},i}[k]| \leq |N_i[k]|. \]

Thus, we can apply the law of total probability to obtain that

\[
\Pr \left[ X_i[k+1] | X_i[k] = c, M[k] \right] = \sum_{m=0}^{|N_i[k]|} \Pr \left[ X_i[k+1] | X_i[k] = c, M[k], |N_{\text{inf},i}[k]| = m \right] \\
\quad \cdot \Pr \left[ |N_{\text{inf},i}[k]| = m | X_i[k] = c, M[k] \right],
\]

which simplifies to

\[
\Pr \left[ X_i[k+1] | X_i[k] = c, M[k] \right] = \sum_{m=0}^{|N_i[k]|} \Pr \left[ X_i[k+1] | X_i[k] = c, |N_{\text{inf},i}[k]| = m \right] \\
\quad \cdot \Pr \left[ |N_{\text{inf},i}[k]| = m | M[k] \right].
\]

Subsection 2.1 fully specifies the term \( \Pr \left[ X_i[k+1] | X_i[k] = c, |N_{\text{inf},i}[k]| = m \right] \) in (12). For instance, for the susceptible compartment \( X_i[k] = S \), we obtain that

\[
\Pr \left[ X_i[k+1] = c | X_i[k] = S, |N_{\text{inf},i}[k]| = m \right] = \begin{cases} 
(1 - \beta)^m (1 - \epsilon) & \text{if } c = S, \\
1 - (1 - \beta)^m (1 - \epsilon) & \text{if } c = E, \\
0 & \text{otherwise.}
\end{cases}
\]

To compute (12), it remains to determine the probabilities \( \Pr \left[ |N_{\text{inf},i}[k]| = m | M[k] \right] \) for all cardinalities \( m = 0, 1, ..., |N_i[k]| \). Without loss of generality, we assume that the neighbourhood of individual \( i \) at time \( k \) equals

\[ N_i[k] = \{1, 2, ..., M\}, \]

where \( M = |N_i[k]| \). The law of total probability yields that

\[
\Pr \left[ |N_{\text{inf},i}[k]| = m | M[k] \right] = \sum_{c_1 \in C} ... \sum_{c_M \in C} \Pr \left[ |N_{\text{inf},i}[k]| = m, X_1[k] = c_1, ..., X_M[k] = c_M, M[k] \right] \\
\quad \cdot \Pr \left[ X_1[k] = c_1, ..., X_M[k] = c_M | M[k] \right].
\]

With the definition of the set of all observations \( M[k] \), we obtain that

\[
\Pr \left[ |N_{\text{inf},i}[k]| = m | M[k] \right] = \sum_{c_1 \in C} ... \sum_{c_M \in C} \Pr \left[ |N_{\text{inf},i}[k]| = m | X_1[k] = c_1, ..., X_M[k] = c_M, N_i[k] \right] \\
\quad \cdot \Pr \left[ X_1[k] = c_1, ..., X_M[k] = c_M | X_{\text{rep}}[k], N_{\text{inf},1}[k], ..., N_{\text{inf},N}[k], M[k-1] \right].
\]

\footnote{Otherwise, consider a relabelling of the nodes \( j \) in the set \( N_i[k] \).}
From assumption (7), it follows that

\[
\Pr \left[ |\mathcal{N}_{\text{inf},i}[k]| = m \mid \mathcal{M}[k] \right] = \sum_{c_1 \in \mathcal{C}} \ldots \sum_{c_M \in \mathcal{C}} \Pr \left[ |\mathcal{N}_{\text{inf},i}[k]| = m \mid X_1[k] = c_1, \ldots, X_M[k] = c_M, \mathcal{N}_i[k] \right] \cdot \Pr \left[ X_1[k] = c_1, \ldots, X_M[k] = c_M \mid X_{\text{rep},i}[k], \mathcal{M}[k-1] \right].
\]

With assumption (5), we obtain that

\[
\Pr \left[ |\mathcal{N}_{\text{inf},i}[k]| = m \mid \mathcal{M}[k] \right] = \sum_{c_1 \in \mathcal{C}} \ldots \sum_{c_M \in \mathcal{C}} \Pr \left[ |\mathcal{N}_{\text{inf},i}[k]| = m \mid X_1[k] = c_1, \ldots, X_M[k] = c_M \right] \prod_{j=1}^{M} \Pr \left[ X_j[k] = c_j \mid X_{\text{rep},j}[k], \mathcal{M}[k-1] \right]. \quad (13)
\]

The set \( \mathcal{N}_{\text{inf},i}[k] \) only consists of individuals \( j \) with \( X_j[k] = \mathcal{I} \) or \( X_j[k] = \mathcal{I}_{\text{asym}} \). For \( j = 1, \ldots, M \), we define the Bernoulli random variable \( \psi_j \) as

\[
\psi_j = \begin{cases} 
1 & \text{with probability } p_j, \\
0 & \text{with probability } 1 - p_j,
\end{cases}
\]

with the success probability

\[
p_j = \Pr \left[ X_j[k] = \mathcal{I} \mid X_{\text{rep},j}[k], \mathcal{M}[k-1] \right] + \Pr \left[ X_j[k] = \mathcal{I}_{\text{asym}} \mid X_{\text{rep},j}[k], \mathcal{M}[k-1] \right].
\]

From (13) it follows that the cardinality \( |\mathcal{N}_{\text{inf},i}[k]| \) is the sum of \( M \) Bernoulli random variables \( \psi_j \in \{0, 1\} \) with different success probabilities \( p_j \). Hence, the cardinality \( |\mathcal{N}_{\text{inf},i}[k]| \) follows a Poisson binomial distribution [18]. We obtain the distribution of \( |\mathcal{N}_{\text{inf},i}[k]| \) by convolution of the distributions of the random variables \( \psi_1, \ldots, \psi_M \). If the number \( M \) is large, then the convolution might take long. For large \( M \), there are more efficient algorithms [18] for computing the distribution of the cardinality \( |\mathcal{N}_{\text{inf},i}[k]| \) (based on the discrete Fourier transform).

After the initialisation, the measurement update and the time update are alternated for every time \( k \). Finally, the risk factor (11) is obtained from (11) at the last time step \( k \).

**B Robustness to spreading parameter errors**

We evaluate the robustness of detecting infections if the spreading parameters \( \alpha, \beta, \delta, \epsilon \) and \( \gamma \) are not exactly known. In the computations of the infection risk in Section A, we replace the exact spreading parameters by the respective estimates \( \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\epsilon} \) and \( \hat{\gamma} \). The spreading parameter estimates are subject to random 10% relative errors. For instance, the infection probability estimate \( \hat{\beta} \) is a uniform random number in \([0.9\beta, 1.1\beta]\). Except for the spreading parameters, all parameters are set to the same values as in Section 2. Figures 4 to 6 show that tracing the infections with errors on the spreading parameters performs moderately worse than the simulations without errors in Section 3. Thus, the computation of the infection risk is relatively robust.
Figure 4: Tracing the number of symptomatic infections with spreading parameter errors. The number of symptomatic infections $I_{\text{all}}[k]$ and the estimate $\hat{I}_{\text{all}}[k]$ versus time $k$ for one realisation of the hidden Markov epidemic model and 10% relative errors on the spreading parameters.
Figure 5: **Tracing the number of asymptomatic infections with spreading parameter errors.**
The number of asymptomatic infections $I_{\text{asym, all}}[k]$ and the estimate $\hat{I}_{\text{asym, all}}[k]$ versus time $k$ for one realisation of the hidden Markov epidemic model and 10% relative errors on the spreading parameters.

Figure 6: **Tracing the viral state of single individuals with spreading parameter errors.**
The number of correct viral state estimates $\hat{X}_i[k]$ versus time $k$ and 10% relative errors on the spreading parameters.