

An observation-based quantitative approach to predict the spread of infectious diseases within a cluster

Laura Carnevali¹*, Silvia Dani¹, Niccolò Niccoli¹, Benedetta Picano,
Enrico Vicario¹

Department of Information Engineering, University of Florence, via di Santa Marta 3, Florence, Italy

ARTICLE INFO

Dataset link: <https://doi.org/10.5281/zenodo.21559792>

Keywords:

Infectious diseases
Cluster
Timestamped observations
Quantitative evaluation
Stochastic models

ABSTRACT

Contact tracing, early diagnosis, and social distancing are crucial to contain outbreaks of infectious diseases, especially in the presence of asymptomatic infectives. Therefore, identifying the individuals at the highest risk of infectious disease and prioritizing them for testing is necessary to ensure both the public health and the cost-effectiveness of screening strategies. In this paper, we present an efficient quantitative approach to predict the spread of infectious diseases within a cluster by exploiting a stochastic model of disease evolution in an individual alongside observations of contacts, symptoms, and results of diagnostic tests. Specifically, we present an iterative solution technique to estimate the probability that a subject is infectious and not isolated over time given the observations acquired up to that time, achieving computational efficiency by disregarding dependencies among observations. We generated multiple synthetic data sets of observations, featuring different social network topologies, numbers of individuals, and densities of internal contacts, while accounting for noisy, erroneous, and missing observations. For each data set, we performed extensive experiments by varying the stochastic parameters of our approach from those used to derive a ground truth by performing stochastic simulation. Experimental results demonstrate that our approach effectively ranks subjects according to their probability of being infectious and not isolated, achieving high accuracy with respect to the ground truth. Notably, our approach outperforms an alternative baseline that accounts for dependencies among observations while maintaining a comparable runtime.

1. Introduction

1.1. Motivation

In infectious disease pandemics, a variety of protective measures have been implemented to curtail outbreaks of infection [1–3], notably including contact tracing, early diagnosis, and social distancing. In particular, identifying the individuals at the highest risk of infectious disease, and thus needing to be tested first during screening programs, significantly helps to improve the cost-effectiveness of testing [4,5]. To this end, quantitative evaluation of stochastic models of the spread of infectious diseases within groups of individuals can effectively support early assessment of protective measures and runtime adaptation to changes in behaviors of individuals [6].

* Correspondence to: via di Santa Marta 3, 50139 Firenze, Italy.

E-mail addresses: laura.carnevali@unifi.it (L. Carnevali), silvia.dani@unifi.it (S. Dani), niccolo.niccoli@unifi.it (N. Niccoli), benedetta.picano@unifi.it (B. Picano), enrico.vicario@unifi.it (E. Vicario).

<https://doi.org/10.1016/j.simpat.2026.103332>

Received 24 November 2025; Received in revised form 26 July 2026; Accepted 27 July 2026

Available online 1 August 2026

1569-190X/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1.2. Related works

The literature on epidemiological studies has widely investigated the spread of infectious diseases for different purposes such as disease containment and outbreak reconstruction [7]. A vast body of approaches leverages observational data, spanning from the molecular to the social level, and a variety of computational tools, including time-series forecasting [8], compartmental models [9,10], network-based and agent-based models [11], and Bayesian inference frameworks [12].

Time-series models have been used for epidemic prediction mainly after the COVID-19 emergence [8], accompanied by a significant adoption of Machine Learning (ML) and Deep Learning (DL) methods. Neural approaches are presented in [13], where traditional multilayer perceptrons are employed, and in [14], which exploits classic Long Short-Term Memory (LSTM) networks. Univariate and multivariate approaches are compared in [15] for infectious disease outbreak, whereas convolutional architectures and a combination of recurrent networks with ARIMA-based models are used in the study presented in [16]. While excellent for prediction in the short to medium term, time-series forecasting methods largely lack structural explainability [17].

Compartmental models partition the investigated population into groups corresponding to epidemiological states (compartments) and typically characterize their evolution through ordinary differential equations [9,10]. The authors of Pernice et al. [18] extend the Susceptible-Exposed-Infected-Removed-Susceptible (SEIRS) model to capture the population age structure, and they perform macro-scale stochastic simulation to investigate different scenarios of COVID-19 diffusion in the Piedmont region. Authors in [19] present a comprehensive framework for scaling models from localized epidemics to global pandemics. They emphasize the challenges of multi-scale modeling, where local transmission processes must be integrated into global models that account for inter-region travel and varying public health responses. The paper discusses various computational tools, including hybrid models that blend deterministic and stochastic approaches, to capture the complexity of pandemic spread. The applicability of their model extends beyond infectious diseases, providing a framework for global risk assessments. The authors of Eslahchi and Movahedi [20] present an analytical approach to calculate the transition probabilities in a birth-death process within a Markov epidemic model, showing how fluctuations in the rates of infection and recovery can impact the overall dynamics of an epidemic. Colored Hybrid Petri Nets (CHPNs) are applied in [21] as a flexible modeling approach for multi-scale biological networks, including infectious disease transmission, supporting the analysis of multi-phase epidemic dynamics. Similarly, Colored Stochastic Petri Nets (CSPNs) are used in [22] for modeling the spread of directly transmitted infectious diseases, representing heterogeneity within the population, such as varying susceptibility or recovery rates, by assigning different colors to individuals. While highly interpretable and great for exploring general public health interventions, compartmental models struggle to represent individual heterogeneity and spatial dynamics [23].

Network-based models represent interactions among individuals explicitly and evaluate how the structure of social contacts (whether random, scale-free, or small-world networks) affects transmission patterns, emphasizing the importance of heterogeneous contact structures in epidemic modeling [11]. Network theory in infectious diseases is exploited in [24], focusing on the structure of contact networks and how varying degrees of connectivity, assortativity, and clustering influence disease transmission, providing insights into how different network topologies affect both epidemic size and the effectiveness of control measures. The approach of Bodlaj and Batagelj [25] models complex contagions where the probability of adopting a behavior or contracting an infection depends on multiple exposures rather than a single contact, capturing the effects of strong intra-group connections and weaker inter-group ties on contagion dynamics, and showing that clustering can significantly slow the spread of infections compared to random networks while dense clustering can lead to rapid outbreaks and hotspots of transmission. The impact of recurrent human mobility patterns on the spread of infectious diseases is explored in [26] using a statistical mechanics approach, modeling how frequent movements between communities can accelerate transmission rates and complicate containment efforts, and thus stressing the importance of incorporating mobility data into epidemic models to accurately predict outbreak trajectories and inform intervention strategies. The Covasim (COVID-19 Agent-based Simulator) framework [27] is an example of agent-based simulator that is designed to integrate a heterogeneous set of variables, including demographic information as well as clinical data, with interventions such as physical distancing, vaccinations, and protective equipment. A similar methodology had previously been applied to different diseases, such as the FluTE simulation model for influenza [28]. Overall, network-based and agent-based models are highly realistic and excellent for simulating epidemic scenarios, while being computationally intensive and requiring granular observational data [11,29].

Bayesian analysis is used in [30] for online prediction of epidemics in structured populations, assuming that parts of the epidemic process are unobserved. In [31], Bayesian inference of infectious disease transmission is performed by exploiting genomic data, using both synthetic and real-world data sets for experiments. The approach is extended in [32], supporting simultaneous estimation of transmission trees and phylogenies. Contact, genomic, and symptoms data are exploited in [33] for Bayesian inference of transmission trees, showing that contact data significantly improve the reconstruction of transmission trees, even under realistic limitations on the coverage of contact tracing and the amount of non-infectious mixing between cases. Bayesian inference schemes have also been applied in combination with both genetic and epidemiological data related to Foot-and-Mouth Disease Virus, with the aim to reconstruct most likely transmission patterns and infection dates [34]. Despite the advantages of building on the powerful tools of statistical inference and of easily integrating heterogeneous sets of variables, Bayesian approaches suffer from a computational perspective as the network of individuals grows. They also need to make strong assumptions on the prior distribution of the model parameters.

To overcome the trade-offs of traditional modeling, hybrid approaches have emerged to bridge the gap between macro-level compartmental frameworks and micro-level agent-based simulations. For instance, in [35], a unified modeling and simulation framework is presented to study epidemiological systems from both microscopic and macroscopic perspectives, leveraging Extended Stochastic Symmetric Nets (ESSNs) for high-level modeling, Stochastic Simulation Algorithms (SSAs) to perform macro-scale simulation, and translation of ESSNs into Agent-Based Models (ABMs) to perform micro-scale simulation.

1.3. Contribution

In this paper, we present an efficient analytical method to estimate the probability that an individual in a cluster is infectious and not isolated over time given the observations taken within that time, concerning contacts (with individuals inside or outside the cluster), symptoms, and results of diagnostic tests. To this end, we develop an iterative approach that disregards dependencies among observations, estimating at each iteration n the probability that a subject is infectious and not isolated over time due to an infection sequence consisting of one effective external contact followed by up to n effective internal contacts. We generated multiple synthetic data sets of observations, varying the number of individuals, the topology of the network of social contacts, and the density of the internal contacts, while accounting for noisy, erroneous, and missing observations. For each data set, we performed thousands of experiments by varying the stochastic parameters of our approach from those used to derive a ground truth by performing stochastic simulation. Experimental results show that a few iterations are sufficient for our approach to accurately rank the subjects according to their probability of being infectious and not isolated over time. Notably, our approach achieves high accuracy with respect to the ground truth and outperforms an alternative baseline that accounts for dependencies among observations while maintaining a comparable runtime.

We implemented our approach in the CHITA Java library, available open source under the AGPLv3 licence at <https://doi.org/10.5281/zenodo.21559792>, supporting replication of the experimental results reported in this paper. In particular, CHITA leverages the SIRIO library [36] of the ORIS tool [37] to perform transient analysis of the stochastic model of infection evolution in an individual.

The rest of the paper is organized in four sections. First, we formulate the addressed problem (Section 2) and then present our iterative solution method (Section 3). Then, we illustrate our experimental setup and we present the achieved results (Sections 4 and 5). Finally, we draw our conclusions (Section 6). Implementation details on the derivation of a ground truth by stochastic simulation are reported in the Appendix.

2. Problem statement

In this section, we define the metamodel of an infection chain (Section 2.1) and we illustrate the Stochastic Time Petri Net (STPN) model of the infection evolution starting from an effective contact among subjects (Section 2.2).

2.1. Metamodel of an infection chain

Fig. 1 illustrates the metamodel of the infection chain of a disease (class *Disease*), which is characterized by a probability of developing symptoms and by the probability distribution of: the time from infection to infectiousness, also known as *latent period*; the time from infection to the onset of symptoms, also known as *incubation period*; the time from infectiousness to healing for asymptomatic subjects, also known as *infection duration*; the time from symptoms onset to isolation, also known as *onset-to-isolation duration*; and the time from a positive test to isolation, also known as *notification-to-isolation delay*. We consider a set of subjects (class *Subject*) who may contract the disease, each belonging to a cluster (class *Cluster*). Observations (class *Observation*) are taken on behaviors of subjects, each observation consisting of a timestamp and an observation type (class *ObservationType*). In turn, an observation type can be the result of a diagnostic test (class *Test*), or the annotation of symptoms (class *Symptom*), or the indication of a contact with other subjects (class *Contact*). A contact is characterized by a risk factor (attribute *risk*), intended as the probability of how close the contact is (for each contact, we consider the maximum risk among those annotated by each subject involved in the contact). A contact can be an internal contact among two or more subjects of the cluster (class *InternalContact*) or an external contact between a subject of the cluster and a subject outside the cluster (class *ExternalContact*).

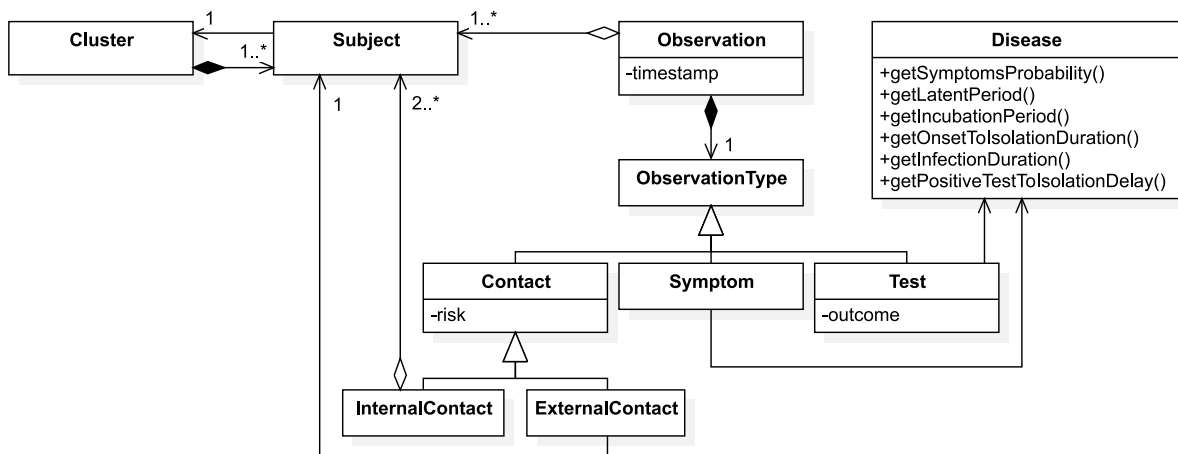


Fig. 1. Metamodel of the infection chain of a disease.

We are interested in computing the likelihood that each subject of a cluster is infectious and not isolated over time given the observations up to that time. These quantitative measures can be used to select the subjects to be tested so as to contain the epidemic. Therefore, the likelihood value of each subject is relevant comparatively to the likelihood values of the other subjects within the cluster, rather than absolutely.

2.2. STPN model of infection evolution from infection to isolation or recovery

Fig. 2(a) shows the STPN model of the evolution of an infection until recovery or isolation. STPNs [38] model concurrent timed systems with stochastic temporal parameters and discrete probabilistic choices. Specifically, immediate (IMM) transitions (depicted as thin bars, e.g., transition `effectiveContact`) model events occurring in zero time; general (GEN) transitions (depicted as thick bars, e.g., transition `infectiousness`) represent activities with stochastic duration, characterized by an absolutely continuous Cumulative Distribution Function (CDF), with corresponding Probability Density Function (PDF); tokens within places (depicted as dots within circles, respectively, e.g., one token within place `EffectiveContact`) represent the discrete logical state of the system; and, directed arcs (depicted as directed arrows) from input places to transitions (e.g., from place `EffectiveContact` to transition `effectiveContact`) and from transitions to output places (e.g., from transition `effectiveContact` to places `Infected` and `Symptomatology`) model preconditions and postconditions, respectively. A marking assigns tokens to places (e.g., the initial marking assigns one token to place `EffectiveContact`). A transition is enabled by a marking if each of its input places contains at least one token (e.g., transition `effectiveContact` is enabled in the initial marking).

In Fig. 2(a), the IMM transition `effectiveContact` models the infection and fires in zero time, removing the token from `EffectiveContact` and adding one token to `Symptomatology` and one token to `Infected`.

- A token in `Symptomatology` enables the IMM transitions `symptoms` and `noSymptoms`, modeling symptomatic and asymptomatic progression, respectively. They fire instantaneously with probability s and $1 - s$, respectively, representing their respective weights. The firing of `symptoms` moves the token from `Symptomatology` to `DevelopingSymptoms`, enabling the GEN transition `symptomsOnset` modeling the duration from infection to symptoms onset. When the time-to-fire sampled from its CDF has elapsed, `symptomsOnset` fires, moving the token from `DevelopingSymptoms` to `Symptomatic`, and enabling the GEN transition `isolating` which models the time from symptoms onset to isolation. Once the sampled time-to-fire has elapsed, `isolating` fires, moving the token from `Symptomatic` to `Isolated`. Conversely, the firing of `noSymptoms` moves the token from `Symptomatology` to `Asymptomatic`.
- A token in `Infected` enables the GEN transition `infectiousness` which models the duration from infection to infectiousness. Once the sampled time-to-fire has elapsed, `infectiousness` fires, moving the token from `Infected` to `Infectious`, and enabling the GEN transition `healing` which models the duration from infectiousness to recovery for asymptomatic subjects. Once the sampled time-to-fire has elapsed, `healing` fires, removing the tokens from `Infectious` and `Asymptomatic` and adding one token to `Healed`.

Note that the sequence of transitions `infectiousness` and `healing` is concurrent with the sequence of transitions `symptomsOnset` and `isolating`. According to this, at the initial time, `infectiousness` and `symptomsOnset` can be concurrently enabled, and the one that samples the shorter duration fires first, thus accounting for possible presymptomatic transmission. Moreover, an asymptomatic infected subject may infect others until healed; conversely, a symptomatic infected subject cannot infect others once isolated (to simplify presentation, the model does not represent the duration from infectiousness to healing for symptomatic subjects, which is typically significantly much larger than the duration from symptomaticity to isolation).

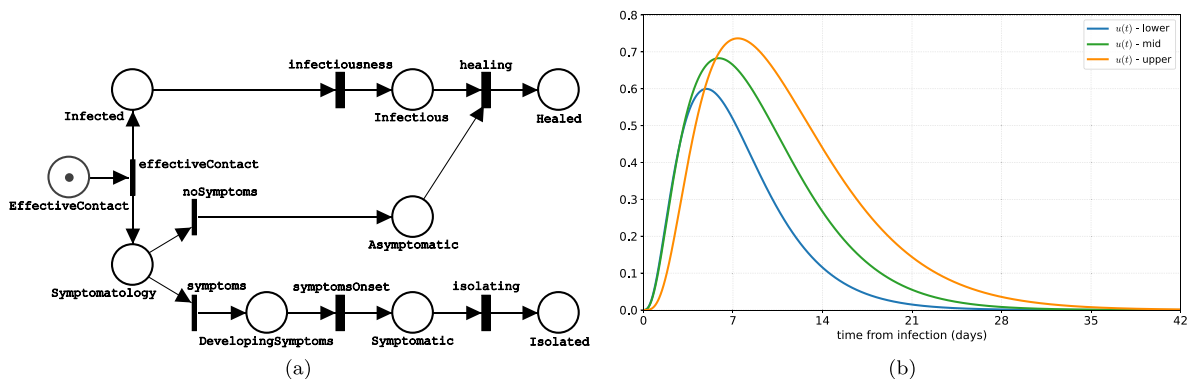


Fig. 2. (a) STPN model of disease evolution. (b) Probability $u(t)$ that a subject is infectious and not isolated at time t given an effective contact at time 0, assuming the lower-, mid-, and upper-case values in Tables 1 and 2 for the STPN parameters.

Given that the STPN model does not account for observations (e.g., the positive result of a test leading an asymptomatic infected subject to isolation), the probability $u(t)$ that a subject is infectious and not isolated at time t given an effective contact at time 0 can be derived by performing forward transient analysis [39] and computing the transient probability that place `Infectious` contains one token and place `Isolated` contains no tokens, e.g., Fig. 2(b) shows $u(t)$ assuming that each stochastic parameter takes its lower-, mid-, and upper-case value in Tables 1 and 2. Note that we consider a time horizon during which each subject can be infected only once; thus, the duration of isolation is not modeled in the STPN, as a subject has no further contacts during isolation and is no longer infectious by the end of the isolation period.

3. Problem formulation and evaluation

In this section, we derive the likelihood that a subject is infectious not isolated over time given a sequence of observations (Section 3.1) and we characterize the complexity of the derivation (Section 3.2).

3.1. Likelihood of infectiousness given the observations

We consider a sequence of observations $\omega(t) := \langle \omega_1, \dots, \omega_m \rangle$ taken within the time interval $[0, t]$, where each observation $\omega_i = \langle \epsilon_i, t_i \rangle$ with $i \in \{1, \dots, m\}$ consists of an event ϵ_i and a timestamp t_i such that $t_1 \leq t_2 \leq \dots \leq t_m \leq t$. For a subject s , we define the *likelihood* $L(s; t; \omega(t))$ as the *approximate probability* $\tilde{P}(s; t; \omega(t))$ that subject s is infectious not isolated at time t given the observation sequence $\omega(t)$:

$$L(s; t; \omega(t)) := \tilde{P}(s; t; \omega(t)) := \tilde{P}\{s \text{ is infectious not isolated at } t \mid \omega(t)\}. \quad (1)$$

Specifically, we introduce the following five *sources of approximation* in the evaluation of $\tilde{P}(s, t, \omega(t))$:

- (A1) We disregard dependencies among observations (i.e., we consider observations independent of each other), thus estimating the probability of an observation sequence $\omega(t) := \langle \omega_1, \dots, \omega_m \rangle$ as the product of the probabilities of the individual observations $\omega_1, \dots, \omega_m$, i.e., $\tilde{P}\{\omega(t)\} = \prod_{i=1}^m P\{\omega_i\}$.
- (A2) In the evaluation of the probability that a contact of subject s is effective given an observation sequence, we disregard the conditioning on observations taken on subjects other than s and on observations of contacts of subject s (i.e., we do not consider the conditioning on these observations).
- (A3) If subject s has had an effective contact γ at time $\tau \leq t$ and the observation sequence $\omega(t)$ does not include the annotation of a positive test of s , we disregard the number ι of internal contacts of the infection sequence that infected s and the conditioning on $\omega(t)$, estimating the probability that s is infectious not isolated at time t and $\iota \leq n$ given the effective contact γ at time τ and given $\omega(t)$ as the probability that s is infectious not isolated at time t given the effective contact γ at time τ , i.e., $\tilde{P}\{s \text{ is infectious not isolated at } t \wedge \iota \leq n \mid \gamma \text{ at } \tau \wedge \omega(t)\} = P\{s \text{ is infectious not isolated at } t \mid \gamma \text{ at } \tau\}$. Conversely, if subject s has had an effective contact γ at time $\tau \leq t$ and the observation sequence $\omega(t)$ includes the annotation of a positive test of s at time $t' \leq t$, we disregard ι , the infectiousness state of s and the conditioning on γ and the remaining observations, estimating $P\{s \text{ is infectious not isolated at } t \wedge \iota \leq n \mid \gamma \text{ at } \tau \wedge \omega(t)\}$ as the probability $1 - \Psi(t - t')$ that s is not isolated at t given a positive test at t' , where $\Psi(t)$ is the distribution of the duration from a positive test to isolation.
- (A4) We disregard all but the most probable infectious subject in multi-subject contacts, thus approximating the probability that a contact involving subject s is effective given an observation sequence $\omega(t)$ as the probability that the most likely infectious subject (among those whom s met) is infectious and not isolated at the time of contact given $\omega(t)$, multiplied by the risk factor of the contact.
- (A5) We disregard the conditioning on observations taken after time t in the evaluation of the probability that a subject is infectious not isolated at time t given a sequence of observations $\omega(t')$ taken within $[0, t']$ with $t' > t$, i.e., $\tilde{P}\{s \text{ is infectious not isolated at } t \mid \omega(t')\} = P\{s \text{ is infectious not isolated at } t \mid \omega(t)\}$.

We refer to any probability derived by introducing these sources of approximation as *approximate probability*.

The infection sequence that may have infected the subject consists of an external effective contact followed by a number of effective internal contacts, with the last contact of the sequence being the primary infection event for s (i.e., the only event of the sequence that involves s). We compute the likelihood $L(s; t; \omega(t))$ as:

$$L(s; t; \omega(t)) = \lim_{n \rightarrow \infty} \tilde{P}(s; t; \omega(t); n) \quad (2)$$

where $\tilde{P}(s; t; \omega(t); n)$ is the approximate probability that subject s is infectious not isolated at t due to an infection sequence made of one effective external contact followed by up to n effective internal contacts, given the observation sequence $\omega(t)$, i.e., $\tilde{P}(s; t; \omega(t); n) := \tilde{P}\{s \text{ is infectious not isolated at } t \wedge \iota \leq n \mid \omega(t)\}$ where ι is the number of effective internal contacts. We iteratively derive $\tilde{P}(s; t; \omega(t); n)$ through Theorem 3.1.

Theorem 3.1. For any subject s , time t , and observation sequence $\omega(t) := \langle \omega_1, \dots, \omega_m \rangle$, the approximate probability $\tilde{P}(s; t; \omega(t); n)$ can be derived as follows $\forall n \in \mathbb{N}$:

$$\tilde{P}(s; t; \omega(t); n) = \begin{cases} \sum_{k=1}^{E_{s,t}} c(t; \tau_{s,k}^{\text{ext}}; \omega(t)) \cdot r(e_{s,k}) \cdot z(s; t; e_{s,k}) \prod_{j=1}^{k-1} (1 - r(e_{s,j}) \cdot z(s; t; e_{s,j})) & \text{if } n = 0 \\ \sum_{l=1}^{I_{s,t}} c(t; \tau_{s,l}^{\text{int}}; \omega(t)) \cdot r(i_{s,l}) \cdot \tilde{P}(q(s; l; n-1); \tau_{s,l}^{\text{int}}; \omega(\tau_{s,l}^{\text{int}}); n-1) \\ \cdot \prod_{j=1}^{l-1} (1 - r(i_{s,j}) \cdot \tilde{P}(q(s; j; n-1); \tau_{s,j}^{\text{int}}; \omega(\tau_{s,j}^{\text{int}}); n-1)) & \text{if } n > 0 \end{cases} \quad (3)$$

where:

- $E_{s,t}$ is the number of external contacts of subject s that occurred within the interval $[0, t]$;
- $I_{s,t}$ is the number of internal contacts of subject s that occurred within the interval $[0, t]$;
- $\tau_{s,k}^{\text{ext}}$ is the time of the k th external contact $e_{s,k}$ of subject s ;
- $\tau_{s,l}^{\text{int}}$ is the time of the l th internal contact $i_{s,l}$ of subject s ;
- $c(t; \tau; \omega(t))$ is either (i) the probability $v(t, \tau)$ that subject s is infectious not isolated at time t given an effective contact γ at time $\tau \leq t$, if the observation sequence $\omega(t)$ does not include a positive test of subject s , or (ii) the probability that subject s has not yet isolated since he/she was tested positive at time $t' \leq t$ with Ψ being the probability distribution of the notification-to-isolation delay, otherwise:

$$c(t; \tau; \omega(t)) := \begin{cases} v(t, \tau) & \text{if } \omega(t) \text{ includes no positive test of } s \\ 1 - \Psi(t - t') & \text{if } \omega(t) \text{ includes a positive test of } s \text{ at time } t' \end{cases} \quad (4)$$

- $r(\gamma)$ is the risk factor of contact γ ;
- $z(s; t; \gamma)$ is the approximate probability that contact γ of subject s is effective given the observations taken within the interval $[0, t]$ (without assuming that an infected subject is part of the contact) divided by the probability that γ is effective, i.e., $z(s; t; \gamma) := \tilde{P}\{\gamma \text{ is effective} \mid \omega(t)\} / P\{\gamma \text{ is effective}\}$:

$$z(s; t; \gamma) = \frac{\prod_{h \in H_{s,t}} P\{\omega_h \mid \gamma \text{ is effective}\}}{\prod_{h \in H_{s,t}} P\{\omega_h\}} \quad (5)$$

where $H_{s,t}$ is the set of indexes of observations of symptoms and test results which were taken on subject s within the interval $[0, t]$;

- $q(s; l; n-1)$ is the subject, met by s during his/her l th internal contact, who has the highest approximate probability of being infectious and not isolated at the contact time $\tau_{s,l}^{\text{int}}$, assuming $l \leq n-1$ effective internal contacts in the infection sequence, i.e., $q(s; l; n-1) := \operatorname{argmax}_{x \in X_{s,l}} \tilde{P}(x; \tau_{s,l}^{\text{int}}; \omega(\tau_{s,l}^{\text{int}}); n-1)$ where $X_{s,l}$ is the set of subjects encountered by s at his/her l th internal contact.

Proof. Base step. We derive $\tilde{P}(s; t; \omega(t); 0) := \tilde{P}\{s \text{ is infectious not isolated at } t \wedge t = 0 \mid \omega(t)\}$. To this end, we consider $P\{s \text{ is infectious not isolated at } t \wedge t = 0 \mid \omega(t)\}$, i.e., the probability that s is infectious not isolated at time t due to an infection sequence made of one effective external contact and $t = 0$ effective internal contacts, given the observation sequence $\omega(t)$. Any external contact of s prior to the primary infection event is ineffective, while the subsequent external contacts of s may or may not be effective. Let $e_{s,1}, \dots, e_{s,E_{s,t}}$ be the temporally ordered external contacts of s that occurred within the interval $[0, t]$, where $E_{s,t}$ is their number. By applying the law of total probability, we express $P\{s \text{ is infectious not isolated at } t \wedge t = 0 \mid \omega(t)\}$ as $\sum_{k=1}^{E_{s,t}} P\{s \text{ is infectious not isolated at } t \wedge t = 0 \wedge e_{s,k} \text{ is effective} \wedge e_{s,1}, \dots, e_{s,k-1} \text{ are ineffective} \mid \omega(t)\}$. By disregarding dependencies among the external contacts of s (source of approximation A1), we obtain:

$$\begin{aligned} \tilde{P}(s; t; \omega(t); 0) &= \sum_{k=1}^{E_{s,t}} P\{s \text{ is infectious not isolated at } t \wedge t = 0 \wedge e_{s,k} \text{ is effective} \mid \omega(t)\} \\ &\cdot \prod_{j=1}^{k-1} P\{e_{s,j} \text{ is not effective} \mid \omega(t)\} \end{aligned}$$

By the definition of conditional probability and the definition of complementary event, we obtain:

$$\begin{aligned} \tilde{P}(s; t; \omega(t); 0) &= \sum_{k=1}^{E_{s,t}} P\{s \text{ is infectious not isolated at } t \wedge t = 0 \mid e_{s,k} \text{ is effective} \wedge \omega(t)\} \\ &\cdot P\{e_{s,k} \text{ is effective} \mid \omega(t)\} \cdot \prod_{j=1}^{k-1} (1 - P\{e_{s,j} \text{ is effective} \mid \omega(t)\}) \end{aligned}$$

By introducing approximation in the first factor (source of approximation A3), disregarding t and either (i) the conditioning on $\omega(t)$ (if $\omega(t)$ does not include the annotation of a positive test of subject s) or (ii) the infectiousness state of s as well as the conditioning

on the effective contact $e_{s,k}$ and the remaining observations (if $\omega(t)$ includes the annotation of a positive test of subject s), we express $\tilde{P}(s; t; \omega(t); 0)$ as:

$$\tilde{P}(s; t; \omega(t); 0) = \sum_{k=1}^{E_{s,t}} c(t; \tau_{s,k}^{\text{ext}}; \omega(t)) \cdot P\{e_{s,k} \text{ is effective} \mid \omega(t)\} \cdot \prod_{j=1}^{k-1} (1 - P\{e_{s,j} \text{ is effective} \mid \omega(t)\}).$$

By the Bayes theorem, $P\{e_{s,k} \text{ is effective} \mid \omega(t)\} = P\{\omega(t) \mid e_{s,k} \text{ is effective}\} \cdot P\{e_{s,k} \text{ is effective}\} / P\{\omega(t)\}$. By disregarding dependencies among observations $\omega_1, \dots, \omega_m$ (source of approximation A1), we obtain $\tilde{P}\{e_{s,k} \text{ is effective} \mid \omega(t)\} = P\{e_{s,k} \text{ is effective}\} \cdot \prod_{h=1}^m P\{\omega_h \mid e_{s,k} \text{ is effective}\} / \prod_{h=1}^m P\{\omega_h\}$. By estimating the probability that the external contact $e_{s,k}$ is effective as the risk factor $r(e_{s,k})$ of the contact, and by disregarding the conditioning on observations taken on subjects other than subject s as well as the conditioning on observations of contacts of subject s (source of approximation A2), we obtain the following expression: $\tilde{P}\{e_{s,k} \text{ is effective} \mid \omega(t)\} = r(e_{s,k}) \cdot \prod_{h \in H_{s,t}} P\{\omega_h \mid e_{s,k} \text{ is effective}\} / \prod_{h \in H_{s,t}} P\{\omega_h\}$, where $H_{s,t}$ is the set of indexes of observations of symptoms and test results taken on subject s within the time interval $[0, t]$. According to this, we can express $\tilde{P}(s; t, \omega(t); 0)$ as in Eq. (3), where $z(s; t; e_{s,k})$ is expressed as in Eq. (5).

Induction step. We derive $\tilde{P}(s; t; \omega(t); n) := \tilde{P}\{s \text{ is infectious not isolated at } t \wedge i \leq n \mid \omega(t)\}$. To this end, we consider $P\{s \text{ is infectious not isolated at } t \wedge i \leq n \mid \omega(t)\}$, i.e., the probability that s is infectious not isolated at time t due to an infection sequence made of one effective external contact and $i \leq n$ effective internal contacts, given the observation sequence $\omega(t)$. Any internal contact of s prior to the primary infection event is ineffective, while the subsequent internal contacts of s may or may not be effective. Let $i_{s,1}, \dots, i_{s,I_{s,t}}$ be the temporally ordered internal contacts of s that occurred within the interval $[0, t]$, where $I_{s,t}$ is their number. By applying the law of total probability, we obtain $P\{s \text{ is infectious not isolated at } t \wedge i \leq n \mid \omega(t)\} = \sum_{l=1}^{I_{s,t}} P\{s \text{ is infectious not isolated at } t \wedge i \leq n-1 \wedge i_{s,l} \text{ is effective} \wedge i_{s,1}, \dots, i_{s,l-1} \text{ are not effective} \mid \omega(t)\}$. By the definition of conditional probability, we obtain $P\{s \text{ is infectious not isolated at } t \wedge i \leq n \mid \omega(t)\} = \sum_{l=1}^{I_{s,t}} P\{s \text{ is infectious not isolated at } t \wedge i \leq n-1 \mid i_{s,l} \text{ is effective} \wedge i_{s,1}, \dots, i_{s,l-1} \text{ are not effective} \wedge \omega(t)\} \cdot P\{i_{s,l} \text{ is effective} \wedge i_{s,1}, \dots, i_{s,l-1} \text{ are not effective} \mid \omega(t)\}$. By disregarding dependencies among the internal contacts $i_{s,1}, \dots, i_{s,l}$ of subject s (source of approximation A1), we obtain:

$$\begin{aligned} & \tilde{P}(s; t; \omega(t); n) \\ &= \sum_{l=1}^{I_{s,t}} P\{s \text{ is infectious not isolated at } t \wedge i \leq n-1 \mid i_{s,l} \text{ is effective} \wedge i_{s,1}, \dots, i_{s,l-1} \text{ are ineffective} \wedge \omega(t)\} \\ & \quad \cdot P\{i_{s,l} \text{ is effective} \mid \omega(t)\} \cdot \prod_{j=1}^{l-1} P\{i_{s,j} \text{ is not effective} \mid \omega(t)\}. \end{aligned}$$

By introducing approximation in the first factor (source of approximation A3), disregarding i and either (i) the conditioning on $\omega(t)$ (if $\omega(t)$ does not include the annotation of a positive test of subject s) or (ii) the infectiousness state of s as well as the conditioning on the effective contact $i_{s,l}$ and the remaining observations (if $\omega(t)$ includes the annotation of a positive test of subject s), we express $\tilde{P}(s; t; \omega(t); n)$ as:

$$\tilde{P}(s; t; \omega(t); n) = \sum_{l=1}^{I_{s,t}} c(t; \tau_{s,l}^{\text{int}}; \omega(t)) \cdot P\{i_{s,l} \text{ is effective} \mid \omega(t)\} \cdot \prod_{j=1}^{l-1} P\{i_{s,j} \text{ is not effective} \mid \omega(t)\}.$$

By considering, among the subjects met by subject s during his/her l th internal contact, only the subject $q(s; l; n-1)$ who has the highest approximate probability of being infectious at the contact time $\tau_{s,l}^{\text{int}}$ (source of approximation A4), we obtain $\tilde{P}\{i_{s,l} \text{ is effective} \mid \omega(t)\} = P\{q(s; l; n-1) \text{ is infectious not isolated at } \tau_{s,l}^{\text{int}} \wedge i \leq n-1 \mid \omega(t)\} \cdot r(i_{s,l})$. To evaluate $P\{q(s; l; n-1) \text{ is infectious not isolated at } \tau_{s,l}^{\text{int}} \wedge i \leq n-1 \mid \omega(t)\}$, we recursively introduce the mentioned sources of approximation and we omit the conditioning on observations taken after time $\tau_{s,l}^{\text{int}}$ (source of approximation A5), obtaining $\tilde{P}(q(s, l, n-1); \tau_{s,l}^{\text{int}}; \omega(\tau_{s,l}^{\text{int}}), n-1) \cdot r(i_{s,l})$. Moreover, by the definition of complementary event, $P\{i_{s,j} \text{ is not effective} \mid \omega(t)\} = 1 - P\{i_{s,j} \text{ is effective} \mid \omega(t)\}$. According to this, we can express $\tilde{P}(s, t, \omega(t), n)$ with $n > 0$ as in Eq. (3). $\square$

3.2. Computational complexity

To characterize the computational complexity, we distinguish the base step and the inductive step.

- **Base step.** According to Eq. (3), for each subject s and time instant t , the derivation of $\tilde{P}(s; t; \omega(t); 0)$ requires a number of operations that is linear in the number $E_{s,t}$ of external contacts of subject s that occurred within time t (as the factors of the product $\prod_{j=1}^{k-1}$ can be calculated incrementally for increasing values of k). Moreover, according to Eq. (5), for each such external contact, the evaluation requires a number of operations proportional to the number $|H_{s,t}|$ of observations of symptoms and test results which were taken on subject s within time t . Furthermore, $c(t; \tau; \omega(t))$ can be precomputed once per disease (by performing transient analysis of the STPN model of Fig. 2(a)) and thus its derivation does not affect the runtime of our iterative approach. Therefore, the computational complexity of the base step is $O(STHE)$, where S is the number of subjects, T is the number of time instants, H is the maximum number of observations of symptoms and test results which were taken on a subject, and E is the maximum number of external contacts of a subject.

- **Inductive step.** According to Eq. (3), for each subject s , time instant t , and iteration n , the derivation of $\tilde{P}(s; t; \omega(t); n)$ requires a number of operations that is linear in the number $I_{s,t}$ of internal contacts of s that occurred within time t (as the factors of the product $\prod_{j=1}^{l-1}$ can be calculated incrementally for increasing values of l). Moreover, for each such internal contact, Eq. (3) requires the identification of the subject $q(s; l; n - 1)$, among those met by s during that contact, who has the highest approximate probability of being infectious and not isolated at the contact time. Furthermore, $c(t; \tau; \omega(t))$ can be precomputed once per disease, as already discussed for the base step. Therefore, the computational complexity of the inductive step is $O(STNMI)$, where N is the number of iterations, I is the maximum number of internal contacts of a subject, and M is the maximum number of subjects met by an individual during an internal contact.

According to this, the computational complexity of our approach is $O(ST(HE + NMI))$.

4. Experimental methodology

In this section, first we define the research objectives of our experimental evaluation (Section 4.1). Then, we describe the stochastic parameters used by our STPN model of disease evolution and by our approach (Section 4.2), and we define the quantitative measures of interest (Section 4.3). Finally, we illustrate how we generate twelve data sets of observations (Section 4.4) and how we exploit three of them to select our hyperparameter, i.e., the number N of iterations performed by our solution method (Section 4.5).

4.1. Aim and scope

We assess the feasibility and effectiveness of our method by a thorough experimental evaluation, considering randomly generated scenarios with an increasing number of subjects and observations. Moreover, we compare our approach with an alternative baseline for ranking the subjects of a cluster according to their probability of being infectious, notably considering scenarios where the stochastic model of infection evolution exploited by our approach (and by the alternative baseline) is disturbed (with respect to the actual stochastic model used to derive a ground truth). Specifically, we aim at answering the following Research Questions (RQs):

- **RQ1:** Is the approach feasible for a large number of subjects and observations?
- **RQ2:** As the number of subjects and observations as well as the structure of the network of social contacts vary, does the approach achieve good or high performance in ranking the subjects according to their probability of being infectious and not isolated? Does it outperform an alternative baseline?
- **RQ3:** If the stochastic parameters are disturbed with respect to the actual ones, is the approach effective at ranking the subjects according to their probability of being infectious and not isolated?

4.2. Derivation of stochastic parameters

4.2.1. Model parameters

We model each stochastic temporal parameter of the STPN of Fig. 2(a) with three distinct PDFs fitting the lower, mid, and upper cases of the variability ranges for the mean value μ and the coefficient of variation cv found in the literature on COVID-19. Specifically, observing $cv < 1$, we consider the PDF of the sum of an Erlang random variable with shape n and rate λ_1 , and an Exponential random variable with rate λ_2 , where $n = \lceil 1/cv^2 - 1 \rceil$, $\lambda_1 = (n\mu + \sqrt{n^2\mu^2 - (\mu^2 - cv^2\mu^2)(n^2 + n)})/(\mu^2 - cv^2\mu^2)$, and $\lambda_2 = 1/(\mu - n/\lambda_1)$, as presented in [40]; note that, if $1/\sqrt{2} < cv < 1$, then the considered PDF takes the form of a hypo-Exponential PDF $f(x) = ((\lambda_1\lambda_2)/(\lambda_1 - \lambda_2))(e^{-\lambda_1x} - e^{-\lambda_2x})$ with $\lambda_1 = (2/\mu)/(1 + \sqrt{2cv^2 - 1})$ and $\lambda_2 = (2/\mu)/(1 - \sqrt{2cv^2 - 1})$, as illustrated in [41]. For the GEN durations in the STPN of Fig. 2(a), we consider the values of the mean μ and standard deviation σ listed in Table 1. This table also reports the coefficients of variation cv and the parameters of the estimated PDFs, which are depicted in Fig. 3.

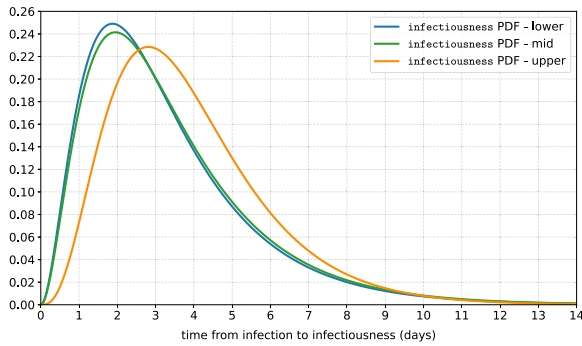
- For transition infectiousness (modeling the time from an effective contact to infectiousness), we consider the range of estimated values reported in Section 3 and Figure 3 of Zhao et al. [42] for μ and σ , i.e., $\langle \mu, \sigma \rangle \in \{ \langle 3.3, 2.20 \rangle, \langle 3.4, 2.25 \rangle, \langle 3.9, 2.05 \rangle \}$ where μ and σ are expressed in days.
- For transition symptomsOnset (modeling the time from an effective contact to the onset of symptoms), we consider the mean value and the 95% Confidence Interval (CI) reported in Section 3 of Zhao et al. [42] for μ , i.e., 6.8 days with 95% CI [6.2, 7.5] days, and σ , i.e., 4.1 days with 95% CI [3.7, 4.8] days. Thus, we consider $\langle \mu, \sigma \rangle \in \{ \langle 6.2, 3.7 \rangle, \langle 6.8, 4.1 \rangle, \langle 7.5, 4.8 \rangle \}$. Note that the mean values estimated for the duration of the latent period (i.e., 3.3, 3.4, and 3.9 days) are nearly half of the mean values estimated for the incubation period (i.e., 6.2, 6.8, and 7.5 days), which is consistent with the observation of presymptomatic transmission for COVID-19 [43–45].
- For transition isolating (modeling the time needed by a symptomatic subject to isolate himself/herself), we consider the median value and the Inter-Quartile Range (IQR) reported in Table 1 of Hellewell et al. [46] for a short delay estimated during the late stages of the 2003 SARS outbreak in Singapore, i.e., median 3.43 days and IQR [2.02, 5.23] days, and a long delay estimated from the early phase of the COVID-19 outbreak in Wuhan, i.e., median 8.09 days and IQR [5.52, 19.93]. Then, we fit each statistics with a Gamma PDF minimizing the sum of the squared differences between the approximate and observed quantiles. Finally, we consider the values μ and σ of the estimated Gamma PDFs, i.e., $\langle \mu, \sigma \rangle \in \{ \langle 3.9, 2.47 \rangle, \langle 8.6, 4.17 \rangle \}$, plus an intermediate case obtained by taking the average of the two mean values and the average of the two standard deviations, i.e., $\langle \mu, \sigma \rangle = \langle 6.25, 3.32 \rangle$.

Table 1
Parameters of the PDFs of the GEN transitions of the STPN in Fig. 2(a).

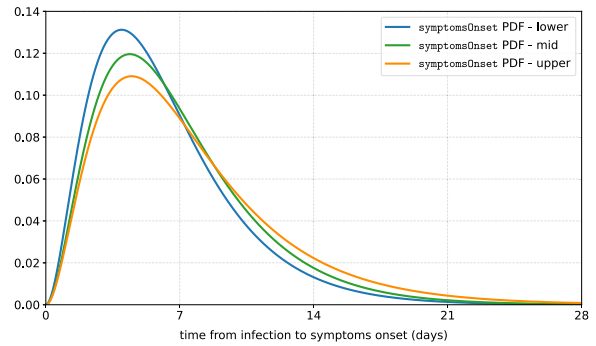
Transition	Case	μ	σ	cv	n	λ_1	λ_2
infectiousness	Lower	3.3	2.20	0.66667	2	1.53627	0.50046
	Mid	3.4	2.25	0.66176	2	1.46110	0.49233
	Upper	3.9	2.05	0.52564	3	1.26195	0.65672
symptomsOnset	Lower	6.2	3.7	0.60294	2	0.61472	0.33939
	Mid	6.8	4.1	0.59677	2	0.54129	0.32205
	Upper	7.5	4.8	0.64000	2	0.60444	0.23860
isolating	Lower	3.9	2.47	0.63333	2	1.12932	0.46970
	Mid	6.25	3.32	0.53120	3	0.80718	0.39473
	Upper	8.6	4.17	0.48488	4	0.73548	0.31632
healing	Lower	9.4	7.01	0.74530	1	0.31902	0.15961
	Mid	11.2	8.35	0.74530	1	0.26775	0.13396
	Upper	13.3	9.91	0.74530	1	0.22547	0.11280

Table 2
Weights of the IMM transitions of the STPN in Fig. 2(a).

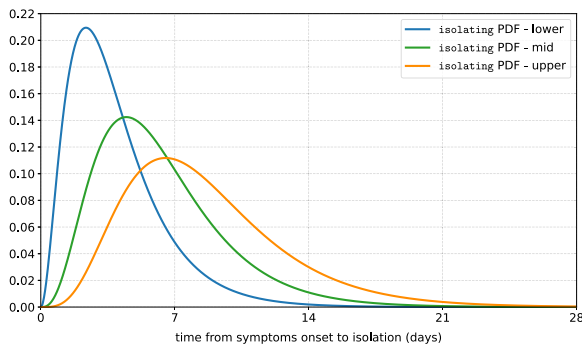
Case	Transition	s	Transition	$1 - s$
Lower	symptoms	0.601	noSymptoms	0.399
Mid	symptoms	0.649	noSymptoms	0.351
Upper	symptoms	0.693	noSymptoms	0.307



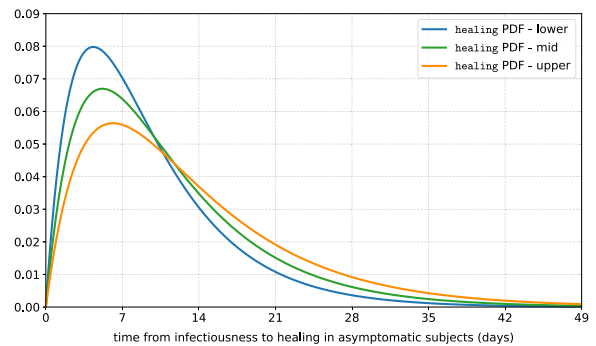
(a) Transition infectiousness.



(b) Transition symptomsOnset.



(c) Transition isolating.



(d) Transition healing.

Fig. 3. PDFs of the GEN transitions of the STPN in Fig. 2(a), whose parameters are reported in Table 1.

- For transition healing (modeling the time from infectiousness to healing for asymptomatic subjects), we consider the median value and the IQR reported in Table 1 of Byrne et al. [47], i.e., median 9.5 days and IQR [3.5, 13] days; we fit them with a Gamma PDF minimizing the sum of the squared differences between the approximate and observed quantiles; and, we obtain mean value 9.937 days and standard deviation 7.406, and thus coefficient of variation 0.745. Moreover, we also take into account the mean value and the 95% CI reported in Table 1 of Kissler et al. [48], i.e., mean 11.2

days and 95% CI [9.4, 13.3] days. Assuming the coefficient of variation estimated from [47], we finally consider $\langle \mu, \sigma \rangle \in \{(9.4, 7.01), (11.2, 8.35), (13.3, 9.91)\}$.

Moreover, for the weights s and $1 - s$ of the IMM transitions symptoms and noSymptoms, modeling the probability of developing symptoms and remaining asymptomatic, respectively, we consider the mean value and the 95% CI of the percentage of infected subjects who never developed clinical symptoms, estimated in Figure 2 of Sah et al. [49] through a meta-analysis involving over 350 studies, i.e., mean value 35.1% and 95% CI [30.7%, 39.9%]. We report the corresponding values of s and $1 - s$ in Table 2.

4.2.2. Analysis parameters

To compute the likelihood that a subject is infectious and not isolated over time given the observations up to that time, we derive the following quantities, appearing in Eqs. (3) to (4). As for the STPN parameters, we consider three distinct values for each quantity of interest, fitting the lower, mid, and upper cases of the variability ranges found in the literature on COVID-19. Specifically:

- We derive the probability $v(t, \tau)$ that a subject is infectious and not isolated at time t given an effective contact at time τ from the probability $u(t)$ that a subject is infectious and not isolated at time t given an effective contact at time 0, computed by analyzing the STPN model of Fig. 2(a) in Section 2.2:

$$v(t, \tau) = \begin{cases} 0 & \text{if } t < \tau \\ u(t - \tau) & \text{if } t \geq \tau \end{cases} \quad (6)$$

- We derive the CDF $\Psi(t)$ of the time to isolation since a positive test from the mean value and the IQR reported in Table 9 of Vogt [50] for 477 subjects in Western Australia from February to June 2020, i.e., $\mu = 0.38$ days and IQR $[0, 1]$ days, in line with the ranges used in [51]. Then, we estimate the standard deviation as $\sigma = \text{IQR}/1.35$, as suggested in [52] for a large sample size, obtaining $\sigma = 0.74074$, and we consider two additional cases obtained by reducing and increasing the mean value by 50%, respectively, i.e., $\langle \mu, \sigma \rangle \in \{(0.19, 0.74074), (0.38, 0.74074), (0.54, 0.74074)\}$. Given that, in each of the three cases, the coefficient of variation cv is larger than 1, we fit μ and cv by a hyper-Exponential PDF $\psi(t) = \sum_{i=1}^2 p_i \lambda_i e^{-\lambda_i t}$ with CDF $\Psi(t) = \sum_{i=1}^2 p_i (1 - e^{-\lambda_i t})$ where $p_i = 0.5 (1 \pm \sqrt{(cv^2 - 1)/(cv^2 + 1)})$ and $\lambda_i = 2p_i/\mu$ for $i \in \{1, 2\}$, as illustrated in [41]. Table 3 shows the values of p_1 , p_2 , λ_1 , and λ_2 for the three cases and Fig. 4(a) plots the resulting PDFs.
- We derive the probability $P\{\omega \mid \gamma \text{ is effective}\}$ of an observation $\omega = \langle \epsilon, t \rangle$ given that contact γ of subject s occurred at time τ is effective, for observations of symptoms and test results. Specifically:

- Let ϵ be the annotation of symptoms and $\phi(t; \tau) := P\{\text{symptoms at time } t \mid \gamma \text{ is effective}\}$. First, we consider the case that the effective contact γ occurred at time $\tau = 0$, and we derive $\phi(t; 0)$ as:

$$\phi(t; 0) = s \cdot (f_{\text{onset}} * G_{\text{symptoms}})(t) \quad (7)$$

where s is the probability of developing symptoms, $f_{\text{onset}}(t)$ is the PDF of the time from infection to symptoms onset (i.e., the PDF of transition symptomsOnset in the STPN of Fig. 2(a)), and $G_{\text{symptoms}}(t) := 1 - F_{\text{symptoms}}(t)$ with $F_{\text{symptoms}}(t)$ being the CDF of the duration of symptoms. For $f_{\text{onset}}(t)$ and s , we consider the values in Tables 1 and 2, respectively. For $F_{\text{symptoms}}(t)$, we estimate symptoms duration from the mean value μ and the standard deviation σ reported in Table 1 of Menni et al. [53] for a population of 4990 subjects in the UK during the periods of dominance of two COVID-19 variants, respectively, i.e., $\langle \mu, \sigma \rangle = \{(6.87, 5.21), (8.89, 5.04)\}$. We also consider an intermediate case obtained by taking the average of the two mean values and the average of the two standard deviations, i.e., $\langle \mu, \sigma \rangle = (7.88, 5.125)$. Given that, in each of the three cases, the coefficient of variation cv is lower than 1, we fit μ and σ by the sum of an Erlang random variable with shape n and rate λ_1 , and an Exponential random variable with rate λ_2 , as done in Section 4.2.1 for the STPN parameters. Table 3 shows the values of n , λ_1 , and λ_2 for the three cases, and Fig. 4(b) plots the resulting PDFs. Moreover, Fig. 4(c) plots $\phi(t; 0)$ for the three cases, obtained considering the lower-, mid-, and upper-case values for s , $f_{\text{onset}}(t)$ and $G_{\text{symptoms}}(t)$, respectively. Finally, we derive $\phi(t; \tau)$ from $\phi(t; 0)$ as follows:

$$\phi(t; \tau) = \begin{cases} 0 & \text{if } t < \tau \\ \phi(t - \tau; 0) & \text{if } t \geq \tau \end{cases} \quad (8)$$

- Let ϵ be the positive result of a diagnostic test and $\theta(t; \tau) := P\{\text{positive test at time } t \mid \gamma \text{ is effective}\}$. First, we consider the case that the effective contact γ occurred at time $\tau = 0$, deriving $\theta(t; 0)$ as:

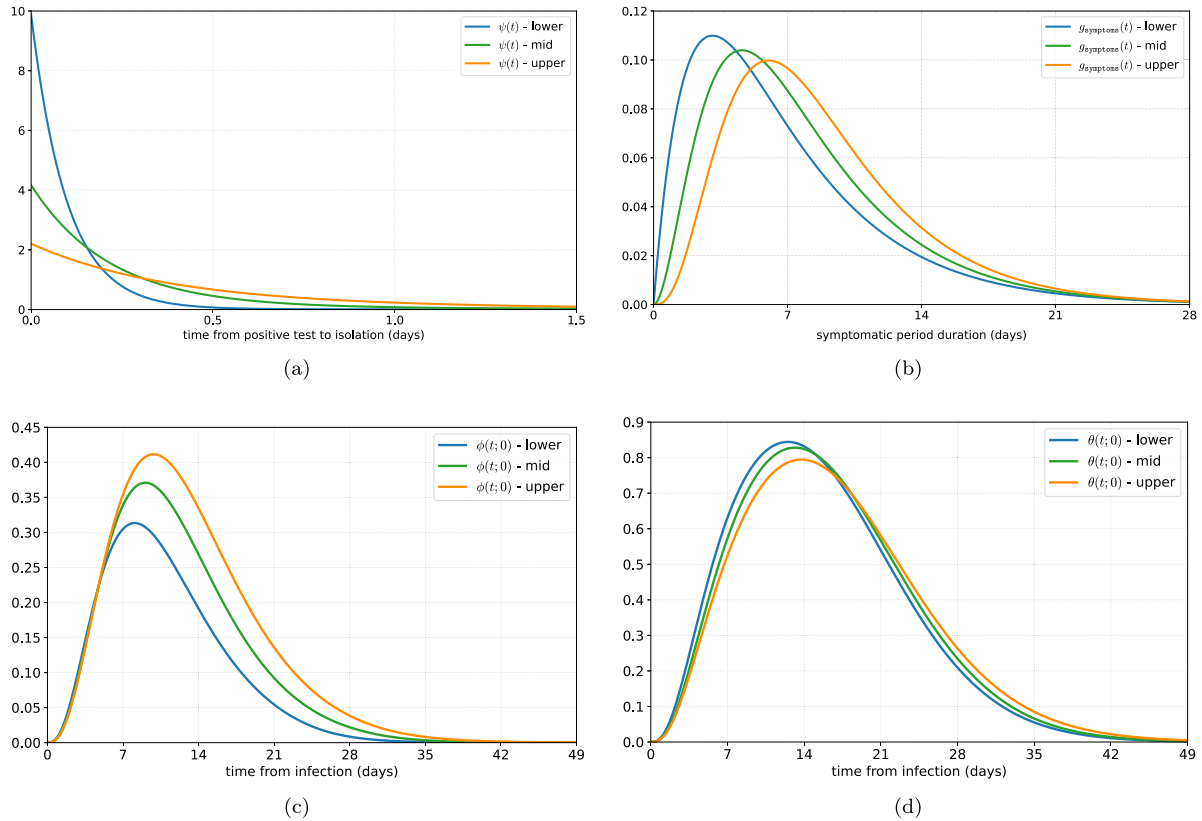
$$\theta(t; 0) = \int_0^t f_{\text{onset}}(\tau) \cdot p(t - \tau) d\tau \quad (9)$$

where $f_{\text{onset}}(t)$ is the PDF of the time from infection to symptoms onset and $p(t) := P\{\text{positive test at time } t \mid \text{symptoms onset at time } 0\}$. For $f_{\text{onset}}(t)$, we consider the values in Table 1. For $p(t)$, we consider Fig. 1 of Wikramaratna et al. [54], estimating $p(t)$ as a Gamma Complementary CDF (CCDF) with shape 7.85, rate 2.14, and scaling factor 0.94. Fig. 4(d) plots $\theta(t; 0)$ obtained using the lower-, mid-, and upper-case values for $f_{\text{onset}}(t)$. Finally, we derive $\theta(t; \tau)$ from $\theta(t; 0)$:

$$\theta(t; \tau) = \begin{cases} 0 & \text{if } t < \tau \\ \theta(t - \tau; 0) & \text{if } t \geq \tau \end{cases} \quad (10)$$

Table 3Parameters for the PDF $\psi(t)$ of time to isolation since a positive test and the PDF $g_{\text{symptoms}}(t)$ of symptoms duration.

Duration	Case	μ	σ	cv	p_1	p_2	n	λ_1	λ_2
Time from positive test to isolation	Lower	0.19	0.74074	3.89864	0.96812	0.03188	–	10.19072	0.33560
	Mid	0.38	0.74074	1.94932	0.88188	0.11812	–	4.64146	0.62170
	Upper	0.54	0.74074	1.29955	0.75307	0.24693	–	2.64235	0.86642
Symptomatic period duration	Lower	6.87	5.21	0.75837	–	–	1	0.47539	0.20980
	Mid	7.88	5.125	0.65038	–	–	2	0.60118	0.21962
	Upper	8.89	5.04	0.56693	–	–	3	0.65074	0.23365

**Fig. 4.** (a) PDF $\psi(t)$ of the time to isolation since a positive test. (b) PDF $g_{\text{symptoms}}(t)$ of symptoms duration. (c) Probability $\phi(t; 0)$ of symptoms at t given an effective contact at 0. (d) Probability $\theta(t; 0)$ of positive test at t given an effective contact at 0.

– Let ϵ be the negative result of a diagnostic test. By the definition of complementary event, we derive $P\{\text{negative test at time } t \mid \gamma \text{ is effective}\} = 1 - P\{\text{positive test at time } t \mid \gamma \text{ is effective}\}$.

- $P\{\omega\}$ is the a-priori probability of observation $\omega = \langle \epsilon, t \rangle$ where the observation type ϵ can be the annotation of symptoms or the (positive or negative) result of a diagnostic test. We consider the same value of $P\{\omega\}$ for every observation type ϵ at any time t , and we evaluate different experimental scenarios by varying this uniform value within the set $\{0.1, 0.3, 0.5, 0.7, 0.9\}$.

Note that the different shape of the PDF of the time to isolation since a positive test (shown in Fig. 4(a)) with respect to the other stochastic parameters shown in Figs. 3 and 4 is due to the fact that the coefficient of variation of the time to isolation since a positive test is larger than 1. Thus, the mean value and coefficient of variation of this duration are fit by a hyper-Exponential PDF, as proposed in [41].

4.3. Quantitative measures of interest

Given a sequence of observations $\omega(t_{\max})$ acquired within the time interval $[0, t_{\max}]$, we use our approach to compute the approximate probability $\tilde{P}(s; t; w(t))$ that each subject $s \in \{s_i\}_{i=1}^S$ is infectious not isolated at time t given $w(t) \forall t \in \{0, \Delta t, 2\Delta t, \dots, t_{\max}\}$, where S is the number of subjects and $\Delta t = 0.1$ h is the considered time step. Then, we average $\tilde{P}(s; t; w(t))$ over a sliding window

of 1 h width, obtaining $\tilde{P}_{\text{avg}}(s; t; \omega(t))$, and we derive the corresponding ranking $z^{\text{ours}}(t)$ of the most infectious not isolated subjects $\forall t \in \{0, \Delta t, 2\Delta t, \dots, t_{\text{max}}\}$, i.e., $z^{\text{ours}}(t) := \langle z_1^{\text{ours}}(t), \dots, z_S^{\text{ours}}(t) \rangle$ where $z_i^{\text{ours}}(t)$ is the subject with the i th highest value of $\tilde{P}_{\text{avg}}(s; t; \omega(t))$ $\forall i \in \{1, \dots, S\}$. Moreover, we also derive the ground truth probability $P(s; t; \omega(t))$ that each subject s is infectious not isolated at time t given $\omega(t)$ $\forall t \in \{0, \Delta t, 2\Delta t, \dots, t_{\text{max}}\}$, as illustrated in [Appendix](#), and we compute the corresponding ground truth ranking $z^{\text{gt}}(t)$ of the most infectious not isolated subjects. We evaluate the accuracy of our approach with respect to the ground truth in terms of *calibration* of the computed probabilities, by computing the following quantitative measures:

- Brier score $B(z^{\text{ours}}(\cdot); z^{\text{gt}}(\cdot); s)$ of each subject s , defined as the mean squared error of the computed probabilities over time with respect to the ground truth probabilities. We compute the average value of $B(z^{\text{ours}}(\cdot); z^{\text{gt}}(\cdot); s)$ over all subjects.
- Expected calibration error $ECE(z^{\text{ours}}(\cdot); z^{\text{gt}}(\cdot); s)$ for each subject s , calculated by partitioning the probabilities into bins (using 10 bins in our experiments) and taking the weighted sum of the absolute difference between confidence and accuracy of all bins. We compute the average value of $ECE(z^{\text{ours}}(\cdot); z^{\text{gt}}(\cdot); s)$ over all subjects.
- Reliability diagrams, which provide a visual representation for predicted and ground truth probabilities.

Moreover, aligning closely with the objectives of our approach, we evaluate its accuracy with respect to the ground truth in terms of *ranking* performance, determining how effectively it identifies subjects with the highest probability of being infectious and not isolated, by computing the following quantitative measures:

- Kendall rank correlation coefficient $\tau(z^{\text{ours}}(t); z^{\text{gt}}(t))$, which measures the ordinal association between the two rankings $z^{\text{ours}}(t)$ and $z^{\text{gt}}(t)$ $\forall t \in \{0, \Delta t, 2\Delta t, \dots, t_{\text{max}}\}$. In particular, $\tau(z^{\text{ours}}(t); z^{\text{gt}}(t))$ takes values within $[-1, 1]$, and it is equal to 1 if the two rankings $z^{\text{ours}}(t)$ and $z^{\text{gt}}(t)$ are the same, and equal to -1 if $z^{\text{ours}}(t)$ is the reverse of $z^{\text{gt}}(t)$. We compute the average value of $\tau(z^{\text{ours}}(t); z^{\text{gt}}(t))$ over time.
- Spearman's rank correlation coefficient $\rho(z^{\text{ours}}(t); z^{\text{gt}}(t))$, which measures the similarity between the two rankings $z^{\text{ours}}(t)$ and $z^{\text{gt}}(t)$ $\forall t \in \{0, \Delta t, 2\Delta t, \dots, t_{\text{max}}\}$, in terms of monotonicity of the relationship between the two rankings and agreement in ordering. In particular, $\rho(z^{\text{ours}}(t); z^{\text{gt}}(t))$ takes values within $[-1, 1]$, and it is equal to 1 if the two rankings $z^{\text{ours}}(t)$ and $z^{\text{gt}}(t)$ have positive monotonic relationship and equal to -1 if they have negative monotonic relationship. We compute the average value of $\rho(z^{\text{ours}}(t); z^{\text{gt}}(t))$ over time.
- Top- k score $\pi(z^{\text{ours}}(t); z^{\text{gt}}(t); k)$, defined $\forall t \in \{0, \Delta t, 2\Delta t, \dots, t_{\text{max}}\}$ as the portion of the k most infectious not isolated subjects according to the ground truth ranking $z^{\text{gt}}(t)$, that ranked within the top k positions by our ranking $z^{\text{ours}}(t)$.¹ For each considered value of k , we compute the average value of $\pi(z^{\text{ours}}(t); z^{\text{gt}}(t); k)$ over time.

Moreover, to compare the accuracy of our approach with that of an alternative baseline, we also perform a number of simulation runs with a total duration not lower than that of the analysis, obtaining the probability $P^{\text{sim}}(s; t; \omega(t))$ that each subject s is infectious not isolated at time t given $\omega(t)$ $\forall t \in \{0, \Delta t, 2\Delta t, \dots, t_{\text{max}}\}$. Then, for each subject s , we average $P^{\text{sim}}(s; t; \omega(t))$ over a sliding window of 1 h width obtaining $\tilde{P}_{\text{avg}}^{\text{sim}}(s; t; \omega(t))$, and we derive the corresponding ranking $z^{\text{sim}}(t)$ of the most infectious not isolated subjects. Note that this baseline takes into account the dependencies among contacts and observations, which are instead discarded by our approach. We evaluate the baseline accuracy with respect to the ground truth by computing the same quantitative measures calculated for our analysis method.

4.4. Generation of data sets of observations

We generate synthetic data sets of observations spanning a time interval of 12 weeks (i.e., 84 days), using three different network topologies for social contacts (i.e., household bubble, scale free, and small world) and three different densities of internal contacts (i.e., low, medium, and high), yielding nine data sets. Moreover, for the selection of the hyperparameter for our approach, we generate an additional data set for each network topology, with a medium number of internal contacts. [Table 4](#) summarizes all the characteristics. Specifically:

- Household bubble data sets HB1, HB2, and HB3 contain 8 individuals and are generated using a complete graph (i.e., 8 nodes and 7 edges per node). They include hundreds of internal contacts (on the order of 200, 400, and 800, respectively, corresponding to nearly 25, 50, and 100 internal contacts per individual on average), tens of external contacts (on the order of 40), a few tens of observations of symptoms (on the order of 20), and several tens of test results (on the order of 100).
- Scale free data sets SF1, SF2, and SF3 contain 100 individuals and are generated via a Barabási–Albert model (i.e., 100 nodes and 3 edges per node). They include thousands of internal contacts (on the order of 1250, 2500, and 5000, respectively, corresponding to nearly 12.5, 25, and 50 internal contacts per individual on average), several hundreds of external contacts (on the order of 1000), and several hundreds of observations of both symptoms and test results (on the order of 1000).
- Small world data sets SW1, SW2, and SW3 contain 100 individuals and are generated via a Watts–Strogatz model (i.e., 100 nodes, each connected to 5 nearest neighbors, with a rewiring probability equal to 0.1). They include thousands of internal contacts (on the order of 1800, 3600, and 7200, respectively, corresponding to nearly 18, 36, and 72 internal contacts per individual on average), several hundreds of external contacts (on the order of 1000), and several hundreds of observations of both symptoms and test results (on the order of 1000).

¹ In our setting, this metric corresponds to both top- k precision and top- k recall, since we select the top k positions both in the ground truth ranking and in the predicted ranking. We use the name top- k score to highlight a crucial difference with respect to standard precision and recall, that is the fact that our ground truth consists of probabilities rather than binary labels.

Table 4

Generated data sets: number of individuals and number of observations of each type.

Data set	# Individuals	# Internal contacts	# External contacts	# Symptoms	# Tests
HB1	8	190	40	18	97
HB2	8	380	40	31	87
HB2-val	8	380	40	31	87
HB3	8	760	40	14	127
SF1	100	1188	950	950	950
SF2	100	2375	950	950	950
SF2-val	100	2375	950	950	950
SF3	100	4750	950	950	950
SW1	100	1710	950	950	950
SW2	100	3420	950	950	950
SW2-val	100	3420	950	950	950
SW3	100	6840	950	950	950

- In addition, data sets HB2-val, SF2-val, and SW2-val, used for the selection of the hyperparameter for our approach, feature the same characteristics as data sets HB2, SF2, and SW2, respectively.

Data sets are generated through the following procedure:

- *Network structure.* The structure of the network of social contacts is created using the NetworkX library.
- *External contacts.* The external contacts (each involving a single subject of the cluster) are created as follows: for data sets HB1, HB2, and HB3, the number of the external contacts for each subject is uniformly sampled from $\{0, \dots, 8\}$, making a total of 40 external contacts; for the remaining data sets (i.e., SF1, SF2, SF3, SW1, SW2, and SW3), each of 1000 external contacts is assigned to a randomly selected individual. Then, each external contact is associated with a timestamp sampled from a uniform distribution over the interval $[0, t_{\max}]$, where the time limit $t_{\max}$ is 84 days (i.e., 2016 h).
- *Internal contacts.* The internal contacts are generated in two steps. First, 1.5% of the external contacts are randomly selected to be effective and, starting from these initial transmission seeds, the spread of the disease is simulated over the social network using the EoN (Epidemics on Networks) library, yielding the effective internal contacts, each associated with a timestamp dynamically determined by the EoN simulation algorithm. Next, the ineffective internal contacts are generated by recurrently and randomly sampling an edge from the graph and designating the two endpoints as the interacting subjects. With a probability of 0.35, neighboring nodes are randomly included in the event to account for multi-person gatherings. The timestamp of each ineffective internal contact is uniformly sampled from $[0, t_{\max}]$.
- *Symptoms and tests.* For data sets HB1, HB2, and HB3, the number of observations of symptoms for each subject is uniformly sampled from $\{0, \dots, 8\}$, yielding a total number of observations of symptoms equal to 18, 31, and 14, respectively, while the number of results of diagnostic tests for each subject is uniformly sampled from $\{0, \dots, 24\}$, resulting in a total number of test results equal to 97, 87, and 127, respectively. For the remaining data sets, each of the 1000 observations of symptoms and each of the 1000 diagnostic tests are assigned to a randomly selected individual. For each data set, the timestamp of each observation of symptoms and each result of a diagnostic test is uniformly sampled from $[0, t_{\max}]$. Then, a simulation run of Algorithm 1 of Appendix (lines 1–58) is performed using the mid-case values of the stochastic parameters² to discard the contacts of the isolated subjects, as well as to maintain/discard the observations of symptoms and to generate the actual test results. Specifically, for each observation of symptoms of subject s at time τ , we maintain the observation if s is symptomatic at time τ , we maintain the observation with probability 0.5 if s is developing symptoms at time τ , and we discard the observation otherwise. Moreover, for each observation of a diagnostic test result for subject s at time τ , we generate a positive result with a probability of $\theta(\tau; \tau')$ if s is infectious at time τ and had an effective contact at time τ' (where $\theta(\tau; \tau')$ is the mid-case probability of a positive test at time τ given an effective contact at time τ' , i.e., the green curve in Fig. 4(d)), with a probability of 0.05 if s has not been infected, and with a probability of 0.025 if s has recovered; in all these cases, we generate a negative result with the complementary probability.
- *Data distortion.* To account for noisy, erroneous, and missing observations in real scenarios, we remove 5% of the observations of each type (i.e., contacts, symptoms, and tests), we perturb the timestamp of 5% of the observations within a time window of ± 6 h, and we flip the result of 5% of the tests.

For each data set, we derived a ground truth as illustrated in Appendix, exploiting Algorithm 1 and the mid-case values of the stochastic parameters, and considering time limit $t_{\max} = 12$ weeks (i.e., the duration of each data set) and time step $\Delta t = 1$ h (yielding $T = 2016$ time instants). Importantly, this derivation was notably performed using the data prior to the perturbations described in the final generation step above.

² This simulation run is used exclusively to discard the contacts of the isolated subjects and to generate the observed test results and symptom logs, and it is executed independently of both ground truth simulation and baseline simulation.

Table 5

Performance measures achieved by our approach on data sets HB2-val, SF2-val, and SW2-val for a different number N of iterations (using $P\{\omega\} = 0.5$ for every observation type and the mid-case values of the other stochastic parameters defined in Section 4.2). The top- k score is evaluated for $k \in \{1, 2, 3, 4\}$ for HB2-val and for $k \in \{10, 20, 30, 40\}$ for SF2-val and SW2-val. For each measure and data set, the best value (achieved by our approach for different values of N) is indicated in bold.

	HB2-val				SF2-val2				SW2-val			
	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 1$	$N = 2$	$N = 3$	$N = 4$
Kendall	0.2150	0.6354	0.6721	0.6704	0.2978	0.4617	0.4785	0.4727	0.2877	0.4820	0.5014	0.4992
Spearman	0.2742	0.8228	0.8538	0.8576	0.3934	0.6300	0.6534	0.6500	0.3847	0.6575	0.6841	0.6858
Top-1/10	0.2937	0.2411	0.2014	0.1622	0.4152	0.4345	0.4425	0.4058	0.2316	0.2173	0.1930	0.2227
Top-2/20	0.3031	0.3713	0.3557	0.3626	0.4511	0.4377	0.4392	0.4499	0.2855	0.2855	0.2617	0.2870
Top-3/30	0.4950	0.5764	0.5728	0.5757	0.4239	0.4049	0.4208	0.4099	0.2566	0.2460	0.2416	0.2531
Top-4/40	0.5046	0.5956	0.5985	0.6055	0.4394	0.4303	0.4237	0.4095	0.2681	0.2767	0.2728	0.2754

4.5. Hyperparameter selection

We applied our approach to data sets HB2-val, SF2-val, and SW2-val executing $N \in \{1, 2, 3, 4\}$ iterations (considering $P\{\omega\} = 0.5$ for every type of observation and using the mid-case values of the stochastic parameters defined in Section 4.2). Table 5 shows the obtained performance measures with respect to the ground truth. Specifically, for all data sets, the Kendall rank correlation coefficient significantly increases when N increases from 1 to 2, and achieves its best value for $N = 3$. Similarly, the Spearman's rank correlation coefficient also significantly increases as N increases from 1 to 2, and achieves its best value for $N = 3$ (for SF2-val) or $N = 4$ (for NB2-val and SW2-val), with very small differences in the values obtained for $N = 3$ and $N = 4$ (on the order of 10^{-3}). For HB2-val, the top-1 score decreases with N , the top-2 score increases as N increases from 1 to 2 and then exhibits minor variations, and the top-3 and top-4 scores increase with N . Conversely, for SF2-val and SW2-val, the top- k scores change slightly as N increases. Therefore, $N = 3$ appears to be a good tradeoff across all data sets and performance measures.

5. Experimental results

In this section, we illustrate the results of our experimental evaluation (Section 5.1) and we discuss the related threats to validity (Section 5.2). All experiments were performed on a virtual machine equipped with 32 vCPUs (distributed across 4 virtual sockets) and 128 GB of RAM, hosted on a dual-socket server featuring Intel Xeon Gold 6238R processors with 2.20 GHz frequency.

5.1. Performance measures

We applied our approach and the baseline defined in Section 4.3 on data sets HB1, HB2, HB3, SF1, SF2, SF3, SW1, SW2, and SW3, using $P\{\omega\} = 0.5$ for every observation type and considering all the combinations of the values of the other stochastic parameters defined in Section 4.2. Specifically, we considered 3 values (i.e., (i) lower, (ii) mid, and (iii) upper values) for 7 parameters (i.e., (i) time from infection to infectiousness, (ii) time from infection to symptoms onset, (iii) time from symptoms onset to isolation, (iv) time from infectiousness to healing in asymptomatic subjects, (v) time from positive test to isolation, (vi) symptomatic period duration, and (vii) probability of developing symptoms), yielding a total of 2187 experiments for each data set. As for the ground truth derivation, for our approach we used time limit $t_{\max} = 12$ weeks (i.e., the duration of each data set) and time step $\Delta t = 1$ h (yielding $T = 2016$ time instants). On average over the 2187 experiments, runtimes are nearly: 0.9 s, 0.9 s, and 1.1 s for data sets HB1, HB2, and HB3, respectively (1.0 s, 1.1 s, and 1.4 s, respectively, for the baseline); 19.6 s, 10.2 s, and 10.6 s for data sets SF1, SF2, and SF3, respectively (24.6 s, 14.0 s, and 17.2 s, respectively, for the baseline); and 20.6 s, 21.7 s, and 22.4 s for data sets SW1, SW2, SW3, respectively (26.0 s, 28.6 s, and 32.9 s, respectively, for the baseline). Conversely, the runtime required to compute the ground truth is nearly: 11 min, 14 min, and 19 min for data sets HB1, HB2, and HB3, respectively; 5.1 h, 6 h, and 8.2 h for data sets SF1, SF2, and SF3, respectively; and, 5.5 h, 7 h, and 10.1 h for data sets SW1, SW2, and SW3, respectively.

Tables 6 to 8 report the performance measures achieved by our approach and the baseline (averages over 2187 experiments) on data sets of type HB, SF, and SW, respectively, indicating whether the difference in the measures is statistically significant (p -value < 0.05) according to the Wilcoxon signed-rank test (results are significant for almost all performance measures and data sets). For a subset of these measures (i.e., Brier score, ECE, Kendall and Spearman's rank correlation coefficients), Figs. 5 to 7 show the Bland-Altman plots comparing the obtained measures for data sets HB, SF, and SW, respectively (the x -axis represents the mean of the measures obtained by our approach and the baseline, while the y -axis represents their difference).

Finally, we performed a sensitivity analysis with respect to the value of the a-priori probability $P\{\omega\}$ of an observation ω , which we assume to be the same for every observation type (i.e., symptoms, positive test, negative test). Specifically, we applied our approach and the baseline defined in Section 4.3 on data sets HB2, SF2, and SW2, considering $P\{\omega\} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ as well as all the combinations of the values of the other stochastic parameters defined in Section 4.2, yielding a total of 2187 experiments for each data set and each value of $P\{\omega\}$. Table 9 reports the performance measures achieved by our approach and the baseline (averages over 2187 experiments for each data set and each value of $P\{\omega\}$), indicating whether the difference in the measures is statistically significant (p -value < 0.05) according to the Wilcoxon signed-rank test (results are significant in almost all the considered cases).

Table 6

Performance measures achieved by our approach (using a-priori observation probability $P\{\omega\} = 0.5$ for every observation type) and the alternative baseline on data sets HB1, HB2, and HB3 (averages over 2187 experiments). For each measure and data set, the best of the two values is indicated in bold, and an asterisk indicates if the difference is statistically significant according to the Wilcoxon signed-rank test (i.e., p -value < 0.05).

	HB1			HB2			HB3		
	Ours	Baseline	sig.	Ours	Baseline	sig.	Ours	Baseline	sig.
Brier	0.0248	0.0194	*	0.0191	0.0197		0.0266	0.0218	*
ECE	0.0901	0.0815	*	0.0787	0.0774	*	0.0993	0.0850	*
Kendall	0.6338	0.4900	*	0.6233	0.5071	*	0.6463	0.4365	*
Spearman	0.8208	0.6090	*	0.8103	0.6207	*	0.8278	0.5352	*
Top-1	0.3752	0.2501	*	0.2543	0.2694		0.2605	0.2151	*
Top-2	0.4478	0.3602	*	0.4012	0.3840	*	0.4263	0.2788	*
Top-3	0.5292	0.5351	*	0.5559	0.5169	*	0.5308	0.4347	*
Top-4	0.6127	0.5979	*	0.6175	0.6057	*	0.5483	0.5284	*

Table 7

Performance measures achieved by our approach (using a-priori observation probability $P\{\omega\} = 0.5$ for every observation type) and the alternative baseline on data sets SF1, SF2, and SF3 (averages over 2187 experiments). For each measure and data set, the best of the two values is indicated in bold, and an asterisk indicates if the difference is statistically significant according to the Wilcoxon signed-rank test (i.e., p -value < 0.05).

	SF1			SF2			SF3		
	Ours	Baseline	sig.	Ours	Baseline	sig.	Ours	Baseline	sig.
Brier	0.0353	0.0329	*	0.0452	0.0392	*	0.0535	0.0706	*
ECE	0.0894	0.1002	*	0.0974	0.1053	*	0.0981	0.1364	*
Kendall	0.6067	0.4148	*	0.6148	0.4010	*	0.5904	0.3303	*
Spearman	0.7827	0.5099	*	0.8017	0.4925	*	0.7751	0.4044	*
Top-10	0.4759	0.2436	*	0.4474	0.1885	*	0.3007	0.1259	*
Top-20	0.4936	0.3372	*	0.4760	0.2547	*	0.3987	0.2314	*
Top-30	0.5509	0.4150	*	0.5188	0.3463	*	0.4924	0.3321	*
Top-40	0.6012	0.4821	*	0.5853	0.4256	*	0.5520	0.4291	*

Table 8

Performance measures achieved by our approach (using a-priori observation probability $P\{\omega\} = 0.5$ for every observation type) and the alternative baseline on data sets SW1, SW2, and SW3 (averages over 2187 experiments). For each measure and data set, the best of the two values is indicated in bold, and an asterisk indicates if the difference is statistically significant according to the Wilcoxon signed-rank test (i.e., p -value < 0.05).

	SW1			SW2			SW3		
	Ours	Baseline	sig.	Ours	Baseline	sig.	Ours	Baseline	sig.
Brier	0.0453	0.0340	*	0.0475	0.0359	*	0.0518	0.0566	*
ECE	0.0991	0.1012	*	0.0919	0.0997	*	0.0940	0.1227	*
Kendall	0.6183	0.4208	*	0.6076	0.4186	*	0.5571	0.3633	*
Spearman	0.7952	0.5168	*	0.7976	0.5142	*	0.7490	0.4457	*
Top-10	0.5062	0.2380	*	0.3138	0.1914	*	0.2235	0.1354	*
Top-20	0.5065	0.3477	*	0.3967	0.3102	*	0.2730	0.3134	*
Top-30	0.5501	0.4087	*	0.4882	0.4198	*	0.3738	0.4089	*
Top-40	0.5915	0.5063	*	0.5296	0.5236	*	0.4573	0.4921	*

Calibration. Our approach achieves good values of calibration global measures on all data sets, with average Brier score in the order of $2 - 5 \cdot 10^{-2}$ and average ECE in the order of $7 - 9 \cdot 10^{-2}$, comparable with values achieved by the baseline. However, these aggregated measures over time mask localized discrepancies, which can be revealed by Bland-Altman plots, which map the mean of the measures obtained by our approach and the baseline against their difference. For the HB data sets, the plots exhibit dense clustering near the $y = 0$ line at low mean values, confirming strong agreement and similar performance between our approach and the baseline. As the mean values increase along the x -axis, the plots display a distinct downward trajectory into y -negative zones, indicating that our approach yields systematically smaller Brier score and ECE than the baseline. Conversely, for the SF and SW data sets, our approach and the baseline expose a pronounced divergence, characterized by highly structured and striped distribution patterns and wider variance.

Further insight into the temporal dynamics driving these calibration metrics is provided by the transient probability curves (i.e., transient probability that each subject is infectious not isolated, computed by our approach and the ground truth) alongside their corresponding reliability diagrams. As an example, Fig. 8 illustrates these plots for the experiment achieving the best Brier score on data set HB1 (where the DKW confidence bands are almost coincident with the ground truth curves, demonstrating its accuracy). Here, our approach tightly mirrors the primary outbreak wave of the ground truth in the first half of the time horizon for most subjects. The localized miscalibration visible in the reliability diagrams manifests primarily as low-amplitude loops, which are driven by a minor overestimation of a late-stage peak in the second half of the time horizon (most visible in subjects 1, 7, and 8).

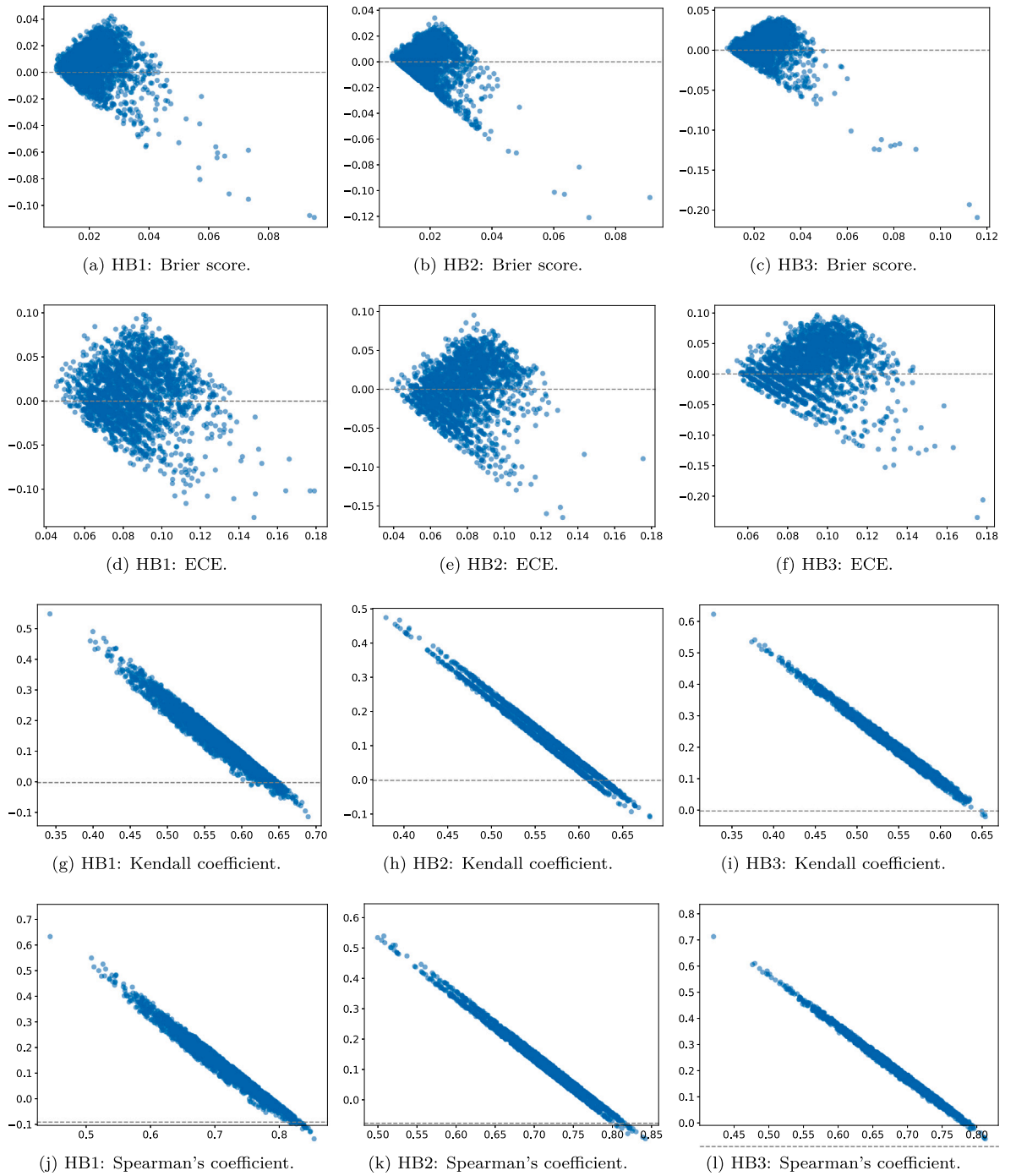


Fig. 5. Bland-Altman plots comparing the performance measures for data sets HB1, HB2, and HB3 (2187 experiments). The x -axis represents the average of the measures achieved by our approach and the baseline; the y -axis displays their difference.

Conversely, Fig. 9 illustrates the plots for the experiment achieving the worst Brier score on data set HB1, where calibration breaks down more severely. In this scenario, our approach significantly overestimates the magnitude of the primary infection wave, with predicted probability peaks doubling the ground truth values for subjects 1, 2, 3, 4, 6, and 8, pushing these transient probabilities far above the true risk levels. Consequently, the sequential time-series paths in the reliability diagrams loop broadly below the ideal identity line, visually confirming a persistent overestimation bias during high-stakes epidemic windows. Similar trends can be observed for the other data sets.

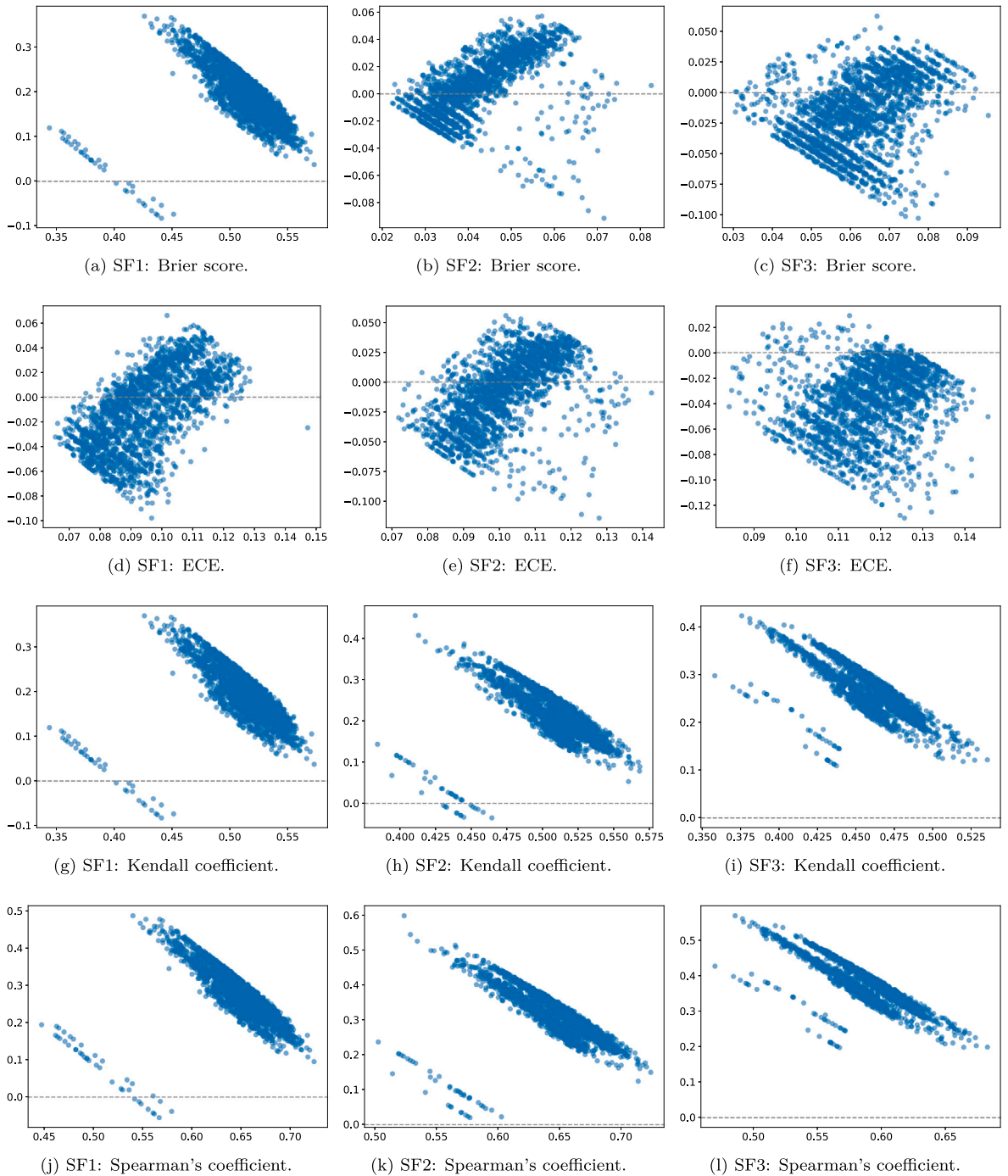


Fig. 6. Bland-Altman plots comparing the performance measures for data sets SF1, SF2, and SF3 (2187 experiments). The x -axis represents the average of the measures achieved by our approach and the baseline; the y -axis displays their difference.

Ranking. Crucially, the primary objective of our approach is not to perfectly estimate the transient probability that each individual is infectious and not isolated, but rather to accurately identify and rank the most infectious individuals over time. In our targeted application scenario, establishing a correct risk hierarchy among subjects is of greater practical significance than precise calibration. The ranking global measures obtained in thousands of experiments for diverse social network topologies and densities of internal contacts (Tables 6 to 8) show that our approach consistently and significantly outperforms the baseline in terms of rank correlation. Specifically, our method achieves remarkably strong monotonic alignment with the ground truth, yielding Kendall and Spearman's

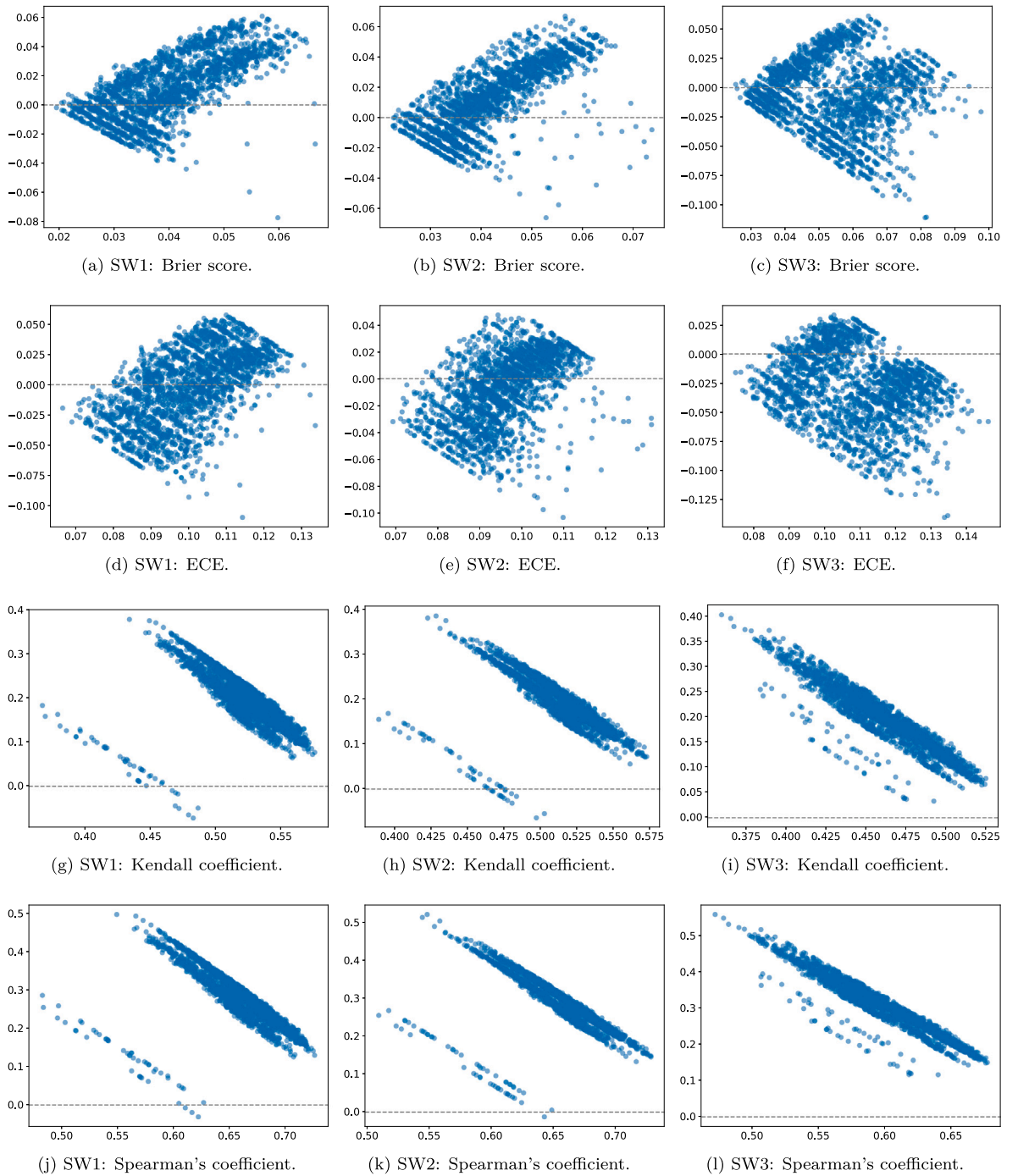


Fig. 7. Bland-Altman plots comparing the performance measures for data sets SW1, SW2, and SW3 (2187 experiments). The x -axis represents the average of the measures achieved by our approach and the baseline; the y -axis displays their difference.

rank correlation coefficients that hover between 0.55–0.65 and 0.74–0.83, respectively. In contrast, the baseline struggles to maintain effective ranking order under complex topologies, with its Kendall and Spearman coefficients dropping as low as 0.33 and 0.40, respectively, for the SF3 data set. This robust ranking capability is further confirmed by the Bland-Altman plots (Figs. 5 to 7), where the data points lay almost entirely above the $y = 0$ line, showing that our approach consistently yields higher correlation coefficients than the baseline across the entire spectrum of experimental configurations. Furthermore, our method shows a very good capacity to isolate the highest-risk individuals, as reflected in the top- k accuracy scores. For instance, for the HB1 data set, our

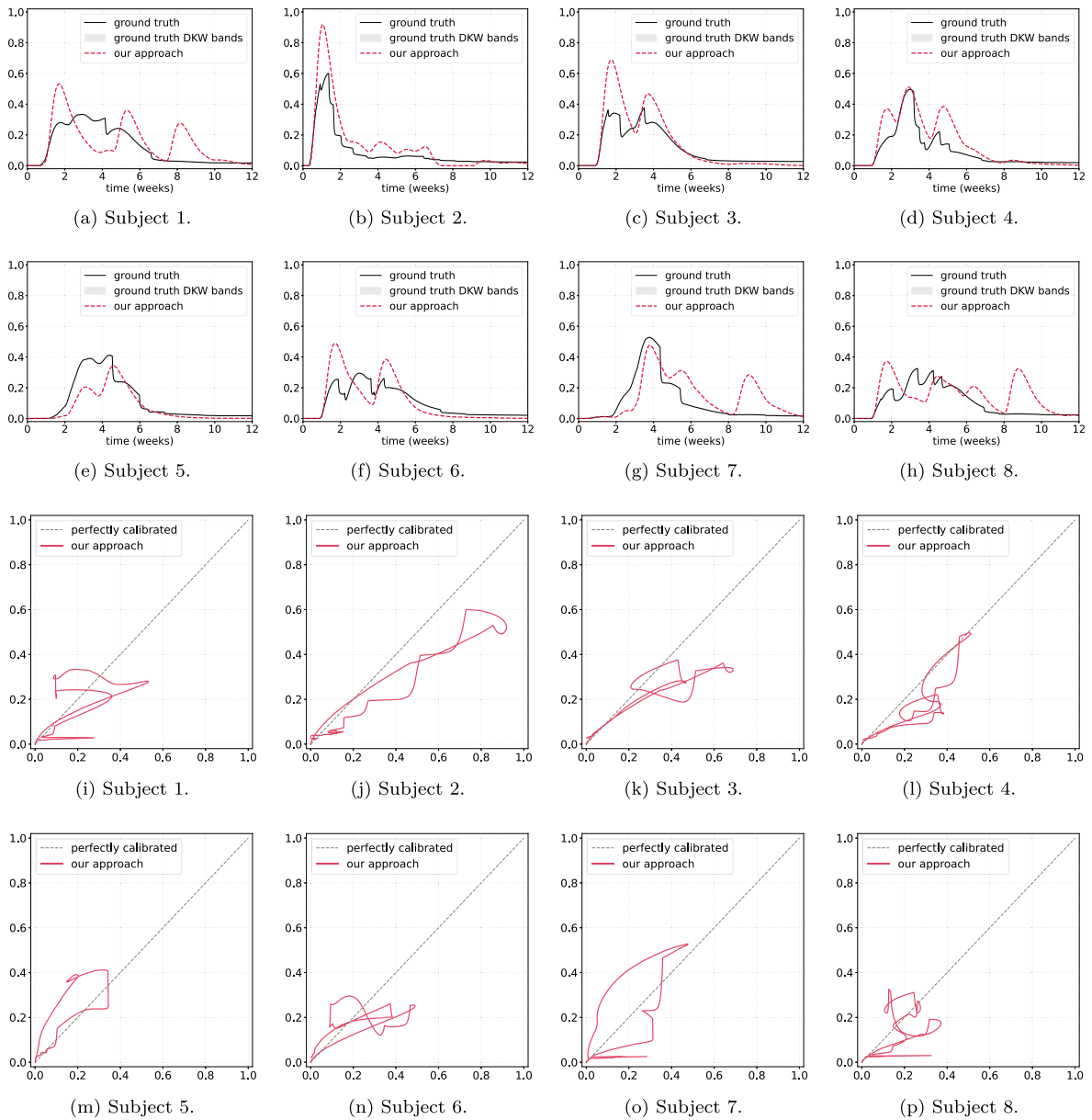


Fig. 8. Experiment achieving the best Brier score on data set HB1 (over 2187 experiments): (a–h) transient probability that each subject is infectious not isolated, computed by our approach and the ground truth; (i–p) corresponding reliability diagrams (the x-axis represents the probabilities computed by our approach and the y-axis represents the ground truth probabilities).

approach achieves a top-1 score of 0.3752 compared to the value of 0.2501 obtained by the baseline. A similarly strong advantage is maintained in the SF1 and SW1 data sets, where our top-10 score reaches 0.4759 and 0.5062, respectively, nearly doubling the performance of the baseline. In a few cases (e.g., for data set SW3), the baseline achieves better top- k scores, though still in the same order of magnitude as those obtained by our approach.

Sensitivity to the α -prior probability of an observation. In the previous experiments, $P\{\omega\} = 0.5$ represents a non-informative maximum-entropy prior, expressing the fact that, in the absence of a history of observations, whether an individual observation will occur or not remains completely uncertain. To address the potential sensitivity to this parameter, we evaluated performance across five evenly spaced values, i.e., $P\{\omega\} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. As shown in Table 9, the risk ranking measures (i.e., Kendall rank correlation coefficient, Spearman's rank correlation coefficient, and top- k scores) prove to be remarkably robust to variations in $P\{\omega\}$, due to the fact that $P\{\omega\}$ serves primarily as a uniform scaling factor across individuals, thus preserving relative risk ordering. Furthermore, this stability holds across all considered network topologies, with our approach consistently outperforming the baseline

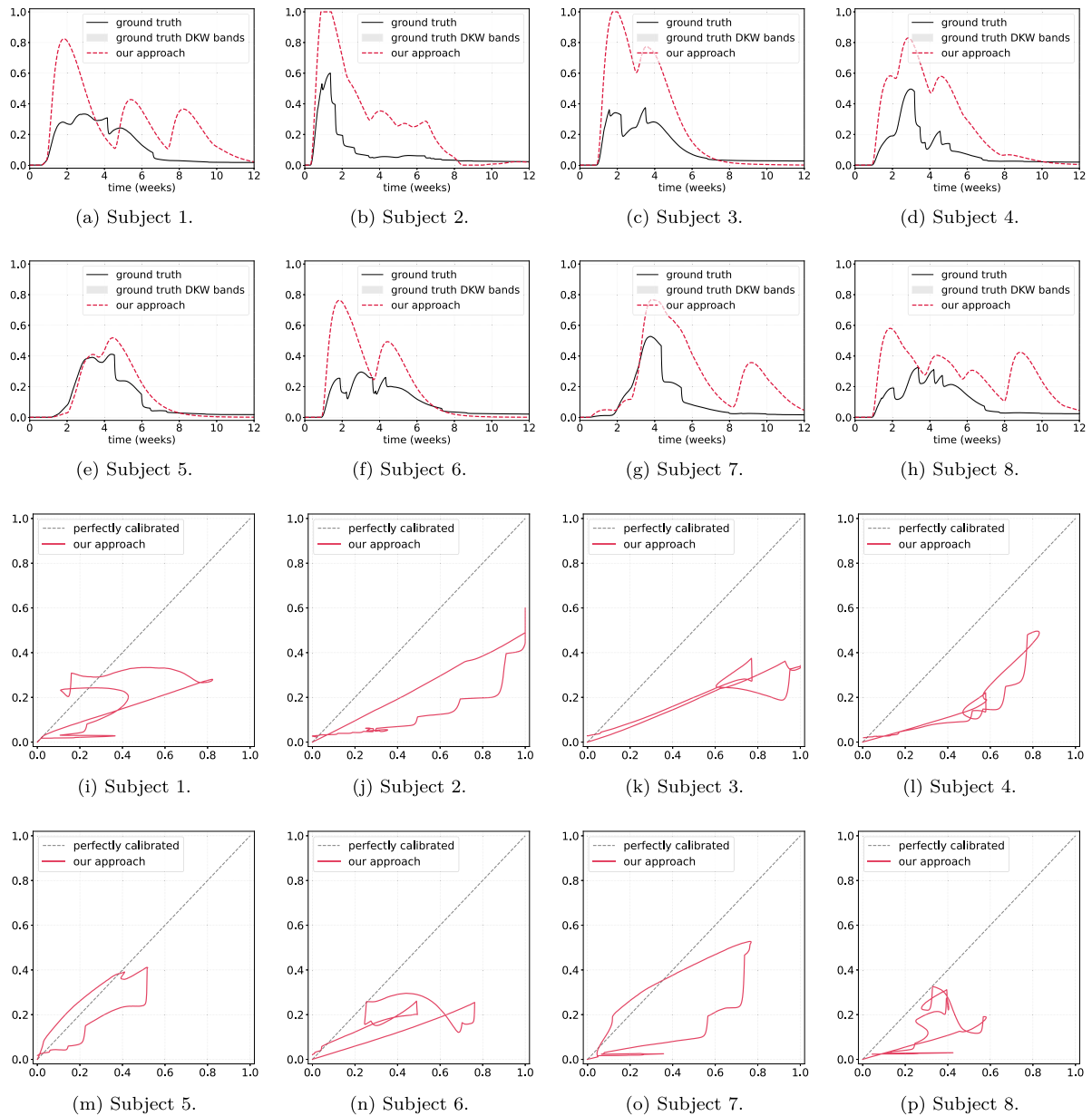


Fig. 9. Experiment achieving the worst Brier score on data set HB1 (over 2187 experiments): (a–h) transient probability that each subject is infectious not isolated, computed by our approach and the ground truth; (i–p) corresponding reliability diagrams (the x-axis represents the probabilities computed by our approach and the y-axis represents the ground truth probabilities).

with statistical significance. Interestingly, while calibration metrics (i.e., Brier score and ECE) remain completely invariant in the structured household bubble settings of the HB2 data set, they exhibit mild variation in the less structured topologies of the SF2 and SW2 data sets, where $P\{\omega\} = 0.5$ achieves the optimal calibration performance.

Remarks. Overall, these results highlight that while our approach may exhibit local calibration errors during transient post-peak phases, it provides a reliable operational tool for prioritizing isolation and intervention strategies targeting the most infectious individuals within a population network. First, we address **RQ1** regarding the computational feasibility of our approach. The extensive evaluation, encompassing 2187 distinct experiments for each of the 9 data sets and processing large-scale tracking over a horizon of more than 2000 time steps, demonstrates that the approach is highly scalable and computationally viable for processing a large number of subjects and prolonged observation windows, making it suitable for real-world deployment.

In response to **RQ2**, our results confirm that, as the number of subjects and observations varies across different social network topologies, the approach maintains high and stable performance in correctly ranking individuals according to their true probability of

Table 9

Performance measures achieved by our approach and the alternative baseline on data sets HB2, SF2, and SW2 (averages over 2187 experiments) for different values of the a-priori probability $P\{\omega\}$ of an observation ω , i.e., $P\{\omega\} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. For each measure and each value of $P\{\omega\}$, the best of the two values is indicated in bold, and labeled with an asterisk if the difference is statistically significant according to the Wilcoxon signed-rank test (i.e., p -value < 0.05).

		$P\{\omega\} = 0.1$		$P\{\omega\} = 0.3$		$P\{\omega\} = 0.5$		$P\{\omega\} = 0.7$		$P\{\omega\} = 0.9$		Baseline
		Ours	sig.	Ours	sig.	Ours	sig.	Ours	sig.	Ours	sig.	
HB2	Brier	0.0191		0.0191		0.0191		0.0191		0.0191		0.0197
	ECE	0.0788	*	0.0788	*	0.0787	*	0.0788	*	0.0788	*	0.0774
	Kendall	0.6233	*	0.6233	*	0.6233	*	0.6233	*	0.6233	*	0.5071
	Spearman	0.8103	*	0.8103	*	0.8103	*	0.8103	*	0.8103	*	0.6207
	Top-1	0.2543		0.2543		0.2543		0.2543		0.2543		0.2694
	Top-2	0.4012	*	0.4012	*	0.4012	*	0.4012	*	0.4012	*	0.3840
	Top-3	0.5559	*	0.5559	*	0.5559	*	0.5559	*	0.5559	*	0.5169
	Top-4	0.6175	*	0.6175	*	0.6175	*	0.6175	*	0.6175	*	0.6057
SF2	Brier	0.1524	*	0.1344	*	0.0452	*	0.1354	*	0.1358	*	0.0394
	ECE	0.1237	*	0.1273	*	0.0974	*	0.1277	*	0.1276	*	0.1063
	Kendall	0.5979	*	0.5921	*	0.6148	*	0.5886	*	0.5873	*	0.4004
	Spearman	0.7890	*	0.7847	*	0.8017	*	0.7805	*	0.7791	*	0.4905
	Top-10	0.4221	*	0.4167	*	0.4474	*	0.4255	*	0.4277	*	0.1894
	Top-20	0.4548	*	0.4672	*	0.4760	*	0.4697	*	0.4685	*	0.2821
	Top-30	0.5017	*	0.5070	*	0.5188	*	0.5077	*	0.5057	*	0.3629
	Top-40	0.5474	*	0.5450	*	0.5853	*	0.5468	*	0.5445	*	0.4481
SW2	Brier	0.2021	*	0.1555	*	0.0475	*	0.1540	*	0.1542	*	0.0359
	ECE	0.1029	*	0.1082	*	0.0919	*	0.1097	*	0.1099	*	0.0997
	Kendall	0.5879	*	0.6088	*	0.6076	*	0.6079	*	0.6061	*	0.4186
	Spearman	0.7789	*	0.8018	*	0.7976	*	0.8006	*	0.7990	*	0.5142
	Top-10	0.3108	*	0.3174	*	0.3138	*	0.3141	*	0.3159	*	0.1914
	Top-20	0.3848	*	0.3881	*	0.3967	*	0.3890	*	0.3882	*	0.3101
	Top-30	0.4788	*	0.4860	*	0.4882	*	0.4899	*	0.4906	*	0.4198
	Top-40	0.5213	*	0.5293	*	0.5296	*	0.5287	*	0.5268	*	0.5236

being infectious and not isolated, achieving robust rank correlation coefficients and top- k scores. In all these scenarios, the approach consistently and significantly outperforms an alternative baseline, even doubling its accuracy measures (e.g., for data sets SF1 and SW1).

Finally, we address **RQ3** concerning the robustness of our approach when its stochastic parameters are intentionally disturbed with respect to the actual ground truth. Notably, the obtained rank-based measures confirm that the approach remains highly effective in these scenarios, and the Bland-Altman plots for the Kendall and Spearman's rank correlation coefficients show that it consistently yields better alignment with the ground truth compared to the baseline. According to this, our framework proves to be highly resilient, ensuring that miscalibrations in absolute probability values do not compromise the model core utility, i.e., the precise prioritization of the most infectious subjects for targeted containment.

5.2. Threats to validity

We discuss the threats to validity by following the classification provided in [55].

- **Construct validity:** The interpretation of the obtained experimental results may be affected by the quantitative measures used to assess the calibration of the computed probabilities (i.e., probability that each subject is infectious and not isolated over time) as well as to compare the ranking of the subjects (according to their probability of being infectious and not isolated) provided by the ground truth with that provided by our approach or the alternative baseline. To mitigate this threat to construct validity, we considered different quantitative measures, capturing different calibration properties (i.e., Brier score and expected calibration error) and different aspects of the comparison between two rankings (i.e., Kendall rank correlation coefficient, Spearman's rank correlation coefficient, top- k scores).
- **Internal validity:** The proposed approach requires the specification of stochastic parameters, mainly concerned with the duration distributions of the evolution phases of an infectious disease in an individual, and with the transient probability that an individual has symptoms or is positively tested given an effective contact at a previous time. To mitigate these threats to internal validity, we estimated each duration with three distinct probability distributions fitting the lower, mid, and upper cases of the variability ranges for the mean value and the coefficient of variation found in the literature on COVID-19. For each data set, we performed a sensitivity analysis using the mid-case parameter values for the ground truth derivation and all the ($3^7 = 2187$) combinations of parameter values for our approach. We performed a sensitivity analysis also with respect to the a-priori probability of an observation.

- **External validity:** The results of our experimental evaluation may be biased by the data set of observations. To mitigate this threat to external validity, we generated nine synthetic data sets of observations (plus three data sets used for hyperparameter selection) by varying the number of individuals, the topology of the network of social contacts, and the density of internal contacts. Moreover, we also perturbed the data sets to account for noisy, erroneous, and missing observations.
- **Reliability:** The independent reproduction and verification of our experimental results may require some effort. To mitigate this threat to reliability, we made a replication package available open-source, including the generated data sets of observations and the stochastic temporal parameters estimated for the steps of disease evolution within an individual, from contact to infectiousness. In particular, for each considered scenario, this package easily supports the derivation of a ground truth, the execution of our analytical approach and the considered baseline, and the plot of results.

6. Conclusions

We presented an efficient quantitative approach to predict how the risk of infectious disease evolves over time for the individuals of a cluster, by exploiting the observations taken up to that time on contacts among individuals, symptoms, and diagnostic test results. The solution method iteratively estimates the probability that a subject is infectious and not isolated over time by considering an increasingly larger number of internal contacts in the sequence that infected the subject, while disregarding dependencies among observations to achieve computational efficiency. Each stochastic temporal parameter used by the approach is carefully estimated by three distinct probability distributions fitting the lower, mid, and upper cases of the variability ranges for the mean value and the coefficient of variation found in the literature on COVID-19.

To validate the approach, we generated nine synthetic data sets of observations, alongside three additional data sets used for hyperparameter selection, varying the number of individuals, the topology of the network of social contacts, and the density of internal contacts, while accounting for noisy, erroneous, and missing observations. For each data set, we used the mid-case values of the stochastic parameters to derive a ground truth, by performing a stochastic simulation that accounts for dependencies among observations. Then, we performed an extensive sensitivity analysis by evaluating thousands of combinations of the values of the stochastic parameters within our approach. The experimental results show that our approach is highly accurate at ranking subjects according to their transient probability of being infectious and not isolated, notably outperforming an alternative baseline that accounts for dependencies among observations (actually disregarded by our solution method) while maintaining a comparable runtime.

Thanks to the very low run times, our approach is well suited for online deployment, where run-time changes in individual behaviors are taken into account as soon as new observations emerge. However, to fully realize this potential in real-world scenarios, future implementation must address the operational delays associated with collecting, curating and digitizing data (e.g., contact logs and test results), which actually represent the primary bottleneck. Concurrently, the approach efficiency makes it immediately viable for offline applications, including the retrospective reconstruction of infection chains.

The availability of finer statistics for the durations characterizing the infectious disease evolution in an individual would support the derivation of more accurate duration distributions, e.g., by fitting the shape of the observed histograms [56,57], enabling full exploitation of the proposed approach. Actually, these data have been increasingly collected and published in epidemiological studies, especially after the COVID-19 pandemic [54,58,59], though typically made available in aggregate form (e.g., moments) rather than as histograms. The approach could also be easily customized to specific infectious diseases by adjusting the structure and the parameters of the stochastic model characterizing the infection evolution in an individual.

Appendix. Stochastic simulation for the derivation of a ground truth

Algorithm 1 simulates the evolution of an infection chain modeled by an instance of the metamodel of Fig. 1. Specifically, given a metamodel instance μ , a time limit $t_{\max}$, a sequence of observations $\omega(t_{\max}) := \langle \langle \epsilon_i, \tau_i \rangle \rangle_{i=1}^m$ acquired within the interval $[0, t_{\max}]$, a time step Δt , and a number R of simulation runs, Algorithm 1 derives, for each subject s , the probability $P(s; t; \omega(t))$ that s is infectious not isolated at time t given $\omega(t)$. To this end, in each simulation run (line 2), the state of each subject is represented as the collection of (line 3): (i) the infectiousness state (vector α), which can be *healthy*, *infected*, *infectious*, *healed*, *isolated*, or *positively_tested*; (ii) the symptomaticity state (vector β), which can be *developing_symptoms*, *symptomatic*, *asymptomatic*, or *N/A*; (iii) the time to the next infectiousness state (vector γ); and, (iv) the time to the next symptomaticity state (vector δ). At the beginning of each run (line 3), the current time is $t = 0$, the index of the next observation is $i = 1$ and, for each subject, the infectiousness state is *healthy*, the symptomaticity state is *N/A*, and the time to the next infectiousness/symptomaticity state is ∞ . In each system state (collecting the states of all subjects), the next event can be: (i) a transition in the infectiousness state of a subject; (ii) a transition in the symptomaticity state of a subject; or, (iii) an observation. Thus, time t is repeatedly advanced to a time t' (lines 4–57) such that the duration $t' - t$ is the minimum among: (i) the minimum time c to next infectiousness state of any subject; (ii) the minimum time d to the next symptomaticity state of any subject; and, (iii) the time $\tau_i - t$ to the next observation. In particular:

- If the next event is a transition from the infectiousness state $\alpha(a)$ of a subject a (i.e., $c' \leq d' \wedge c' \leq \tau_i - t$, lines 6–26), then such state can be *infected*, *infectious* or *positively_tested* (as a transition from *healthy* can be due to an effective contact only, and *healed* and *isolated* are terminating states).

- If $\alpha(a)$ is infected (line 7), then a becomes infectious (line 8). If the symptomaticity state $\beta(a)$ of a is `developing_symptoms`, then the time $\gamma(a)$ to the next infectiousness state is set to ∞ , as no transition is possible from this state until when a becomes symptomatic, and the time $\delta(a)$ to the next symptomaticity state is reduced by c (lines 10–11); otherwise, if $\beta(a)$ is `symptomatic` or `asymptomatic`, then $\gamma(a)$ is sampled from the PDF of the time from symptoms onset to isolation (lines 12–13), i.e., mid (green) curve in Fig. 3(c), or from the PDF of the time from infectiousness to healing (lines 14–15), i.e., mid (green) curve in Fig. 3(d), respectively.
- If $\alpha(a)$ is infectious (line 16), then it becomes either isolated, if $\beta(a)$ is symptomatic (lines 17–18), or healed, if $\beta(a)$ is asymptomatic (lines 19–20). In both cases, the time $\gamma(a)$ to the next infectiousness state becomes ∞ , as healed and isolated are terminating states (line 21).
- If $\alpha(a)$ is `positively_tested` (line 22), then a becomes isolated (line 23) and, as in the previous case, the time $\gamma(a)$ to the next infectiousness state becomes ∞ (line 24).
- In all three cases, the time to the next infectiousness/symptomaticity state of any subject other than a is reduced by c (line 25) and time is advanced by c (line 26).

Algorithm 1: Ground truth simulation of the evolution of an infection chain

```

Simulation( $\mu, t_{\max}, \omega(t_{\max}), \Delta t, R_{\max}, R_{\text{batch}}, BS$ )
input : metamodel instance  $\mu$ ; time limit  $t_{\max}$ ; observation sequence  $\omega(t_{\max}) := \langle \langle e_i, \tau_i \rangle \rangle_{i=1}^m$  taken within  $[0, t_{\max}]$ ;
time step  $\Delta t$ ; maximum number of simulation runs  $R_{\max}$ , batch size  $R_{\text{batch}}$ , threshold  $BS$  on the Brier score
output: for each subject  $s$ , probability  $P(s, t, \omega(t))$  that  $s$  is infectious not isolated at time  $t$  given  $\omega(t)$ 
1  $S \leftarrow$  number of subjects,  $T \leftarrow t_{\max}/\Delta t + 1$  (number of time points),  $\text{counters} \leftarrow [0]_{S \times T}$ 
2 do
3    $\text{run} \leftarrow 1$  (current run),  $t \leftarrow 0$  (current time),  $i \leftarrow 1$  (current observation index),  $\alpha \leftarrow [\text{healthy}]_{S \times 1}$  (current infectiousness states),
    $\beta \leftarrow [N/A]_{S \times 1}$  (current symptomaticity states),  $\gamma, \delta \leftarrow [\infty]_{S \times 1}$  (current times to the next infectiousness/symptomaticity states)
4   do
5      $a \leftarrow \arg \min_s \{\gamma(s)\}$ ,  $b \leftarrow \arg \min_s \{\delta(s)\}$ ,  $c \leftarrow \gamma(a)$ ,  $d \leftarrow \delta(b)$ ,  $\text{counters}' \leftarrow [0]_{S \times T}$ 
6     if  $c \leq d \wedge c \leq \tau_i - t$  then
7       if  $\alpha(a)$  is infected then
8          $\alpha(a) \leftarrow$  infectious
9         if  $\beta(a)$  is developing_symptoms then
10           $\gamma(a) \leftarrow \infty$ 
11          reduce  $\delta(a)$  by  $c$ 
12          if  $\beta(a)$  is symptomatic then
13             $\gamma(a) \leftarrow$  sample of the PDF of the time from symptoms onset to isolation
14            if  $\beta(a)$  is asymptomatic then
15               $\gamma(a) \leftarrow$  sample of the PDF of the time from infectiousness to healing in asymptomatic subjects
16          if  $\alpha(a)$  is infectious then
17            if  $\beta(a)$  is symptomatic then
18               $\alpha(a) \leftarrow$  isolated
19            if  $\beta(a)$  is asymptomatic then
20               $\alpha(a) \leftarrow$  healed
21             $\gamma(a) \leftarrow \infty$ 
22          if  $\alpha(a)$  is positively_tested then
23             $\alpha(a) \leftarrow$  isolated
24             $\gamma(a) \leftarrow \infty$ 
25          reduce the time to the next infectiousness/symptomaticity state of any subject  $\neq a$  by  $c$ 
26           $t' \leftarrow t + c$ 
27       else
28         if  $d \leq c \wedge d \leq \tau_i - t$  then
29            $\beta(b) \leftarrow$  symptomatic
30            $\delta(b) \leftarrow \infty$ 
31           if  $\alpha(b)$  is infected then
32             reduce  $\gamma(b)$  by  $d'$ 
33           if  $\alpha(b)$  is infectious then
34              $\gamma(b) \leftarrow$  sample of the PDF of the time from symptoms onset to isolation
35             reduce the time to the next infectiousness/symptomaticity state of any subject  $\neq b$  by  $d$ 
36              $t' \leftarrow t + d$ 
37         else
38           if  $e_i$  is a contact with an infectious or positively_tested subject then
39             foreach healthy subject  $s$  involved in the contact do
40               if  $\text{rand}(0, 1) \leq \text{risk factor } r(e_i)$  then
41                  $\alpha(s) \leftarrow$  infected
42                  $\gamma(s) \leftarrow$  sample of the PDF of the time from infection to infectiousness
43                 if  $\text{rand}(0, 1) \leq \text{probability of developing symptoms}$  then
44                    $\beta(s) \leftarrow$  developing_symptoms
45                    $\delta(s) \leftarrow$  sample of the PDF of the time from infection to symptoms onset
46                 else
47                    $\beta(s) \leftarrow$  asymptomatic
48             if  $e_i$  is the positive outcome of a test of a subject  $s$  then
49                $\alpha(s) \leftarrow$  positively_tested
50                $\gamma(s) \leftarrow$  sample of the PDF of the time from positive test to isolation
51                $\beta(s) \leftarrow N/A$ 
52                $\delta(s) \leftarrow \infty$ 
53             reduce the time to the next (infectiousness/symptomaticity) state of any subject by  $\tau_i - t$ 
54              $t' \leftarrow \tau_i$ 
55              $i \leftarrow i + 1$ 
56           foreach infectious subject  $s$  in  $[t, t']$  do
57              $\text{counters}'(s, z) \leftarrow 1 \ \forall z \in [t/\Delta t + 1, t'/\Delta t + 1]$ 
58            $t \leftarrow t'$ 
59   while  $t \leq t_{\max} \wedge i \leq m$ ;
60    $\text{run} \leftarrow \text{run} + 1$ 
61    $\text{counters} \leftarrow \text{counters} + \text{counters}'$ 
62   if  $\text{run} \% R_{\text{batch}} == 0$  then
63     foreach subject  $s$ , time point  $t$  do
64        $\text{previous\_probs}(s, t) \leftarrow \text{probs}(s, t)$ 
65        $\text{probs}(s, t) \leftarrow \text{counters}(s, t) / \text{run}$ 
66   while  $\text{run} < R_{\max} \wedge (\text{run} \% R_{\text{batch}} == 0 \wedge \text{Brier score of } \text{probs and previous\_probs} > BS \text{ for some subject})$ ;
67   return  $\text{probs}$ 

```

- If the next event is a transition from the symptomaticity state of a subject b (i.e., $d \leq c \wedge d \leq \tau_i - t$, lines 28–36), then such state is `developing_symptoms` and becomes `symptomatic` (line 29). As no transition is possible from this state, the time $\delta(b)$ to the next symptomaticity state is set to ∞ (line 30). If the infectiousness state $\alpha(b)$ of b is `infected`, then the time $\gamma(b)$ to the next infectiousness state is reduced by d (lines 31–32); otherwise, if $\alpha(b)$ is `infectious`, then $\gamma(b)$ is sampled from the PDF of the time from symptoms onset to isolation (lines 33–34), i.e., mid (green) curve in Fig. 3(c). In both cases, the time to the next infectiousness/symptomaticity state of any subject other than b is reduced by d (line 35) and time is advanced by we advance time by d (line 36).
- If the next event is an observation $\langle \epsilon_i, \tau_i \rangle$ (i.e., $\tau_i - t \leq c \wedge \tau_i - t \leq d$, lines 38–55), then we distinguish the observation type. If the observation is a contact ϵ_i involving an `infectious` or `positively_tested` subject (lines 38–47), then each involved healthy subject s may become infected with a probability equal to the risk factor $r(\epsilon_i)$ of the contact (lines 39–40), annotated within the observation (as specified by the metamodel of Fig. 1). In this case, the infectiousness state of subject s becomes `infected` (line 41), and the time $\gamma(s)$ to the next infectiousness is sampled from the PDF of the time from infection to infectiousness (line 42), i.e., mid (green) curve in Fig. 3(a); moreover, with probability equal to the probability that an infected subject develops symptoms, subject s becomes `symptomatic` and the time $\delta(s)$ to the next symptomaticity state is sampled from the PDF of the time from infection to symptoms onset (lines 43–45); otherwise, subject s becomes `asymptomatic` (lines 46–47).
Otherwise, if the observation is the positive outcome of a test of a subject s (lines 48–52), then the infectiousness state $\alpha(s)$ of s becomes `positively_tested` (line 49), the time $\gamma(s)$ to the next infectiousness state is sampled from the PDF of the time from positive test to isolation (line 50), i.e., mid (green) curve in Fig. 3(c), the symptomaticity state $\beta(s)$ is set to `N/A` (line 51), and $\delta(s)$ is set to ∞ (line 52) as the symptomaticity state is not relevant given the positive test result.
In both cases, the time to the infectiousness/symptomaticity state of any subject is reduced by the time $\tau_i - t$ elapsed since the last event (line 53), time is advanced by $\tau_i - t$ (line 54), and the index of the next observation becomes $i + 1$ (line 55).

Next, we take note of which subjects are `infectious` and not `isolated` within the time interval $[t, t']$ between the last two events, by setting the corresponding entries of a matrix *counters'* equal to 1 (lines 56–57). At the end of each simulation run (line 60), this matrix is added to a matrix *counters* (line 61).

When the number of runs is a multiple of the batch size R_{batch} , then, for each subject s , we obtain the probability $P(s; t; \omega(t))$ that s is `infectious` not `isolated` at time t given the observation sequence $\omega(t)$, by dividing each entry of matrix *counters* by the number of executed runs (lines 62–65). Finally, simulation ends (line 65) when the number of executed runs is equal to R_{max} or when, for each subject s , the Brier score of the probabilities $P(s, t, \omega(t))$ computed at the end of the last two batches is lower than the threshold BS .

For each data set considered in our experiments, we derive a ground truth by performing Algorithm 1 with maximum number of runs $R_{\text{max}} = 100\,000$, batch size $R_{\text{batch}} = 1000$, and threshold $BS = 10^{-8}$.

Data availability

We have provided a replication package, which is publicly available open source under the AGPLv3 licence at <https://doi.org/10.5281/zenodo.21559792>.

References

- [1] P. Saunders-Hastings, J.A. Crispo, L. Sikora, D. Krewski, Effectiveness of personal protective measures in reducing pandemic influenza transmission: A systematic review and meta-analysis, *Epidemics* 20 (2017) 1–20.
- [2] K. Meier, T. Glatz, M.C. Guijt, M. Piccininni, M. van der Meulen, K. Atmar, A.-T.C. Jolink, T. Kurth, J.L. Rohmann, A.H. Zamanipoor Najafabadi, et al., Public perspectives on protective measures during the COVID-19 pandemic in the Netherlands, Germany and Italy: A survey study, *PLoS One* 15 (8) (2020) e0236917.
- [3] L. Cirrincione, F. Plescia, C. Ledda, V. Rapisarda, D. Martorana, R.E. Moldovan, K. Theodoridou, E. Cannizzaro, COVID-19 pandemic: Prevention and protection measures to be adopted at the workplace, *Sustainability* 12 (9) (2020) 3603.
- [4] S. Vandepitte, T. Alleman, I. Nopens, J. Baetens, S. Coenen, D. De Smedt, Cost-effectiveness of COVID-19 policy measures: a systematic review, *Value Health* 24 (11) (2021) 1551–1569.
- [5] B. González López-Valcárcel, L. Vallejo-Torres, The costs of COVID-19 and the cost-effectiveness of testing, *Appl. Econ. Anal.* 29 (85) (2021) 77–89.
- [6] T. Britton, Stochastic epidemic models: a survey, *Math. Biosci.* 225 (1) (2010) 24–35.
- [7] P.H. Guzzi, F. Petrizzi, T. Mazza, Disease spreading modeling and analysis: A survey, *Brief. Bioinform.* 23 (4) (2022) bbac230.
- [8] A.B. Bamana, M.S. Kamalabad, D.L. Oberski, A systematic literature review of time series methods applied to epidemic prediction, *Inform. Med. Unlocked* 50 (2024) 101571.
- [9] F. Brauer, Compartmental models in epidemiology, in: *Mathematical Epidemiology*, Springer, 2008, pp. 19–79.
- [10] J. Tolles, T. Luong, Modeling epidemics with compartmental models, *Jama* 323 (24) (2020) 2515–2516.
- [11] M.J. Keeling, K.T. Eames, Networks and epidemic models, *J. Royal Soc. Interface* 2 (4) (2005) 295–307, <http://dx.doi.org/10.1098/rsif.2005.0051>, arXiv:<https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2005.0051>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2005.0051>.
- [12] S. Greenland, Bayesian perspectives for epidemiological research: I. Foundations and basic methods, *Int. J. Epidemiol.* 35 (3) (2006) 765–775.
- [13] P.H. Borghi, O. Zakordonets, J.P. Teixeira, A COVID-19 time series forecasting model based on MLP ANN, *Procedia Comput. Sci.* 181 (2021) 940–947.
- [14] Z.E. Rasjid, R. Setiawan, A. Effendi, A comparison: prediction of death and infected COVID-19 cases in Indonesia using time series smoothing and LSTM neural network, *Procedia Comput. Sci.* 179 (2021) 982–988.
- [15] D.B.N. Assad, J. Cara, M. Ortega-Mier, Comparing short-term univariate and multivariate time-series forecasting models in infectious disease outbreak, *Bull. Math. Biol.* 85 (1) (2022) 9.

- [16] Y.-C. Jin, Q. Cao, Q. Sun, Y. Lin, D.-M. Liu, C.-X. Wang, X.-L. Wang, X.-Y. Wang, et al., Models for COVID-19 data prediction based on improved LSTM-ARIMA algorithms, *IEEE Access* 12 (2023) 3981–3991.
- [17] M.U. Kraemer, J.L.-H. Tsui, S.Y. Chang, S. Lytras, M.P. Khurana, S. Vanderslott, S. Bajaj, N. Scheidwasser, J.L. Curran-Sebastian, E. Semenova, et al., Artificial intelligence for modelling infectious disease epidemics, *Nature* 638 (8051) (2025) 623–635.
- [18] S. Pernice, P. Castagno, L. Marcotulli, M.M. Maule, L. Richiardi, G. Moirano, M. Sereno, F. Cordero, M. Beccuti, Impacts of reopening strategies for COVID-19 epidemic: a modeling study in Piedmont region, *BMC Infect. Dis.* 20 (1) (2020) 798.
- [19] H. Cintra, F. Fontinele, Estimative of real number of infections by COVID-19 in Brazil and possible scenarios, *Infect. Dis. Model.* 5 (2020) 720–736, <http://dx.doi.org/10.1016/j.idm.2020.09.004>, URL: <https://www.sciencedirect.com/science/article/pii/S2468042720300506>.
- [20] C. Eslahchi, F. Movahedi, Calculation of transition probabilities in the birth and death Markov process in the epidemic model, *Math. Comput. Modelling* 55 (3–4) (2012) 810–815.
- [21] M. Herajy, F. Liu, C. Rohr, M. Heiner, Coloured hybrid Petri nets: An adaptable modelling approach for multi-scale biological networks, *Comput. Biol. Chem.* 76 (2018) 87–100.
- [22] N. Bahi-Jaber, D. Pontier, Modeling transmission of directly transmitted infectious diseases using colored stochastic Petri nets, *Math. Biosci.* 185 (1) (2003) 1–13.
- [23] B. Reyné, N. Saby, M.T. Sofonea, Principles of mathematical epidemiology and compartmental modelling application to COVID-19, *Anaesth. Crit. Care Pain Med.* 41 (1) (2021) 101017.
- [24] L. Danon, A.P. Ford, T. House, C.P. Jewell, M.J. Keeling, G.O. Roberts, J.V. Ross, M.C. Vernon, Networks and the epidemiology of infectious disease, *Interdiscip. Perspect. Infect. Dis.* 2011 (1) (2011) 284909, <http://dx.doi.org/10.1155/2011/284909>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1155/2011/284909>, URL: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2011/284909>.
- [25] J. Bodlaj, V. Batagelj, Hierarchical link clustering algorithm in networks, *Phys. Rev. E* 91 (2015) 062814, <http://dx.doi.org/10.1103/PhysRevE.91.062814>, URL: <https://link.aps.org/doi/10.1103/PhysRevE.91.062814>.
- [26] J. Yang, The spreading of infectious diseases with recurrent mobility of community population, *Phys. A* 541 (2020) 123316.
- [27] C.C. Kerr, R.M. Stuart, D. Mistry, R.G. Abeysuriya, K. Rosenfeld, G.R. Hart, R.C. Núñez, J.A. Cohen, P. Selvaraj, B. Hagedorn, et al., Covasim: an agent-based model of COVID-19 dynamics and interventions, *PLoS Comput. Biol.* 17 (7) (2021) e1009149.
- [28] D.L. Chao, M.E. Halloran, V.J. Obenchain, I.M. Longini Jr., FluTE, a publicly available stochastic influenza epidemic simulation model, *PLoS Comput. Biol.* 6 (1) (2010) e1000656.
- [29] E. Hunter, B. Mac Namee, J.D. Kelleher, A taxonomy for agent-based models in human infectious disease epidemiology, *J. Artif. Soc. Soc. Simul.* 20 (3) (2017).
- [30] C.P. Jewell, T. Kypraios, P. Neal, G.O. Roberts, Bayesian analysis for emerging infectious diseases, 2009.
- [31] X. Didelot, J. Gardy, C. Colijn, Bayesian inference of infectious disease transmission from whole-genome sequence data, *Mol. Biol. Evol.* 31 (7) (2014) 1869–1879.
- [32] M. Hall, M. Woolhouse, A. Rambaut, Epidemic reconstruction in a phylogenetics framework: transmission trees as partitions of the node set, *PLoS Comput. Biol.* 11 (12) (2015) e1004613.
- [33] F. Campbell, A. Cori, N. Ferguson, T. Jombart, Bayesian inference of transmission chains using timing of symptoms, pathogen genomes and contact data, *PLoS Comput. Biol.* 15 (3) (2019) e1006930.
- [34] M.J. Morelli, G. Thébaud, J. Chadeuf, D.P. King, D.T. Haydon, S. Soubeyrand, A Bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data, *PLoS Comput. Biol.* 8 (11) (2012) e1002768.
- [35] E. Amparore, M. Beccuti, P. Castagno, S. Pernice, G. Franceschinis, M. Pennisi, From compositional Petri net modeling to macro and micro simulation by means of stochastic simulation and agent-based models, *ACM Trans. Model. Perform. Eval. Comput. Syst.* 9 (1) (2023) 1–30.
- [36] SIRIO Library, 2023, <https://github.com/oris-tool/sirio>.
- [37] M. Paolieri, M. Biagi, L. Carnevali, E. Vicario, The ORIS tool: quantitative evaluation of non-Markovian systems, *IEEE Trans. Software Engineering* 47 (6) (2021) 1211–1225.
- [38] E. Vicario, L. Sassoli, L. Carnevali, Using stochastic state classes in quantitative evaluation of dense-time reactive systems, *IEEE Trans. Software Engineering* 35 (5) (2009) 703–719.
- [39] A. Horváth, M. Paolieri, L. Ridi, E. Vicario, Transient analysis of non-Markovian models using stochastic state classes, *Elsevier Performance Evaluation* 69 (7–8) (2012) 315–335, <http://dx.doi.org/10.1016/j.peva.2011.11.002>.
- [40] M. Biagi, L. Carnevali, M. Paolieri, F. Patara, E. Vicario, A continuous-time model-based approach for activity recognition in pervasive environments, *IEEE Trans. Hum. Mach. Syst.* 49 (4) (2019) 293–303.
- [41] W. Whitt, Approximating a point process by a renewal process, I: Two basic methods, *Oper. Res.* 30 (1) (1982) 125–147.
- [42] S. Zhao, B. Tang, S.S. Musa, S. Ma, J. Zhang, M. Zeng, Q. Yun, W. Guo, Y. Zheng, Z. Yang, et al., Estimating the generation interval and inferring the latent period of COVID-19 from the contact tracing data, *Epidemics* 36 (2021) 100482.
- [43] L.C. Tindale, J.E. Stockdale, M. Coombe, E.S. Garlock, W.Y.V. Lau, M. Saraswat, L. Zhang, D. Chen, J. Wallinga, C. Colijn, Evidence for transmission of COVID-19 prior to symptom onset, *Elife* 9 (2020) e57149.
- [44] X. He, E.H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y.C. Lau, J.Y. Wong, Y. Guan, X. Tan, et al., Temporal dynamics in viral shedding and transmissibility of COVID-19, *Nature Med.* 26 (5) (2020) 672–675.
- [45] D. Kong, Y. Zheng, H. Wu, H. Pan, A.L. Wagner, Y. Zheng, X. Gong, Y. Zhu, B. Jin, W. Xiao, et al., Pre-symptomatic transmission of novel coronavirus in community settings, *Influ. Other Respir. Viruses* 14 (6) (2020) 610–614.
- [46] J. Hellewell, S. Abbott, A. Gimma, N.I. Bosse, C.I. Jarvis, T.W. Russell, J.D. Munday, A.J. Kucharski, W.J. Edmunds, F. Sun, et al., Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts, *Lancet Glob. Health* 8 (4) (2020) e488–e496.
- [47] A.W. Byrne, D. McEvoy, A.B. Collins, K. Hunt, M. Casey, A. Barber, F. Butler, J. Griffin, E.A. Lane, C. McAloon, et al., Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases, *BMJ Open* 10 (8) (2020) e039856.
- [48] S.M. Kissler, J.R. Fauver, C. Mack, S.W. Olesen, C. Tai, K.Y. Shiue, C.C. Kalinich, S. Jednak, I.M. Ott, C.B. Vogels, et al., Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies, *PLoS Biol.* 19 (7) (2021) e3001333.
- [49] P. Sah, M.C. Fitzpatrick, C.F. Zimmer, E. Abdollahi, L. Juden-Kelly, S.M. Moghadas, B.H. Singer, A.P. Galvani, Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis, *Proc. Natl. Acad. Sci.* 118 (34) (2021) e2109229118.
- [50] H. Vogt, Applied Epidemiology of Infectious Diseases in Western Australia (Master's thesis), The Australian National University, (Australia), 2021.
- [51] M.E. Kretzschmar, G. Rozhnova, M.C. Bootsma, M. Van Boven, J.H. van de Wijgert, M.J. Bonten, Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study, *Lancet Public Health* 5 (8) (2020) e452–e459.
- [52] X. Wan, W. Wang, J. Liu, T. Tong, Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range, *BMC Med. Res. Methodol.* 14 (1) (2014) 135.
- [53] C. Menni, A.M. Valdes, L. Polidori, M. Antonelli, S. Penamakuri, A. Nogal, P. Louca, A. May, J.C. Figueiredo, C. Hu, et al., Symptom prevalence, duration, and risk of hospital admission in individuals infected with SARS-CoV-2 during periods of omicron and delta variant dominance: a prospective observational study from the ZOE COVID Study, *Lancet* 399 (10335) (2022) 1618–1624.

- [54] P.S. Wikramaratna, R.S. Paton, M. Ghafari, J. Lourenço, Estimating the false-negative test probability of SARS-CoV-2 by RT-PCR, *Eurosurveillance* 25 (50) (2020) 2000568.
- [55] P. Runeson, M. Höst, Guidelines for conducting and reporting case study research in software engineering, *Empir. Softw. Eng.* 14 (2) (2009) 131–164.
- [56] P. Reinecke, T. Krauß, K. Wolter, Phase-type fitting using HyperStar, in: *European Workshop on Computer Performance Engineering*, Springer, 2013, pp. 164–175.
- [57] A. Horváth, M. Telek, Phfit: A general phase-type fitting tool, in: *Int. Conf. on Modelling Techniques and Tools for Computer Performance Evaluation*, Springer, 2002, pp. 82–91.
- [58] J.Y. Chun, G. Baek, Y. Kim, Transmission onset distribution of COVID-19, *Int. J. Infect. Dis.* 99 (2020) 403–407.
- [59] S. Fantaw, D.D. Debeke, Time to recovery and determinant factors of COVID-19 patients under treatment in Sidama region, Ethiopia: a retrospective cohort study, *Heliyon* 10 (1) (2024).