



6-17-2021

The Impact of Contact Tracing on the Spread of COVID-19: An Egocentric Agent-Based Model

Andrew Pilny

University of Kentucky, andy.pilny@uky.edu

Lin Xiang

University of Kentucky, lin.xiang@uky.edu

Corey Huber

University of Kentucky, Joseph.Huber@uky.edu

Will Silberman

University of Kentucky, wrsi227@g.uky.edu

Sean Goatley-Soan

University of Kentucky, sgoatleysoan@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/comm_facpub



Part of the [Communication Commons](#)

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Repository Citation

Pilny, Andrew; Xiang, Lin; Huber, Corey; Silberman, Will; and Goatley-Soan, Sean, "The Impact of Contact Tracing on the Spread of COVID-19: An Egocentric Agent-Based Model" (2021). *Communication Faculty Publications*. 25.

https://uknowledge.uky.edu/comm_facpub/25

This Article is brought to you for free and open access by the Communication at UKnowledge. It has been accepted for inclusion in Communication Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

The Impact of Contact Tracing on the Spread of COVID-19: An Egocentric Agent-Based Model

Digital Object Identifier (DOI)

<https://doi.org/10.21307/connections-2021.022>

Notes/Citation Information

Published in *Connections*, v. 41, issue 1.

© 2021 Authors

This work is licensed under the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).

The impact of contact tracing on the spread of COVID-19: an egocentric agent-based model

Andrew Pilny^{1,*}, Lin Xiang²,
Corey Huber¹, Will Silberman¹
and Sean Goatley-Soan¹

¹Department of Communication,
University of Kentucky, Lexington,
KY.

²Department of STEM Education,
University of Kentucky, Lexington,
KY.

*E-mail: andy.pilny@uky.edu

Abstract

At its core, contact tracing is a form of egocentric network analysis (ENA). One of the biggest obstacles for ENA is informant accuracy (i.e., amount of true contacts identified), which is even more prominent for interaction-based network ties because they often represent episodic relational events, rather than enduring relational states. This research examines the effect of informant accuracy on the spread of COVID-19 through an egocentric, agent-based model. Overall when the average person transmits COVID-19 to 1.62 other people (i.e., the R_0), they must be, on average, 75% accurate with naming their contacts. In higher transmission contexts (i.e., transmitting to at least two other people), the results show that multi-level tracing (i.e., contact tracing the contacts) is the only viable strategy. Finally, sensitivity analysis shows that the effectiveness of contact tracing is negatively impacted by the timing and overall percent of asymptomatic cases. Overall, the results suggest that if contact tracing is to be effective, it must be fast, accurate, and accompanied by other interventions like mask-wearing to drive down the average R_0 .

Keywords

Egocentric network analysis, Agent-based model, COVID-19, Contagion.

Issues regarding the reliability and validity of relational data have been a long concern for social network researchers (Perry et al., 2018), especially when such data are supposed to represent observable behavior (i.e., A interacted with B this week). And although there has been considerable research demonstrating various biases and cognitive limitations participants have when reporting their network ties (see, Smith et al., 2020, for a review), less work has demonstrated some of the more applied consequences of such issues. In other words, in what applied contexts should researchers take a deeper look at the outcomes of informant accuracy?

One such applicable area where the reliability and validity of relational data are important is the practice of contact tracing. Contact tracing is a strategy used to help contain various infectious diseases that

spread through interpersonal contact. Put simply, contact tracing is the process of retrospectively identifying persons who may have made interpersonal contact with a confirmed infectious individual (Eames and Keeling, 2003). The logic of contact tracing is straightforward: identify contacts who have had interactions with an infected patient and remove them from the social system to (hopefully) prohibit the further spread of the disease. Such a method has been credited as effective because it allows for the testing of at-risk contacts and this method also allows for the identifying of hotspots and clusters of diseases (Klinkenberg et al., 2006).

At its core, contact tracing is simply a form of egocentric network analysis (ENA). ENA is the study of individuals (i.e., egos) and people in the ego's immediate social environment (i.e., alters; Perry et al.,

2018). Conducting a reliable and valid ENA study is challenging, and there is a rich academic/applied literature dedicated to improving the rigor behind these efforts (Crossley et al., 2015; McCarty et al., 2019; Perry et al., 2018). However, there has been scant research determining how varying levels of reliability and validity might influence how effective contact tracing can be at containing infectious diseases.

This research is organized as follows. First, we review the literature behind contact tracing and our key factor of interest: informant accuracy. Second, we set up the details behind the egocentric agent-based model, examining the effect of informant accuracy and multi-level tracing on the spread of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the biologic strain of coronavirus that causes the illness known as COVID-19. Third, we conduct a sensitivity analysis to examine the impact of the timing of the tracing and the percent of asymptomatic cases in the population. Finally, we discuss the results in terms of the wider literature on contact tracing and ENA.

A brief review of contact tracing

Contact tracing is a well-known type of social network intervention for understanding how infectious diseases spread (Valente, 2010). Indeed, rudimentary versions were even used during waves of the bubonic plague in the sixteenth century (Cohn and O'Brien, 2020). Contact tracing has been employed in infectious diseases caused by various pathogens (e.g., STIs, Macke and Maher, 1999; SARS: Donnelly et al., 2003; TB: Mandalakas et al., 2017). Consequently, it is considered efficient at reducing the prevalence of infections, especially when dealing with an isolated number of cases (e.g., STIs) or toward novel forms of viruses (Eames et al., 2010).

Although contact tracing is generally considered a useful means of gathering potential transmission data, it does not always suffice, by itself, as a control measure to contain a given epidemic (Eames et al., 2010). Moreover, the effectiveness of contact tracing is dependent on the transmission dynamics of the outbreak (Klinkenberg et al., 2006) and on the timing of the tracing itself (Kretzschmar et al., 2020). For example, Cheng et al. (2020) conducted a contact tracing assessment toward COVID-19. They found that, due to the high transmission rate before and near an individual's symptom onset, contact tracing would be inadequate on its own. Instead, contact tracing should be at least implemented with other interventional strategies, such as social distancing

and mask-wearing. Likewise, a recent stochastic transmission model by Hellewell et al. (2020) sought to investigate the potential efficacy of contact tracing and isolation of cases toward COVID-19. They did so through a variety of simulated outbreaks and ranging their R_0 's and transmission before symptom onset percentage. They found that contact tracing probability (i.e., odds that the tracing happens) must be high (i.e., 80%+) to control COVID-19 transmission.

Additionally, agent-based models (ABMs) have been used to simulate contact tracing data. Kucharski et al. (2020a, b) employed ABM to simulate a variety of scenarios involving responses to COVID-19 (e.g. no control measures, self-isolation away from the household, quarantining, self-isolation in the household). The authors found that a combined approach of isolating symptomatic cases and contact tracing the contacts of positive cases reduced the spread of COVID-19 when compared to individually implementing measures. Furthermore, the authors note that in an instance where asymptomatic cases were high, many contacts would need to be traced and tested to consider transmission at a higher network level (i.e., transmission from secondhand contact).

At the end of the day, contact tracing is theoretically useful for understanding contagion dynamics and practically useful for mitigating the spread of infectious diseases. Additionally, as Eames et al. (2010) point out, collecting contact tracing data is not an easy feat. It requires significant material resources to compensate tracers and it requires informational resources to train tracers to overcome some of the many obstacles to collect reliable and valid contact network data. Overcoming these obstacles is important because the effectiveness of contact tracing is only as good as the data collected.

Informant accuracy: how do people remember their contacts?

Because contact tracing is a form of ENA, it is prone to a host of issues regarding reliability and validity (for a review, see Perry et al., 2018). For contact tracing, perhaps the most sensitive issue for data collection is informant accuracy, the difference between an ego's perceptions of their network and their actual network (Bernard et al., 1984). When researchers compare and contrast between perceived and actual networks, these analyses are commonly known as accuracy studies, perhaps most notably popularized by Bernard, Kilworth, and Sailer, otherwise known as the BKS studies (for a brief review, see Bernard et al., 1981). The general theme of the BKS studies is that individuals' self-reports of their network

behaviors do not align very well with their actual behaviors. Moreover, the network structures of self-reported and actual behavioral networks differ from one another as well (Quintane, 2012). Although self-reported data on perceived network contacts may be useful as cognitive social structures in its pure form (Krackhardt, 1987), for contact tracing, it is clear that the minimization of the error between the two is a central concern. Indeed, since accumulation of a handful of accuracy studies, there has been debate, research, and theoretical moves to try and figure out why there is such a discrepancy between perceived and actual contacts (e.g., Corman and Scott, 1994; Pilny et al., 2017), how to mitigate some of these biases in data collection (Kogovšek et al., 2002), and how to account for informant (in) accuracy in inferential network analyses (Butts, 2003).

Nevertheless, it is important to understand why individuals have trouble recalling network ties in the first place. For starters, humans do not store information like network contacts in a vacuum, they are primarily organized through cognitive schemas that cluster information based on how ties are related to one another (McCarty et al., 2019). For instance, Brashears and Quintane (2015) found that alters tend to be better remembered in terms of common group membership (e.g., role relations like family, work, neighbors) and ‘chunked’ structures like triads. Moreover, contacts are better remembered if they are repeated and represent long-term, consistent interactions (Freeman et al., 1987). Likewise, contacts that are not very popular themselves (e.g., less central) are often difficult to remember than popular ones (Marin, 2004). Perhaps more importantly for contact tracing, individuals whom the ego feels less close and interacts with less frequently are more prone to recall issues (Brewer, 2000). The key implication here is that there are certain network ties that are more likely to be forgotten.

For extracting contact tracing data, all of the above issues may be present, but may be even further complicated by the fact that such data represents the extraction of relational events, not relational states. A relational event can be defined as a ‘discrete event generated by a social actor and directed toward one or more targets’ (Butts, 2008, p. 159), while relational states can be viewed as ‘continuously persistent relationships between nodes’ (Borgatti et al., 2013, p. 3). The transmission of COVID-19 does not require an established relational state between dyads, a simple relational event will suffice.

The key difference is the nature of the tie: relational events are episodic, while relational states are more enduring. For instance, some psychological theories

of recall generally find that some information (e.g., interpersonal contacts) are better remembered if they are encoded in a meaningful way. For instance, if a relational event occurs with somebody where there is no meaningful relational state (e.g., friendship, work relationship) or other cue of information, such an event may be more difficult to recall because the episodic event may be harder to encode in an elaborative (i.e., meaningful) fashion (e.g., Craik and Lockhart, 1972). Relational events divorced from relational states or other helpful memory schemas like elevator conversations, interactions with waiters/bartenders, fellow parent at the park, etc. might lend themselves more prone to recall problems.

Moreover, individual-level differences can also exacerbate similar informant accuracy problems. In other words, it could be characteristics about the ego, rather than the alter, that can influence informant accuracy. A substantial amount of research has demonstrated that factors like gender (Brashears et al., 2016), age (Hsieh, 2014), mood (Hlebec and Ferligoj, 2001), and occupation (Marineau et al., 2018) can influence network recall as well.

As such, previous ENA research has demonstrated that informant accuracy problems exist and influenced by characteristics of both the ego and the alter. Poised with this phenomenon, we consider how informant accuracy impacts the spread of COVID-19 when contact tracing is implemented. Important questions like ‘how accurate does the contact tracing need to be to have any effect?’ and ‘at what point do we start to see diminishing returns?’ can be answered by analyzing the relationship between informant accuracy and viral spread. Moreover, contact tracing requires participants to engage in the unpleasant task of going into quarantine. Here, there is also a challenge to be economical and try to require as few as people as possible to go under quarantine. As such, we ask:

R1: What is the relationship between informant accuracy during contact tracing and (a) the spread of COVID-19 and (b) number of quarantines?

Multi-level tracing using snowball sampling

In addition to the effect of informant accuracy, it is important to ask if there are any probable and useful alternative strategies for collecting contact tracing data. Because contact tracing is a respondent-driven technique, we consider how a snowball sample may be a useful strategy for finding potential infectious contacts (Johnson, 1990). In the context of social

network research, Borgatti et al. (2013) defined snowball sampling as the process of gathering network data ‘on any qualifying actor with a tie to any actor already selected, up to K waves or until quotas or cost limits are reached’ (p. 34). Such a technique has been used in other studies attempting to gain access to at-risk populations. For instance, Kendall et al. (2008) were able to get up to five waves of data for an HIV at-risk population in Brazil.

In the context of COVID-19, a qualified actor is someone who has tested positive for the virus. When the actor is interviewed for contact tracing data, the practitioner will elicit from the actor a set of alters and will inform those alters to get tested and quarantine. This traditional method can be described as first-level contact tracing as it revolves around making use of the first wave of data from the initial infected individual. Second-level contact tracing would simply repeat the process for the set of initial alters, which is to say it would contact trace the contacts. A practitioner would then elicit alters from these contacts and require them to get tested and quarantine for the time being, a strategy used in the early stages of the COVID-19 outbreak in South Korea (Schneider et al., 2020). Third-level contact tracing would repeat the same process for the new alters and so on.

One of the key advantages of snowball sampling is accessing hidden and ‘hard to reach’ populations (Browne, 2005), such as those individuals who are spreading COVID-19 but may not yet know it. In this sense, it allows researchers to work backward to catch up on the viral diffusion process. As such, we ask:

R2: What is the relationship between multi-level tracing and (a) the spread of COVID-19 and (b) the number of quarantines?

Finally, informant accuracy and multi-level tracing may be interdependent on their effectiveness at mitigating the spread of COVID-19. That is, while they may have main effects by themselves, they can be dependent in complex ways. For instance, different levels tracing may have different thresholds of informant accuracy to contain the spread. To our knowledge, there is no empirical work analyzing the relationship between informant accuracy and different levels of tracing. As such, we ask our third research question:

R3: How do informant accuracy and multi-level contact tracing interact to influence the (a) the spread of COVID-19 and (b) the number of quarantines?

Methods

An egocentric agent-based model

Agent-based models (ABMs) are constructed to ‘simulate simultaneously multiple agents, or actors, who behave in ways to that impact one another’ (Larson, 2012, p. 84). They are particularly useful for understanding how simple rules guiding agents’ behavior or other manipulations of input factors can influence the emergence of complex social structures (Corman, 1996). ABMs are very common in public health and epidemiology, where the goal is usually to create models to understand infectious disease dynamics that can help inform policy and responses to epidemics (e.g., Epstein, 2009).

An ‘egocentric’ ABM does not radically depart from the basic mechanics of social simulation because ABMs already typically focus on interactions between agents. However, the current egocentric approach puts special emphasis on the ENA portion as factors that can be manipulated (i.e., informant accuracy and level of tracing). In the current model, egocentric interaction networks are collected during each iteration (i.e., a day), allowing to user to inspect any network-interaction history of any agent.

What follows is a description of the key building blocks of the current ABM, which we call ConTrace (see Fig. 1 for visualization). Following Hammond’s (2015) suggestions for best practices of reporting ABMs, we specify the following: Properties, Actions, Rules, Time, and Environment (PARTE). Documenting the details of any ABM can sometimes be overwhelming. As such, what follows is an abbreviated summary. More fine-grained details can be found in an online appendix¹, including additional robustness checks, model verifications, and the ConTrace model in NetLogo.

Basic overview

The developed ABM of infectious disease tracing and quarantine is based on a slightly modified susceptible-exposed-infectious-removed (SEIR) model with an expansion by adding the procedures of contact tracing and quarantine. The SEIR model, an expansion of SIR model, divided the population into four groups: (i) susceptible, (ii) exposed, (iii) infectious, and (iv) recovered. Individuals in a population could go through all the four phases during an epidemic outbreak.

¹Anonymous link to Appendix: https://osf.io/493n7/?view_only=dc6bbca1533f4341ad188e5fe08b32c1

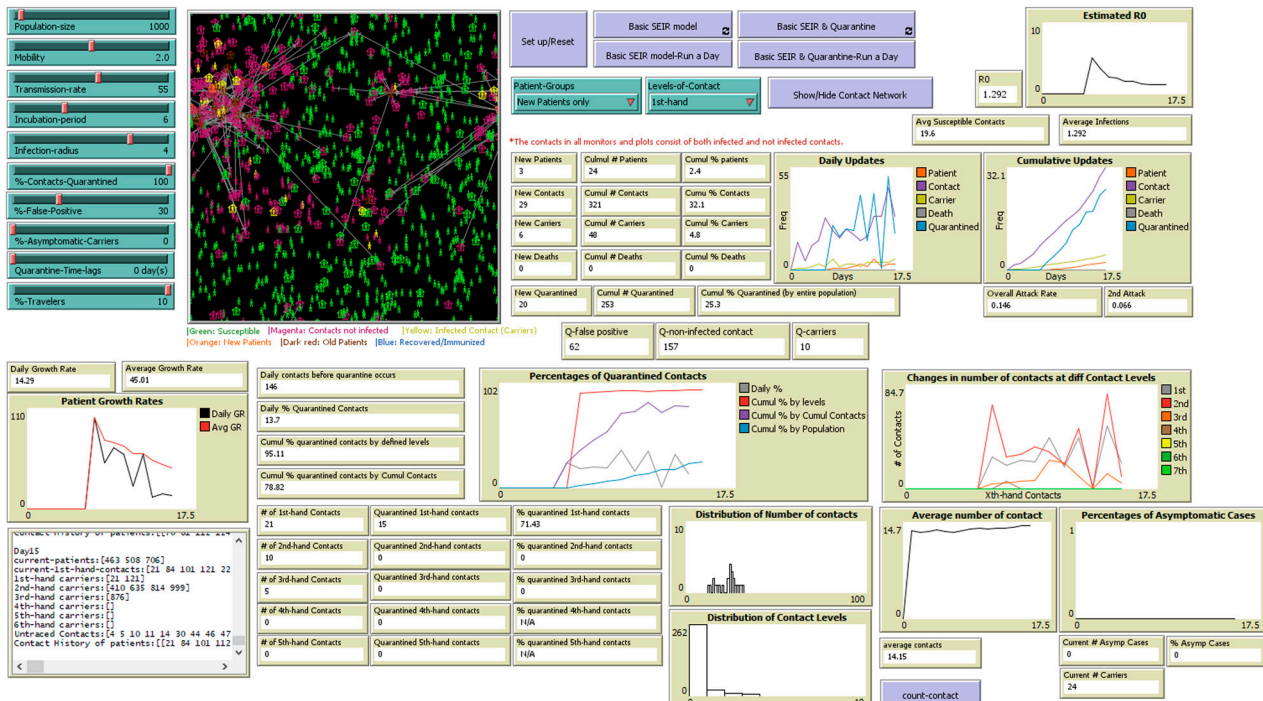


Figure 1: NetLogo interface of ConTrace.

Figure 2 illustrates the structure and flow of the model. Ovals represent human agents and rhombuses represent model decisions. A ‘susceptible’ contact moves into ‘exposed’ when they have made contact with somebody that has COVID-19. Exposed contacts now have the ability to be traced when the previous infectious agent becomes symptomatic, goes to the doctor, and gets contact traced. The odds that the exposed contact must go under quarantine will depend on the level of informant accuracy. If the exposed contact does not get traced, they remain in the system. All the while and depending on the transmission and asymptomatic rate, they may continue to spread COVID-19 if infected, and undergo contact tracing when symptoms develop (i.e., incubation period is over).

Properties

Properties represent the mutable and immutable attributes of agents in the system, which can also be observable and unobservable to other agents. In the present ABM, only one type of agent, people, is used. However, the agents are classified into six groups – (i) susceptible, (ii) non-infected contacts, (iii) presymptomatic contacts, (iv) asymptomatic contacts, (v) patients, and (vi) recovered – based on

four attributes (see Table 1). The classification is made by considering both the traditional grouping in the previous studies and current modeling functions. For example, it is necessary to separate people who have never been exposed to the disease and who have been exposed but not infected. We regard the former as ‘susceptible people’ and the latter as ‘non-infected contacts’. They both remain susceptible to the infectious disease but only the exposed contacts may be traced and quarantined. We define the ‘contact’ state of the ‘patients’ and ‘recovered people’ as false even though they were exposed to the disease. This arbitrary state setting allows us to focus on tracing and quarantining the contacts in the model. The deaths of the disease do not have a category as they disappear from the model.

The agents are created in the setup procedure of the model. When setting up the model, a defined number of agents are created and randomly distributed in the simulation window. All of them are susceptible people except one presymptomatic individual. To help visualize the process, we color code the susceptible people as green, the non-infected contacts as magenta, the presymptomatic and asymptomatic individuals as yellow, the patients as orange and dark red, and the recovered people as blue (see Fig. 3). When an agent is isolated, i.e.,

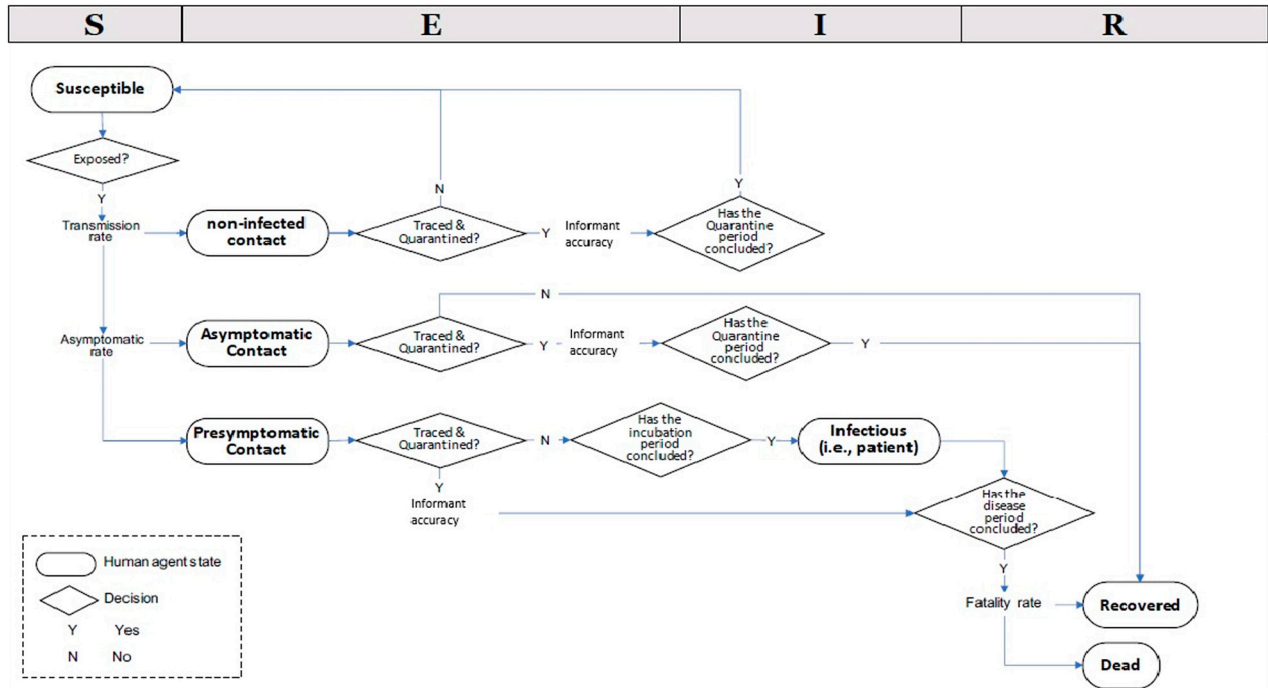


Figure 2: Conceptual flow chart of ConTrace.

a patient, or quarantined, i.e., a contact, its shape changes from a person to a sheltered person.

Actions

The catalog of behaviors that each agent performs within the simulation is actions. There are three basic actions that the agents take in the simulation. The first is mobility, which refers to how agents move around in the environments. Like previous SEIR simulations, we assume agents move a certain distance in a random direction. The second is interaction, which

assumes that agents are interacting with one another in a way that allows for potential transmission of COVID-19. For instance, John Hopkins University defines contacts as sustained interaction within six feet for at least 15 min (Gurley, 2020). Through being mobile and interacting, agents can change the properties of other agents to represent the viral transmission. Consequentially, this has a change on the environment because infected agents will eventually remove themselves from the system and quarantine, where they will either become immune or die.

Table 1. Six groups of agents.

	Susceptible?	Contact?	Infected?	Symptom?
Susceptible	✓	X	X	X
Non-infected contacts	✓	✓	X	X
Presymptomatic and asymptomatic contacts	X	✓	✓	X
Patients	X	X	✓	✓
Recovered	X	X	X	X

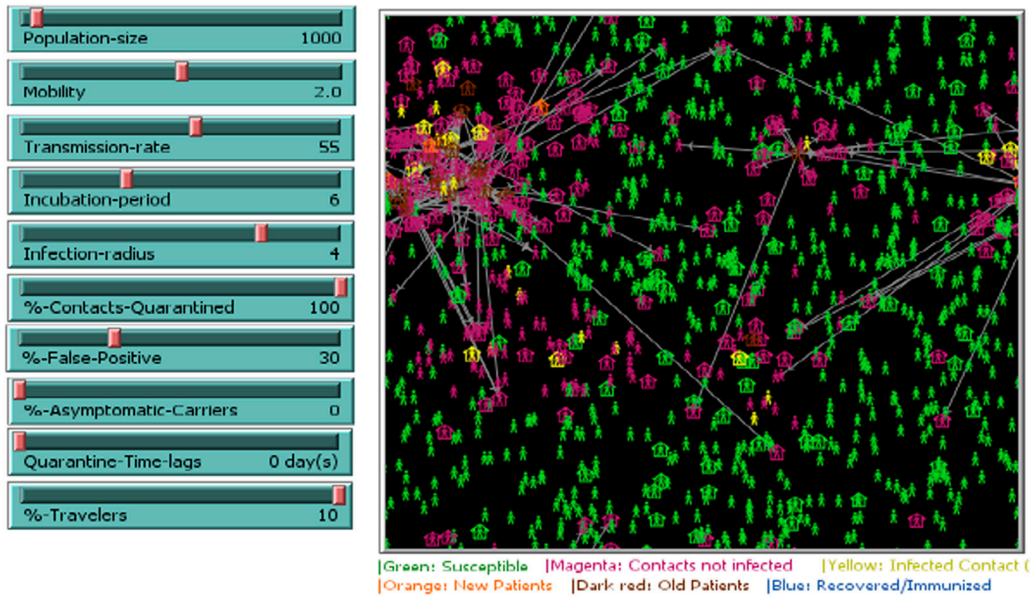


Figure 3: Properties of agents in ConTrace.

Contact network

The resulting contact network is fixed, meaning that interactions do not follow a stochastic model for new selection of ties as in the case of inferential network models (e.g., a stochastic actor-oriented model); the placement of the agents is random. However, because the random placements of objects in a fixed space tends to produce clusters in what is commonly known as the clustering illusion (Gilovich, 1991, pp. 19-20), what emerges is a typical ‘small-world’ world network with an above-average clustering coefficient. Users can also specify a ‘traveler’ percentage to manipulate the amount of agents that travel to random portion of the space, unrestricted by their current mobility settings. The purpose of this function is to introduce potential super-spreaders that travel more across the space to make more contacts across clusters (i.e., brokering). Indeed, the current default setting (mobility=2 and travelers=10%) tended to produce a contact network that could not reject the assumption that it comes from a power-law distribution (GOF=0.049, $p=0.67$) according to Clauset et al. (2009) distribution comparisons.

In general, the lower the mobility settings, the more clustered the contact network will be because the agents are more restricted to their initial local placements. Likewise, the higher the traveler settings, the more centralized the network will be until that value reaches 50%. That is, after 50%, the majority of

agents will have unrestricted mobility. For instance, if set at 100%, a Bernoulli random contact network will be extracted because 100% of the agents are placed at random areas of the map after each interaction (Table 2).

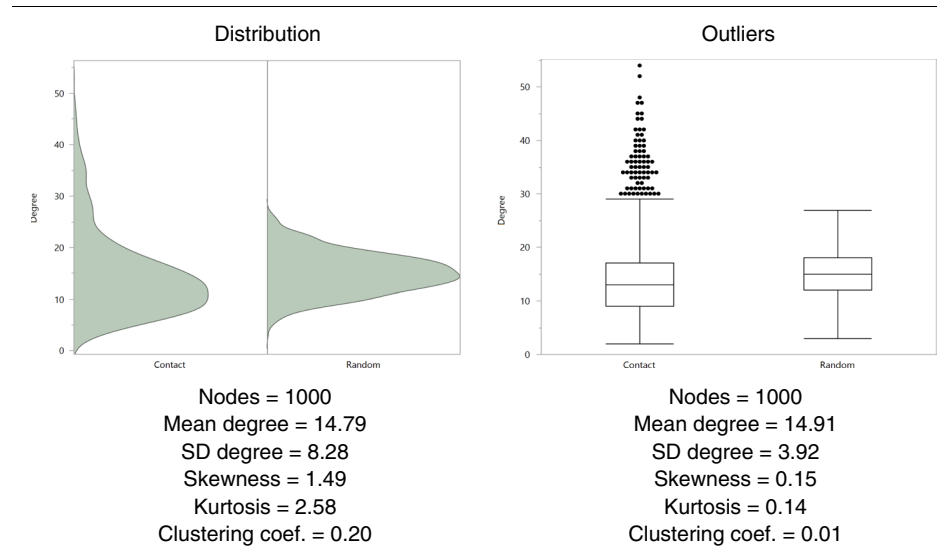
Rules

Rules are the heart of any ABM. They define ‘how agents choose an action, update properties, and interact with each other and their environment’ (Hammond, 2015, p. 176). Given we aim to investigate the effect of contact tracing, we begin with the basic SEIR rules followed by the inclusion of contact tracing rules.

Basic SEIR rules Before articulating specific rules, the following assumptions are made:

- We assume agents’ backgrounds, such as age, gender, occupation, health history, etc., are uniform as these features should not determine whether they should be traced or quarantined.
- We assume agents’ mobilities are uniform to simplify the model.
- We assume patients, when under quarantine, are so well isolated that they do not infect other people in this model.

Table 2. Typical contact network from the simulation.



- We assume the recovered people are fully immune to the disease as we focus on the effect of quarantine during one epidemic outbreak event.

The basic SEIR model rules are set as below in each tick, which we interpret as a full day:

- All agents move a certain distance in a random direction, unless a traveler percentage is set.
- Each presymptomatic or asymptomatic ego defines all susceptible people and unquarantined non-infected alters within its infection radius as its contacts (Fig. 4) and record these agents' IDs in its egocentric contact-history list. At a transmission rate, the presymptomatic ego infects one of these contacts and records the infected agent's ID in its infection-history list.
- If the presymptomatic individuals pass the incubation period, they become patients and are immediately isolated.
- If the patients pass the disease period, they either recover and become immune, or die at the fatality rate.
- The asymptomatic individuals have no symptoms. Unless they are traced and quarantined, they infect the susceptible people and non-infected contacts within their infection radius. They become recovered and immune after 14 days.

The main rule differences between our models and other SEIR models are that the presymptomatic

and asymptomatic individuals are the only agents who infect others.

Contact tracing rules Since all presymptomatic individuals have a constantly updated contact-history and infection-history in this model, when they enter the patient phase, we can trace all their contacts and then quarantine part of or all the contacts. The tracing rules in each day are set as below:

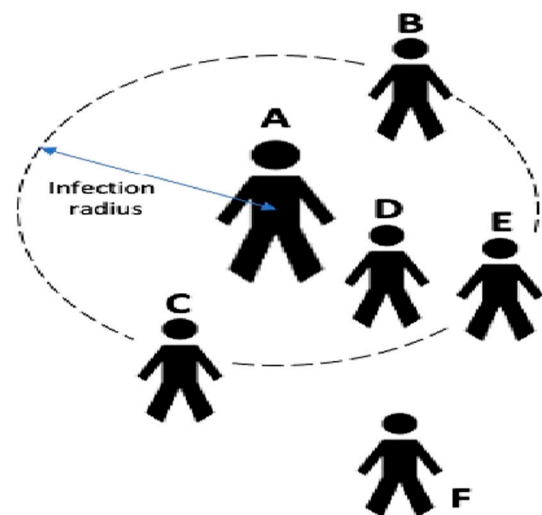


Figure 4: Visualization of an infection radius as an indicator of an ego's contacts.

- Identify the new patients and generate a full contact list based on the contact-history of all these new patients. The contacts in this list are regarded as the 1st-level contact.
- Examine the ‘infected’ status of the 1st-level contacts, if any presymptomatic or asymptomatic individuals are found, trace their contacts using their contact-history and generate the next level contact list, regarded as the 2nd-level contacts. We found some contacts could have been included in the 1st-level contact list, because a person may be counted as the contact by more than one patient. If this happens, we exclude the contacts who have been included in the 1st-level contact list.
- Examine the ‘infected’ status of 2nd-level contacts, if any presymptomatic or asymptomatic individuals are found, trace their contacts using their contact-history and generate the 3rd-level contact list. Exclude the contacts who have been included in the 2nd-level contact list.
- Repeat the above procedures until all contacts are traced.

- Calculate the total number of contacts at a certain level based on the particular contact list.
- Use the total number and the defined quarantine rate to calculate the number of contacts to be quarantined. If the number has a decimal, round down the number.
- Randomly choose and quarantine the number of contacts in the list.

When quarantining beyond 1st-level contacts, we identify the presymptomatic and asymptomatic individuals from the quarantined agents, trace their contacts, calculate the number of contacts to be quarantined in the same way as stated above, and then quarantine that number of the contacts.

Time

Time refers to the unit of analysis representing a passage of time. Operationally, it is usually represented by terms such as ‘ticks’, ‘iterations’, or ‘rounds’. In the current simulation, a ‘tick’, for interpretation, represents a 24-hr day. Moving from day 1 to day 2, the model is put into action. What happens on day 2 will have implications for what happens on day 3 and so on. For instance, if a contact is infected on day 33, the

In the model, we can test the effect of quarantine up to different contact levels. The quarantine rules are specified below:

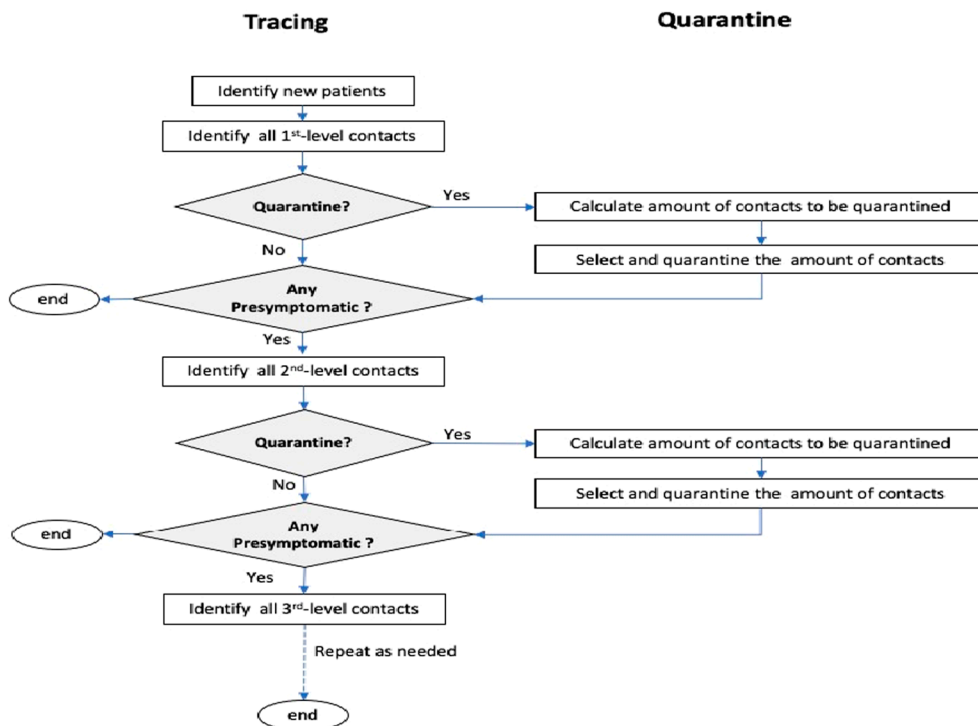


Figure 5: Contact tracing and quarantine procedure flow.

incubation period of the contact will be cumulated when day 34 begins.

Environment

Finally, the environment represents the geometric space and the components in the background. The current simulation is defined by (i) size of the space (i.e., world resolution) and (ii) number of agents. In other words, agents are moving and interacting with one another in an open space, defined by how big it is and how many other agents are also included.

Model verification

The developed model is verified in three ways. First, we carefully go through the model procedures to ensure the conceptual rules are properly translated into the programming codes. Secondly, we compare our data on susceptible, infected, and recovered groups with the data from the classic epidemic models and confirm that our model can produce a typical epidemic data pattern for these people groups. Third, we consult with experts in Public Health to ensure the model elements and assumptions are appropriate.

Once the model is verified, we run a series of one-factor-at-a-time (OFAT) tests (Ten Broeke et al., 2016) to validate the model and calibrate the following two emerging parameters: (i) number of contacts per person and (ii) the basic reproduction number (R_0) of COVID-19. Additional details on these tests are included in the Online Appendix², the main goal is to find settings by exploring reasonable values with respect to the transmission rate, population size, and world resolution that would reproduce typical outcomes related to interaction contacts in the social network analysis literature and R_0 (i.e., average number of people an infected person infects) in the epidemiological literature.

After calibrating our model, we found that, when the world resolution is 30×30 with a population size of 1,000, an agent with an infection radius of 4 may have 14 to 15 contacts, which somewhat reflects the average contacts in the USA and European countries reported by social network and tracing studies (e.g., Del Valle et al., 2007; Mossong et al., 2008; Rothwell, 2020). However, there is uncertainty

regarding the true R_0 value of COVID-19 (Liu et al., 2020a, b). As such, we report our testing results for three situations.

R_0 manipulation

Based on a meta-analysis of COVID-19 R_0 estimation studies, Liu et al. (2020a, b) conclude that the best-guessed estimation seems to be a number between two and three. Several studies seem to vary on the lower and higher end of that range. For instance, a number of studies have found estimated the typical R_0 to be somewhere near 2.20 (see Table 3). Thus, in situation one, we proceed with settings that produce an average R_0 of 2.11. However, some studies have estimated the R_0 to be slightly higher at 2.50 (Imai et al., 2020), 2.55 (Majumder and Mandl, 2020), and 2.68 (Wu et al., 2020). As such, we produce a second situation that produces R_0 values at an average of 2.56 and can interpret these settings as the higher-end R_0 context.

However, in these two contexts, it is assumed that COVID-19 is spreading with little public health interventions to slow down the spread. Given the rise of measures like mask-wearing, we include a third situation which assumes a similar type of intervention. For instance, Li et al.'s (2020b) simulation suggests that if about half the population is regularly wearing masks, the R_0 can drop to between 1.60 and 1.70. As such, we modified the transmission rate to produce a third situation of an average R_0 of 1.62 (see Table 3).

Manipulated input variables

Informant accuracy

To manipulate informant accuracy, we simply adjust the ‘%-contacts-quarantined rate’. For instance, when a presymptomatic patient (i.e., ego) finally feels their symptoms and becomes sick, we assume they get contact traced and their contacts (i.e., alters) are told to immediately quarantine for 14 days. To add heterogeneity to this manipulation, the set percentage is the mean of a normal distribution (SD=5%) and a value is drawn from this distribution.

The user can also manipulate false positives. In this case false positive mean that an agent being contact traced names an agent not on their contact list to go under quarantine. Because people tend to error less with false positives (Bernard et al., 1982), we set the rate to be drawn at a normal distribution with a mean of 10% (SD=5%). The online appendix reports results ranging from 0 to 30%.

²The online appendix also details a typical infection network, which closely mirrors a core-periphery infection structure. The over-dispersion parameter (k) is 0.14, in line with typical estimations of the k -value for COVID-19 (Endo, 2020).

Table 3. Three contexts of model testing and selected previous estimations.

Current model situations		
Situation	Average R_0	Context
S1	1.62	Public health intervention like mask-wearing
S2	2.11	R_0 near lower end of best estimations, no public health intervention
S3	2.56	R_0 near higher end of best estimations, no public health intervention

Previous R_0 estimates		
R_0	Citation	Context
2.20	Li et al. (2020)	Wuhan; January 22, 2020
2.20	Riou and Althaus (2020)	China and overseas; January 18, 2020
2.24	Zhao et al. (2020)	China; January 10-14, 2020
2.28	Zhang et al. (2020)	Diamond Cruise Ship, February, 2020
2.35	Kucharski et al. (2020a)	Wuhan; January-February, 2020
2.50	Imai et al. (2020)	Wuhan, January 18, 2020
2.55	Majumder and Mandl (2020)	Wuhan, December 8, and January 26, 2020
2.68	Wu et al. (2020)	Wuhan; December 31-January, 28, 2020
2.90	Liu et al. (2020)	China and overseas; January 23, 2020

Multi-level tracing

To manipulate the level of contact tracing, we analyze 1st, 2nd, and 3rd-level waves of a snowball sample for each ego that is infectious after their incubation period. 1st-level tracing represents the traditional

baseline approach. Infectious egos are contact traced and their alters are quarantined. In 2nd-level tracing, the alters of initial set of alters are traced as well and they are correspondingly quarantined. Finally, in the 3rd-level tracing, the alters from the second wave are traced as well and then quarantined (see Fig. 6).

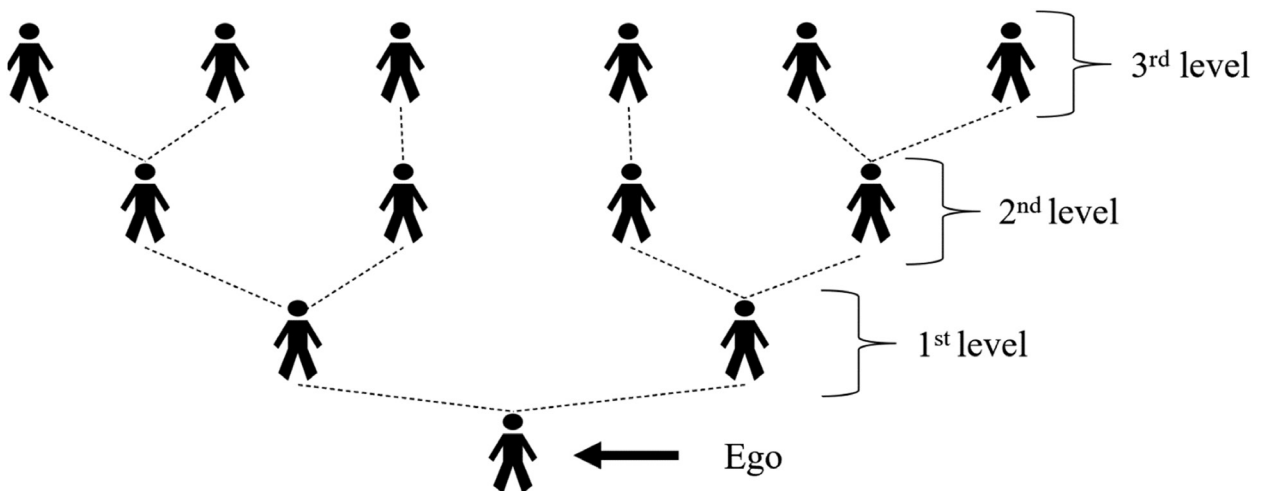


Figure 6: Levels of contact tracing.

Outcome of interest: prevalence and quarantines

There are a variety of outcomes researchers can measure to gauge the severity of an epidemiological outbreak (Rainwater-Lovett et al., 2016). For our purposes, we are most interested in strategies for containment that would prevent the need for herd immunity. As such, prevalence, which is simply the percentage of total infections in a population, is used as the key outcome. There is no precise rule-of-thumb for a critical value of prevalence but obviously, the lower, the better³.

However, low prevalence may come at the cost of excess quarantine orders. To account for this, we also look at the number of quarantines that were administered because of the contact tracing. This could roughly be interpreted as a measure of efficiency because the ideal strategy would be to have the lowest infection prevalence paired with the least amount of quarantine orders. For instance, if everybody just quarantined all the time, it would prohibit the spread of the disease. However, in real-life, that would not be an ideal policy as excessive quarantines would come at drastic social (e.g., isolation) and economic costs (e.g., halt in economic output) (e.g., Elmer and Stadtfeld, 2020). Indeed, this number can even go over 100% because somebody can be quarantined more than once if they are listed as a contact multiple times over the duration of the simulation.

Analysis procedures

The model (ConTrace) is created using NetLogo 6.1.1 (Wilensky, 1999) and is available for download in the online appendix. For informant accuracy, we run from 0 (i.e. basic SEIR model) to 100 percent informant accuracy, increasing the value in increments of five

³Fine et al. (2011), for example, define critical level of infection prevalence for herd immunity to develop (H_c) as a function of the R_0 value of the infection:

$$H_c = \frac{R_0 - 1}{R_0}$$

Based on this equation, the critical prevalence level (H_c) for the average-context spread condition ($R_0=2.11$) would be 52.60%. Likewise, the crucial prevalence level for the high-context spread condition ($R_0=2.56$) would be 60.93%. Finally, for the context assuming mask-wearing, we would assume a critical level ($R_0=1.62$) of 38.27%.

(e.g., 0, 5, 10, etc.). For each value, the simulation is run 30 times. Finally, we repeat the process for 1st, 2nd, and 3rd-level contact tracing and for when the R_0 values are 1.62, 2.11, and 2.56. This results in 5,400 simulations to answer R1, R2, and R3. The median values of prevalence and number of quarantines are the key outcomes of interest.

Results

The three situations (S1, S2, and S3) represent estimates on the spread of COVID-19 using three different R_0 values. S1 represents the lowest spread condition where the average R_0 is 1.62. This R_0 assumes that some sort of public health intervention like mask-wearing is in place (Li et al., 2020a, b). On the other hand, we manipulate two different R_0 values between two and three, the best range estimates of the typical R_0 of COVID-19 (Liu et al., 2020a, b). In the lower-end situation (S2), the average R_0 value is 2.11. In the higher-end situation (S3), the average R_0 value is 2.56. The main effects of informant accuracy are plotted in Figure 7 and the main effects of contact tracing level are plotted in Figure 8. To understand how these two factors interact, the three-way interaction (informant accuracy, level of contact tracing, and situation) is plotted in Figure 9.

R1: Impact of informant accuracy

Because collecting egocentric data is challenging, R1 asked how informant accuracy would influence the spread of COVID-19. Figure 7 plots the main effect of informant accuracy across each situation. Two key trends emerge. First, there seems to be a clear linear trend: the higher the informant accuracy, the less prevalent COVID-19 is and the less people have to quarantine. Second, the steepness of the slope depends on R_0 in each situation. That is, the lower the R_0 , the less the accurate the informant needs to be.

For instance, to get at under 10% prevalence, when the $R_0=1.62$, a patient only needs to be 45% accurate with their contacts. However, when the $R_0=2.11$, that patient needs to be at least 75% accurate. Moreover, this context comes at a cost of quarantining more people as well (near 40%). When the $R_0=2.56$, informant accuracy has less of an impact. For instance, even with 100% informant accuracy, about half will have to go under quarantine.

R2: Impact of contact tracing level

R2 asked what impact multi-level contact tracing would have assuming a traditional snowball design.

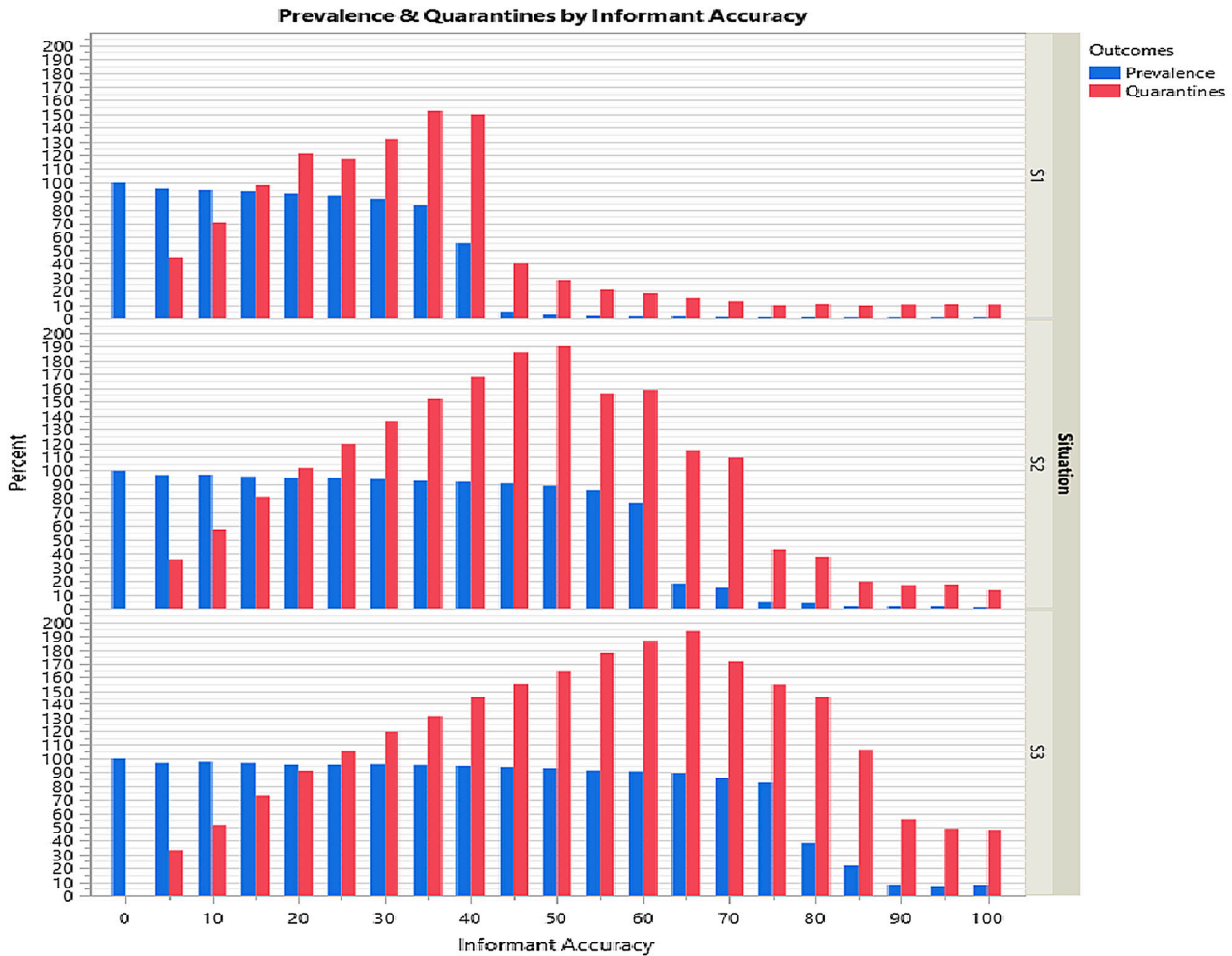


Figure 7: Main effect of informant accuracy.

To look at the main effect of level of contact tracing, Figure 8 plots the average prevalence and quarantine values across each of the three situations. The clear theme that emerges is that there are significant differences in prevalence and quarantines between 1st and 2nd-level tracing, but smaller differences between 2nd and 3rd.

However, there is one caveat. There appears to be only small differences in prevalence, but significant differences in quarantines when comparing situations where the $R_0=2.11$ and where the $R_0=2.56$. That is, overall, the main effect of multi-level tracing appears in the context of S1, when the $R_0=1.62$.

R3: Interaction

Figure 9 plots the main results across the interaction between informant accuracy and level of contact tracing. Overall, this graph tells a much more

complete story. The three situations are separated by color: blue lines represent 1st-level tracing, red lines represent 2nd-level tracing, and green lines represent 3rd-level tracing. The two outcomes are distinguished by line type: solid lines represent prevalence and dashed lines represent quarantines. The Y-axis represents percent of the population with respect to these two outcomes. Quarantines can go over 100% because an agent can be quarantined more than once (e.g., after their 14-day quarantine is over, they can be traced again if they made contact with a different infected agent). Finally, informant accuracy is plotted on the X-axis.

To begin at answering R3, we will look at 1st level contact tracing across each context and see how prevalence and quarantines (Y-axis) differs across levels of network accuracy (X-axis). Overall, the results are contingent on the transmission dynamics of COVID-19. When the $R_0=1.62$, a key inflection point emerges at

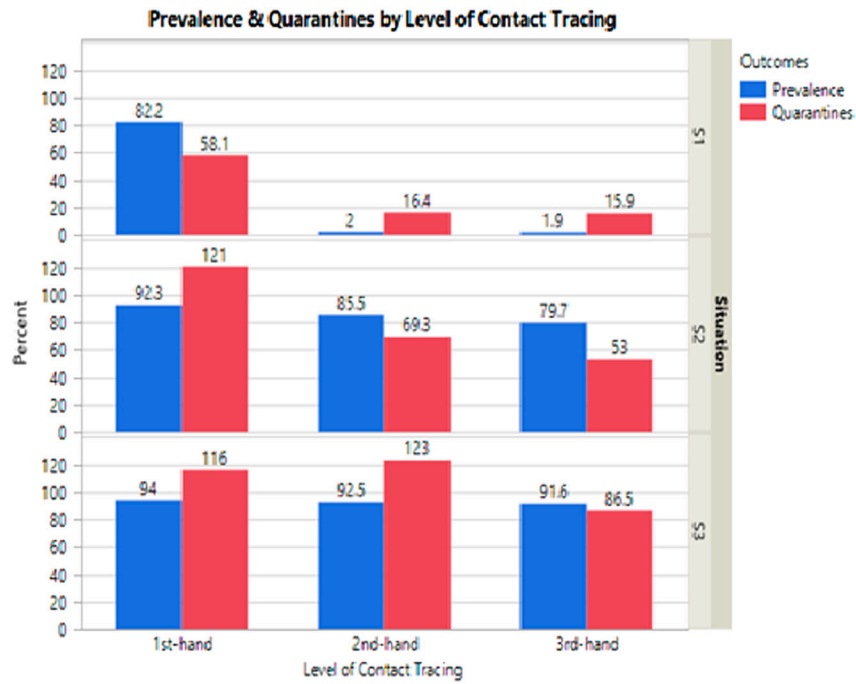


Figure 8: Main effect of level of contact tracing.

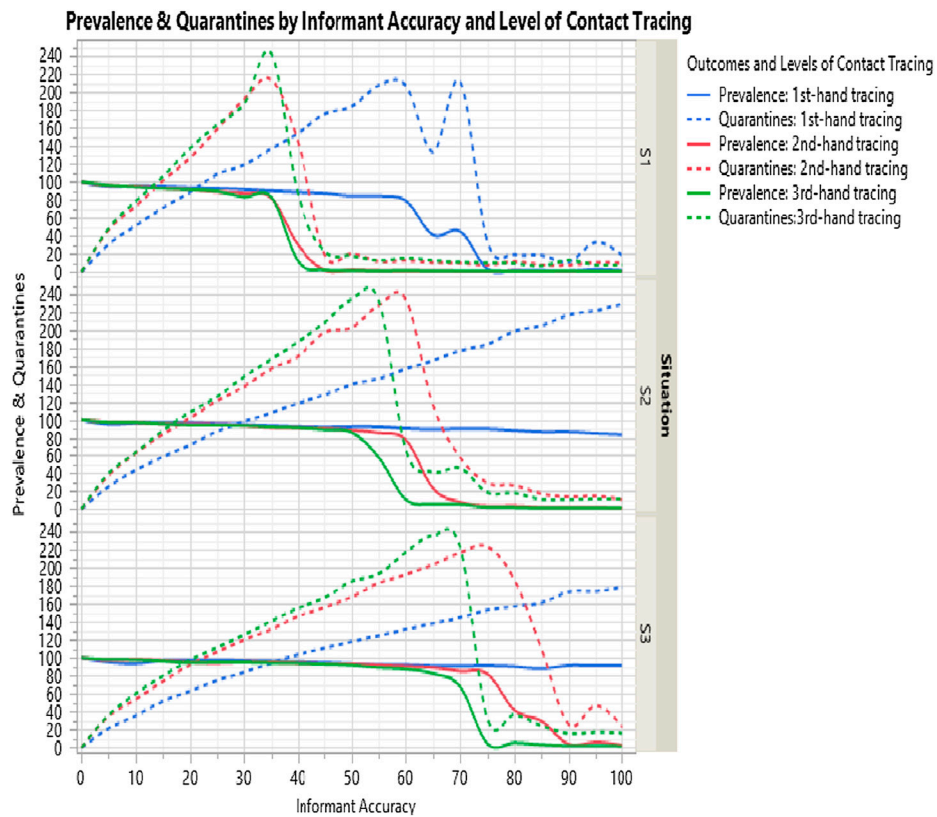


Figure 9: Interaction between informant accuracy and level of contact tracing.

about 75% informant accuracy with prevalence at under 5% with less than 20% of the population having to go under quarantine. When the $R_0=2.11$ and the $R_0=2.56$, there really exists no viable strategy for 1st-level contact tracing. Even at 100% informant accuracy, the outbreak is not contained as vast majority of agents end up contracting COVID-19, even when most agents have to go under quarantine at least once. In other words, it is the worst of both worlds: heavy infection rates and lots of quarantine mandates.

Overall, 2nd and 3rd-level tracing look strikingly similar. For instance, when $R_0=1.62$ for 2nd and 3rd-level tracing, the key inflection point seems to drop to 45% informant accuracy to obtain levels of prevalence under 5% with less than 20% of the population having to go under quarantine. In other words, patients only have to only be about half right about their contacts if those contacts are traced as well. When $R_0=2.11$, similar critical informant accuracy level rise to about 75% for both strategies. However, when $R_0=2.56$, 2nd-level tracing needs to be about 90% accurate for any viable results, but can dip to about 75% informant accuracy without needing too many excessive quarantines (~20%) if 3rd-level tracing is used.

Summary of results

In light of these results, we see three general themes, all revolving around a critical value of 75%, but in different contexts:

1. Traditional 1st-level contract tracing is only really effective when other interventions are taking place like mask-wearing to drive down the R_0 near 1.62. In that case, patients need to be, on average, 75% accurate in naming their contacts.
2. When the R_0 is at the lower end of between two and three (e.g., 2.11), 2nd-level tracing may be a viable strategy. In this case, patients will have to be, on average, 75% accurate in naming their contacts.
3. If the R_0 is at the higher end of between two and three (e.g., 2.56), contact tracing alone will not be effective at mitigating the spread of COVID-19. The only viable strategy is 3rd-level contact tracing in which respondents will still need to be at least 75% accurate.

Sensitivity analysis

Sensitivity analysis is 'a method that measures how the impact of uncertainties of one or more input variables

can lead to uncertainties on the output variables' (Pichery, 2014). Here, we explore how two input factors not initially included in the original model might influence prevalence rates: (i) timing of contact tracing and (ii) asymptomatic cases. We view these as crucial factors because, in real-world settings, it is unclear when contact tracing is implemented because of fluctuations in receiving back test results and because of resources available to do the tracing. Likewise, it is unclear what percentage of COVID-19 cases are asymptomatic, meaning that asymptomatic individuals are highly unlikely to quarantine. In these tests, we manipulate the timing of the contact tracing and percent of asymptomatic cases while keeping the ideal setting of network accuracy at an average of 100% as a constant to demonstrate general trends (see Fig. 10).

The results show that the impact of contact tracing on the spread of COVID-19 is sensitive to both the timing of contact tracing and the percentage of symptomatic cases. When $R_0=1.62$, there is more time to spare at increases in prevalence only take shape after the third day. However, this is still dependent on the amount of the population that is asymptomatic. Once the population is at least 50% asymptomatic, the timing nor the method of tracing has little impact as the majority of the population will become infected.

In 2nd-level tracing, the impact of timing and the R_0 can be significantly relaxed, but only if the asymptomatic rate is not over 50%. For instance, when $R_0=1.62$, noticeable differences in prevalence do not emerge until the fourth day. In higher R_0 situations (e.g., 2.11), that begins to change to the third day. Finally, in even higher R_0 situations (e.g., 2.56), contact tracing seems to be ineffective even if it occurs on the same day.

Discussion

Because contact tracing is essentially a form of ENA, the current research sought to investigate the impact of contact tracing on the spread of COVID-19 through an egocentric, agent-based model. After creating a basic SEIR model of COVID-19 for three different contexts, we manipulated two aspects of contact tracing: (i) informant accuracy (i.e., how accurate the contact tracing was in eliciting alters) and (ii) level of tracing (i.e., the number of snowball waves used). Finally, we did a post hoc sensitivity analysis to see how the (i) timing of contact tracing and (ii) amount of asymptomatic cases influenced the spread of COVID-19 as well. Below, we will discuss the results in line with research on informant accuracy, best practices for collecting egocentric data collection, and wider literature regarding contact tracing.

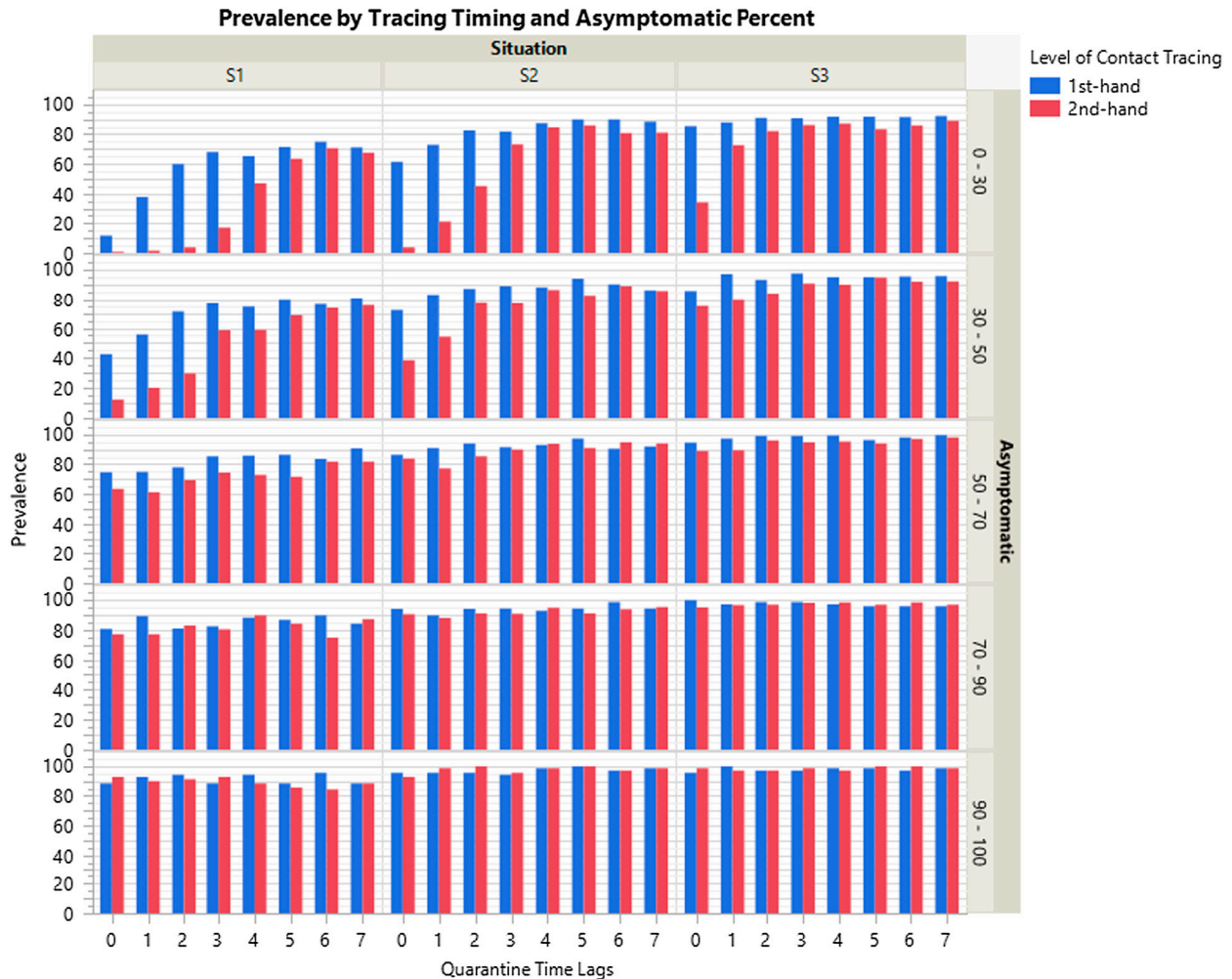


Figure 10: Sensitivity analysis on the timing of contact tracing and asymptomatic percent.

General informant accuracy levels and contact tracing

One general theme of the current results is that higher levels of informant accuracy results in better mitigation of COVID-19 and less quarantines. However, how accurate are informants with respect to their contacts? This question may be difficult to answer because of all the different ways informant accuracy is measured, especially in the seminal BKS studies. For instance, Bernard et al. (1982) provide 15 different informant accuracy measures. In the context of contact tracing and the current study, the most relevant measures are omission errors, which Bernard et al. (1982) label as T2: the number people not recalled who were actually communicated with (i.e., number of true positives). In their electronic information exchange system (EIES) study of 57 users

(i.e., students and scientists) of EIES, the average T2 percentage was 66%. In other words, the informant accuracy rate in line with the current study would only be 34%. Contrastingly, in a study of teletype workers (Killworth and Bernard, 1976), informant accuracy was at about 50%. For reasonable outcomes of COVID-19 prevalence, this would mean that there would need to be mask-wearing and second-level tracing to be effective.

However, some contextual variables give reason to suspect that informants might be slightly better able to remember their contacts during contact tracing. Most notably, the BKS studies are often during longer periods of time ranging from several weeks to months. It can be extremely easy to forget who you may have bumped into several months past. On the other hand, contact tracing poses short-term, rather than long-term memory challenges. According to John Hopkins

University, typical contact tracing only requires patients to remember contacts from the past five to seven days, the average incubation period (Gurley, 2020). In any case, more research may be needed to gauge typical baseline T2 values of informant accuracy for contacts from the past several days.

Nevertheless, the current research has demonstrated that informant accuracy plays a big role COVID-19 prevalence and the number of quarantines issued. For instance, when the $R_0=1.62$, a decrease from 75 to 70% informant accuracy jumps from under 5% to just over 40% and moves the number of quarantines from just under 20% to an unfeasible percentage of over 100%. As such, what are some ways in which informant accuracy can be improved? The next section discusses the current results in line with these important efforts.

Strategies to improve informant accuracy: moving from states to events

Improving informant accuracy for ENA has long been a concern for social network researchers. For instance, Hsieh (2014) formalized the retrieval cue approach for ENA. This approach assumes that the ‘successful recall of an event depends primarily on how well the retrieval cues match the event’s representations in one’s memory organization’ (p. 3). A retrieval cue is any additional piece of information that helps an ego remembers a past event. The retrieval cue approach mirrors Tulving’s (1974) theory of cue-dependent forgetting. This theory assumes that forgetting (i.e., the inability to recall something in the present that could be recalled in the past) does not mean that the memory is lost, but only temporarily inaccessible. In Hsieh’s study, participants were randomly assigned to a retrieval cue condition in which they were instructed to look at their (i) cell phone contact list, (ii) last 30 emails, and (iii) friend list on Facebook and Twitter as retrieval cues. Hsieh’s results found that this approach yielded more contacts than a baseline approach.

However, although Hsieh’s (2014) recall aid might be a good strategy for remembering relational states, it may not be as useful for certain relational events. For instance, it is unlikely that a random elevator conversational contact would turn up in somebody’s cell phone, social media, or e-mail. As such, are there certain recall aids that may be better tailored for recalling relational events, rather than states?

One possibility may be the context-based recall aid designed by Bidart and Charbonneau (2011). The context-based name generator begins by asking individuals about social context cues in everyday life

(e.g., work activity, shopping, home life). Then, once relevant contexts have been triggered, contact names corresponding to those contexts are generated. As Bidart and Charbonneau explain, the context-based approach is motivated by field theory (Feld, 1981). Field theory regards situated action (e.g., activity foci) as a key unit of analysis. For instance, if certain activity foci can be activated (e.g., transportation), more nonchalant relational events might be better able to be recalled. Indeed, Pilny and Huber (2021) tested three different contact tracing aids and found that the context-based instrument significantly elicited more contacts and places visited. If informant accuracy significantly influences the efficacy of contact tracing, then more work is indeed needed to develop recall aids more tailored toward relational events, rather than relational states.

Second, the timing of contact tracing matters greatly. The current results largely replicate the models put forth by Kretzschmar et al. (2020), who find significant delays in the effectiveness in contact tracing even after one day. After three days, contact tracing essentially has no effect unless in contexts of mask-wearing. This is the case because contact tracing is inherently dependent on not just quarantining contacts, but when those contacts are quarantined. If there is a significant delay, then it simply allows those contacts to remain in the system and further transmit COVID-19. In other words, the damage has already been done.

For applied implications, the results of this study strongly suggest that practitioners in charge of training contact tracers spend time working with techniques to improve informant accuracy. We suggest the following techniques to help improve informant accuracy:

- Encouraging the use of contextual recall aids (e.g., activating relevant foci that puts somewhat random interactions in context).
- Improving basic qualitative interview techniques (e.g., the use of probing questions, establishing rapport and trust).
- Implementing contact tracing first to reduce interviewee fatigue.
- Clearly establishing the definition of a contact to reduce differential interpretation (e.g., interactions within six feet for at least 10 min).

Exploring the potential of multi-level tracing

The results of the simulations found that multi-level tracing was more effective at reducing the prevalence

of COVID-19 than traditional single-level tracing. The reasoning seems quite simple. Not only were more contacts traced, but these contacts were not just random; they were the contacts of the initial contacts. That is, in second-level tracing, we contact traced the contacts who are at much more risk of contracting COVID-19 because they were in contact with someone who made contact with an infectious person.

Moreover, multi-level contact tracing was less sensitive to informant accuracy, timing, and percent of asymptomatic cases. However, is multi-level contact tracing practical in real-life? Though some countries like South Korea have used versions of multi-level tracing (Schneider et al., 2020), they require fast implementation in order to be effective. For instance, consider if the amount of average contacts is 14 in the simulations. Not only does a tracer have to contact to those 14 contacts to notify them to quarantine and get tested, but the 14 contacts also need to be traced as well. Then, the contacts of those 14 contacts need to be notified to quarantine and get tested, resulting in an additional 196 notifications (i.e., 14^2).

The growing needs for quicker contact tracing have sparked interest in using advanced information and communication technologies. For instance, smartphones and their applications have been utilized to trace contacts and inform individuals if they have been near an infected individual with Ebola (see Danquah et al., 2019). Contemporarily, similar technology is in development by Apple and Google to trace the spread of COVID-19 via a phone's Bluetooth (Greenberg, 2020), with other countries, such as Singapore, distributing wearable dongles to do so (Asher, 2020). While these novel approaches are exciting, they are still fraught with complications and privacy issues (Danquah et al., 2019). Indeed, a recent survey done by Avira found that 71% of Americans would not be willing to use a contact tracing smartphone application.

Another approach to quickly administer contact tracing may be to move the procedure from in-person phone calls to electronic surveys. For instance, there is a considerable amount of ENA research using surveys to collect data on alters going back to the General Social Survey (Burt, 1984). However, there are still questions as to how reliable surveys are compared to the more 'gold-standard' method of personal interviews. Nevertheless, some recent research may be beginning to make progress in this vein.

For instance, Hogan et al. (2019) report on the efficacy of Network Canvas, a digital egocentric network data collection tool designed to help ease the burden of collecting such egocentric data. In a

similar vein, Hollstein et al. (2020) tested four different egocentric tools that emphasized visualization. They found that most participants preferred concentric circles over funnel tools and free designs, even though these instruments did not significantly influence network size or composition (see also Eddens and Fagan, 2018). For multi-level tracing, we recommend that researchers continue to explore designing electric instruments for egocentric networks, but perhaps also pay more attention to completion times and effort needed to complete the tracing. As the current results show, multi-level tracing is more robust against informant accuracy, so accuracy can be sacrificed a bit if the instrument can be deployed quickly and to many people. For instance, when the $R_0=1.62$, respondents only need to be about 50% accurate in second-level tracing.

Limitations

There are several limitations worth noting. The most obvious is that the contact network is largely 'fixed' and does not follow a stochastic selection process that includes endogenous factors like closure/centralization or exogenous tendencies like mixing (e.g., attribute homophily). This is especially notable because the social network structure underlying any infectious disease is just as important as a disease's biological properties when determining how contagious that disease is with metrics like the R_0 value (Hébert-Dufresne et al., 2020). This suggests that the results of the current simulations cannot be generalized beyond similar network structures reported in Table 2. This is important because different contexts, communities, and cities may have varying structures of interpersonal contact. For instance, de Anda-Jáuregui et al. (2020), using cell phone data, report on the contact network of Mexico City. And although their results show similar levels of centralization and clustering, it was much more fragmented (i.e., had lots of separate components). Future work may consider how manipulating the contact network structure may impact the spread of COVID-19 and how contact tracing can be leveraged with such general network selection tendencies (e.g., Prem et al., 2017).

Finally, beyond the dynamics of the contact network, there are still unknowns related to the dynamic spread of COVID-19. Various studies have reported different R_0 , secondary transmission rates, and growth rates related to COVID-19. Without precise values on these transmission dynamics, researchers are taking educated guesses as to how COVID-19 spreads. Indeed, as the current results

suggest, different transmission dynamics, such as varying secondary transmission rates, are likely to return different results.

Concluding remarks

By now, informant accuracy is a well-known and well-established issue regarding social egocentric network data collection (Corman et al., 2021). Contact tracing is a one applied technique of ENA where researchers can move beyond merely demonstrating that informant accuracy exists, but can begin to investigate the consequences of varying levels of informant accuracy. The current research shows that informant accuracy is critical to controlling the diffusion of an infectious disease. Overall, the results suggest that if contact tracing is to be effective, it must be fast, accurate, and accompanied by other interventions like mask-wearing to drive down the average R_0 . Moreover, the results show the promise of multi-level tracing because it is more robust to lower levels of informant accuracy. How researchers and practitioners can deploy fast and accurate contact tracing instruments could be a vital next step in helping control outbreaks of various infectious diseases.

Acknowledgments

The first two authors contributed equally. This research was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1TR001998. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

Asher, S. 2020. TraceTogether: Singapore turns to wearable contact-tracing Covid tech. *BBC*, available at: <https://www.bbc.com/news/technology-53146360>.

Bernard, H. R., Killworth, P. and Sailer, L. 1981. Summary of research on informant accuracy in network data and the reverse small world problem. *Connections* 4: 11–25.

Bernard, H. R., Killworth, P. D. and Sailer, L. 1982. Informant accuracy in social-network data V. An experimental attempt to predict actual communication from recall data. *Social Science Research* 11: 30–66.

Bernard, H. R., Killworth, P. D., Kronenfeld, D. and Sailer, L. 1984. On the validity of retrospective data: the problem of informant accuracy. *Annual Review of Anthropology* 13: 17.

Bidart, C. and Charbonneau, J. 2011. How to generate personal networks: issues and tools for a sociological perspective. *Field Methods* 23: 266–286.

Borgatti, S. E., Everett, M. M. and Johnson, J. 2013. *Analyzing Social Networks* Sage, Thousand Oaks, CA.

Brashears, M. E. and Quintane, E. 2015. The microstructures of network recall: how social networks are encoded and represented in human memory. *Social Networks* 41: 113–126.

Breashears, M. E., Hoagland, E. and Quintane, E. 2016. Sex and network recall accuracy. *Social Networks* 44: 74–84.

Brewer, D. D. 2000. Forgetting in the recall-based elicitation of personal and social networks. *Social Networks* 22: 29–43.

Browne, K. 2005. Snowball sampling: using social networks to research non-heterosexual women. *International Journal of Social Research Methodology* 8: 47–60.

Burt, R. S. 1984. Network items and the general social survey. *Social Networks* 6: 293–339.

Butts, C. T. 2003. Network inference, error, and informant (in) accuracy: a Bayesian approach. *Social Networks* 25: 103–140.

Butts, C. T. 2008. 4. A relational event framework for social action. *Sociological Methodology* 38: 155–200.

Cheng, H. Y., Jian, S. W., Liu, D. P., Ng, T. C., Huang, W. T., Lin, H. -H. and Team, F. T. T. C. -O. I. 2020. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. *JAMA Internal Medicine* 180(9): 1156–1163, doi: 10.1001/jamainternmed.2020.2020.

Clauset, A., Shalizi, C. R. and Newman, M. E. J. 2009. Power-Law distributions in empirical data. *SIAM Review*, 51(4): 661–703, doi: 10.1137/070710111.

Cohn, S. and O'Brien, M. 2020. How physicians used contact tracing 500 years ago to control the bubonic plague, available at: <https://blogs.ed.ac.uk/covid19perspectives/2020/06/10/how-physicians-used-contact-tracing-500-years-ago-to-control-the-bubonic-plague-by-samuel-cohn-and-mona-obrien/>.

Corman, S. R. 1996. “Cellular automata as models of unintended consequences of organizational communication”, In Watts, J. C. and VanLear, A. (Eds), *Dynamic Patterns in Communication Processes* Sage, Thousand Oaks, CA, pp. 191–212.

Corman, S. R. and Scott, C. R. 1994. Perceived networks, activity foci, and observable communication in social collectives. *Communication Theory* 4: 171–190.

Corman, S. R., Steiner, E., Proulx, J. D., Dutta, A., Yahja, A., Poole, M. S. and Bliss, D. W. B. 2021. Revisiting the accuracy problem in network analysis using a unique dataset. *Social Networks* 66: 1–9.

Craik, F. I. M. and Lockhart, R. S. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6): 671–684, doi: 10.1016/S0022-5371(72)80001-X.

- Crossley, N., Bellotti, E., Edwards, G., Everett, M. G., Koskinen, J. and Tranmer, M. 2015. *Social Network Analysis for Ego-nets: Social Network Analysis for Actor-centred Networks* Sage, Thousand Oaks, CA.
- Danquah, L. O., Hasham, N., MacFarlane, M., Conteh, F. E., Momoh, F., Tedesco, A. A. and Weiss, H. A. 2019. Use of a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: a proof-of-concept study. *BMC Infectious Diseases* 19: 810.
- de Anda-Jáuregui, G., Guzmán, P. and Hernández-Rosales, M. 2020. The contact network of Mexico City. arXiv preprint arXiv:2007.14596.
- Del Valle, S. Y., Hyman, J. M., Hethcote, H. W. and Eubank, S. G. 2007. Mixing patterns between age groups in social networks. *Social Networks* 29: 539–54.
- Donnelly, C. A., Ghani, A. C., Leung, G. M., Hedley, A. J., Fraser, C., Riley, S. and Chan, K. P. 2003. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *The Lancet* 361: 1761–1766.
- Eames, K. T. and Keeling, M. J. 2003. Contact tracing and disease control. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270: 2565–2571.
- Eames, K. T., Webb, C., Thomas, K., Smith, J., Salmon, R. and Temple, J. M. F. 2010. Assessing the role of contact tracing in a suspected H7N2 influenza A outbreak in humans in Wales. *BMC Infectious Diseases* 10: 141.
- Eddens, K. and Fagan, J. M. 2018. Comparing nascent approaches for gathering alter-tie data for egocentric studies. *Social Networks*, 55: 130–141, doi: 10.1016/j.socnet.2018.05.009.
- Elmer, T. and Stadtfeld, C. 2020. Depressive symptoms are associated with social isolation in face-to-face interaction networks. *Scientific Reports* 10: 1–12.
- Endo, A. 2020. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research* 5(67): 1–17, doi: 10.12688/wellcomeopenres.15842.3.
- Epstein, J. M. 2009. Modelling to contain pandemics. *Nature* 460: 687–707.
- Feld, S. L. 1981. The focused organization of social ties. *American Journal of Sociology* 86: 1015–1035.
- Fine, P., Eames, K. and Heymann, D. L. 2011. “Herd Immunity”: A rough guide. *Clinical Infectious Diseases*, 52(7): 911–916, doi: 10.1093/cid/cir007.
- Freeman, L. C., Romney, A. K. and Freeman, S. C. 1987. Cognitive structure and informant accuracy. *American Anthropologist* 89: 310–325.
- Gilovich, T. 1991. *How We Know What isn't so: the Fallibility of Human Reason in Everyday Life* The Free Press, New York.
- Greenberg, A. 2020. How Apple and Google are enabling Covid-19 contact-tracing. *WIRED*, available at: <https://www.wired.com/story/apple-google-bluetooth-contact-tracing-covid-19/>.
- Gurley, E. 2020. COVID-19 contact tracing. Online Coursera lecture. John Hopkins University, available at: <https://www.coursera.org/learn/covid-19-contact-tracing>.
- Hammond, R. A. 2015. “Considerations and best practices in agent-based modeling to inform policy”, In Wallace, R., Geller, A. and Ogawa, A. (Eds), *Assessing the Use of Agent-based Models for Tobacco Regulation*, National Academies Press, Washington, DC, pp. 164–191.
- Hébert-Dufresne, L., Althouse, B. M., Scarpino, S. V. and Allard, A. 2020. Beyond R0: the importance of contact tracing when predicting epidemics. arXiv preprint arXiv:2002.04004.
- Hellewell, J., Abbott, S., Gimma, A., Bosse, N. I., Jarvis, C. I., Russell, T. W. and Flasche, S. 2020. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health* 8: 488–496.
- Hlebec, V. and Ferligoj, A. 2001. Respondent mood and the instability of survey network measurements. *Social Networks* 23: 125–140.
- Hogan, B., Janulis, P., Phillips, G., Melville, J., Mustanski, B., Contractor, N. and Birkett, M. 2020. Assessing the stability of egocentric networks over time using the digital participant-aided sociogram tool Network Canvas. *Network Science* 8(2): 204–222, doi: 10.1017/nws.2019.27.
- Hollstein, B., Töpfer, T. and Pfeffer, J. 2020. Collecting egocentric network data with visual tools: A comparative study. *Network Science* 8(2): 223–250, doi: 10.1017/nws.2020.4.
- Hsieh, Y. P. 2014. Check the phone book: testing information and communication technology (ICT) recall aids for personal network surveys. *Social Networks* 41: 101–12.
- Imai, N., Cori, A., Dorigatti, I., Baguelin, M., Donnelly, C. A., Riley, S. and Ferguson, N. M. 2020. *Report 3: Transmissibility of 2019-nCoV* Imperial College, London.
- Johnson, J. C. 1990. *Selecting Ethnographic Informants* Sage, Newbury Park, CA.
- Kendall, C., Kerr, L. R., Gondim, R. C., Werneck, G. L., Macena, R. H. M., Pontes, M. K. and McFarland, W. 2008. An empirical comparison of respondent-driven sampling, time location sampling, and snowball sampling for behavioral surveillance in men who have sex with men, Fortaleza, Brazil. *AIDS and Behavior* 12: 97.
- Killworth, P. and Bernard, H. 1976. Informant accuracy in social network data. *Human Organization* 35: 269–286.
- Klinkenberg, D., Fraser, C. and Heesterbeek, H. 2006. The effectiveness of contact tracing in emerging epidemics. *PLoS ONE* 1: e12.
- Kogovšek, T., Ferligoj, A., Coenders, G. and Saris, W. E. 2002. Estimating the reliability and validity of personal support measures: full information ML estimation with planned incomplete data. *Social Networks* 24: 1–20.

- Krackhardt, D. 1987. Cognitive social structures. *Social Networks* 9: 109–134.
- Kretzschmar, M. E., Rozhnova, G., Bootsma, M. C. J., van Boven, M., van de Wijkert, J. H. H. M. and Bonten, M. J. M. 2020. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *The Lancet Public Health* 5(8): e452–e459, doi: 10.1016/S2468-2667(20)30157-2.
- Kucharski, A. J., Klepac, P., Conlan, A., Kissler, S. M., Tang, M. and Fry, H. 2020a. Effectiveness of isolation, testing, contact tracing and physical distancing on reducing transmission of SARS-CoV-2 in different settings. MedRxiv, 2004-2023.
- Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S. and Flasche, S. 2020b. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases* 20: 553–558.
- Larson, J. Jr 2012. “Computer simulation methods for groups”, In Hollingshead, A. and Poole, M. S. (Eds), *Research Methods for Studying Groups and Teams* Routledge, New York, NY, pp. 329–357.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y. and Feng, Z. 2020a. Early transmission dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *New England Journal of Medicine* 382: 1199–1207.
- Li, T., Liu, Y., Li, M., Qian, X. and Dai, S. Y. 2020b. Mask or no mask for COVID-19: a public health and market study. *PLoS ONE* 15(8): e0237691, doi: 10.1371/journal.pone.0237691.
- Liu, T., Hu, J., Xiao, J., He, G., Kang, M., Rong, Z. and Ma, W. 2020b. Time-varying transmission dynamics of Novel Coronavirus Pneumonia in China. *bioRxiv*. doi: 10.1101/2020.01.25.919787.
- Liu, Y., Gayle, A. A., Wilder-Smith, A. and Rocklöv, J. 2020a. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*.
- McCarty, C., Lubbers, M. J., Vacca, R. and Molina, J. L. 2019. *Conducting Personal Network Research: A Practical Guide* Guilford Publications, New York.
- Macke, B. A. and Maher, J. E. 1999. Partner notification in the United States: an evidence-based review. *American Journal of Preventive Medicine* 17: 230–242.
- Majumder, M. and Mandl, K. D. 2020. Early transmissibility assessment of a Novel Coronavirus in Wuhan, China (January 26), available at: <http://dx.doi.org/10.2139/ssrn.3524675>.
- Mandalakas, A. M., Ngo, K., Alonso Ustero, P., Golin, R., Anabwani, F., Mzileni, B. and Stevens, R. 2017. BUTIMBA: Intensifying the hunt for child TB in Swaziland through household contact tracing. *PLoS ONE* 12: e0169769.
- Marin, A. 2004. Are respondents more likely to list alters with certain characteristics?: implications for name generator data. *Social Networks* 26: 289–307.
- Marineau, J. E., Labianca, G. J., Brass, D. J., Borgatti, S. P. and Vecchi, P. 2018. Individuals’ power and their social network accuracy: a situated cognition perspective. *Social Networks* 54: 145–161.
- Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R. and Heijne, J. 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 5: e74.
- Omodei, E., Brashears, M. E. and Arenas, A. 2017. A mechanistic model of human recall of social network structure and relationship affect. *Scientific Reports* 7: 17133.
- Perry, B. L., Pescosolido, B. A. and Borgatti, S. P. 2018. *Egocentric Network Analysis: Foundations, Methods, and Models, Vol. 44* Cambridge University Press.
- Pichery, C. 2014. “Sensitivity analysis”, In Wexler, P. (Ed.), *Encyclopedia of Toxicology* Springer, New York.
- Pilny, A. and Huber, C. J. 2021. An egocentric network contact tracing experiment: testing different procedures to elicit contacts and places. *International Journal of Environmental Research and Public Health* 18: 1466.
- Pilny, A., Proulx, J. D., Dinh, L. and Bryan, A. L. 2017. An adapted structural framework for the emergence of communication networks. *Communication Studies* 68: 72–94.
- Prem, K., Cook, A. R. and Jit, M. 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Computational Biology* 13: e1005697.
- Quintane, E. 2012. “Comparing networks: a structural examination of the correspondence between behavioral and recall networks”, In Lusher, D., Koskinen, J. and Robins, G. L. (Eds), *Exponential Random Graph Models for Social Networks: Theories, Methods and Applications, Vol. 35* Cambridge University Press, Cambridge, pp. 272–284.
- Rainwater-Lovett, K., Rodriguez-Barraquer, I. and Moss, W. J. 2016. “Chapter 18 - viral epidemiology: tracking viruses with smartphones and social media”, In Katze, M. G., Korth, M. J., Law, G. L. and Nathanson, N. (Eds), *Viral Pathogenesis*, 3rd ed., Academic Press, Boston, pp. 241–252.
- Riou, J. and Althaus, C. L. 2020. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* 25.
- Rothwell, J. 2020. Americans’ social contacts during the COVID-19 pandemic. GALLUP BLOG, APRIL 21, 2020, available at: <https://news.gallup.com/opinion/gallup/308444/americans-social-contacts-during-covid-pandemic.aspx>.
- Schneider, M., Thornell, C., Parvaneh, D. and Chakraborty, R. 2020. The big lesson from South Korea’s coronavirus response. Vox, available at: <https://www.vox.com/videos/2020/8/6/21356265/south-korea-coronavirus-response-testing>.

Smith, E. B., Brands, R. A., Brashears, M. E. and Kleinbaum, A. M. 2020. Social networks and cognition. *Annual Review of Sociology*, 46(1): 159–174, doi: 10.1146/annurev-soc-121919-054736.

Ten Broeke, G., Van Voorn, G. and Ligtenberg, A. 2016. Which sensitivity analysis method should I use for my agent-based model?. *Journal of Artificial Societies and Social Simulation* 19(1): 5.

Tulving, E. 1974. Cue-dependent forgetting: when we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. *American Scientist* 62: 74–82.

Valente, T. W. 2010. *Social Networks and Health: Models, Methods, and Applications* Oxford University Press, Oxford.

Wilensky, U. 1999. Center for connected learning and computer-based modeling. NetLogo, Northwestern University.

Wu, J. T., Leung, K. and Leung, G. M. 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 395: 689–697.

Zhang, S., Diao, M., Yu, W., Pei, L., Lin, Z. and Chen, D. 2020. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: a data-driven analysis. *International Journal of Infectious Diseases* 93: 201–204.

Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W. and Wang, M. H. 2020. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases* 92: 214–217.