Graph Learning for Bidirectional Disease Contact Tracing on Real Human Mobility Data

Sofia Hurtado and Radu Marculescu

Abstract For rapidly spreading diseases where many cases show no symptoms, swift and effective contact tracing is essential. While exposure notification applications provide alerts on potential exposures, a fully automated system is needed to track the infectious transmission routes. To this end, our research leverages large-scale contact networks from real human mobility data to identify the path of transmission. More precisely, we introduce a new Infectious Path Centrality network metric that informs a graph learning edge classifier to identify important transmission events, achieving an F1-score of 94%. Additionally, we explore bidirectional contact tracing, which quarantines individuals both retroactively and proactively, and compare its effectiveness against traditional forward tracing, which only isolates individuals after testing positive. Our results indicate that when only 30% of symptomatic individuals are tested, bidirectional tracing can reduce infectious effective reproduction rate by 71%, thus significantly controlling the outbreak.

Keywords Graph neural networks · Infection dynamics · COVID-19

1 Introduction

Contact tracing has long been a cornerstone public health strategy for managing outbreaks of highly traceable diseases such as Ebola and Rabies [16, 23]. In such instances, the transmission routes are typically clear, as infectees can recall specific events like animal bites or direct contact with infectious individuals. However, managing pathogens with aerosol transmission and asymptomatic cases like SARS-CoV-2 [25], RSV [9], or Influenza [26] presents greater disease containment challenges, as seen during the COVID-19 pandemic.

S. Hurtado (✉) · R. Marculescu
University of Texas at Austin, Austin, TX, USA
e-mail: slhurtad@utexas.edu

R. Marculescu
e-mail: radum@utexas.edu

I.-H. Ting et al. (eds.), *Advances in Social Networks Analysis and Mining*, Lecture Notes in Social Networks, https://doi.org/10.1007/978-3-031-85386-9_15

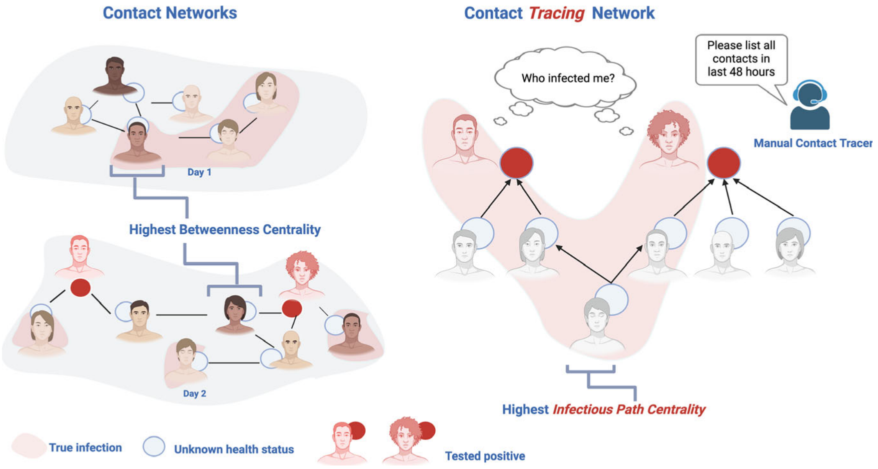

Fig. 1 Manual contact tracing involves collecting past interactions to construct a directed acyclic graph (DAG), where parent nodes are potential sources of infection for their child nodes (forming a contact tracing network). When identifying superspreaders, potential infections, or vaccination candidates, most studies use network analysis techniques such as betweenness centrality on networks with static interactions but dynamic node labels (i.e., health status). However, as illustrated in the contact networks from Day 1 and Day 2 (left), nodes with the highest betweenness centrality [19] do not necessarily hold significant roles in the contact tracing network (right). Instead, nodes with the highest value of our proposed metric, Infectious Path Centrality-which measures the number of paths connecting two positive leaf nodes-are often the most recent common ancestors, making them (and their offspring) crucial for targeted quarantines. We evaluate our metric by comparing its effectiveness in a bidirectional graph learning mitigation framework, which uses this new transmission network metric to identify and quarantine unseen branches of the disease, against traditional forward contact tracing that quarantines those who test positive

To contain a disease outbreak, manual contact tracers need to contact each person who tests positive to identify their recent contacts from the past 24–48 h [28]. However, the effectiveness of manual tracing depends on the accuracy of people's memories, their awareness of their surroundings, and the tracer's ability to reach out to these contacts [17].

In the beginning of an outbreak, the goal of manual contact tracers is to determine the initial source, often referred to as "patient zero." However, as an outbreak escalates into an epidemic, manual tracing can become overwhelmed by rapidly spreading pathogens, especially when community spread occurs [20, 29]. This is because numerous undetected transmission events within the community can make it difficult for individuals to trace their infections to the original source. The goal of contact tracers then pivots to identifying untested paths of transmission to notify people who may be infectious.

In response to the rapidly evolving SARS-CoV-2 virus, companies raced to digitize disease contact tracing. By monitoring interactions among individuals, modern systems can function as continuous disease surveillance tools. Exposure notification applications, for instance, alert users to recent contacts with infected individuals, leveraging advancements in location-tracking technology [18]. Despite these advances, a critical question remains: *Can we effectively backtrack the chain of transmission, particularly for rapidly spreading diseases where many carriers are asymptomatic?*

With access to the Foursquare Mobility dataset [1] that contains visits at various Points of Interest (POIs), we present an *always-ready* disease surveillance system that tracks the past interactions and performs retroactive disease path detection. In addition, we propose a new propagation network metric, *Infectious Path Centrality*, that characterizes the centrality of a node along a *path of a transmission* on a contact tracing network; this is in contrast to the static betweenness centrality on the instantaneous contact networks [7] (Fig. 1).

In this research, we have developed a graph learning framework that utilizes our newly proposed metric (Infectious Path Centrality) as node features. This framework automates the bidirectional contact tracing process, allowing it to identify and quarantine more individuals along the suspected transmission paths. To this end, our contributions are as follows:

1. We introduce a novel Infectious Path Centrality metric, which measures how central a person is among individuals in a contact tracing network.
2. We provide an automated bidirectional disease contact tracing graph learning framework using real mobility data that identifies transmission events along infectious paths.
3. We show that bidirectional contact tracing is more effective than forward contact tracing at reducing the disease's effective reproduction rate.

Taken together, our contributions can help build the technology needed to mitigate the next unknown disease outbreak. The remainder of this paper is organized as follows: Sect. 2 discusses prior work, Sect. 3 describes our approach, Sect. 4 presents our experimental results. Finally, Sect. 5 summarizes our contributions.

2 Prior Work and Novel Contribution

In this section we present the relevant prior work in disease contact tracing, targeting strategies, and graph learning.

2.1 Disease Contact Tracing

Traditionally, contact tracing has been a manual process, relying heavily on interviews conducted by trained health workers to identify individuals who might have been exposed to an infectious person. This method was instrumental during disease outbreaks like smallpox [11] and tuberculosis [3]. In these cases, meticulous record-keeping and follow-ups were key strategies for slowing the spread.

The COVID-19 pandemic spurred unprecedented advancements and the adoption of digital contact tracing tools. Various countries developed mobile applications that utilized GPS and Bluetooth technologies to automate the detection of close contacts [2]. For example, the TraceTogether [6] application in Singapore and Apple and Google's joint exposure notification application GAEN [27] were pivotal in scaling up contact tracing to large populations. These applications could notify users if they had been near someone who tested positive for COVID-19, thereby facilitating a quicker self-isolation or testing response. Though useful in theory, in reality too many false positive alerts resulted in less overall public uptake.

2.2 Targeted Mitigation Strategies

Researchers aimed to measure the efficacy of targeted mitigation strategies such as enacting targeted testing [13], quarantines [15], and vaccinations [5]. To do so, they put effort into identifying significant players within the transmission dynamics, such as superspreaders and transmission bottlenecks, through analyzing network metrics [30]. Most similar to our proposed network metric, Lev et al. introduced an Infectious Betweenness Centrality that charts the betweenness centrality of a node along the path from an infectious node to a susceptible node [24]. In contrast, instead of node betweenness, we propose a Infectious Path Centrality metric that describes the nodes that have high betweenness on *paths* connecting one infectious person to another thus identifying the likely path of disease transmission.

Furthermore, researchers have compared the effectiveness of forward contact tracing, where a person isolates upon testing positive, to bidirectional contact tracing, where an infectious person also retraces their interactions to identify potential unseen infections [4]. They found that isolating these potential cases can prevent more infections than just isolating those who test positive. Our work aims to provide the first machine learning-based solution that charts the paths of disease transmission.

2.3 Graph Learning

Graph learning has emerged as a crucial discipline within machine learning, primarily due to the ubiquity of graph-structured data across various domains such as social networks, biological networks, or communication systems. The introduction of Graph Neural Networks (GNNs) marked a significant shift towards using deep learning techniques for graph data [22]. GNNs iteratively update node representations by aggregating features from neighboring nodes, effectively capturing the local graph structure. This paradigm was further extended into Graph Convolutional Networks

(GCNs) [12], simplifying the graph convolution process and significantly improving the computational efficiency. GCNs have become a cornerstone for numerous applications in link prediction and classification on nodes, edges, and graphs.

Recent work has addressed the challenges of graph learning in the presence of significant class imbalances [14]. However, the extent of class imbalance in certain applications, such as disease contact tracing, is particularly severe. For instance, in scenarios where the goal is to determine the transmission path of an infection, typically only one edge (or interaction) out of many possible interactions is responsible for the transmission. This results in a class imbalance ratio that is inversely proportional to each node degree when contact tracing.

Given the limitations of handling graph imbalance in such extreme cases, our approach seeks to provide a solution against such profound imbalances. Moreover, disease contact tracing introduces a novel challenge to graph learning: the task of identifying specific paths (or graph traversals) based on transmission dynamics. This necessitates not only managing the severe class imbalance, but also developing techniques capable of accurately tracing the paths of transmission in complex network structures.

In this paper, we build on prior work in disease contact tracing and graph learning to (1) develop a new graph learning framework that learns how to chart the path on infection during a large-scale outbreak, and (2) show that bidirectional contact tracing outperforms standard forward contact tracing when trying to mitigate transmission.

3 Proposed Approach

In Fig. 2, we provide a comprehensive overview of our approach. First, we convert individual dwell times from Foursquare mobility data into person-to-person networks. We then apply a network-based compartmental epidemiological model (SEIR) via simulation (step 2). Following this, we map out the path of transmission, creating an accumulative contact tracing network, and compute the Infectious Path Centrality for each neighbor connected to an infectious individual (steps 3 and 4). Using these centrality measures as features, we train a graph edge classifier to identify edges that represent transmission events (step 5). Finally, to evaluate the experimental efficacy, we conduct new simulations that quarantine nodes based on standard forward tracing and our bidirectional contact tracing method (step 6).

In this section, we describe in detail our approach for network construction, epidemic model, Infectious Path Centrality metric, and mitigation evaluation.

3.1 Person-to-Person Network

Given the visits from Foursquare mobility data with venues and dwell times, we construct the initial person-to-person graph $G = (V, E)$ where V is the set of nodes

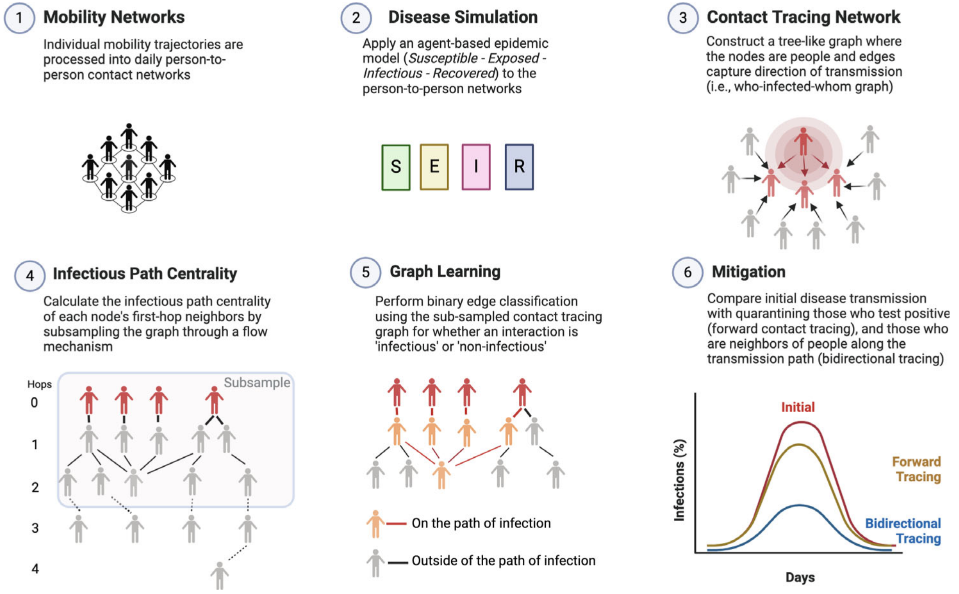
1 Mobility Networks
Individual mobility trajectories are processed into daily person-to-person contact networks
2 Disease Simulation
Apply an agent-based epidemic model (*Susceptible - Exposed - Infectious - Recovered*) to the person-to-person networks
S
E
I
R
3 Contact Tracing Network
Construct a tree-like graph where the nodes are people and edges capture direction of transmission (i.e., who-infected-whom graph)
4 Infectious Path Centrality
Calculate the infectious path centrality of each node's first-hop neighbors by subsampling the graph through a flow mechanism
Hops
0
1
2
3
4
Subsample
5 Graph Learning
Perform binary edge classification using the sub-sampled contact tracing graph for whether an interaction is 'infectious' or 'non-infectious'
On the path of infection
Outside of the path of infection
6 Mitigation
Compare initial disease transmission with quarantining those who test positive (forward contact tracing), and those who are neighbors of people along the transmission path (bidirectional tracing)
Initial
Forward Tracing
Bidirectional Tracing
Infections (%)
Days

Fig. 2 Our experiment starts by processing Foursquare mobility data containing device dwell times at POIs into person-to-person contact networks (step 1). Two devices (i.e., people) are connected when they visit the same location within the same hour. We then apply an agent-based epidemiological SEIR (*Susceptible-Exposed-Infectious-Recovered*) model [10] on the dynamic contact networks (step 2). To keep track of transmission events, we form a contact *tracing* network where the parent is a potential source of infection to the child (step 3). For every transmission event (infection in step (2)), we add the infectious interaction to the contact tracing network, as well as all other interactions the infectee has on the day of infection. This mimics the manual contact tracing where someone recalls all of their interactions on the day they got infected. Step 4 consists of calculating our proposed Infectious Path Centrality metric that is then used as features in edge classification (step 5). The graph learning module classifies the edges in the contact tracing network as being 'infectious' (i.e., a true transmission event), or 'non-infectious'). Finally, we test the efficacy of our approach by comparing a population seeded with the same infectious individuals that undergoes no mitigation, forward contact tracing, and bidirectional contact tracing using our mitigation framework (step 6)

(i.e., persons), and E is the set of edges (i.e., interactions), between them. We define an interaction if two people are at the same venue within the same hour. We construct these graphs for each day in the Foursquare mobility dataset with the time granularity of one hour. We assume an equal chance for any two people to interact within a venue, therefore, within each venue at a given hour the subgraph is fully connected.

3.2 Epidemic Model

Because we lack the ground truth for health labels on each individual within the dataset, we rely on simulating the disease spreading. More precisely, we deploy an agent-based SEIR model [10] on the population where a node on the contact networks can be one of the four health states, i.e., {*Susceptible, Exposed, Infectious, Recovered*}. To introduce heterogeneity, we assign each individual with immunity $\delta \in [0, 1]$ and virality $v \in [0, 1]$ values. All nodes start off as *Susceptible* and move to *Incubating* when they interact (i.e., edge) with an infectious node that has a virality v greater than their immunity δ (Eq. 1):

$$S_j \rightarrow I_j : v_i > \delta_j \tag{1}$$

After the incubation period (5 d), the individual is considered to be *Infectious* and is assigned a v virality value. After a sickness period (7 d), the individual is considered *Recovered* and cannot be infected again within the testing span of the experiment (30 d). Note that the immunity threshold, virality value, incubation period, and illness period, are all tunable parameters that could be chosen to simulate a different infectious disease.

3.3 Contact Tracing Network Construction

After simulating a disease outbreak on the daily person-to-person networks, we construct a contact tracing graph. This graph aims to emulate the contact tracing performed by a manual contact tracer who calls each infectious individual and queries about their past interactions. To achieve this, for every node that transitions from *Susceptible* to *Incubating*, we add the node's neighborhood to the contact tracing graph where all edges point to the most recently infected node. This forms a directed acyclic graph (tree) where the leaves are all *Infectious* nodes (Fig. 2 Step 3).

3.4 Infectious Path Centrality

We abstract the question of *'who infected whom'* to a binary edge classification task to determine whether an interaction contains a transmission event. Because we assume that a person gets infected by one other entity (i.e., person or place), there is a large class imbalance between infectious and non-infectious interactions.

To account for class imbalance, we take advantage of the transmission dynamics of a disease. As shown in Fig. 3b, we start with the contact tracing network's leaf node u and apply an attenuating signal that has a decay $\alpha \in [0, 1]$ with each hop h to act as the edge weights $w_{(u,v)}$ (Eq. 2).[1] Note that α gives the importance of each interaction in the h-hop neighborhoods and reduces overvaluation of superspreader neighbors. We traverse a maximum of H hops and accumulate the weights of all incoming edges at each node y to get ϕ_y (Eq. 3). Finally, we reverse the edges to travel from H hops back to the 1-hop neighbors of the infectious node u to accumulate all ϕ for all incoming edges resulting in the Infectious Path Centrality value π_v (Eq. 4)

$$w_{(u,v)} = \alpha^{h-1} \tag{2}$$

$$\phi_y = \sum_{\forall (x,y) \in N_{h=1}(y)} w_{(x,y)} \tag{3}$$

$$\pi_v = w_{(u,v)} + \sum_{\forall (v,x) \in N_{h=2}(v)}^{N_{H-1}(v)} \phi_x \tag{4}$$

3.5 Mitigation

We evaluate our mitigation strategies by setting up side-by-side comparison between no-mitigation, forward contact tracing, and bidirectional contact tracing in an 'online' simulation. Starting from interactions in July's person-to-person networks and seed infections, we then test the population of sick individuals with a virality $v > 0.5$ to simulate testing only symptomatic cases. For each day in the simulation, we then update the contact tracing network and Infectious Path Centralities for each 1-hop neighbor of those who recently test positive. We then use a trained infectious edge classifier using May and June contact tracing network to identify who infected the leaf nodes. In the forward contact tracing, we simply quarantine the individuals who tested positive for the infectious period (7 d). For the bidirectional contact tracing, we go back 5 d (incubation period) and quarantine all of the potential new infections. Finally, we compare the effective reproduction number R_t [8] (that averages the number of new infections caused by one person) of the outbreaks resulting from no mitigation, forward contact tracing mitigation, bidirectional contact tracing mitigation.

[1] All notations relevant to this section are explained in Fig. 3a.

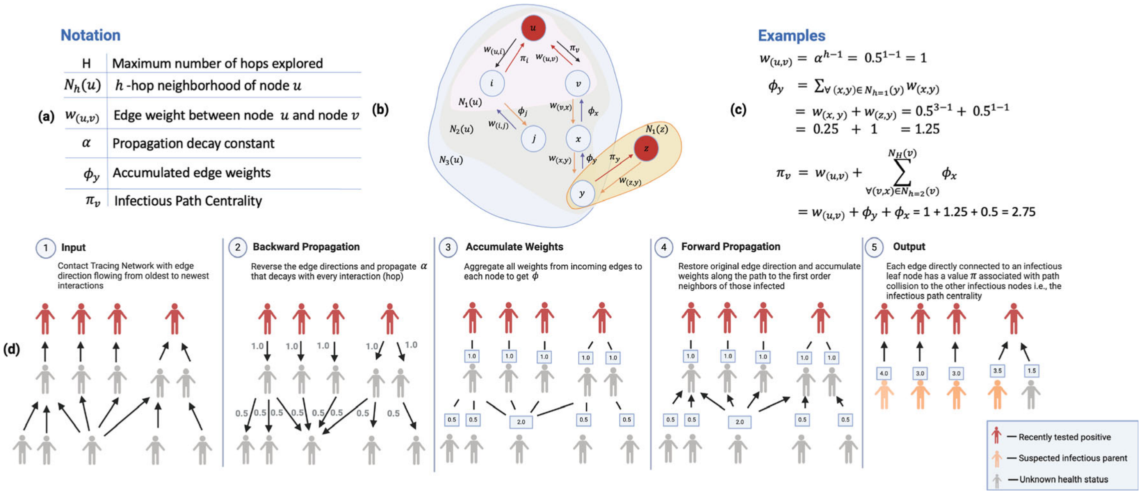
Notation
(a)
H Maximum number of hops explored
$N_h(u)$ h-hop neighborhood of node u
$w_{(u,v)}$ Edge weight between node u and node v
α Propagation decay constant
ϕ_y Accumulated edge weights
π_v Infectious Path Centrality
(b)
u
i
v
j
x
y
z
$w_{(u,i)}$
π_i
$w_{(u,v)}$
π_v
ϕ_j
$w_{(i,j)}$
$w_{(v,x)}$
ϕ_x
$w_{(x,y)}$
ϕ_y
π_y
$w_{(z,y)}$
$N_1(u)$
$N_2(u)$
$N_3(u)$
$N_1(z)$
Examples
(c)
$w_{(u,v)} = \alpha^{h-1} = 0.5^{1-1} = 1$
$\phi_y = \sum_{\forall (x,y) \in N_{h=1}(y)} w_{(x,y)}$
$= w_{(x,y)} + w_{(z,y)} = 0.5^{3-1} + 0.5^{1-1}$
$= 0.25 + 1 = 1.25$
$\pi_v = w_{(u,v)} + \sum_{\forall (v,x) \in N_{h=2}(v)}^{N_H(v)} \phi_x$
$= w_{(u,v)} + \phi_y + \phi_x = 1 + 1.25 + 0.5 = 2.75$
(d)
1 Input
Contact Tracing Network with edge direction flowing from oldest to newest interactions
2 Backward Propagation
Reverse the edge directions and propagate α that decays with every interaction (hop)
1.0
0.5
3 Accumulate Weights
Aggregate all weights from incoming edges to each node to get ϕ
1.0
0.5
2.0
4 Forward Propagation
Restore original edge direction and accumulate weights along the path to the first order neighbors of those infected
1.0
0.5
2.0
5 Output
Each edge directly connected to an infectious leaf node has a value π associated with path collision to the other infectious nodes i.e., the infectious path centrality
4.0
3.0
3.0
3.5
1.5
— Recently tested positive
— Suspected infectious parent
— Unknown health status

Fig. 3 **a** Notation for equations 2–4. **b** Toy example of a contact tracing network using the notation for our proposed Infectious Path Centrality metric. Nodes u and z just tested positive and are tracing past interactions to identify who is the likely source of infection between nodes v and i; we assume node y infected node z. H denotes the number of hops (i.e., depth), the Infectious Path Centrality encompasses. α denotes a propagation decay constant that facilitates calculating the weight w of each edge in the N_h (h-hop neighborhood). Note that w acts as an attenuating signal originating from node u that sub-samples the larger contact tracing network. The term ϕ_y represents the total weights of incoming edges of node y from all paths originating from infectious leaf nodes (i.e., nodes u and z). The Infectious Path Centrality π_v then quantifies the forward accumulation of all ϕ values leading back to u's immediate neighbors (i.e., $N_1(u)$). **c** Example calculations for $w_{(u,v)}$, ϕ_y, and π_v. **d** Step-by-step process of calculating Infectious centrality where (1) we input the contact tracing network and (2) reverse the edges to attenuate the weights down to the maximum hop level H. (3) Each node accumulates weights w from all incoming edges to get ϕ. (4) We then restore original edge directions and accumulate ϕ back to the first-hop neighbors of the leaf nodes. (5) Next, we add all the ϕ from the incoming edges into the first-hop neighbors to create the their respective Infectious Path Centrality value π. (6) Finally, we normalize all π values and use them as features in the edge classification module. By design, if a node exists on paths leading to multiple infectious leafs, the π value will be greater. We hypothesize that the (orange) node with the maximum π is the infectious source

4 Results

In this section, we present our experimental set up, investigate the H hops for the Infectious Path Centrality, perform an ablation study on α, and evaluate a few mitigation strategies.

4.1 Experimental Set Up

We built our framework using four months (i.e., May, June, July, August) in 2020, of the Foursquare mobility dataset during various stages of COVID-19 lockdown and reopening in Austin, Texas. For each month, we seed 1% of the person-to-person networks with infections and form the resulting disease contact tracing network. The statistics for devices captured, number of infections, and contact tracing network nodes and edges are shown in Table 1.

We visualize the contact tracing network in Fig. 4a–c to show the relationship between node infections and edge infections. The nodes in red have contracted the virus, and the edges in red chart the path of infection. Every node has only one red incoming edge Fig. 4c which signifies that every node is only infected by one person. Figure 5 shows the node degree distribution for all the months captured in the experiment. Given the heavy tailed distribution, we can conclude that the directed contact tracing network is scale-free graph where there exists few superspreaders.

4.2 Infectious Path Centrality

After constructing the contact tracing, we calculate the Infectious Path Centralities for each positive node's 1-hop neighbor, and analyze the maximum number of hops H (i.e., depth) needed in order train the edge classifier well. We first investigate the amount of graph covered by traversing each hop on different graph topologies. In Fig. 6, we compare the contact tracing graph pulled from a disease simulation on Austin in May of 2020, to scale free, random, and mesh networks of equal size (20, 000 nodes).

Table 1 Foursquare sample size statistics for 2020

Month	Devices	Infections	Nodes	Edges
May	37,049	17,075	20,095	116,615
June	37,039	20,036	21,974	154,548
July	36,347	17,644	19,899	124,821
August	47,598	28,082	31,073	244,357

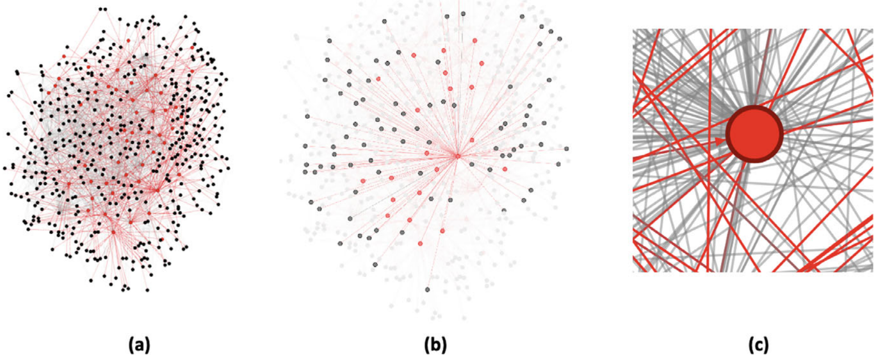

Fig. 4 **a** 500 node sub-sample of a contact tracing network. Red edges signify a transmission event where the parent node infects the child node. **b** portrays an ego-network view of a leaf node that has recently tested positive. Their ego-network consists of all contacts made on the day of infection (incoming edges), as well as those they have infected (outgoing red edges). **c** depicts a zoomed in view of an infectious node where there is only one incoming red edge (i.e., source of infection), and many outgoing red edges (parent of infections). There are also many incoming grey edges that signify interactions on the day of infection that were not transmission events

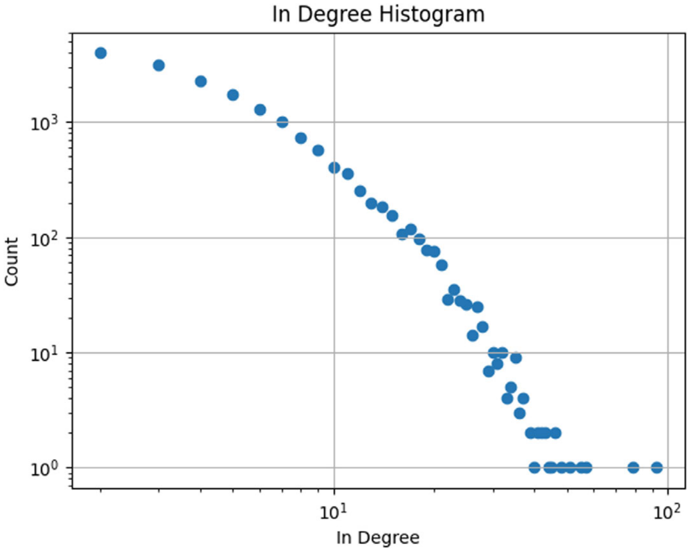

Fig. 5 The in-degree histogram in log-log scale shows that the contact tracing network is scale-free. This means that few nodes have many incoming edges, while most nodes have few. The class imbalance for identifying the incoming transmission edge is proportional to the in-degree which makes training an edge classifier largely unbalanced

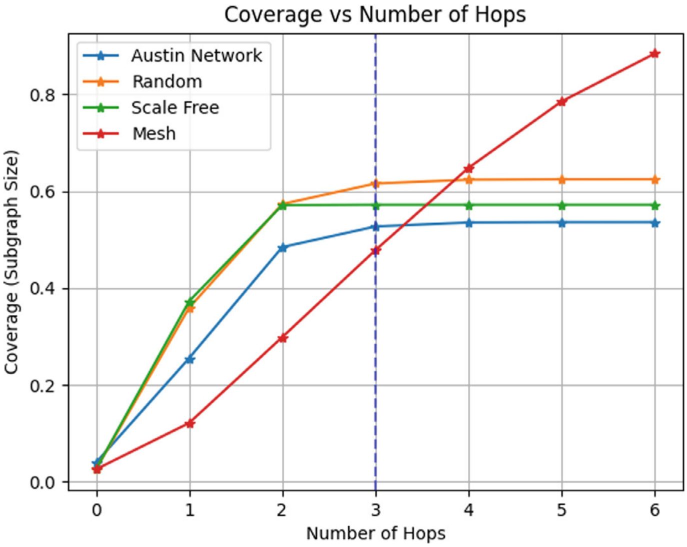

Fig. 6 Comparison of percentage of nodes explored between topologies according to hop-depth. Each network consists of 20,000 nodes and 116,000 edges to be comparable to the Austin contact tracing network. Of note, the contact tracing network is a DAG structure whereas, the other graphs are undirected. The scale free network coverage plateaus at 2 hops whereas the Austin contact tracing and random networks plateau at 3 hops. In contrast, the mesh network coverage is proportional to the number of hops

We start from a sample of 500 nodes on each graph, traverse to each h-hop neighborhood and keep track of how many unique nodes are visited. We find that for all cases except the mesh graph, traversing to the 3-hop neighborhood samples the largest subset of the graph (Fig. 6). Though the random, scale free, and Austin contact tracing graphs all have an average path length of around 6, this coverage suggests that the sample nodes are roughly 3 hops away from a superspreader. In contrast, the mesh covers more nodes with every hop.

To further investigate, we train the edge classifier on different maximum hop depths H to compare the F1-scores in Fig. 7. We use the F1 metric [21] to take into account the class imbalance rather than accuracy since most of the edges are not transmission events. We can see the edge classification model achieves the highest F1-score when traversing to 2-hop neighbors (orange curve). Note that this is one-hop less than achieving the largest coverage (3-hops). Perhaps this is because at 3-hops the Infectious Path Centrality metric has a hard time saturating as there are too many collisions to differentiate the paths.

[Ablation study] In addition, we investigate various values of α to determine how important we should weigh each h-hop neighborhood; we find that $\alpha = 0.5$ yields the highest F1-score (Fig. 8).

As such, we train all of the months using $h = 2$ with $\alpha = 0.5$ in Fig. 9 where F1-scores range between 0.81 and 0.94. After determining the number of hops that the

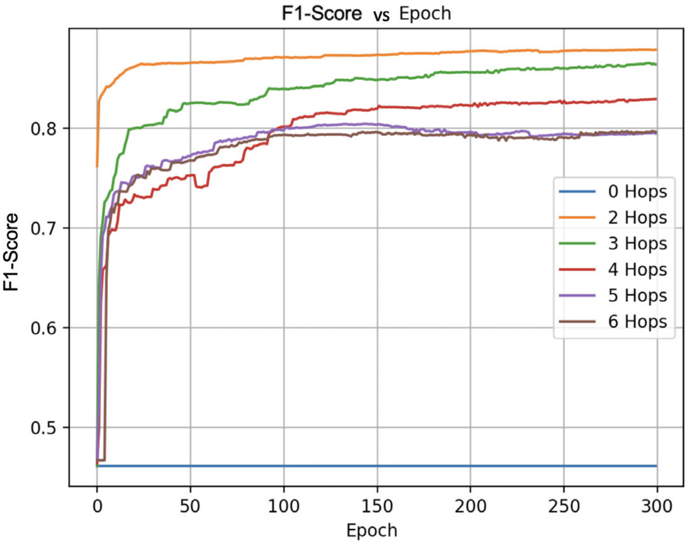

Fig. 7 Comparison of the effect of hops on F1-score for an edge classifier trained over 300 epochs. 0 hops signifies performing edge classification without the Infectious Path Centrality metric. The contact tracing graph is taken from Austin mobility interactions from May, 2020. The Infectious Path Centrality metric calculated by traveling 2-hops away from the leaf node gets the highest F1-score of 0.87 after 300 epochs

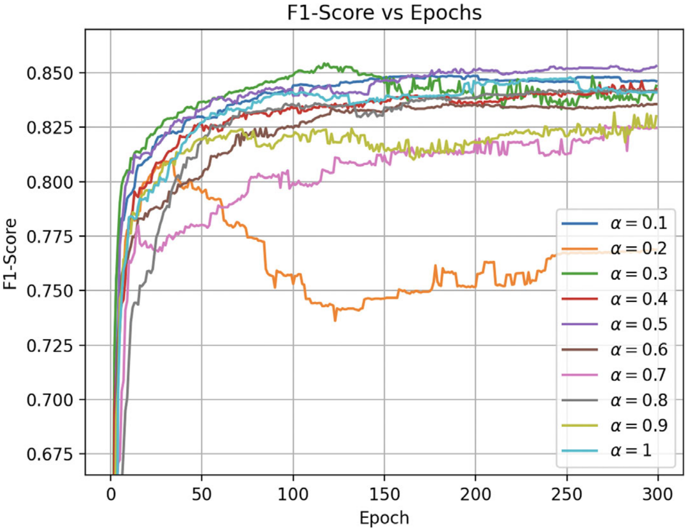

Fig. 8 **[Ablation study]** Comparison of α value used on Austin contact tracing network for May of 2020 using $h = 2$. For example, an $\alpha = 0.1$, means that the propagation signal diminishes quickly between h hops. Decaying the signal by $\alpha = 0.5$ yields the highest F1-score of 0.85

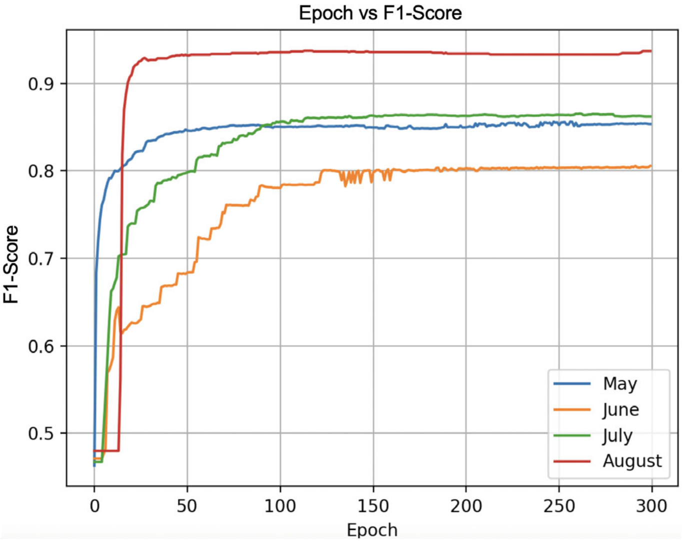

Fig. 9 Comparing training edge classifier on contact tracing networks from Austin in May, June, July, and August in 2020. August has the most nodes (31,073) and edges (244,357) resulting in the largest sub-sample for training. The edge classifier achieves an F1-score of 0.94 after training for 300 epochs on August. In contrast, July has the smallest network with a 0.81 F1-score

flow metric should traverse, and training the edge classification model, we investigate the efficacy of bidirectional contact tracing.

4.3 Mitigation

We utilize an edge classifier trained on contact tracing networks from May and June and apply ‘online’ mitigation to July and August. As described in the approach, we test newly symptomatic individuals and calculate the Infectious Path Centralities by traversing to their $H = 2$-hop neighbors. We then update the contact tracing network and classify the edges of the 1-hop neighbors to gather the parent infections. From there, we test whether quarantining those who tested positive (forward contact tracing), or retroactively quarantining all neighbors of those who infected the positive individuals (bidirectional contact tracing) lowers the disease reproduction rate compared to the unmitigated population.

In each case, we seed the infections with the same 1% of individuals so that the infection comes from the same starting point. We implement the mitigation techniques starting on the eighth day to ensure that the disease has propagated. Figure 10 shows a comparison of the effective reproduction number at the end of each simulation for different percent of population tested. For example, when only 1% of symptomatic cases (i.e., virality $v > 0.5$) is tested, the unmitigated population

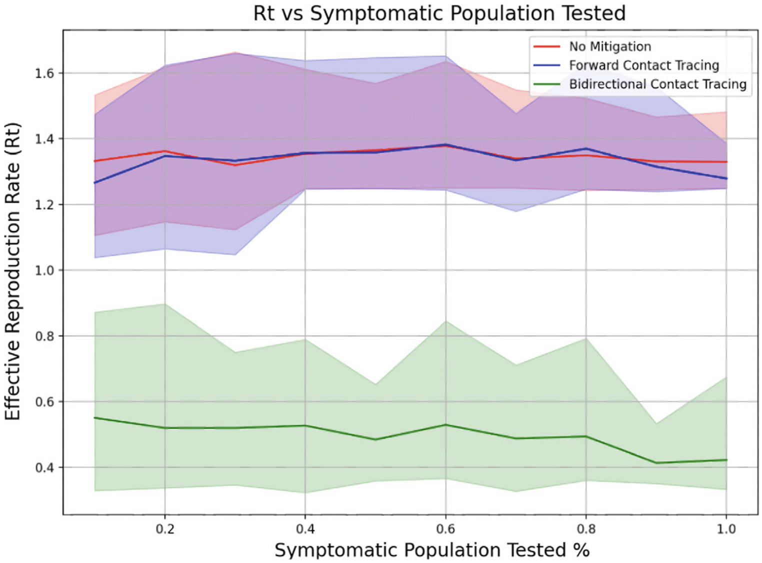

Fig. 10 Comparison of effective reproduction number R_t across 20 simulation runs for three scenarios: no mitigation measures, forward contact tracing, and bidirectional contact tracing, varying the percentage of symptomatic cases tested. The mitigation strategies are implemented starting on the eighth day, and R_t represents the average number of secondary infections per case by the end of each simulation. The shades with lighter colors represent the minimum and maximum R_t among the 20 simulations

has an average effective reproduction rate of 1.33 while the forward tracing has a rate of 1.27 and bidirectional contact tracing has 0.54. Most notably, regardless of the percent of population tested, we see a dramatic decrease in effective reproduction rate between bidirectional contact tracing and unmitigated populations (71%) and forward contact tracing (54%). This is in contrast to only using forward contact tracing which decreases the effective reproduction rate by only 16%. Intuitively, this is because when performing forward contact tracing while only testing a percentage of the symptomatic cases, many transmission paths will go untested which results in the continuation of viral propagation. When 100% of the symptomatic individuals are tested, the bidirectional contact tracing mitigation strategy results in a effective reproductive rate of 0.42 which means that, on average, every ten infected people produce roughly four infectious offspring, thereby significantly slowing down the outbreak.

5 Conclusion

In this paper, we have presented a framework to automate bidirectional contact tracing using Foursquare mobility data. We have formulated the transmission path identification problem on contact tracing networks as graph learning edge classification that determines whether an edge is a transmission event. We have also proposed a new network metric, Infectious Path Centrality, that describes the centrality of a node along the path from two infectious nodes.

Our proposed metric solves the class imbalance problem on transmission path identification, as it sub-samples the nodes along the H-hop neighborhoods. Moreover, unlike manual contact tracing, our metric performs better in scenarios of widespread community transmission. Through training, our edge classification model achieves an F1-score of 0.94; when used to perform bidirectional contact tracing, the R_t decreases by 71% compared to unmitigated populations.

Our work is limited by the fact that we do not have the ground truth for the true transmission dynamics, as there is no public dataset that contains large scale interactions, as well as health labels. Future work includes extending our analysis to other cities and communities of various size (i.e., New York City vs St. Louis) to evaluate the viability of our proposed metric. Furthermore, we intend to investigate the robustness to imperfect quarantines with varying degree of compliance, and inaccurate testing (i.e., false positives and/or false negatives).

Taken together, our work is an important step towards automating disease contact tracing to help mitigate the next unknown viral outbreak.

Acknowledgements This work is supported, in part, by NSF grant CCF 2107085.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Foursquare visits dataset. https://foursquare.com/products/visits/
2. Bannister-Tyrrell M, Chen M, Choi V, Miglietta A, Galea G (2023) Systematic scoping review of the implementation, adoption, use, and effectiveness of digital contact tracing interventions for covid-19 in the western pacific region. Lancet 34
3. Baxter S, Goyder E, Chambers D, Johnson M, Preston L, Booth A (2017) Interventions to improve contact tracing for tuberculosis in specific groups and in wider populations: an evidence synthesis. Health Services and Delivery Research
4. Bradshaw W, Alley E, Huggins J, Lloyd A, Esvelt K (2021) Bidirectional contact tracing could dramatically improve covid-19 control. Nat Commun 12(232)
5. Chen J, Hoops S, Marathe A, Mortveit H, Lewis B, Venkatramanan S, Haddadan A, Bhattacharya P, Adiga A, Vullikanti A, Srinivasan A, Wilson M, Ehrlich G, Fenster M, Eubank

S, Barrett C, Marathe M (2022) Effective social network-based allocation of covid-19 vaccines. In: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, pp 4675–4683

6. Chow B, Lim YD, Poh R, Ko A, Hong GH, Zou S, Cheah J, Ho S, Lee V, Ho M (2023) Use of a digital contact tracing system in Singapore to mitigate covid-19 spread. BMC Public Health 23(2253)
7. Christley R, Pinchbeck G, Bowers R, Clancy D, French N, Bennett R (2005) Infection in social networks: using network analysis to identify high-risk individuals. Am J Epidemiol
8. Gostic K, McGough L, Baskerville E, Abbott S, Joshi K, Tedijanto C, Kahn R, Niehus R, Hay J, Salazar P, Hellewell J, Meakin S, Munday J (2023) Practical considerations for measuring the effective reproductive number, RT. arXiv:2304.04300
9. Kulkarni H, Smith C, Lee D (2016) Evidence of respiratory syncytial virus spread by aerosol: time to revisit infection control strategies? Am J Respir Crit Care Med
10. He S, Peng Y, Sun K (2020) SEIR modeling of the covid-19 and its dynamics. Nonlinear Dyn
11. Kerrod E, Geddes AM, Regan M, Leach S (2005) Surveillance and control measures during smallpox outbreaks. Emerg Infect Dis 11(2)
12. Kipf T, Welling M (2017) Semi-supervised classification with graph convolutional networks. In: International conference on learning representations (ICLR)
13. Laval F, Grosset-Janin A, Delon F, Allonneau A, Tong C, Letois F, Couderc A, Sanchez M, Destanque C, Biot F, Raynaud F, Bigaillon C, Ferraris O, Simon-Loriere E, Enouf V, Andriamanantena D, de San9 VP, Javelle E, Merens A (2021) Lessons learned from the investigation of a covid-19 cluster in Creil, France: effectiveness of targeting symptomatic cases and conducting contact tracing around them. BMC Infect Dis 21(457)
14. Ma Y, Tian Y, Moniz N, Chawla N (2023) Class-imbalanced learning on graphs: a survey. arXiv:2304.04300
15. Manzo G, Rijt A (2020) Halting SARS-CoV-2 by targeting high-contact individuals. J Artif Soc Soc Simul 23(4)
16. Masthi R, Swamygowda P (2018) An exploratory study on rabies exposure through contact tracing in a rural area near Bengaluru, Karnataka, India. PLOS Negl Trop Dis
17. Megnin-Viggars O, Carter P, Melendrez-Torres GJ, Weston D, Rubin GJ (2020) Facilitators and barriers to engagement with contact tracing during infectious disease outbreaks: a rapid review of the evidence. PLoS ONE
18. Nebeker, C., Kareem, D., Yong, A., Kunowski, R., Malekinejad, M., Aronoff-Spencer, E.: Digital exposure notification tools: A global landscape analysis. PLoS Digital Health (2023)
19. Newman M, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69(2)
20. Pei S, Kandula S, Vega JC, Yang W, Foerster S, Thompson C, Baumgartner J, Ahuia SD, Blanev K, Varma JK, Long T, Shaman J (2022) Contact tracing reveals community transmission of covid-19 in New York city. Nat Commun
21. van Rijsbergen CJ (1979) Information retrieval. Buttersworth
22. Scaselli F, Gori M, Tsoi A, Hagenbuchner M, Monfardini G (2009) The graph neural network model. IEEE Trans Neural Netw 20(1)
23. Shrivastava S, Shrivastava P (2017) Role of contact tracing the Ebola 2014 outbreak: a review. Afr Health Sci
24. Lev T, Shmueli E (2021) State-based targeted vaccination. Appl Netw Sci 6(6)
25. Tang S, Mao Y, Jones R, Tan Q, Ji J, Li N, Shen J, Lv Y, Pan L, Ding P, Wang X, Wang Y, MacIntyre CR, Shi X (2020) Aerosol transmission of SARS-CoV-2? evidence, prevention and control. Environ Int
26. Tellier R (2006) Review of aerosol transmission of influenza a virus. Emerg Infect Dis
27. Wilson A, Aviles N, Petrie J, Beamer P, Szabo Z, Xie M, Mclllece J, Anad Y, Son YC, Halai S, Ernst TW, Masel J (2022) Quantifying SARS-CoV-2 infection risk within the google/apple exposure notification framework to inform quarantine recommendations. Risk Anal 42(1)
28. World Health Organization: Contact tracing and quarantine in the context of covid-19 (2022)

29. Wu F, Xiao A, Zhang J, Moniz K, Endo N, Armas F, Bushman M, Chai P, Duvallet C, Erickson T, Foppe K, Ghaeli N, Gu X, Hanage W, Huang K, Lee W, McElroy K, Rhode S, Matus M, Wuertz S, Thompson J, Alm E (2021) Wastewater surveillance of SARS-CoV-2 across 40 u.s. states from February to June 2020. Water Research
30. Yang S, Senapati P, Wang D, Bauch C, Fountoulakis K (2021) Targeted pandemic containment through identifying local contact network bottlenecks. PLoS Comput Biol