

Network-Side Digital Contact Tracing on a Large University Campus

Matthew L. Malloy, Lance Hartung, Steve Wangen, Suman Banerjee
University of Wisconsin, Madison, United States

ABSTRACT

We describe a study conducted at a large public university campus in the United States which shows the efficacy of network log information for digital contact tracing and prediction of COVID-19 cases. Over the period of January 18, 2021 to May 7, 2021, more than 216 million client-access-point associations were logged across more than 11,000 wireless access points (APs). The association information was used to find potential contacts for approximately 30,000 individuals. Contacts are determined using an AP colocation algorithm, which supposes contact when two individuals connect to the same WiFi AP at approximately the same time. The approach was validated with a truth set of 350 positive COVID-19 cases inferred from the network log data by observing associations with APs in isolation residence halls reserved for individuals with a confirmed (clinical) positive COVID-19 test result. The network log data and AP-colocation have a predictive value of greater than 10%; more precisely, the contacts of an individual with a confirmed positive COVID-19 test have greater than a 10% chance of testing positive in the following 7 days (compared with a 0.79% chance when chosen at random, a relative risk ratio of 12.6). Moreover, a cumulative *exposure score* is computed to account for exposure to multiple individuals that test positive. Over the duration of the study, the cumulative exposure score predicts positive cases with a true positive rate of 16.5% and missed detection rate of 79% at a specified operating point.

ACM Reference Format:

Matthew L. Malloy, Lance Hartung, Steve Wangen, Suman Banerjee. 2022. Network-Side Digital Contact Tracing on a Large University Campus. In *Proceedings of The 28th Annual International Conference on Mobile Computing and Networking (MobiCom '22)*. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/1122445.1122456>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MobiCom '22, September 2022, Planet Earth

© 2022 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . \$15.00
<https://doi.org/10.1145/1122445.1122456>

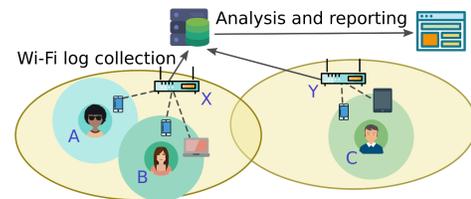


Figure 1: Network-based digital contact tracing. Two WiFi APs (denoted X and Y) are shown. Two users (denoted A and B) connect simultaneously (i.e., colocate) to access point X, supposing epidemiological contact.

1 INTRODUCTION

Digital contact tracing – the use of digital devices such as mobile phones to establish potential epidemiological contacts – has received significant attention as a potential tool in the COVID-19 pandemic. Most proposed digital contact tracing approaches require installation of a mobile app or operating system level access which can be significant hurdles in countries where high public compliance is expected [29, 37] and a fatal flaw elsewhere [14, 29].

Conversely, network-side approaches to digital contact tracing do not require installation of a mobile application, operating system customization (i.e. Google and Apple’s exposure notification system[13]), or any information collected on a client device. Instead, they rely on network-side information such as connection logs to trace potential contacts. As network-side approaches can be enabled by network operators without burden on end-users, they are attractive for large-scale, automated contact tracing.

Access point (AP) colocation is the co-occurrence of users on the same WiFi access point (AP) at the same time. As WiFi APs have limited range, AP-colocation is a proxy for physical colocation, which supposes epidemiological contact. Like physical colocation, prolonged AP-colocation with an infected individual may correlate with increased risk of contracting the infectious disease. To infer contacts and predict future infections, we present an *AP-colocation algorithm*. The algorithm generates a confidence score between two

individuals that increases as the duration and number of AP-colocations increase. More precisely, with an input corpus of connection logs, the algorithm outputs a time-varying weighted *contact graph*. The nodes in the graph correspond to individuals (or proxies for individuals, such as digital devices), and edges in the graph represent contacts between individuals. The weights of the edges are the confidence scores, which depend on the number of times two individuals colocate on a single AP and the number of other individuals connected to that AP. The approach also assigns a cumulative *exposure score* to individuals, which increases as multiple neighboring nodes in the contact graph test positive.

While approaches for digital contact tracing based on colocation have been proposed in the past [25, 43, 44], the primary contribution of this paper is a study of AP-colocation for digital contact tracing of COVID-19 at a large public university campus in the United States. The campus wireless network consists of more than 11,000 WiFi APs covering approximately one square mile and under normal circumstances serves around 50,000 students, employees, and visitors on a typical day. The APs are located primarily inside residence halls, classroom buildings, libraries, dining halls, research and administrative buildings, shared outdoor spaces, and other spaces typical of a large university campus. As students and employees move throughout the campus, their mobile devices connect and disconnect from the APs. The APs log connections and disconnections, creating a record of the approximate location of the user and others in their proximity. We apply the AP-colocation algorithm to a dataset of over 216 million WiFi association records collected over the duration of the Spring 2021 semester. The resulting contact graph exhibits immense scale, and we report on its statistics.

To validate our approach, a truth set of positive (350) and negative (6,101) COVID-19 cases is inferred from the WiFi association dataset. Positive cases are inferred by observing client-AP associations in dormitories reserved for COVID-19 isolation of individuals with a confirmed clinical positive. Likewise, negative cases are inferred from associations in residence halls not reserved for isolation, which require a twice-a-week negative test result. The ground truth dataset enables validation of the utility of the contact graph and the exposure scores for prediction of positive COVID-19 cases.

Results indicate that the use of network log data and AP-colocation has a predictive value of greater than 10% over the course of the study (above 16% under some parameter choices resulting in limited scale). More precisely, when tuned to return 2 contacts per positive case (on average), the returned contacts have greater than 10% chance of having a confirmed positive COVID-19 result in the following 7 days. This is contrasted with the 0.79% chance of a positive result in the next 7-days when a contact is selected at random. To exploit when an individual is exposed to multiple positive

cases, an exposure score is described and computed for each individual in the study. For particular algorithm parameters, the cumulative exposure score predicts positive cases with a true positive rate of 16.5% and missed detection rate of 79%.

While the approach shows promise in settings such as a university or large corporate campus, there are significant shortcomings to using WiFi log data for digital contact tracing. First, individuals must carry on their person a digital device that associates with the network. Estimation of the percentage of individuals on the campus that do not associate with the enterprise WiFi was outside the scope of this study. Second, it is possible and likely common for individuals that connect to the same AP to never come within a distance that supposes disease transmission. As such, there are inevitable false positives (and missed detections), and the approach is best suited to establishing contacts associated with repeated and long term interaction between individuals. As with any digital contact tracing, there are significant privacy considerations that must be addressed, and potential privacy risk must be contrasted with the benefit of such a system. We discuss these trade-offs and note this study is meant to be a starting point for further conversation. In light of these limitations, while the study suggests that digital contact tracing using network-side information can be effective, we recommend that it augment traditional contact tracing.

Lastly, although this study was conducted at a large university campus, the ideas and techniques can be extrapolated to other settings in which network log data is collected. In particular, both cellular network operators and entities in the digital advertising ecosystem collect the information required to implement network-side digital contact tracing in some form. We refer the reader to [25].

In summary, this paper proposes an approach for digital contact tracing based on network log data and describes a study conducted at a large public university campus in the United States. To the best of our knowledge, this is the first study in which ground truth COVID-19 cases are used for validation of network-side contact tracing.

2 DATA

Data was collected at a large university campus in the United States during the Spring 2021 semester from January 18th 2021 through May 7th, 2021. The data was collected from 11,964 physical Wi-Fi APs and 16 Aruba Networks enterprise network controllers. Automated log files containing association and disassociation event notifications were collected from the network controllers on a nightly basis. On average, a single day's log files contain 2,390,087 association events and 1,252,451 disassociation events during February 2021 (the first full month of the study in which students were present on campus and classes were in session).

Our study relies on the association records, which take the form (*anonymized MAC address i, AP ID k, timestamp t*), approximately localizing a user’s device at a specific time. Associations are only logged if the user also successfully authenticates with the network. This excludes records corresponding to randomized MAC addresses (i.e, those present in datasets collected from sniffing WiFi traffic such as probe requests outside an enterprise network). For security and auditing purposes, the network also records logs of authentication events. A sanitized dataset comprising 1-to-many mappings of securely hashed user IDs to MAC addresses was collected. Using the authentication dataset, we are able to group devices owned by the same user, which is crucial to interpreting our results. The anonymized mapping was used to convert the association records to the form (*anonymized user ID i, AP ID k, timestamp t*).

One limitation of the logging of the Aruba Network controllers is that approximately half of all recorded WiFi sessions ended without an explicit disassociation message, likely due to client devices roaming out of the AP’s range. The absence of a clear session duration makes it challenging to implement contact tracing based on a precise calculation of the duration of AP colocation. Instead, in a method that also facilitates faster calculation, we discretize time into 15-minute epochs and record instances of (*anonymized user ID i, AP ID k, epoch timestamp t*). Repeated associations during an epoch are merged into a single record, and disassociation events are disregarded for this study.

Despite the scale and complexity, the data pipeline was remarkably reliable. Nonetheless, some level of attrition in collection was experienced, as software issues gradually prevented a subset of APs from reporting log messages. On three separate occasions, the APs from entire buildings were removed from our collection pool permanently as summarized in the table below. Affected buildings included dormitories, teaching spaces, and dining halls, but none of the isolation dorms were impacted. Visibility into 66% of the campus was retained through the end of the Spring semester.

Date	APs Reporting	Buildings Included
January 18, 2021	11,964 (100%)	207 (100%)
March 24, 2021	10,011 (84%)	174 (84%)
April 9, 2021	8,066 (67%)	170 (82%)
April 25, 2021	7,927 (66%)	164 (79%)

Table 1: APs and buildings during the study duration.

2.1 Truth Set

A truth set of positive and negative COVID-19 cases of residents of on-campus housing was inferred from the network

traffic. In the Spring 2021 semester, on-campus students were required to take twice-weekly rapid saliva-based COVID-19 tests. Students who tested positive (and were residents of university housing) were required to move into one of five designated isolation dormitories. Since all campus dormitories including the isolation dorms are covered by campus WiFi infrastructure, this allowed inference of positive cases based on extended and repeated observation of MAC address association to WiFi APs in the isolation dormitories.

Likewise, an assumed negative was inferred by extended observation of a MAC address in campus residence halls not reserved for isolation. Since students were tested twice weekly and positive cases were quickly moved to the isolation dormitories, an individual in a residence hall not reserved for isolation was an assumed negative. Full details of the approach are included in the Appendix.

Ultimately, the inferred ground truth data-set consisted of anonymized user IDs, an indicator if the user had an inferred positive test, and if so, the date and time at which they were observed to connect to an AP in the isolation dorm.

3 METHODOLOGY

In this section we discuss the methodology used to predict potential exposure to infected individuals. The approach requires first constructing a *contact graph* followed by using the graph to predict contacts and ultimately new positive cases.

3.1 Contact Graph

To analyze and predict future cases from the WiFi association data, we construct a weighted, undirected, time-varying graph $G(t)$. Following standard notation, a graph (or network) consists of a set of nodes V and set of edges E . An edge $e(t) \in E$ is a two element subset of the node set with an associated weight, $e(t) = (i, j, w) \in V \times V \times \mathbb{R}$.

In an epidemiological contact graph, nodes correspond to an individuals (or surrogate for an individual, such as device identifier or a MAC address). An edge represents a potential epidemiological contact between two individuals. To control precision and recall, confidence scores – denoted $w_{i,j}$ – are assigned to edges. Larger weights represent a high potential for epidemiological exposure and higher likelihood for disease transmission. In the presentation of the algorithm we assume time has been discretized into epochs $t = 1, 2, \dots$. The algorithm is described as follows.

Algorithm 1 takes client-AP associations as input, and computes a weighted (undirected) graph, similar to the approach proposed in [25]. The algorithm has two parameters: an optional scaling parameter $\alpha \geq 0$, and a look-back duration τ_g . Colocation between IDs prior to time $t - \tau_g$ are excluded from the contact graph $G(t)$. τ_g is chosen to be

Algorithm 1 AP Colocation Contact Graph

-
- 1: **parameters:** look-back duration $\tau_g > 0$, scaling parameter $\alpha \geq 0$ (default $\tau_g = 7$ days, $\alpha = 1$)
 - 2: **input:** AP associations: (device/user ID i , AP ID k , epoch $t' \leq t$)
 - 3: $V =$ set of unique device/user IDs
 - 4: **for** each AP k , each epoch $t' \in [t - \tau_g, t]$
 - 5: $N_{k,t'}$ = number of IDs on AP k on epoch t'
 - 6: **for** all pairs of IDs (i, j) on AP k
 - 7: $w_{i,j,k,t'} = \frac{1}{N_{k,t'}^\alpha}$
 - 8: $w_{i,j} = \sum_{k,t'} w_{i,j,k,t'}$ for all (i, j)
 - 9: $E = \{(i, j, w_{i,j})\}$
 - 10: **return** $G(t) = (V, E)$
-

longer than the incubation period for the disease, and excludes contacts that happened sufficiently far into the past. The parameter α captures how the confidence score scales with number of devices connected to a common AP. If $\alpha = 1$, the score is inversely proportional with the number of devices on the AP, while when $\alpha = 0$, the edge weight between two users is the count of epochs and APs for which the users collocated. A large value of α dilutes the effect of high volume APs, such as those found in dining halls.

Alg. 1 assumes that time has been discretized into epochs. The algorithm proceeds as follows: for each epoch t' and access point k , the algorithm computes a corresponding weight between all IDs collocated on the AP. Two users collocate on an AP if they both associate during a single epoch. Weights are summed over valid epochs and APs to create edge weights. The pairs of users that collocate on at least one epoch and their associated weights define the (time varying) graph $G(t)$.

3.2 Predicting Positive Cases

After construction of the contact graph, positive cases can be predicted. If user i has a (clinical) positive result, all neighbors of user i (at the time of the positive result) with an edge weight above γ are returned as predicted positives.

When a disease is highly prevalent in a population, multiple contacts of an individual may test positive, increasing the chances of transmission to that individual. In general, this is not captured by traditional contact tracing, as the contacts of each positive case are identified independently.

To capture the potential predictive power of knowledge that multiple contacts have tested positive, we introduce the notion of an exposure score. The exposure score is cumulative: for a single individual, the confidence scores associated with all positive contacts are summed. More precisely, let t_j be the time at which user j tests positive. The exposure

score for user i is defined as

$$s_i(t) = \sum_{j \in N_p: t \in [t_j, t_j + \tau_s]} w_{i,j}(t_j)$$

where N_p is the set of nodes that test positive during the study. Since $w_{i,j} = 0$ for non-neighboring nodes, the sum is taken over neighbors of node i that have a clinical positive test result.

4 RESULTS AND VALIDATION

4.1 AP-Colocation Graph

The methodology of Sec. 3 was applied to a campus-wide dataset of more than 216 million client-AP associations over the duration of the study. As the graph is time-varying, we first report on the characteristics of the contact graph with $\tau_g = 7$ during the time period February 1, 2021 through February 8, 2021. We note that since a mapping between a MAC address and a user was only available for a subset of this data, the campus-wide contact graph was generated such that each node corresponds to a MAC address as opposed to an individual.

Campus-wide contact graph	count
client-AP associations	16,124,734
nodes, $ V(t) $	47,415
edge count, $ E(t) $	2,242,934
average degree ($w_{i,j} > 0$)	94.6

Table 2: Statistics of the campus-wide contact graph $G(t)$, $\tau_g = 7$ days, $t =$ February 8, 2021.

4.2 Validation

While the methodology was applied to a campus-wide dataset, only a subset of these associations correspond to individuals for which ground truth data was available. The subset of data corresponding to these individuals was used to create a *labeled* contact graph. The labeled contact graph G consists of 6,451 nodes (individuals), of which 350 have a ground truth positive result during the course of the study. The remaining 6,101 individuals are assumed to be negative as described in Sec. 2 and the Appendix, for a positivity rate of 5.7% over the duration of the study.

For each individual that tests positive in the labeled dataset, the graph $G(t_i)$ was used to predict contacts, where t_i denotes the time at which the individual tests positive. A predicted positive contact is a neighbor of i with a confidence score above a threshold γ ; i.e., a predicted positive contact is a node j such that $w_{i,j}(t_i) \geq \gamma$. If such a neighbor tests positive in the following $\tau_p \in \{7, 28\}$ days, a true positive (TP) event is

recorded. If such a neighbor does not test positive, a false positive (FP) event is recorded.

Let P denote the set of individuals with a ground truth positive over the course of the study and $V(t)$ denote the set of individuals in the study at time t (an individual is excluded from the study after testing positive). The positive predictive value (PPV) is defined as

$$\begin{aligned} \text{PPV} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ &= \frac{\sum_{i \in P} \sum_{j \in P} \mathbb{I}\{w_{i,j}(t_i) \geq \gamma \cap t_i < t_j \leq t_i + \tau_p\}}{\sum_{i \in P} \sum_{j \in V(t_i)} \mathbb{I}\{w_{i,j}(t_i) \geq \gamma\}}. \end{aligned} \quad (1)$$

where TP and FP represent the count of false positives and true positives over the duration of the study, and $\mathbb{I}\{\cdot\}$ is the indicator function.

Validation Set	count
individuals (nodes) at start, $ V(0) $	6,451
individuals (nodes) at end, $ V(t_e) $	6,101
positive cases (nodes), $ P $	350

Table 3: Statistics of the validation set.

For comparison, the positive predictive value of contacts chosen at random was calculated. Again let i and j index the individuals (nodes), and t_i, t_j denote the respective time at which the individual tests positive, then

$$\text{PPV}_{\text{rand}} = \frac{1}{|P|} \sum_{i \in P} \frac{\sum_{j \in P} \mathbb{I}\{t_i < t_j \leq t_i + \tau_p\}}{|V(t_i)|}.$$

	parameters (days)	PPV	scale
Alg. 1	$\tau_p = 7, \tau_g = 7$	5.0%	5.2
Alg. 1	$\tau_p = 7, \tau_g = 7$	10.0%	1.7
Alg. 1	$\tau_p = 7, \tau_g = 7$	12.5%	0.9
Alg. 1	$\tau_p = 28, \tau_g = 7$	5.0%	8.4
Alg. 1	$\tau_p = 28, \tau_g = 7$	10.0%	2.1
Alg. 1	$\tau_p = 28, \tau_g = 7$	12.5%	1.1
PPV _{rand}	$\tau_p = 7$	0.79%	NA
PPV _{rand}	$\tau_p = 28$	2.12%	NA

Table 4: Validation results at various operating points, compared with predicted contacts chosen at random.

Scale is defined as the average number of contacts that are returned for a given contact score threshold. More specifically,

$$\text{scale} = \frac{\text{TP} + \text{FP}}{|P|} = \frac{\sum_{i \in P} \sum_{j \in V(t_i)} \mathbb{I}\{w_{i,j}(t_i) \geq \gamma\}}{|P|}.$$

Scale and PPV are shown in Fig. 2 and Fig. 3 for the following parameter settings: $\alpha = 1, \gamma \in [0, 20], \tau_g \in \{7, 28\}$

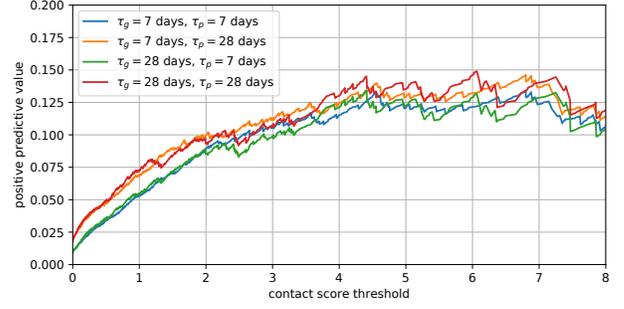


Figure 2: Percent of predicted contacts that test positive in the τ_p -days ($\tau_p \in \{7, 28\}$) after a positive case as a function of the threshold γ , denoted PPV. Contact graph with $\tau_g \in \{7, 28\}$. $\alpha = 1$.

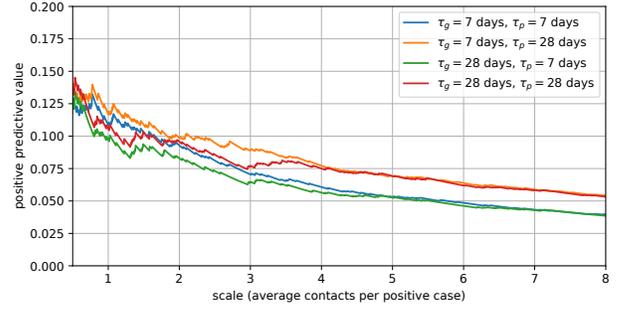


Figure 3: Scale (average number of predicted contacts) vs. positive predictive value for $\tau_p \in \{7, 28\}$ and $\tau_g \in \{7, 28\}$. $\alpha = 1$.

days, $\tau_p \in \{7, 28\}$ days. Additional results for a variety of parameter choices are shown in Fig. 7 through Fig. 9.

Fig. 10 show analysis of the sensitivity to participation in the study. In particular, of the 6,451 individuals, 25% and 50% were excluded at random, and positive predictive value and scale analysis was repeated. This resulted in a study size of 4,840 users and 256 positives for 75% participation, and 3,226 users and 159 for 50% participation (contrasted with 6,454 and 350 for full participation). The positive predictive value does not change significantly since both the numerator and denominator of (1) are reduced by approximately the same amount. The scale (average returned contacts per positive) decreases proportional to the number of devices in the study.

Fig. 4 shows a histogram of the number of plausible transmissions, denoted \hat{R} . A *plausible transmission* is the number of contacts of a positive case (with edge weight $w \geq \gamma$) that test positive in the following τ_p days. More precisely, let i

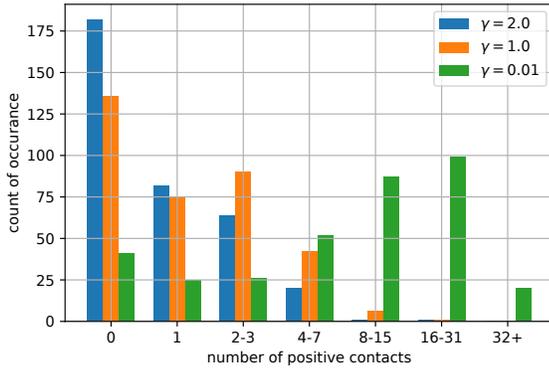


Figure 4: Histogram showing the number of plausible transmissions from each positive case. The x-axis indicates number of contacts with an edge weight $w \geq \gamma$ and a confirmed positive in the following $\tau_p = 7$ days, and the y-axis shows the count of occurrences out of the 350 positive cases. $\tau_g = 7$ and $\alpha = 1$.

index a user with a confirmed positive at time t_i . Then the plausible transmissions for case i is given as

$$\widehat{R}_i = \sum_{j \in P} \mathbb{I}\{w_{i,j}(t_i) \geq \gamma \cap t_i < t_j \leq t_i + \tau_p\}.$$

Fig. 4 shows histogram of \widehat{R}_i over the 350 positive cases.

In addition to the validation of the confidence score produced by the graph, the exposure score as a predictor of a future positive COVID-19 test was validated. The *exposure score* aims to capture the potential predictive power of knowledge that multiple contacts of an individual have tested positive. The exposure score $s_i(t)$ of ten individuals (chosen randomly) over the course of the study are shown in Fig. 5. The left figure in Fig. 5 shows the score corresponding to individuals that test positive. The date of the positive test is shown on the plot. In all cases, the exposure score is elevated prior to positive result. The right figure shows example traces of individuals that do not test positive. Note that the y-axis is scaled for each subplot.

The cumulative exposure score was also used to predict positive cases. For each individual, a positive prediction was declared at the first time the exposure score exceeded a threshold γ . If the time of the prediction was *after* the individual had a ground truth positive, the individual was excluded. If the time of the prediction preceded a ground truth positive test by less than τ_p days, a true positive (TP) was recorded; if the time of the prediction preceded a ground truth positive test by more than τ_p days, or the individual did not have a ground truth positive over the course of the study, a false positive (FP) was recorded. Likewise, if the confidence score

of the individual was below γ for the duration of the study and the individual did not test positive, a true negative (TN) was recorded. If the confidence score of the individual was below the threshold for the duration of the study and the individual tested positive, a false negative (FN) was recorded. The true positive rate is:

$$\text{true positive rate} = \frac{TP}{TP + FP}.$$

Likewise, the missed detection rate is the percentage of total positive cases that are not predicted

$$\text{missed detection rate} = \frac{FN}{TP + FN}.$$

Fig. 6 show the true positive rate and missed detection rate for a number of values of γ over the duration of the study. Note that γ can be used to select an appropriate operating point: the approach can operate with a true positive rate of 16.5% and missed detection rate of less than 80%, or, for example, a missed detection rate below 20% and true positive rate above 5%.

Fig. 11 and Fig. 12 show the true positive rate, missed detection rate, and risk ratio when positive cases are predicted by an exposure score that exceeds a threshold γ for a variety of algorithm parameters, *as a function of time* during the study. Note that Fig. 11 and Fig. 12 show results on a *per individual basis*. In other words, the figures show the percent of the predicted *individuals* that later test positive. This is in contrast to Fig. 2 and Fig. 3, which show the average number of contacts of individuals who test positive.

For reference, Fig. 13 show the count of new positive cases in the $\tau_p \in \{7, 28\}$ days following the date on the x-axis.

Lastly, we highlight the power of the dataset with a table that shows potential ‘high-spread’ events. Table 5 shows the top (AP, hour) pairs when sorted by the highest percentage of positive users in the following 14 days. We stress that the table does not imply that transmission occurred during these events.

5 DISCUSSION

The results indicate that the use of network log data and the AP-colocation algorithm has a positive predictive value of greater than 10% (and above 16% under some parameter choices resulting in limited scale). More specifically, at a scale of 1.7 contacts per positive case (on average), those returned contacts have greater than a 10% chance of having a clinical positive COVID-19 result in the following 7 days. At a higher scale, of 5.2 predicted contacts per positive case, the predicted contacts have a 5.0% chance of a clinical positive in the following 7 days (see Table 4 and Fig. 3). This is contrasted with the 0.79% chance of a positive result in the next 7-days when contacts are selected at random from the ground truth dataset. Comparing the 10% predictive value of Alg. 1 to

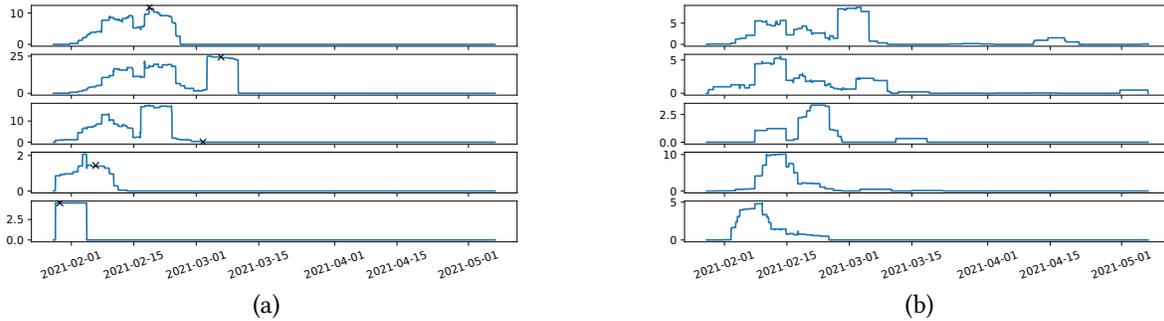


Figure 5: Exposure scores $s_i(t)$ of five users that test positive (Figure a) and negative (Figure b) during the course of the study. The date of the positive test result is indicated by an ‘x’. Note the scale of the y-axis is different for each sub-plot. $\tau_s = 7$ days.

AP	AP location	date and time	total users	positive users $\tau_p = 7$ days	positive users $\tau_p = 14$ days
A	residence hall	2021-02-05, 20:00 - 21:00	14	8	10
B	residence hall	2021-02-10, 02:00 - 03:00	10	5	6
C	residence hall	2021-02-11, 12:00 - 13:00	13	7	7
D	residence hall	2021-02-11, 09:00 - 10:00	12	4	6
E	residence hall	2021-02-10, 13:00 - 14:00	11	3	5
F	dining hall	2021-02-02, 19:00 - 20:00	499	17	25
G	residence hall	2021-02-04, 19:00 - 20:00	165	6	20
H	dining hall	2021-02-02, 19:00 - 20:00	361	12	20
I	residence hall	2021-02-07, 20:00 - 21:00	69	1	13
J	residence hall	2021-02-01, 19:00 - 20:00	53	8	12

Table 5: Potential high-spread events, February 2021. Rows A-E, top (AP, hour) pairs with 10 or more users, sorted by highest percentage of users that test positive in the following $\tau_p = 14$ days. Rows F-J, top (AP, hour) sorted by largest number of positive users in following $\tau_p = 14$ days. Each AP is only listed once.

contacts chosen at random, the WiFi colocation contacts have a more than 12-fold increase in chances of a clinical positive test results in the following 7 days.

To account for exposure to multiple individuals that test positive, the *exposure score* can be employed, providing a stronger predictive power. Note that the false positive rate and false negative rates associated with a fixed threshold vary as the prevalence of the disease in the population varies. Depending on the overall prevalence of the disease, users with an exposure score above a threshold exhibit a more than five-fold increase in odds of testing positive for COVID-19 over those with a score below the threshold (see Fig. 11).

6 ETHICAL CONSIDERATIONS

With the collection of any network data there are natural privacy concerns, and there must be effort to balance the tradeoffs between these concerns and any potential benefit

of such a system to the community. During a global pandemic, the potential benefits of better contact tracing are extremely high. Positive cases of *reportable medical diseases* (such as COVID-19) must be reported to public health officials, who have broad authority to collect data related to contact tracing, whether those records are provided through oral history or electronically. Outside a pandemic, it is unlikely the risk/benefit to such a system is viable. During a pandemic, the specific risk/benefit depends on the prevalence of the disease, the severity of the disease, and the efficacy of other contact tracing approaches. All factors must be considered to establish if such a system is within an acceptable risk tolerance as it pertains to privacy, and we reference work that focuses on the privacy aspects of digital contact tracing [2, 5, 17–19, 21, 31, 36, 39–41].

Graph datasets are attractive from a privacy perspective as their edges capture pair-wise relationships between individuals, and do not disclose the location (i.e, the AP or

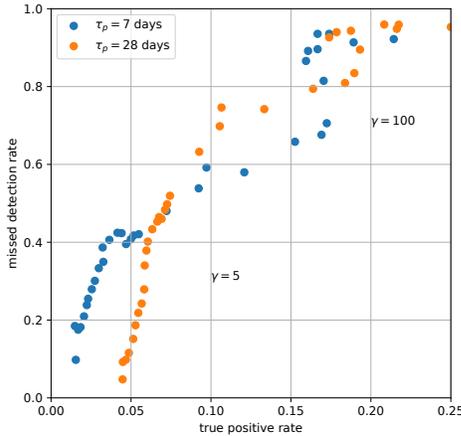


Figure 6: Receiver operating characteristic curve for the exposure score $s_i(t)$ over the duration of the study. The true positive rate plotted against missed the detection rate for values of $\gamma \in [1, 200]$, for $\tau_p = \{7, 28\}$ days. $\tau_g = 7$, $\tau_s = 7$.

physical location). Even if a contact graph is constructed from information that may be considered sensitive (i.e., physical location), it can be stored and analyzed without any of the sensitive information. Specific to our study, additional measures were taken to mitigate any disclosure of sensitive information. All identifiers – MAC addresses, authentication information, AP identifiers – were anonymized via a one way hash before analysis was completed, and the authors of the study never had access to plain text user identifiers.

Lastly, this work has been determined by the Minimal Risk IRB at the University of Wisconsin as not research involving human subjects as defined by the United States Department of Health and Human Services (DHHS) and United States Food and Drug Administration (FDA). We reference the Menlo Report [10] as it establishes ethical principles and provides context for these principles in computing and communication research. The Menlo Report provides guidance with respect to Institutional Review Board (IRB) and self evaluation.

7 RELATED WORK

Digital contact tracing has received a surge of attention since the start of the COVID-19 pandemic. Many of the initial approaches required installation of a mobile application for data collection and rely on location services (i.e., GPS) or proximity sensing using technologies such as Bluetooth low energy (BLE). Similar to app-based approaches, operating system level approaches use the operating system (OS) to

collect data after a user ‘opts-in’, but in both cases, data is collected on the client device. Notably, in a joint venture, Google and Apple developed and released contact tracing technology known as GAEN (Google/Apple Exposure Notification) [13], which relies on a BLE protocol to sense surrounding devices. Both app-based and OS based technologies have been adopted by governments with varying degrees of success. In some instances, where state level governments *highly recommended* contact tracing applications, installation was noted to be under 3% of adults [14]. Furthermore, while the relatively short range of Bluetooth would initially seem an advantage in identifying close contacts, growing evidence of aerosol transmission of SARS-CoV-2 at distances of up to 60 feet [6], which more closely matches indoor WiFi transmission range, suggests that tools based on WiFi would be a useful addition to the contact tracing tool set.

In contrast to approaches based on client-side data collection, network-side contact tracing, such as the approach proposed in this paper, can be implemented without the installation of an app. Closely related to the work in this paper is that of [44, 49]. In [44], the authors propose use of WiFi network association logs gathered by enterprise networks to create a graph data structure which can be used for contact tracing. The approach (including Alg. 1 of [44]) is similar to proposed co-location graph algorithms for other applications, including [12, 24]. The authors of [44] implement their approach and demonstrate its efficacy with WiFi datasets but simulated disease outbreak data, as they do not have access to ground truth COVID test results. Other campus-wide studies related to the COVID-19 pandemic using network log data include [51] which studies super-spreader events using data collected on a university campus. Many additional papers focused on cellular or WiFi networks have surfaced during the COVID-19 pandemic [1, 9, 11, 15, 22, 23, 26–28, 30, 32, 42, 45–48, 50, 52] which do not include ground truth datasets. Overview articles that explore WiFi technologies include [4, 7, 8, 33–35, 38].

Another closely related work is the approach proposed in [25], which relies on data collected by entities in the digital advertising ecosystem, and *IP-colocation* (as opposed to *AP-colocation*). The techniques and construction of the co-location graph are similar. IP-colocation and AP-colocation are likely to exhibit significant overlap. In many homes and small businesses, WiFi users spend significant time in close proximity and share a single public IP address through NAT. Enterprise networks, such as the WiFi network in this study, have more flexibility in managing their IP address space with policy options ranging from assigning each device a unique public address to arbitrarily distributing devices across a pool of public addresses through NAT.

Finally, network-side contact tracing has some overlap with the problem of WiFi localization, particularly from the

vantage point of locating transmitters within a network. Approaches based on received signal strength [3], angle of arrival [20], and time of flight [16] all have the potential to improve contact tracing accuracy by providing an improved measure of physical proximity. We speculate that narrowing the spatial resolution to the room level is likely to be of epidemiological value. Additionally, if the wireless network facilitates tracking of user devices with finer granularity than association and disassociation events, it would enable a more precise calculation of AP colocation duration despite devices roaming without explicit disassociation messages.

8 SUMMARY

Our results suggest that network-side contact tracing can help identify individuals at increased risk of disease plausibly through exposure in a shared environment. In contexts where available, e.g. academic and corporate campuses, Wi-Fi networks can be effectively leveraged to help control disease spread. The approaches discussed here can be extrapolated to other settings in which network log data is collected.

While the confidence scores defined by algorithm 1 are predictive of positive cases, determining actionable and effective criteria to implement interventions is a sizable outstanding effort. In the future, more work must be done before such a system is viable, in particular working with health and administrative officials to determine if notification of possible exposure to SARS-CoV-19 (or other communicable diseases) is warranted.

APPENDIX

Ground Truth

At the start of the Spring 2021 semester, students returned to campus housing facilities but required frequent testing with rapid saliva-based COVID-19 tests. Students who were residents of university housing and tested positive were required to move into one of five designated isolation dormitories or self-isolate in off-campus housing. Access to the isolation facilities was controlled by the university administration, which closely monitored and recorded the daily occupancy of each dormitory. Since all of the campus dormitories including the isolation dorms are also covered by campus WiFi infrastructure, we hypothesized that the WiFi logs would offer insight into devices that moved from regular dorms to isolation dorms without revealing personally identifiable information about the device owner. To this end, we assessed the validity of our hypothesis by comparing the number of devices detected in isolation dorms with the true count of admitted residents supplied by the university.

While visualizing device activity over time, we quickly realized that time of day has a large impact on observed WiFi

activity. During busy hours of the day, we would overestimate the number of residents by a factor of two or more because of the presence of staff. A prior study of device type and user behavior [43] suggests differences in the types of devices most likely to be found at varying hours of the day. In the early hours of the morning when dormitory residents are likely to be present and sleeping and staff or visitors are likely absent, we expect any devices still connected to the WiFi network are most likely to be the personal phones belonging to residents. Whereas phones may maintain a network connection in order to receive updates, most other devices such as laptops and entertainment devices are likely to be in a low power, non-transmitting state. If true, this tendency could not only assist us inferring residence but could also help identify the devices most likely to be carried with the person for contact tracing.

To test the feasibility of estimating building occupancy, we compare the number of unique devices (anonymized MAC addresses) detected during different times of the day with the daily resident counts from the campus health authorities. We frame this as an optimization problem of choosing the best hour of the day to use WiFi device counts to predict the number of residents in a dormitory. More specifically, we count the number of devices detected in the five isolation dorms during a given one hour window for each day over the course of the study and compute the mean squared error (MSE) against the true counts in the respective dorms. We repeat this for each of the 24 hours of the day and find that the one hour window of 4:00-5:00 AM minimizes the prediction error, as shown in figure 16. We find that the resulting device counts are a good approximation of the true number of residents. Figure 14 shows the five isolation dorms with inferred and true resident counts. Compared to the general problem of predicting a building's occupancy from WiFi activity, highly predictable device and user behavior make dorm residency a relatively easier problem.

Inferring Dorm Residents

After finding a correlation between the number of devices detected in the early morning and the number of isolation dorm residents, we turn our attention to identifying devices and their corresponding owners who reside in any of the campus dormitories. Recall that for the purpose of evaluating the effectiveness of our contact tracing approach, we must limit our population to the users who live in campus housing because this is the subset of users for whom we can infer COVID-19 cases based on presence in the isolation dorms. Although we do not have ground truth numbers for Spring semester residents, we take the reported capacity from university housing web pages as an estimate of the true numbers. We find that simply considering any device detected during

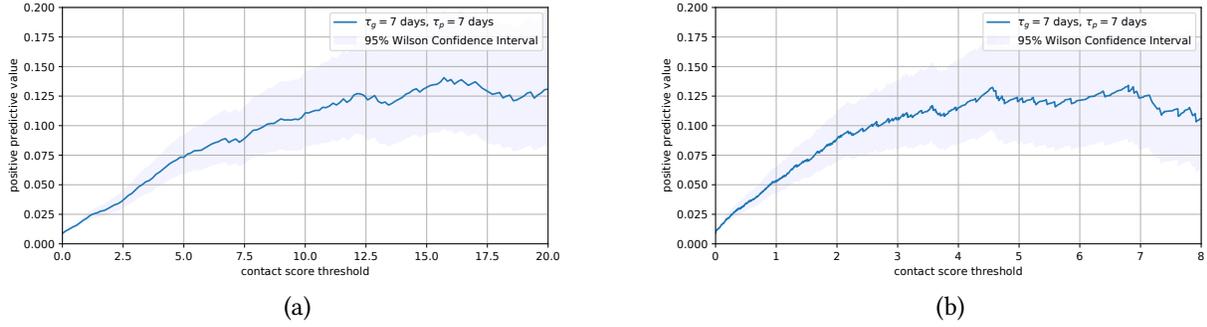


Figure 7: Percent of predicted contacts that test positive in the $\tau_p = 7$ days after a positive case, as a function of the threshold γ . $\tau_g = 7$ days. A 95%-Wilson confidence interval is shown. Figure a, $\alpha = 0$. Figure b, $\alpha = 1$.

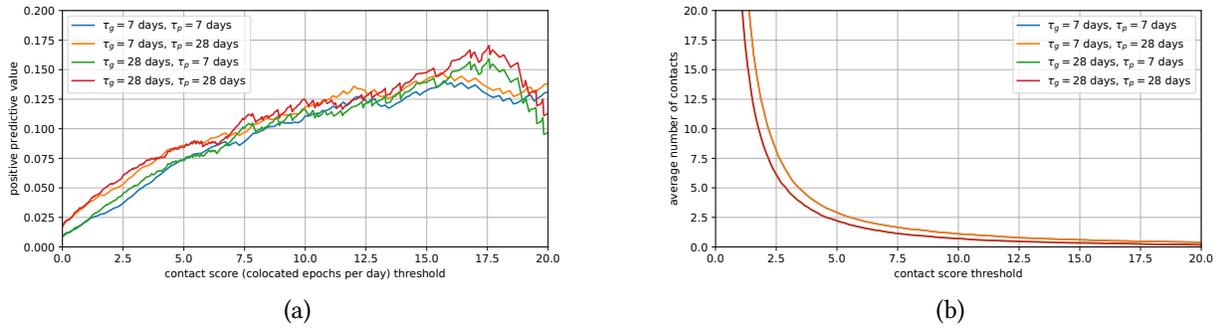


Figure 8: Figure a. $\alpha = 0$. Percent of predicted contacts that test positive in the τ_p -days ($\tau_p \in \{7, 28\}$) after a positive case as a function of the threshold γ , denoted PPV. Contact graph with $\tau_g \in \{7, 28\}$. Figure b. Average number of predicted contacts as a function of contact score γ for $\tau_g \in \{7, 28\}$ and $\tau_p \in \{7, 28\}$. Note that the plots do not depend on τ_p ; two traces are obscured.

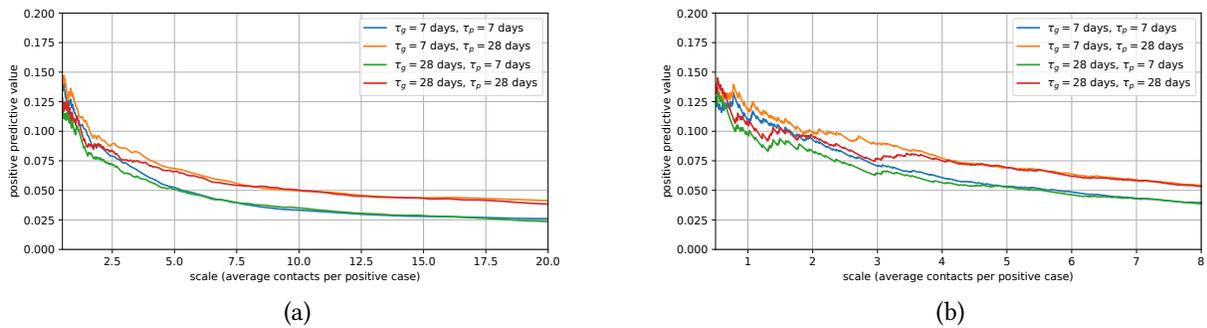


Figure 9: Scale (average number of predicted contacts) vs. positive predictive value for $\tau_p \in \{7, 28\}$ and $\tau_g \in \{7, 28\}$. Figure a, $\alpha = 0$. Figure b, $\alpha = 1$. Note that at some scale (i.e, 5), $\alpha = 1$ has a higher positive predictive value than $\alpha = 0$.

the hour of 4:00-5:00 AM as belonging to a resident leads to a large overcount, perhaps explained in part by visitors, so it is necessary to filter the more probable residents.

This time operating over the range of user IDs, we count the number of days in which any device belonging to a user is detected in a building during the hour of 4:00-5:00 AM.

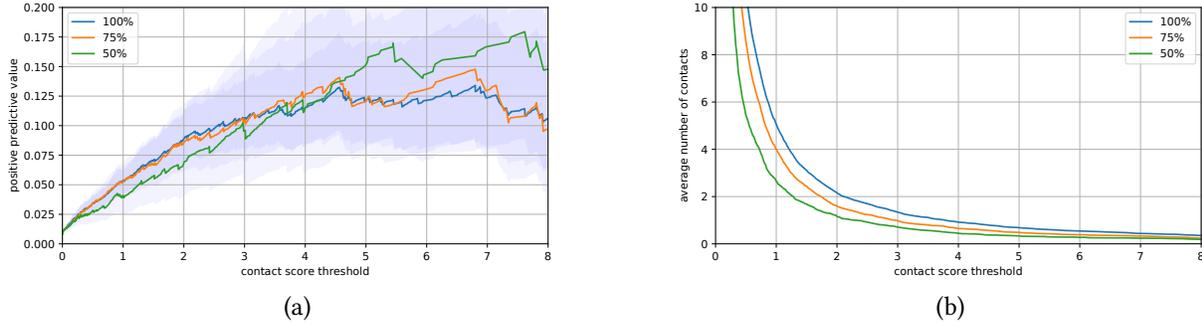


Figure 10: Sensitivity analysis when 100%, 75%, and 50% of the original 6,451 users participate in the study. 95%-Wilson confidence intervals are shown. Note the positive predicted value is largely unaffected by decreased participation, but the average number of contacts (per positive case) decreases accordingly. $\tau_p = 7$ and $\tau_g = 7$. $\alpha = 1$.

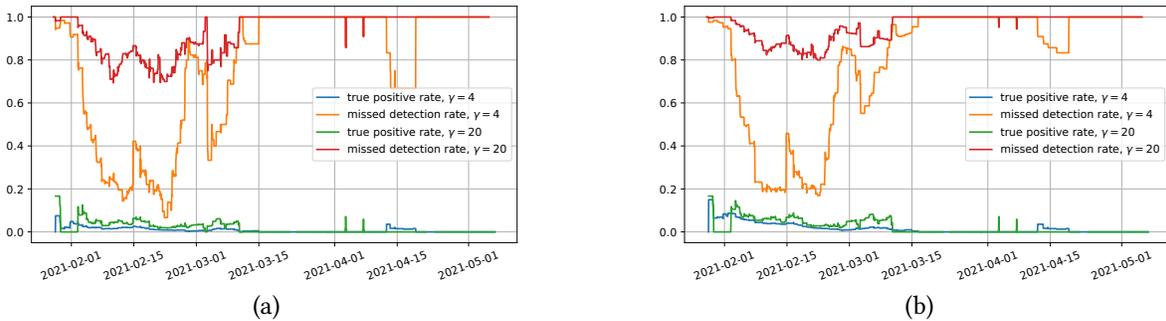


Figure 11: True positive rate – predicted positive individuals that test positive in τ_p days after date, divided by the total number of predicted positive individuals. Missed detection rate – predicted negatives that test positive in τ_p days after the date, divided by count of positive tests in τ_p days after the date. A predicted positive is an individual with a cumulative exposure score above $\gamma \in \{4, 20\}$, and a predicted negative is a an individual with a cumulative exposure score below $\gamma \in \{4, 20\}$. $\tau_g = 7$, $\tau_s = 7$. Figure a, $\tau_p = 7$. Figure b $\tau_p = 28$.

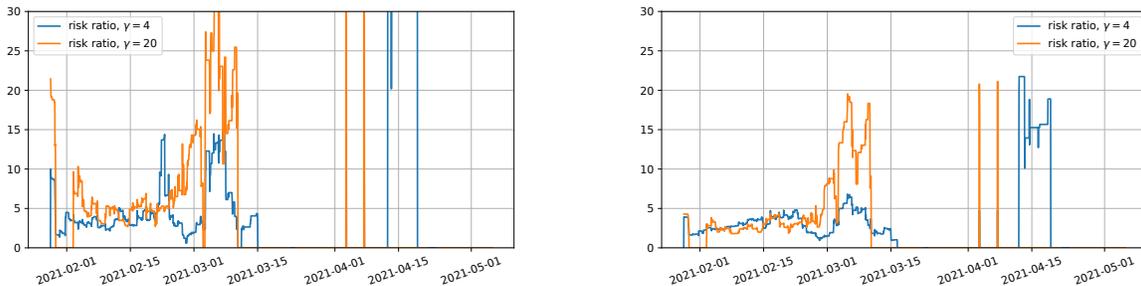


Figure 12: Risk ratio. The risk ratio is defined as the odds of testing positive (in the following τ_p days) given a cumulative exposure score above $\gamma \in \{4, 20\}$, divided by the odds of testing positive (in the following τ_p days) given a cumulative exposure score below $\gamma \in \{4, 20\}$. $\tau_g = 7$, $\tau_s = 7$. Figure a, $\tau_p = 7$. Figure b $\tau_p = 28$.

If multiple devices belonging to the same user are detected on the same morning, we only attribute one detection to

that building. If the number of times the user is detected in a given building meets or exceeds a threshold, τ_r , then

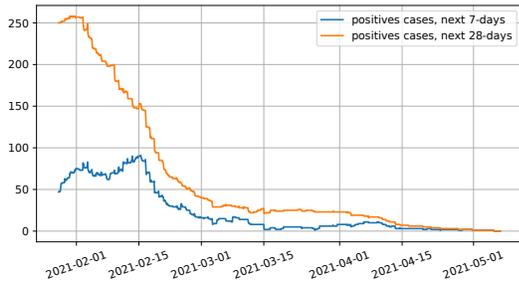


Figure 13: Count of new positive cases in the $\tau_p = 7$ and $\tau_p = 28$ day period after date on the x-axis.

we consider the user a probable resident of that building. For a given value of τ_r , we count the number of inferred residents in each building, which strictly decreases as τ_r increases. Iterating over $\tau_r = 0, 1, 2, \dots, 30$, we find that $\tau_r = 3$ minimizes the MSE against reported dorm capacity numbers, as seen in figure 16. We take the users meeting this threshold as the population of dorm residents for the remainder of our study and disregard non-residents. Although we are able to compute AP colocation metrics for non-resident devices and perform digital contact tracing, we are only able to evaluate efficacy for contact tracing in the COVID-19 pandemic on the subset of devices that belong to residents.

We briefly note two potential issues with this approach. First, this choice of threshold is still expected to result in overestimation because distancing measures and more students choosing to live off campus likely meant that dorms were under capacity during the semester. However, we felt it important not to falsely exclude too many users from our evaluation, and furthermore, overestimating the population of residents is unlikely to bias the results in our favor. Second, under this approach a small number of users may be considered as probable residents of more than one building. Since our goal is to identify the set of dorm residents, we do not consider that a problem, but if we needed to infer a user’s building of residence, some refinements could be made such as selecting the building with the highest detection count.

Inferring Isolation Residents

Finally, we apply a similar approach from the previous section to identify dorm residents who temporarily move into one of the five isolation dorms. Using the same threshold, $\tau_r = 3$, we scan the association logs for users whose device(s) appears in an isolation dorm for at least τ_r mornings and label these as positive cases. Starting from the first time one of the user’s devices is detected in isolation, we find the longest uninterrupted length of time until any of the user’s devices is detected in a different building, and consider that the user’s

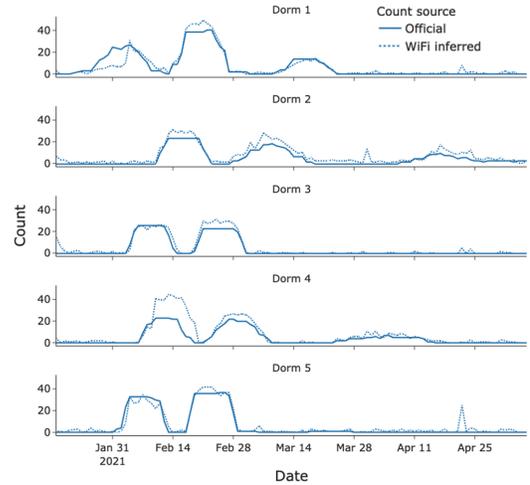


Figure 14: Device counts detected in the early morning hour of 4:00-5:00 AM compared with true number of residents in five isolation dorms.

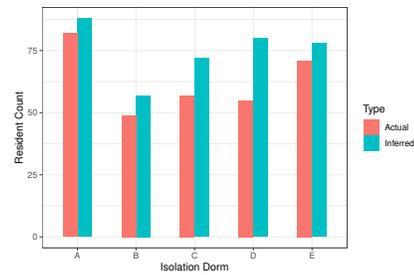


Figure 15: Number of inferred users in each isolation dorm compared with true counts.

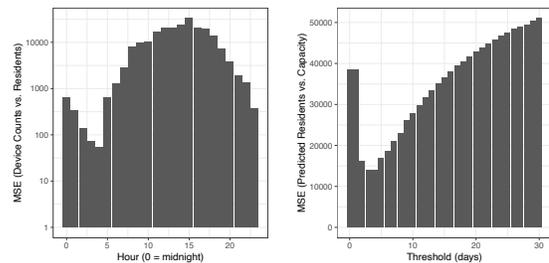


Figure 16: MSE of device count at different times of the day as a predictor of isolation dorm residents. (right). MSE of inferred resident counts compared to reported building capacities for 19 dormitories.

length of stay in isolation. Figure 15 shows the number of users inferred as residents in each isolation dorm compared to the ground truth data for the semester, which we received from campus health authorities.

ACKNOWLEDGEMENTS

Suman Banerjee and Lance Hartung are supported in part through the following US National Science Foundation grants: CNS-1719336, CNS-1647152, CNS-1629833, and CNS-2003129 and an award from the US Department of Commerce with award number 70NANB21H043.

REFERENCES

- [1] Gregory D Abowd. 2020. Using digital technologies to support pandemic response on campus: A case study in the opportunities and challenges of WiFi. (2020).
- [2] Nadeem Ahmed, Regio A Michelin, Wanli Xue, Guntur Dharma Putra, Sushmita Ruj, Salil S Kanhere, and Sanjay Jha. 2021. DIMY: Enabling privacy-preserving contact tracing. *arXiv preprint arXiv:2103.05873* (2021).
- [3] Victor Bahl and Venkat Padmanabhan. 2000. RADAR: An In-Building RF-based User Location and Tracking System. In *Proceedings of IEEE INFOCOM 2000*. Institute of Electrical and Electronics Engineers, Inc. <https://www.microsoft.com/en-us/research/publication/radar-an-in-building-rf-based-user-location-and-tracking-system/> ACM SIGMOBILE Test-of-Time Paper Award, 2016.
- [4] Syed Mohammed Basheeruddin Asdaq, N Raghavendra Naveen, Lakshmi Narasimha Gunturu, Kalpana Pamayyagari, Ibrahim Abdullah, Nagaraja Sreeharsha, Mohd Imran, Abdulkhalik J Alsalmán, Maitham A Al Hawaj, and Abdullah A Alsubaie. 2021. Wireless Networking-Driven Healthcare Approaches in Combating COVID-19. *BioMed Research International* 2021 (2021).
- [5] Lars Baumgärtner, Alexandra Dmitrienko, Bernd Freisleben, Alexander Gruler, Jonas Höchst, Joshua Kühlberg, Mira Mezini, Richard Mitev, Markus Miettinen, Anel Muhamedagic, et al. 2020. Mind the GAP: Security & privacy risks of contact tracing apps. In *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 458–467.
- [6] Martin Z. Bazant and John W. M. Bush. 2021. A guideline to limit indoor airborne transmission of COVID-19. *Proceedings of the National Academy of Sciences* 118, 17 (2021). <https://doi.org/10.1073/pnas.2018995118> arXiv:<https://www.pnas.org/content/118/17/e2018995118.full.pdf>
- [7] Isobel Braithwaite, Tom Callender, Miriam Bullock, and Robert W Aldridge. 2020. Automated and partially-automated contact tracing: a rapid systematic review to inform the control of COVID-19. *medRxiv* (2020).
- [8] Isobel Braithwaite, Tom Callender, Miriam Bullock, Robert W Aldridge, and Isobel Braithwaite. 2020. Automated and partially-automated contact tracing: a rapid systematic review to inform the control of COVID-19. *medRxiv* (2020).
- [9] Stéphane Bressan. 2021. A Data Warehouse of Wi-Fi Sessions for Contact Tracing and Outbreak Investigation. *Transactions on Large-Scale Data-and Knowledge-Centered Systems XLVIII: Special Issue In Memory of Univ. Prof. Dr. Roland Wagner* (2021), 85.
- [10] D. Dittrich and E. Kenneally. 2012. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. In *US Department of Homeland Security Tech Report*.
- [11] Mikhail Dmitrienko, Abhishek Singh, Patrick Erichsen, and Ramesh Raskar. 2020. Proximity Inference with Wifi-Colocation during the COVID-19 Pandemic. *arXiv preprint arXiv:2009.12699* (2020).
- [12] Keith Funkhouser, Matthew Malloy, Enis Ceyhun Alp, Phillip Poon, and Paul Barford. 2018. Device Graphing by Example. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 273–282.
- [13] Alejandro De La Garza. 2020. Apple and Google Partner on COVID-19 Contact Tracing Technology. <https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/> (2020).
- [14] Alejandro De La Garza. 2020. Contact Tracing Apps Were Big Tech’s Best Idea for Fighting COVID-19. Why Haven’t They Helped? *Time Magazine* (November 2020).
- [15] Domenico Giustiniano, Giuseppe Bianchi, Andrea Conti, Stefania Bartoletti, and Nicola Blefari Melazzi. 2021. 5G and Beyond for Contact Tracing. *IEEE Communications Magazine* 59, 9 (2021), 36–41.
- [16] Domenico Giustiniano and Stefan Mangold. 2011. CAESAR: Carrier Sense-Based Ranging in off-the-Shelf 802.11 Wireless LAN. In *Proceedings of the Seventh Conference on Emerging Networking Experiments and Technologies* (Tokyo, Japan) (CoNEXT ’11). Association for Computing Machinery, New York, NY, USA, Article 10, 12 pages. <https://doi.org/10.1145/2079296.2079306>
- [17] Majid Hatamian, Samuel Wairimu, Nurul Momen, and Lothar Fritsch. 2021. A privacy and security analysis of early-deployed COVID-19 contact tracing Android apps. *Empirical Software Engineering* 26, 3 (2021), 1–51.
- [18] Arvin Hekmati, Gowri Ramachandran, and Bhaskar Krishnamachari. 2021. CONTAIN: Privacy-oriented contact tracing protocols for epidemics. In *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 872–877.
- [19] Kilian Holzapfel, Martina Karl, Linus Lotz, Georg Carle, Christian Djefal, Christian Fruck, Christian Haack, Dirk Heckmann, Philipp H Kindt, Michael Köppl, et al. 2020. Digital Contact Tracing Service: An improved decentralized design for privacy and effectiveness. *arXiv preprint arXiv:2006.16960* (2020).
- [20] Kiran Joshi, Steven Hong, and Sachin Katti. 2013. PinPoint: Localizing Interfering Radios. In *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*. USENIX Association, Lombard, IL, 241–253. <https://www.usenix.org/conference/nsdi13/technical-sessions/presentation/joshi>
- [21] Franck Legendre, Mathias Humbert, Alain Merroud, and Vincent Lenders. 2020. Contact tracing: An overview of technologies and cyber risks. *arXiv preprint arXiv:2007.02806* (2020).
- [22] Guanyao Li, Siyan Hu, Shuhan Zhong, Wai Lun Tsui, and S-H Gary Chan. 2021. VContact: Private WiFi-based IoT contact tracing with virus lifespan. *IEEE Internet of Things Journal* (2021).
- [23] Ruofeng Liu, Zhimeng Yin, Wenchao Jiang, and Tian He. 2021. WiBeacon: expanding BLE location-based services via wifi. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 83–96.
- [24] Matthew Malloy, Paul Barford, Enis Ceyhun Alp, Jonathan Koller, and Adria Jewell. 2017. Internet device graphs. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1913–1921.
- [25] Matthew Malloy, Aaron Cahn, and Jon Koller. 2021. Digital Contact Tracing Using IP Colocation. In *2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoT&IS)*. IEEE, 73–78.
- [26] S Yoshita Manavi, Vinuthna Nekkanti, Ram Shankar Choudhary, and N Jayapandian. 2020. Review on emerging Internet of Things technologies to fight the COVID-19. In *2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. IEEE, 202–208.
- [27] Hamid Mcheick, Houssam Hajj Hassan, and Hanane Kteich. 2021. D2D Communication: COVID-19 Contact Tracing Application Using Wi-Fi Direct. In *2021 International Conference on Smart Applications, Communications and Networking (SmartNets)*. IEEE, 1–6.
- [28] Callie Monroe, Faiza Tazi, and Sanchari Das. 2021. Location Data and COVID-19 Contact Tracing: How Data Privacy Regulations and Cell

- Service Providers Work In Tandem. In *Proceedings of the Workshop on Usable Security and Privacy (USEC)*.
- [29] Simon Munzert, Peter Selb, Anita Gohdes, Lukas F Stoetzer, and Will Lowe. 2021. Tracking and promoting the usage of a COVID-19 contact tracing app. *Nature Human Behaviour* (2021), 1–9.
- [30] Khuong An Nguyen, Zhiyuan Luo, and Chris Watkins. 2020. Epidemic contact tracing with smartphone sensors. *Journal of Location Based Services* 14, 2 (2020), 92–128.
- [31] Patrick Ocheja, Yang Cao, Shiyao Ding, and Masatoshi Yoshikawa. 2020. Quantifying the Privacy-Utility Trade-offs in COVID-19 Contact Tracing Apps. *arXiv preprint arXiv:2012.13061* (2020).
- [32] Theodoros Oikonomidis, Konstantinos Fouskas, and Maro Vlachopoulou. 2021. The Role of Wi-Fi Positioning Systems in Safety Against COVID-19. In *Strategic Innovative Marketing and Tourism in the COVID-19 Era: 9th ICSMAT Conference 2020*. Springer, 111–119.
- [33] N Petrović and Đ Kocić. 2021. IoT for COVID-19 Indoor Spread Prevention: Cough Detection, Air Quality Control and Contact Tracing. In *2021 IEEE 32nd International Conference on Microelectronics (MIEL)*. IEEE, 297–300.
- [34] Abhishek Roy, Farooque H Kumbhar, Harpreet S Dhillon, Navrati Saxena, Soo Young Shin, and Sukhdeep Singh. 2020. Efficient monitoring and contact tracing for COVID-19: A smart IoT-based framework. *IEEE Internet of Things Magazine* 3, 3 (2020), 17–23.
- [35] Yesin Sahraoui, Ludovica De Lucia, Anna Maria Vegni, Chaker Abdelaziz Kerrache, Marica Amadeo, and Ahmed Korichi. 2021. TraceMe: Real-Time Contact Tracing and Early Prevention of COVID-19 based on Online Social Networks. *Academia.edu* (2021).
- [36] Pollyanna Sanderson. 2021. Balancing public health and civil liberties: Privacy aspects of contact-tracing technologies. *IEEE Security & Privacy* 19, 4 (2021), 65–69.
- [37] Young Ern Saw, Edina Yi-Qin Tan, Jessica Shijia Liu, and Jean CJ Liu. 2021. Predicting Public Uptake of Digital Contact Tracing During the COVID-19 Pandemic: Results From a Nationwide Survey in Singapore. *Journal of medical Internet research* 23, 2 (2021), e24730.
- [38] Hao-Chen Sun, Xiao-Fan Liu, Zhan-Wei Du, Xiao-Ke Xu, and Ye Wu. 2021. Mitigating COVID-19 Transmission in Schools With Digital Contact Tracing. *IEEE Transactions on Computational Social Systems* (2021).
- [39] Aman Tahiliani, Vikas Hassija, Vinay Chamola, Salil S Kanhere, Mohsen Guizani, et al. 2021. Privacy-Preserving and Incentivized Contact Tracing for COVID-19 Using Blockchain. *IEEE Internet of Things Magazine* 4, 3 (2021), 72–79.
- [40] Qiang Tang. 2020. Privacy-preserving contact tracing: current solutions and open questions. *arXiv preprint arXiv:2004.06818* (2020).
- [41] Qiang Tang. 2021. Another look at privacy-preserving automated contact tracing. *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 8, 2 (2021), 1–27.
- [42] Laxmi Thakare, Ruchi Dohare, and Nadeem Akhtar. 2021. P-Tracer: Proximity Detection for Contact Tracing. In *2021 International Conference on COMMunication Systems & NETWORKS (COMSNETS)*. IEEE, 116–118.
- [43] Ameer Trivedi, Jeremy Gummeson, and Prashant Shenoy. 2020. Empirical Characterization of Mobility of Multi-Device Internet Users. *arXiv:2003.08512 [cs.CY]*
- [44] Ameer Trivedi, Camellia Zakaria, Rajesh Balan, Ann Becker, George Corey, and Prashant Shenoy. 2021. WiFiTrace: Network-based Contact Tracing for Infectious Diseases Using Passive WiFi Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–26.
- [45] Pengjia Tu, Junhuai Li, Huaijun Wang, Kan Wanga, and Yuan Yuan. 2021. Epidemic Contact Tracing with Campus WiFi Network and Smartphone-based Pedestrian Dead Reckoning. *IEEE Sensors Journal* (2021).
- [46] Fan Yi, Yaxiong Xie, and Kyle Jamieson. 2021. Cellular-assisted COVID-19 contact tracing. In *Proceedings of the 2nd Workshop on Deep Learning for Wellbeing Applications Leveraging Mobile Devices and Edge Computing*. 1–6.
- [47] Wonjae Yoo, Hyoungsub Kim, Minjae Shin, and Mark J Clayton. 2020. BIM-Based Automatic Contact Tracing System Using Wi-Fi. (2020).
- [48] Guilherme Augusto Zagatti, Tingfeng Wu, See-Kiong Ng, and Stéphane Bressan. 2021. A Large-scale Disease Outbreak Analytics System based on Wi-Fi Session Logs. In *2021 22nd IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 236–239.
- [49] Camellia Zakaria, Ameer Trivedi, Emmanuel Cecchet, Michael Chee, Prashant Shenoy, and Rajesh Balan. 2020. Analyzing the impact of covid-19 control policies on campus occupancy and mobility via passive wifi sensing. *arXiv preprint arXiv:2005.12050* (2020).
- [50] Jinyan Zang and Latanya Sweeney. 2021. Building A Collocation Detection System Using A Wi-Fi Sensor Array for COVID-19 Contact Tracing in A University Setting. *Case Studies in Public Interest Technology* (2021), 168.
- [51] Cheng Zhang, Yunze Pan, Yunqi Zhang, Adam C Champion, Zhaohui Shen, Dong Xuan, Zhiqiang Lin, and Ness B Shroff. 2021. WLAN-Log-Based Superspreader Detection in the COVID-19 Pandemic. *High-Confidence Computing* (2021), 100005.
- [52] Qingchuan Zhao, Haohuang Wen, Zhiqiang Lin, Dong Xuan, and Ness Shroff. 2020. On the accuracy of measured proximity of Bluetooth-based contact tracing apps. In *International Conference on Security and Privacy in Communication Systems*. Springer, 49–60.