ComMit: Blind Community-based Early Mitigation Strategy against Viral Spread

Pegah Hozhabrierdi
Syracuse University
phozhabr@syr.edu

Sucheta Soundarajan
Syracuse University
susounda@syr.edu

***Abstract*—In the early stages of a pandemic, epidemiological knowledge of the disease is limited and no vaccination is available. This poses the problem of determining an *Early Mitigation Strategy*. Previous studies have tackled this problem through finding globally influential nodes that contribute the most to the spread. These methods are often not practical due to their assumptions that (1) accessing the full contact social network is possible; (2) there is an unlimited budget for the mitigation strategy; (3) healthy individuals can be isolated for indefinite amount of time, which in practice can have serious mental health and economic consequences.**

In this work, we study the problem of developing an early mitigation strategy from a *community perspective* and propose a dynamic Community-based Mitigation strategy, *ComMit*. The distinguishing features of *ComMit* are: (1) It is agnostic to the dynamics of the spread; (2) does not require prior knowledge of contact network; (3) it works within a limited budget; and (4) it enforces bursts of short-term restriction on small communities instead of long-term isolation of healthy individuals. *ComMit* relies on updated data from test-trace reports and its strategy evolves over time. We have tested *ComMit* on several real-world social networks. The results of our experiments show that, within a small budget, *ComMit* can reduce the peak of infection by 73% and shorten the duration of infection by 90%, even for spreads that would reach a steady state of non-zero infections otherwise (e.g., SIS contagion model).

***Index Terms*—community, contagion, spread, mitigation**

I. INTRODUCTION

In response to a viral spread, multiple factors determine the efficacy of different mitigation strategies, namely the epidemiological knowledge of the spread dynamics, the possibility of medical intervention (i.e., vaccination), and the existence of mobility and interaction data [1, 2, 3, 4].

In the early stages of a pandemic, the disease dynamic is unknown, the contact network is partially known at best, and no vaccination is available. These are the challenges against an *Early Mitigation Strategy*. The objective of such a strategy is to minimize the peak and/or total number of infections with the least possible perturbations introduced to the social network (through, e.g., quarantine and isolation approaches) [3, 5, 6]. In contrast, once a vaccine is available, the objective of the *Immunization* problem is to minimize the amount of time it takes to halt the spread effectively with the least amount of vaccine. Despite the similarity of approaches, immunization problem and early mitigation strategy optimize different objective functions and the former does not introduce perturbations to the network; the candidate nodes chosen in an early mitigation setting will be isolated for the duration of the disease (removal of edges in the contact network), but in immunization problem they are vaccinated and no network perturbation is introduced.

The majority of studies on the two mentioned strategies are based on detecting globally influential nodes (e.g., degree centrality [7] and betweenness centrality [8]) that contribute the most to the spread (*targeted strategies*). Despite promising theoretical results, these methods are generally difficult to implement due to their assumption of full knowledge of the contact network [9]. Additionally, complex social networks demonstrate high clustering and individuals tend to form groups (communities) [10], which can both alleviate and aggravate viral expansion [1, 11]. The global centrality measures do not consider the local influence of the node in their respective communities [3]. We argue that for tackling these shortcomings, a practical mitigation strategy should not assume a prior knowledge of the contact network structure and the dynamics of the spread. It also should consider the cost of a certain intervention scheme and avoid isolating healthy members of the population.

In this work, we study the problem of developing an early mitigation strategy from a *community perspective* and propose a dynamic Community-based Mitigation strategy, *ComMit*, that only utilizes geographical information to infer community membership and data from test-trace to update its knowledge of the spread, without enforcing any assumptions about the nature of the disease. Because *ComMit* relies on updated data from test-trace reports, it is dynamic and the mitigation strategy can evolve over time. Unlike previous works, we have designed *ComMit* with two important assumptions: (1) there is no global information on the social network contacts; (2) the candidates for isolation are small clusters instead of single healthy individuals. The second condition aims to minimize the economic and psychological damage ([12], [13]).

Using the information from the test-trace step, *ComMit* introduces appropriate network perturbations to combat the magnitude of the spread based on the current knowledge of the underlying network and testing outcome. These perturbations are aimed to fragment the network communities. *ComMit* achieves that through the *divider* block that forms small clusters of nodes (sub-clusters) that are to be temporarily

This work was supported in part by NSF grant 1908048.

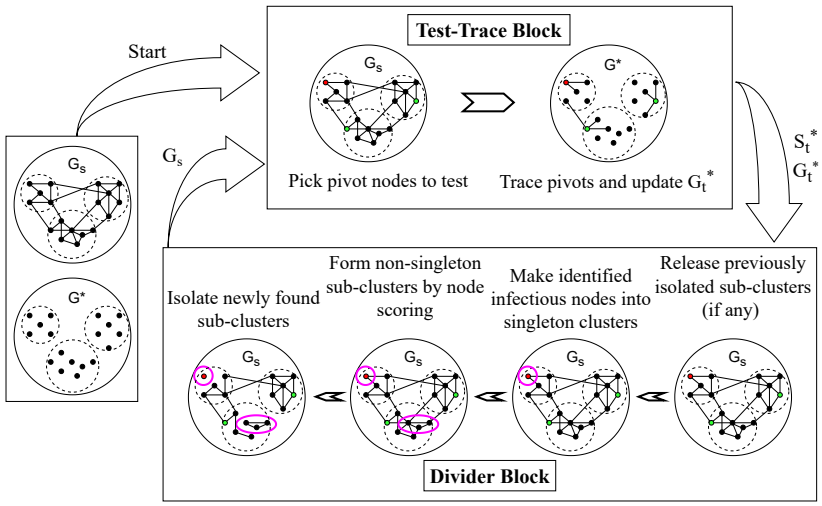

Fig. 1: *ComMit pipeline.* ***Start.*** *Contact network G_s is unknown and the known graph to the algorithm, G^*, is empty. The dashed lines show the communities known to the algorithm. The figure shows the first iteration of the algorithm.* ***Test-Trace.*** *The coloring of the pivot indicates the result of the test (red is infectious). Tracing of the pivots updates the edges of G^*.* ***Divider.*** *The block uses the updated information on G_t^* and identified infections from S_t^* to form sub-clusters (in purple), whose isolation fragments the communities, reducing the magnitude of the spread. The network perturbations by divider updates G_s on which the spread runs. The iteration continues until the termination condition is met (see Section IV-C).*

isolated as a community from the rest of the network. After a certain time has passed these sub-clusters are released back to the network and will not be isolated until some time has passed from their last isolation. The pipeline for *ComMit* is shown in Figure 1.

Our contributions can be summarized as follows,

- We formulate the early mitigation problem based on real-world constraints.
- To the best of our knowledge, we are the first to propose an early mitigation strategy that (1) works with no knowledge of either the social network structure and the spread dynamic; (2) considers the practical cost of the strategy and operates within a limited budget.
- We validate our mitigation strategy, *ComMit*, on five real-world datasets that are obtained from national address database and Copenhagen project (see Section V).
- The results of our experiments show that within its limited budget, *ComMit* is very effective in reducing the peak and duration of infection, reducing them up to 73% and 90%, respectively. In all of our case studies, *ComMit* successfully turns a steady state spread process[1], such as SIS contagion model, into a dying process with a relatively short absorption time[2].
- We open-source the repository containing our processed datasets and code for reproducing the result of the experiments in this paper[3].

[1]A spread dynamic that reaches a steady state of maintaining a non-zero number of infectious nodes.

[2]The time that it takes for the number of infectious nodes to become zero.

[3]https://github.com/Pegayus/ComMit

II. Related Work

Our work involves three bodies of research; targeted intervention strategies against viral spread, the impact of community structure on dynamics of such a spread, and community-based intervention strategies.

A. Targeted Intervention Strategies

Early Mitigation Strategy

Gross et al. [14] is the closest study to ours in modeling a contact network based on geo-spatial data. Their contact network model is a modular 2D lattice in which each module represents a city and each city can only connect to its immediate neighboring module. Their proposed mitigation strategies are social distancing and reducing degrees in and outside of the communities by isolating individuals. Our approaches differ in that *ComMit* (1) does not limit contact network to a 2D lattice; (2) considers a mixture of sampling (testing) and isolation; (3) does not isolate any healthy individuals; (4) does not assume complete knowledge of the contact data.

Immunization Strategy

Rosenblat et al. [2] challenge the popular target-based immunization strategies by studying different immunization methods in the presence of partially observed network data. They conclude that popular targeted methods, such as degree and betweenness centralities, only perform well with little to no missing data, but self-reported local information from sampled individuals compensates for a large volume of missing data. Salathé et al. [15] place a similar emphasis on partially observed networks and propose a heuristic method to find community bridge nodes (CBF that stands for Community

Bridge Finder). They show how targeting bridge nodes for immunization outperforms acquaintance immunization (the only other network structure agnostic method) [16], in which the most frequently visited acquaintances of randomly selected nodes are vaccinated first. CBF relies on random walks and constant path finding between current and visited nodes. This limits its value in a practical dynamic setting where these computations need to run iteratively. In our method, we show how we can avoid such costly (and impractical) computations by leveraging the known geo-spatial communities that are shown to be predictors of contact-based communities [17, 18].

B. Community Structure and Dynamics of Spread

Topîrceano [19] shows the importance of geo-spatial information in predicting the dynamics of an outbreak. This paper offers a Geo-spatial Population Model (GPM) that estimates the predictors of mobility between different regions in a country based on the region's population density. Their results suggest that changing the number of regions and their population density directly impacts the size and duration of the outbreak.

C. Community-based Intervention Strategies

The definition of community in these studies is diverse: from subgraphs with the highest $\frac{\text{number of subgraph intra edges}}{\text{number of subgraph inter edges}}$ and k-cores to geo-spatial and ground truth communities. Serafino et al. [6] showed that disconnecting bridge nodes that connect super-spreader k-cores considerably reduces the radius of the spread. However, they rely on betweenness centrality which is a global measure that requires full knowledge of the contact network.

Yang et al. [20] propose a flow-based edge betweenness measure to minimize the p-norm of the flow between communities in the network. They show that the bridge-based methods are superior to degree-based intervention methods.

Block et al. [12] consider social behavior patterns within communities and propose to (1) limit interaction to few repeated contacts; (2) choose those contacts based on some similarity (e.g., homophily); (3) strengthen contact with those pairs that interact in more than one community. We leverage their finding and that of Topîrceano [19] to design the *fragmentation* step in *ComMit* (see IV-B). The main problem with their strategy is the assumption of full network knowledge.

Yuan et al. [3] emphasize the local importance of the nodes to their community in contrast with their global centrality. They measure how important nodes are to their communities and how important their communities are to the overall network. Their scores are based on eigenvalue and eigenvector pairs obtained from the spectral clustering of the neighborhood matrix. This method is susceptible to edge percolation and loses its performance with partial network data.

III. Problem Statement

Inspired by the findings of Block et al. [12] and Topîrceano [19], we argue that fragmenting network communities into small clusters (*sub-clusters*) and isolating these sub-clusters, rather than isolating individuals, is the best strategy during the early stages of a contagion. We refer to the problem of finding such sub-clusters as the *Network Fragmentation Problem* and it is the backbone of the *ComMit* algorithm.

TABLE I: *Notations.*

Symbol	Definition
Input to blind network fragmentation problem	
$G_t^* = (V, E*_t)$	Learned contact social network at time t
$S_t^*(.)$	Partially known outcome of $S_t(.)$ at time t
$\mathcal{I}_t^*$	Set of identified infected nodes at time t
b_f	Budget for forming sub-clusters (divider block)
Additional input to ComMit algorithm	
$G_g = (V, E_g, W_g)$	Geo-proximity network
$\mathcal{C}_g$	Geo communities inferred from G_g
b_t	Budget for testing (test-trace block)
a_t	Accuracy of self-reports in contact tracing
ϵ_t	Value of ϵ for testing at time t
d_ϵ	Decaying factor for updating ϵ_t
t_r	Restriction period of isolating sub-clusters
Other notations	
$\mathfrak{C}_t$	Set of sub-clusters at time t
$G_s = (V, E_s)$	Contact social network
$S_t(.) = \{s_t^v \mid v \in V\}$	Outcome of spread at time t
$\mathcal{I}_t = \{v \in V \mid s_t^v = I\}$	Set of infected nodes at time t

Assume the beginning of an unknown viral infection within an unknown contact network with a known underlying geo-spatial structure (e.g., the geographic coordinates of domiciles). Suppose we have the power of restraining individuals to limit their interactions within a certain group in exchange for a compensation. This introduces perturbations in the underlying unknown contact network that changes the dynamic of the spread. The main question is how to choose groups of individuals such that isolating them as a group from the rest of the network, while maintaining their inner-group interaction, most efficiently inhibits the spread. This is the network fragmentation problem that we formally define in Section III-D, but first, we discuss the population model, contagion model, and assumptions on network perturbations, as follows. The notations used in this and next section are summarized in Table I.

A. Population Model

Empirical studies on human contact have shown geo-spatial distance to be the most important factor in forming connections [17]. More recent studies on online social networks show the geo-spatial distance also influences the presence of online contacts and they are inversely correlated by a power-law [18]. This observation can be used to compensate for having no knowledge of the contact network structure.

We model our population as a two-layer network consisting of the contact network $G_s = (V, E_s)$ and its underlying geo-location graph $G_g = (V, E_g, W_g)$. Both layers are undirected and share the same set of nodes, V. G_g is a complete graph and a weighted edge $(i, j, {w_g}^{ij}) \in E_g$ indicates a geo-distance of ${w_g}^{ij}$ between nodes i and j. G_s, however, is sparse and an edge $(i, j) \in E_s$ implies the existence of contact between nodes i and j. We assume the distance between individual domiciles and their contact patterns do not change.

The community membership of each node is inferred from G_g, while the infection spreads through the links in G_s. The key underlying assumption is the inverse relationship between

W_g and E_s, as demonstrated in [17, 18]. More specifically, the empirical results in [18] suggest a Zipf's law:[4] i.e., the probability of an edge between nodes (i, j) in G_s is

$$p((i,j) \in E_s) \approx \frac{b}{w_g^{ij}}, \quad 0 < b \leq 1, \quad 1 \leq w_g^{ij} \tag{1}$$

for a constant b. We use this rule in building our datasets in Section V. From the perspective of a mitigation strategy, E_s is partially known. We represent this partially known network at time t by $G_t^* = (V, E_t^*)$. In each iteration, G_t^* is updated by the information from test-trace (Figure 1). If nothing about E_s is known (i.e., in the start of the algorithm, or in the absence of test-trace block), E_t^* is empty.

B. Contagion Model

Consider a viral spread with unknown dynamics, $S_t(G_s) = \{s_t^v | v \in V\}$, that impacts the contact network G_s by changing the state of nodes in V at each timestamp. In this definition, s_t^v denotes the state of node $v \in V$ at time t, and $S_t(.)$ is a graph function whose domain and range are V and a pre-defined set of possible states, respectively. The only known facts about $S_t(G_s)$ from the perspective of an early mitigation strategy are (1) infectious ($s_t^v = I$) is one of the possible states, and (2) the infection spreads through direct contact.

C. Network Perturbations

The only network perturbations required for the network fragmentation problem are edge deletion and edge addition. The edge addition is only limited to the edges that have been previously deleted by the algorithm (isolation process) and are to be released. Since one of the criteria for the early mitigation strategy is to minimize the isolation of healthy individuals, the selection of edges for perturbation is performed through selection of sub-clusters of nodes. The healthy individuals are restricted through isolation of these sub-cluster; i.e., the members of a sub-cluster can only contact others within the sub-cluster and not outside of it. This means the intra-cluster edges of the sub-cluster will be preserved while the inter-cluster edges are removed.

To limit the amount of network disturbance (e.g., due to economic cost), there is a budget for the selection of sets of nodes to form sub-clusters. This budget, which we refer to as b_f, represents the cost of restricting the movement of individuals in a network (e.g., daily monetary compensation). As such, it is logical to consider b_f in terms of the number of restricted nodes per timestamp rather than the number of edges that are perturbed (e.g., we pay restricted individuals the same compensation regardless of their number of contacts).

D. Network Fragmentation Problem Statement

Given the contact network $G_s(V, E_s)$, the outcome of a temporal spreading process $S_t(.)$, and a fragmentation budget b_f, the network fragmentation problem is to find a set of sub-clusters $\mathfrak{C}_t(G_s, b_f)$ at time t whose isolation minimizes the total number of infectious nodes at time $t+1$. In formation of these sub-clusters, only *known* infectious nodes are allowed to form singleton sub-clusters. Formally,

$$\begin{aligned} \mathfrak{C}_t(G_s, b_f) = \min_{G_s, S_t(.)} \ & |\mathcal{I}_{t+1}| \\ \text{s.t} \ & \sum_{C \in \mathfrak{C}_t(G_s, b_f)} |C| \leq b_f \\ & |C| > 1 \ , \forall C \in \mathfrak{C}_t(G_s, b_f) \quad \text{if} \quad s_t^v \neq I \ , \forall v \in C \\ & |C| = 1 \ , \forall C \in \mathfrak{C}_t(G_s, b_f) \quad \text{if} \quad s_t^v = I \ , \forall v \in C. \end{aligned} \tag{2}$$

With known G_s and $S_t(.)$, and $\mathcal{I}_t \leq b_f$, the answer to this problem is trivial: putting all infectious nodes in $\mathcal{I}_t$ in singleton sub-clusters and isolating them gives the optimal solution.

The problem is non-trivial once we add the assumptions of the early mitigation strategy: partially known G_s and $S_t(G_s)$ at time t, which are shown as G_t^* and S_t^* in Table I, respectively. This problem, which we will refer to as *Blind Network Fragmentation Problem*, is then formulated as follows,

$$\begin{aligned} \mathfrak{C}_t(G_s, b_f) = \min_{G_t^*, S_t^*(G_s)} \ & |\mathcal{I}_{t+1}| \\ \text{s.t} \ & \sum_{C \in \mathfrak{C}_t(G_s, b_f)} |C| \leq b_f \\ & |C| > 1 \ , \forall C \in \mathfrak{C}_t(G_s, b_f) \quad \text{if} \quad s_t^{*v} \neq I \ , \forall v \in C \\ & |C| = 1 \ , \forall C \in \mathfrak{C}_t(G_s, b_f) \quad \text{if} \quad s_t^{*v} = I \ , \forall v \in C, \end{aligned} \tag{3}$$

in which $s_t^{*v} \in S_t^*$. Note that the difference between 2 and 3 is that 3 uses the information from partial observations, G_t^* and S_t^*, to minimize $\mathcal{I}_{t+1}$. *ComMit* is a heuristic algorithm that aims to minimize 3. The next section outlines its details.

IV. Method

Here, we introduce the *ComMit* algorithm for dynamically perturbing a network to inhibit the progress of a viral spread, as defined in 3. *ComMit* does not require a priori knowledge of the contact network structure. Other methods with a similar assumption (which mainly deal with immunizations), overcome this limitation by relying on extensive sampling from the contact network (in the form of random walks and/or random node sampling) [2, 15, 16]. In practice, assuming there is an unlimited budget for sampling is unrealistic.

Another assumption of *ComMit* is blindness to the dynamic of the spread, which in turn calls for an efficient testing strategy to identify as many infectious nodes as possible. Although the intuition behind sampling and testing is different (one tries to learn about the network structure whereas the other aims to locate the infectious nodes), the mechanism by which they operate is the same: they select candidates from the pool of nodes in the network based on certain criteria and both within a limited budget in real-world scenarios. Considering their similarity, we combine the sampling and testing into one temporal algorithm. At each timestamp, the goal of this algorithm is to update *ComMit*'s knowledge about the network structure and the infectious hubs simultaneously. We refer to

[4] In [18], the exponent of the best power-law fit if sound to be -1.03 with a standard error of 0.03, which can be approximated by a Zipf's law.

this algorithm in the *ComMit*'s pipeline as *test-trace block*. Iteratively, the output of this block is fed into the *divider block* in which the fragmentation-based mitigation strategy of *ComMit* perturbs the network to inhibit the spread (Figure 1). Below, we discus the details of these two blocks.

A. Test-Trace Block

As evident from the name, the test-trace block consists of two steps: **Testing.** The selection of candidates (pivots) from the population to be tested. This step determines whether these candidates are infectious or not. **Tracing.** Contact tracing of pivots in order to update the known contact network, G^*. Note that the traced contacts will not be tested.

Consider a temporal **testing strategy**, $T_t(G_s, b_t) = \{s_t^v | s_t^v \in S_t(G_s)\}$ with limited budget b_t, whose purpose is two-fold: (1) finding as many infectious nodes in $\mathcal{I}_t$ as possible; (2) gathering information about unknown G_s network to update the known G_t^* network. More formally, an optimal testing strategy would minimize the following,

$$T_t(G_s, b_t) = \min_{G_s, S_t(G_s)} \text{dist}(G_s, G_t^*) + \text{dist}(S_t(G_s), S_t^*(G_s)) \quad \text{s.t.} \quad |T_t(G_s)| \leq b_t, \tag{4}$$

in which $dist(a, b)$ denotes the distance between a and b. This problem is similar to the exploration-exploitation scenario.

A well-known algorithm to address the exploration-exploitation problem in machine learning is ϵ-*greedy*. This algorithm selects an action from a set of possible actions based on a given reward function; the action that maximizes the reward function is selected with probability $p = 1 - \epsilon$ and a random action is chosen with probability $p = \epsilon$. We adapt this idea to our graph-based exploration-exploitation problem and select the pivots as follows,

$$\text{pivot}_t = \begin{cases} \text{randomly choose from } \mathcal{I}^*{}_{t-1}, & p = 1 - \epsilon_t \\ \text{randomly choose from } V, & p = \epsilon_t \end{cases} \tag{5}$$

in which $\mathcal{I}_{t-1}^*$ is the set of infected nodes identified in the previous timestamp. At each time, *ComMit* selects as many pivot nodes as allowed by b_t. As time progresses, we have more knowledge about the network and can rely on exploitation more than exploration. To make that possible, the value of ϵ is updated through a decaying factor d_ϵ as $\epsilon_t = \max(\epsilon_{t-1} - \frac{\epsilon_{t-1}}{d_\epsilon}, 0), \quad d_\epsilon > 0$.

Once the pivots are tested and $\mathcal{I}_t^*$ is updated, the **tracing strategy** is straightforward: the pivots are asked to provide the information about their immediate neighborhood. This information may have less than 100% accuracy (i.e., missing edges). We denote this accuracy by a_t and study its impact in Section V-F. The new edges obtained from tracing update G_t^* which will be used by the divider block.

B. Divider Block

The divider is the main building block of *ComMit* that handles the network perturbations aimed at decreasing the magnitude of the spread. The intuition behind divider is to fragment the bigger communities by reducing the density of its inter-connections. Using the updated $\mathcal{I}_t^*$ and G_t^* from test-trace, the divider identifies a new set of sub-clusters, $\mathfrak{C}_t$, to be temporarily isolated from the network. It does so by attributing a score to each candidate node for forming a sub-cluster. The score is calculated for the community $C \in \mathcal{C}_g$ of each node, where $\mathcal{C}_g$ is the geo-communities inferred from G_g.The scoring function has three components:

$$\text{score} = \frac{1}{3}(\text{norm-size} + \text{inf-rate} + \text{density}), \tag{6}$$

$$\text{norm-size} = \frac{|C|}{|V|}, \tag{7}$$

$$\text{inf-rate} = \frac{|\{v \in C | s_t^{*v} = I\}|}{|C|}, \tag{8}$$

$$\text{density} = \frac{1}{|E_t^*|}(|\{(v_1, v_2) \in E_t^* | v_1, v_2 \in C\}| - |\{(v_1, v_2) \in E_t^* | v_1 \in C, v_2 \notin C\}|), \tag{9}$$

which, in order, are: (1) normalized community size, (2) proportion of nodes within community that are known to be infectious, and (3) community density as the proportion of edges in the known contact network that are inside of the community (i.e., excluding the outgoing edges). The nodes within a community all have the same score. The divider randomly picks b_{fs} candidates from the top 20^{th} percentile of the scores as the seed for the sub-cluster. It ensures the sub-cluster is not singleton by randomly adding b_{fn} neighbors of each seed that are available (e.g., not isolated with another sub-cluster) to the sub-cluster. Hence, the overall budget of the divider is $b_f = b_{fs} \times b_{fn}$. The isolation of a sub-cluster refers to cutting all the outgoing edges from a sub-cluster while maintaining the edges inside.

The divider is also responsible for releasing the currently isolated sub-clusters that have served their isolation time (t_r). To assure these released sub-clusters do not get restricted again and indefinitely, divider places the members of these sub-clusters in a *banned list* that inhibits these nodes from forming another isolated sub-cluster for at least t_r time. The steps for the divider algorithm at time t are,

1) Release sub-clusters isolated at time $t - t_r$.
2) Add the members of the released sub-clusters in the banned list and remove those who have been in the list for t_r time.
3) Put recently identified infectious nodes ($\mathcal{I}_{t-1}^*$) into singleton sub-clusters.
4) Calculate the community score according to 6 for nodes that are neither in an isolated sub-cluster nor in the banned list.
5) Pick b_{fs} seed nodes with the score in the top 20^{th} percentile of scores and form their sub-clusters by selecting b_{fn} of their neighbors at random. If the neighboring list is empty, we remove the corresponding seed node from the candidates.
6) Remove outgoing edges of the new sub-clusters to isolate them.

C. ComMit Algorithm

Combining the *test-trace* and *divider* blocks into a pipeline that iteratively perturbs the contact network yields the final *ComMit* algorithm (see Algorithm 1). An important note is when ComMit terminates. Ideally, it would terminate when either there is no more budget allocated from time t onward, or the spread has dies out (i.e., $|\mathcal{I}_t| = 0$). Since $\mathcal{I}_t$ is unknown, we set the latter **termination condition** such that if for T consecutive timestamps no new infectious node is found (i.e. $|\mathcal{I}^*_{t_i}| = 0, \quad \text{for} \quad t_i \in \{t-T, t-T+1, ..., t\}$), the spread is considered eradicated.

Algorithm 1: `ComMit()`

Input: V, G_g, t_r, b_{fn}, b_{fs}, b_t, ϵ, d_ϵ

$C_g \leftarrow$ `ExtractCommunities`(G_g)

$S^*_0 \leftarrow \{s^{*v}_0 = \bar{I} | v \in V\}$ // $\bar{I}$ = non-infectious

$G^*_0 \leftarrow (V, \{\})$; $\mathcal{I}^*_0 \leftarrow \{\}$; $t \leftarrow 1$

/* iterate while termination condition not met */

while *NotTerminated()* **do**

  /* TestTrace() operates on unknown network, G_s */

  $\mathcal{I}^*_t, S^*_t, G^*_t \leftarrow$ `TestTrace`$(\mathcal{I}^*_{t-1}, S^*_{t-1}, G^*_{t-1}, \epsilon, d_\epsilon, b_t)$

  /* Divider() updates G_s on which spread runs */

  `Divider`$(\mathcal{I}^*_t, S^*_t, G^*_t, C_g, t_r, b_{fn}, b_{fs})$

  $t \leftarrow t+1$

end

D. Budget Analysis

ComMit has two budgets: the testing budget (b_t as the number of nodes tested at time t) and the fragmentation budget (b_f as the number of non-infectious nodes that are members of restricted sub-clusters t). The latter is divided into two separate budgets; one for choosing the sub-cluster seed nodes (b_{fs}) and the other for selecting a certain number of known immediate neighbors of each seed node (b_{fn}). Empirically, we have witnessed that, for b_{fn} values greater than two, no significant performance gain is achieved. The other two budgets are expressed as a proportion of $|V|$: $b_{fs} = \alpha|V|$ and $b_t = \beta|V|$. Thus, the total budget for *ComMit* becomes $(b_{fn} \times \alpha + \beta)|V|$, which can be tuned by setting α and β accordingly (see V-F).

V. Experiments & Discussion

A. Contagion Model for Simulation

The majority of previous studies use SIR model in their simulations as the permanent immunity condition facilitates the analytical tractability [9]. To explore a less investigated direction, we consider the SIS model as the underlying dynamic of the spread. The SIS contagion model models viruses such as common cold, influenza, and COVID-19 [21]. In the SIS model, each node at time stamp t can either be susceptible (S) or infectious (I). The transition from S to I is controlled by the infection rate β. The infected nodes transition back to S once they pass the disease duration t_d. The default values for β and t_d in our experiments are 0.5 and 3, unless otherwise is specified. We initialize an infection by selecting $0.01 \times |V|$ nodes form the population uniformly at random.

TABLE II: *Datasets general information (G_s).*

	Albany	Syracuse	Rochester	Copenhagen	Ithaca
$\lvert V\rvert$	2,858	2,385	1,312	512	127
$\lvert E_s\rvert$	4,641	1,756	4,742	1,416	315
$\lvert \mathcal{C}_g\rvert$	4	4	5	16	3

B. Dataset

The ideal real-world dataset for testing our geo-social network model should contain the information on both the geo-locations and social interactions between the nodes. To the best of our knowledge, due to privacy concerns, such datasets are not available. To navigate this problem, we use equation 1 and consider two types of data: *(A)* data with real-world geo-locations and their pairwise distance, and *(B)* the data with real-world social interactions and their pairwise probability of contact. We use the following strategies to process each category.

- **Constructing social network from geo network:** We first map the pairwise distances to $[1, \inf)$ interval. Then, using equation 1 with $b = 1$, we obtain the probability of contact between each pair. We keep the edges with non-zero probability values (rounded to one decimal). Community membership is obtained via k-means clustering [22] with optimal k that minimizes the inertia.
- **Constructing geo network from social network:** We first form the social network from mobility data with edge weights (w^{ij}_s) in $(0, 1]$. Equation 1 for $b = 1$ gives the Geo network weights w^{ij}_g. This is a partially constructed geo network as some edges in the social network are non-existent. To complete the geo network, we use the weighted shortest-path length between two nodes in the partially constructed geo network. The community memberships are obtained using Louvain algorithm [23] on the constructed geo-network.

NAD Dataset. For datatype A, we use the U.S National Address Database (NAD)[5] (see II) and build four different geo-networks: Syracuse, Albany, Rochester, and Ithaca. pairwise distance is computed using latitude and longitude.

Copenhagen Dataset. For datatype B, we consider the Copenhagen Network Study Interaction Data [24] (see II). In this study, students were followed through their Bluetooth devices across the campus for 28 days. Every five minutes, the Bluetooth devices detected in their vicinity are recorded. Following the definition of close contact by CDC[6], we translate these recordings as close contact if at least 15 minutes of contact is observed within a 24-hour interval for each pair of students. The social network weights are defined as average daily frequency of each close contact and are mapped into $(0, 1]$ interval to represent the pairwise probability of contact.

[5] https://www.transportation.gov/gis/national-address-database/national-address-database-nad-disclaimer

[6] https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/appendix.html

TABLE III: *Performance of various mitigation strategies. Community-based and degree-based ComMit consistently reduce the peak of infection and the absorption time with limited budget, whereas the other methods do not give consistent performance gain across all datasets. The results are averaged among* 10 *runs of the simulation.*

	Albany			Syracuse			Rochester			Copenhage			Ithaca		
	max_bud	**duration**	**inf_peak**	**max_bud**	**duration**	**inf_peak**	**max_bud**	**duration**	**inf_peak**	**max_bud**	**duration**	**inf_peak**	**max_bud**	**duration**	**inf_peak**
commit_cscore	0.056	43.0	0.051	0.029	19.0	0.022	0.120	61.7	0.167	0.209	49.2	0.510	0.117	16.9	0.121
commit_dscore	0.064	36.2	0.045	0.053	12.9	0.020	0.112	69.0	0.152	0.204	58.4	0.487	0.098	21.6	0.113
commit_iscore	0.043	46.2	0.047	0.018	24.3	0.024	0.085	88.4	0.274	0.088	200.0	0.753	0.057	30.5	0.138
acq_imm	0.086	200.0	0.136	0.066	178.1	0.025	0.087	200.0	0.356	0.088	200.0	0.790	0.069	146.3	0.190
com_isolation	0.830	200.0	0.180	0.018	200.0	0.025	0.785	200.0	0.374	0.952	200.0	0.727	0.476	146.4	0.211
no_mitigation	NA	200.0	0.186	NA	200.0	0.025	NA	200.0	0.373	NA	200.0	0.856	NA	146.0	0.209

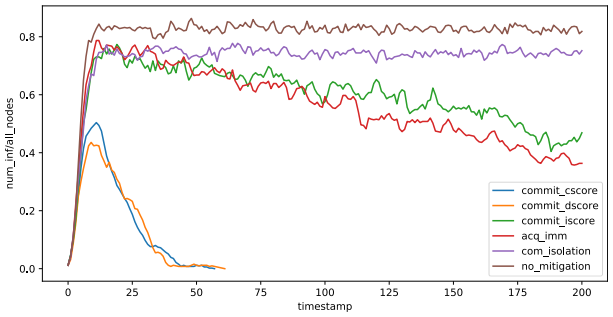

Fig. 2: *The change in the dynamic of the spread due to mitigation strategies for Copenhagen dataset. The community-based and degree-based ComMit have the best performance in terms of lowering the peak of infection and shortening the absorption time.*

C. Evaluation Metric

A spread can be described by its (1) absorption time (the time it takes until no infectious node exists in the population, i.e., $|\mathcal{I}_t| = 0$); and (2) the peak of infection. In the absence of a vaccine, SIS spreads often reach a steady state of maintaining non-zero infection rather than absorption state. Hence, we limit the simulation time to 200 steps. We will show that *ComMit* effectively absorbs the steady-state SIS infection in a short time for all of our datasets.

D. Benchmarks

To the best of our knowledge, there are no temporal mitigation strategies that consider all the limitations of the early-stages in a viral spread (i.e., no knowledge of the network structure and dynamics of the spread, and limitation of the sampling and network perturbation budgets). For a fair comparison, we build our benchmarks by using the same test-trace method as in *ComMit* to give the advantage of efficiently probing the network within a limited budget. Our benchmarks for the divider block of *ComMit* are: **ComMit_CScore.** The original *ComMit* pipeline discussed in Section IV. **ComMit_DScore.** Similar to *ComMit_CScore*, but uses the degree centrality in G_t^* to score and choose seed nodes. **ComMit_IScore.** Inspired by test-based strategies whose goal is to find the most number of infectious nodes to isolate, we change the divider such that it selects the seed nodes from the known infectious by their degree centrality in G_t^*. **Acq_Imm.** Similar to acquaintance immunization method in [16], we randomly select seed nodes and their neighbors to form sub-clusters. Note that in this method there are no singleton sub-clusters and identified infectious nodes may or may not be included in the sub-clusters. **Com_Isolation.** Considering the good performance of community-based isolation (with known contact network) in our previous study [5], we use the information from test-trace to decide whether to isolate the entirety of a community. This method does not form sub-clusters. Once the ratio of the infectious nodes within the community surpasses a certain threshold, the community is isolated for the duration of t_r (the same value across all baselines). Our experiments show a threshold of 0.1 gives the best reasonable trade-off between the budget and performance. **No_Mitigation.** The baseline without any mitigation strategy.

E. Results

The results of our simulations are shown in Table III. In addition to the evaluation metrics, we also report the maximum divider budget for each strategy (the test budget is the same for all). The default values for hyperparameters are: $a_t = 1$, $b_t = 0.1 \times |V|$, $b_{fs} = 0.01$, $b_{fn} = 2$, and $t_r = 3$. The results show that *ComMit* variants, `commit_cscore` and `commit_dscore`, yield similar performance with the exception that the former has a shorter absorption time on average. The other two variants, `commit_infscore` and `com_isolation`, do not have guaranteed performance as in some cases they either do not terminate the spread or use an unrealistically large budget. Acquaintance immunization (`acq_imm`) consistently yields a poor performance across all datasets. Figure 2 is an example of changing spread dynamic for each strategy. At its best, *ComMit* reduces the peak of infection by 73% and the absorption time by 80% (see the first row for Albany). At its worst, it reduces the peak by 6% and the absorption time by 90% (see the first row for Rochester); a trade-off that still beats the other baselines.

F. Ablation Studies & Final Remarks

Figure 3 depicts the results of our ablation study on *ComMit*'s hyperparameters. The experiments are run with the default parameters as above and reported for `commit_cscore`. Figure (a) and (b) show that by increasing the divider's budget, b_{fs} and b_{fn}, no significant performance boost is observed. In Figure (c), we keep the duration of infection, t_d, as 3 and change the divider's restriction time, t_r. The result shows that choosing a value closer to the actual infection time yields a shorter absorption time. The impact of self-reports accuracy, a_t, is tested in Figure (d). Higher accuracy results in discovering more edges quickly, but does not change the performance of *ComMit* drastically. This result suggests that *ComMit* does not rely on full knowledge of the graph to reach

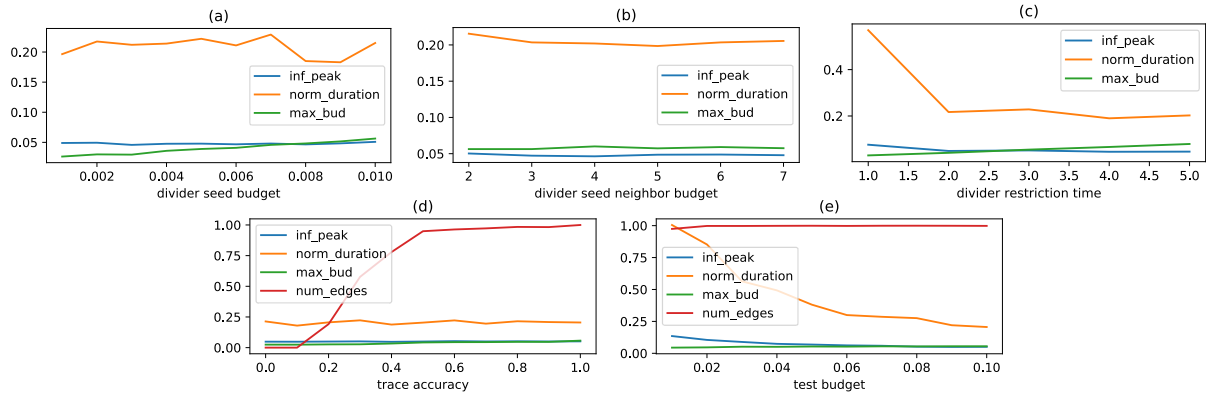

Fig. 3: *Ablation studies. Figures (a) to (e), in order, study b_{fs}, b_{fn}, t_r, a_t, b_t. The* `inf_peak`*,* `norm_duration`*,* `max_bud`*, and* `num_edges` *signify the peak of infection, the duration of infection normalized by the duration of simulation, the maximum divider budget in terms of number of restricted nodes normalized by $|V|$, and the number of edges discovered by the test strategy normalized by the number of edges in G_s, respectively.*

its best performance. In Figure (e), we see that increasing the test budget b_t, for $a_t = 1$, can drastically shorten the absorption time. However, small values of b_t still do well at probing the full graph.

ComMit relies on geo-network for estimating the community structure in the contact network. If these communities are known through other means (e.g., government survey data), no geo information is required. **Limitation & Ethical Issue:** ComMit relies on commitment of individuals to follow the isolation instruction and is not tested on the disobedience scenario. Moreover, owing to the exploration component, it is possible to test a candidate with low infection probability. There are ethical issues involved with violating one's privacy by requiring their social information when they are not likely to put others in danger.

VI. Conclusion

We formally defined the problem of early mitigation strategy and offered a dynamic algorithm, *ComMit*, that incorporates the realistic assumptions of blindness towards network and spread dynamics. *ComMit* relies on an exploration-exploitation test-trace strategy to gain more information about both network and status of the spread, and introduces network perturbations that control the magnitude of the spread by following a community fragmentation strategy. Our experiments showed effectiveness of *ComMit* in reducing the peak and duration of infection.

References

[1] C. Stegehuis, R. Van Der Hofstad, and J. S. Van Leeuwaarden, "Epidemic spreading on complex networks with community structures," *Scientific reports*, 2016.

[2] S. F. Rosenblatt, J. A. Smith, G. R. Gauthier, and L. Hébert-Dufresne, "Immunization strategies in networks with missing data," *PLoS computational biology*, 2020.

[3] P. Yuan and S. Tang, "Community-based immunization in opportunistic social networks," *Statistical Mechanics and its Applications*, 2015.

[4] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos, "Rise and fall patterns of information diffusion: model and implications," in *ACM SIGKDD*, 2012.

[5] P. Hozhabrierdi, R. Zhu, M. Onyewu, and S. Soundarajan, "Network-based analysis of early pandemic mitigation strategies: Solutions, and future directions," *Northeast Journal of Complex Systems*, 2021.

[6] M. Serafino, H. Monteiro, S. Luo, S. Reis, A. Igual, Carles Neto, M. Travizano, J. Andrade, and H. Makse, "Superspreading k-cores at the center of covid-19 pandemic persistence," *arXiv*, 2021.

[7] P. Holme, "Efficient local strategies for vaccination and network attack," *EPL (Europhysics Letters)*, 2004.

[8] C. Schneider, T. Mihaljev, S. Havlin, and H. Herrmann, "Suppressing epidemics with a limited amount of immunization units," *Physical Review*, 2011.

[9] L. Pellis, F. Ball, S. Bansal, K. Eames, T. House, V. Isham, and P. Trapman, "Eight challenges for network epidemic models," *Epidemics*, 2015.

[10] S. Wasserman, K. Faust *et al.*, "Social network analysis: Methods and applications," 1994.

[11] A. Galstyan and P. Cohen, "Cascading dynamics in modular networks," *Physical Review E*, 2007.

[12] P. Block, M. Hoffman, I. Raabe, B. Dowd, C. Rahal, R. Kashyap, and M. Mills, "Social network-based distancing strategies to flatten the covid-19 curve in a post-lockdown world," *Nature Human Behaviour*, 2020.

[13] M. I. Meltzer, N. J. Cox, and K. Fukuda, "The economic impact of pandemic influenza in the united states: priorities for intervention." *Emerging infectious diseases*, 1999.

[14] B. Gross and S. Havlin, "Epidemic spreading and control strategies in spatial modular network," *Applied Network Science*, 2020.

[15] M. Salathé and J. H. Jones, "Dynamics and control of diseases in networks with community structure," *PLoS Comput Biol*, 2010.

[16] R. Cohen, S. Havlin, and D. Avraham, "Efficient immunization strategies for computer networks and populations," *Physical review*, 2003.

[17] R. J. Johnston, "Social distance, proximity and social contact: Eleven cul-de-sacs in christchurch, new zealand," *Geografiska Annaler: Series B, Human Geography*, 1974.

[18] J. Goldenberg and M. Levy, "Distance is not dead: Social interaction and geographical distance in the internet era," *arXiv*, 2009.

[19] A. Topîrceanu, "Analyzing the impact of geo-spatial organization of real-world communities on epidemic spreading dynamics," in *International Conference on Complex Networks and Their Applications*, 2020.

[20] S. Yang, P. Senapati, D. Wang, C. T. Bauch, and K. Fountoulakis, "Targeted pandemic containment through identifying local contact network bottlenecks," *arXiv*, 2020.

[21] A. Radbruch and H. D. Chang, "A long-term perspective on immunity to covid," *Nature News and Views*, 2021.

[22] P. S. Bradley, K. P. Bennett, and A. Demiriz, "Constrained k-means clustering," *Microsoft Research, Redmond*, 2000.

[23] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *JSTAT*, 2008.

[24] P. Sapiezynski, A. Stopczynski, D. Lassen, and S. Jørgensen, "The copenhagen networks study interaction data. figshare," 2019.