Building a Sybil-Resilient Digital Community Utilizing Trust-Graph Connectivity

Ouri Poupko, Gal Shahaf, Ehud Shapiro, and Nimrod Talmon

***Abstract*— Preventing fake or duplicate digital identities (aka *sybils*) from joining a digital community may be crucial to its survival, especially if it utilizes a consensus protocol among its members or employs democratic governance, where sybils can undermine consensus, tilt decisions, or even take over. Here, we explore the use of a trust-graph of identities, with edges representing trust among identity owners, to allow a community to grow indefinitely without increasing its sybil penetration. Since identities are admitted to the digital community based on their trust by existing digital community members, *corrupt* identities, which may trust sybils, also pose a threat to the digital community. Sybils and their corrupt perpetrators are together referred to as *byzantines*, and the overarching aim is to limit their penetration into a digital community. We propose two alternative tools to achieve this goal. One is graph conductance, which works under the assumption that honest people are averse to corrupt ones and tend to distrust them. The second is vertex expansion, which relies on the assumption that there are not too many corrupt identities in the community. Of particular interest is keeping the fraction of byzantines below one third, as it would allow the use of Byzantine Agreement (Lamport *et al.*, 1982) for consensus as well as for sybil-resilient social choice (Shahaf *et al.*, 2019). This paper considers incrementally growing a trust graph and shows that, under its key assumptions and additional requirements, including keeping the conductance or vertex expansion of the community trust graph sufficiently high, a community may grow safely, indefinitely.**

***Index Terms*— Network theory (graphs).**

I. Introduction

THE goal of this paper is to identify conditions under which a digital community of predominantly *genuine* (singular and unique) digital identities [4] may grow without increasing the penetration of *sybil* (fake or duplicate) digital identities. Our particular context of interest is digital democracy [5], [6], where a sovereign digital community conducts its affairs via egalitarian decision processes; another motivation is the task of growing a permissioned distributed system. In contrast to other works in this field, where the goal is to discover sybils in existing social networks, we assume a dedicated network where the members are actively engaged in authenticating each other towards a democratic digital community aiming for one person one vote. For example, an application for digital democracy may require people to authenticate each other, while registering for using the application, and specify d other people which they know and trust, and also verified that their profile image in the application is authentic. The application can than construct a graph out of these specifications and that will be the trust graph used by the algorithms in this paper to bound the number of sybils in the community of registered users. Consider an initial digital community with low sybil penetration that wishes to admit new members without admitting too many sybils. As it is not realistic to expect that no sybils will be admitted, the goal is to keep the fraction of sybils below a certain threshold. In a separate paper [3], we show that a digital democracy can tolerate up to one-third sybil penetration and still function democratically. Still, the fewer the sybils, the smaller the supermajority needed to defend against them. Also, keeping the fraction of sybils and corrupt identities that control them below one third will enable the community to safely conduct a shared distributed ledger, using Byzantine agreement [2].

We model a digital community via a trust graph with a vertex for each identity and with edges representing trust relations between the owners of the corresponding identities (formal definitions are given in Section II). The model considers *genuine* and *sybil* identities (cf. [4]), and refers to the genuine identities that do not trust sybils as *honest* and those that do as *corrupt*. Furthermore, to describe an admission process that facilitates incremental community growth, the model presents sequences of trust graphs that may result from such a process. While we assume some underlying social graph, our trust graph grows with different properties that make it more sybil resilient than the initial graph. For example, the underlying social graph may naturally have a power-low degree distribution, while the generated trust graph may be d-regular, by design.

The goal is to identify sufficient conditions on such graphs, for example, the type of identities in the graph, their relative fractions, and their trust relations, under which a community may grow while keeping the fraction of sybils in it low. To achieve this, we use two similar approaches, which differ in the assumptions made on the power of the adversary: The first approach assumes that honest identities tend to trust honest identities rather than corrupt ones, therefore it is hard for the corrupt ones (the adversary) to create trust edges with honest identities. In this case graph conductance bounds the ratio of sybils in the graph. The second approach assumes that there

Manuscript received February 12, 2020; revised August 20, 2020, December 14, 2020, and April 18, 2021; accepted May 17, 2021; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor C. W. Tan. Date of publication June 3, 2021; date of current version October 15, 2021. The work of Nimrod Talmon was supported by the Israel Science Foundation under Grant 630/19. A preliminary version of this paper was presented at the 14th International Computer Science Symposium in Russia, July 1–5, 2019, Novosibirsk, Russia [1]. *(Corresponding author: Ouri Poupko.)*

Ouri Poupko, Gal Shahaf, and Ehud Shapiro are with the Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: ouri.poupko@weizmann.ac.il; gal.shahaf@weizmann.ac.il; ehud.shapiro@weizmann.ac.il).

Nimrod Talmon is with the Department of Industrial Engineering and Management, Ben-Gurion University, Beersheba 8410501, Israel (e-mail: talmonn@bgu.ac.il).

Digital Object Identifier 10.1109/TNET.2021.3084303

are not too many corrupt identities, therefore the adversary power is limited by its own size. In this case vertex expansion bounds the ratio of sybils in the graph.

A. Related Work

This section reviews existing work, particularly work that helps clarifying the differences in our proposed model. A large portion of the literature on sybil attacks (see, for example, [7]–[9] and their citations) is focused on identifying the sybil agents from the honest ones. Of particular interest is the approach initiated by Yu *et al.* [10], which relies on structural properties of the underlying social network. Yu *et al.* show how to separate the honest and sybil regions by leveraging the assumption that there are, relatively, few number of edges between them. This framework was studied further [11]–[16]. As pointed out by Alvisi *et al.* [17], however, such attempts to differentiate between sybils and honests on a global scale is efficient only in instances where the honest region is sufficiently connected, which is rarely the case in actual social networks. Consequently, Alvisi *et al.* suggest a more modest goal of producing a whitelist of honest vertices in the graph with respect to a given agent; that is, securely identifying a *local* safe region, in contrast to the *global* algorithms proposed before.

Boshmaf *et al.* [18] show an impressive improvement over SybilRank [16], by studying the social features of a third tier between honests and sybils, which they identify as victims – real accounts that have befriended fakes. They use learning algorithms to pick the effective features, then use these features to assign different weights to different nodes in the graph, then apply sybil detection algorithms, like SybilRank, on the resulting weighted graph. We adopt the same separation into three categories, though we regard the middle tier as the offenders (corrupt), regardless if they are hostile, victims, or their accounts were hijacked. We do not try to improve over Boshmaf *et al.*, but rather propose a different approach that relies less on existing properties of social networks and ask the participants instead to cooperate towards the goal of sybil resilience. This will probably be a downside for the proposed algorithm, when looking for sybils in existing social networks, but we believe it to be a better starting position when building a trust network towards democratic governance.

Liu *et al.* [19] study the effect of temporal dynamics. In particular, they show that by modifying the trust graph over time, by replacing existing sybils with new sybils, replacing attack edges between sybils and honests, and even exploiting changes in the honest region, they can successfully bypass several existing sybil defense algorithms, as they were designed to analyze a fixed graph in a single instant of time. Our work fits very well exactly in this spot, as we look for the conditions that will allow a community to safely grow over time. As we rely only on the graph structure in every time step, we believe our algorithm to be robust against temporal dynamics attacks.

Friebe *et al.* [20] present *Detasyr*, which is a ticket-based algorithm to fend off sybils. The algorithm starts with a group of ticket sources (nodes in the graph) that generates cryptographic tickets and flood them in the graph. A node that wishes to get authorized and join the graph collects such tickets from its already authorized friends, and if it collects enough tickets it gets authorized. The set of ticket sources for the next round are then selected randomly, again by traversing the graph. This work has interesting similarities to our work. It also assumes active participation of the individuals in selecting the friends they trust. It grows the graph (the authorized community) in rounds. Its sybil defence is based on graph connectivity, which is assumed to be higher between genuine nodes, and, like this paper, their goal is to bound the number of sybils rather then detecting them. While Detasyr assumes that a sybil node will have a bounded number of authorized friends, this paper assumes a wider assumption that applies to the community as a whole, rather than individuals. We do not rely on leaders such as the ticket sources, and based on our assumptions, we are able to specify an analytic bound on the number of sybils, which is missing in the Detasyr paper.

A problem of a similar flavor is that of *corruption detection* in networks, posed by Alon *et al.* [21] and later refined by Jin *et al.* [22]. This setting, inspired by auditing networks, consists of a graph with each of its vertices being either truthful or corrupt, where the overall goal is to detect the corrupt region. In contrast to the sybil detection problem, the corrupt agents are assumed to be immersed throughout the network, and the setting assumes a very restrictive assumption, namely that each agent may accurately determine the true label of its neighbors and report it to a central authority. The authors show how good connectivity properties of the graph allows an approximate recovery of the truthful and corrupt regions.

Note that social networks have some special structure, for example, having low diameters (a.k.a., the *small world phenomena* [23]) or fragmented to highly-connected clusters with low connectivity between different clusters. Moreover, as observed by some researchers [12], [17], [24], the attacker's inability to maintain sufficiently many attack edges typically results in certain "bottlenecks", which can be utilized to pin-point the sybil regions.

In our line of work we consider a duplicate identity also as being a sybil. We believe that, within an existing social network, a genuine being with many acquaintances can and should be able to conduct multiple identities within the social network, all being a genuine representation of herself, with her genuine acquaintances divided between the different identities. For example, one can have one facebook user with which she shares information with her family, and another facebook user, with which she shares information with her friends. Prior work does not consider duplicate identities as being sybils, as long as the different duplicates are all valid. When considering digital identities for digital democratic governance, duplicates are not allowed, else we compromise the concept of one person one vote. This paper does not handle duplicate identities directly, but it assumes that the trust edge information also identifies these cases. An honest identity will not trust any but the first identity of another person. This can be achieved either with a very high level of trust, as shown in a related work [4], or by using some biometric identification, and requiring the participants to verify each others proof of uniqueness. For this reason we lean in this work more on the participation of the

members of the community, rather than relying on more social features of the graph, like Boshmaf *et al.* [18] and others.

B. Informal Model

While the problem addressed is related to sybil detection, and indeed we incorporate some of the insights of the works discussed above, here the main goal is different: Safe community growth. This work aims to find conditions under which a community may grow without increasing the fraction of hostile members within it; but without necessarily identifying explicitly who is hostile and who is not. An additional difference from existing literature is the notions of identity and trust. Specifically, most existing works consider identities or agents of only two types, "good" and "bad", with various names for the two categories. In this work the notion of identities [4], is more refined and, we believe, may be closer to reality.

In particular, this work considers genuine and sybil identities, with the intention that in a real-world scenario these would be characterized by the nature of their *representation*: genuine identities are singular and unique, else are sybil (duplicate or fake, namely not corresponding to a single real person). It further distinguishes between two types of genuine identities, based on their *behavior*: honest, which do not form trust relations with sybils, and corrupt, which do. This behavioral distinction is captured formally in the proposed model. We naturally assume that the owners of corrupt identities are the creators and operators of the sybils and that, in the worst case, all sybils and their corrupt perpetrators may cooperate, hence the model labels them together as *byzantines*, and aims to limit their fraction within the community.

We thus begin with a unified formal model of such identities and their *trust graph*, consisting of vertices that represent identities and edges that represent trust relations among the owners of such identities. The exact definition of these trust relations are outside the scope of this paper, but in a related work [4] we consider a spectrum of such trust relations, expressed as mutual sureties among identity owners, and inspect their applicability also to the work presented here. Considering the task of sybil-resilient community growth, the model defines the *community history* that aims to capture the incremental changes a community trust graph undergoes in discrete steps. In order to properly characterize identities, the model first employs the basic distinction between genuine and sybil identities. Then, using the community history, it makes a further delicate distinction within genuine identities between *honest* identities, which never trust sybils, and *corrupt* identities, which may trust sybils and, furthermore, may cooperate with other corrupt or sybil community members to introduce sybils into the community.

Some assumptions on the power of the sybils and their perpetrators are needed; otherwise there is no hope in achieving our goal. We present two possible alternative assumptions: The first intuitive assumption is that honest identities are averse to corrupt identities, and hence are not likely to trust them. Trust edges that connect honest and corrupt identities are referred to as *attack edges*. So, loosely speaking, the assumption is that there are not too many attack edges. We view this assumption as more realistic than the assumption made in related works [21], [22], that truthful agents can identify *precisely* whether a neighbor is corrupt or not. Figure 1 illustrates the general setting. The second assumption is that there are not too many corrupt identities in the community. This assumption could be realized, for example, by an incentive mechanism that penalizes for trusting sybils and rewards honest identities.

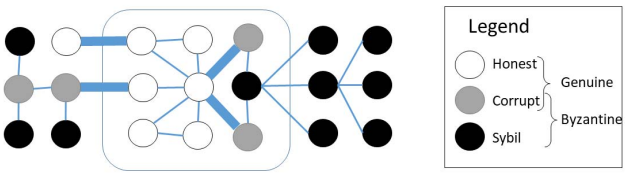

Fig. 1. Illustration of the general setting: The white vertices (honest identities) **and** grey vertices (corrupt identities) form the set of genuine identities, while grey vertices (corrupt identities) **and** black vertices (sybil identities) form the set of byzantines. Bold edges represent attack edges. The circled area contains the current community that wishes to grow. Notice that the nine identities in the community contain one sybil and two corrupt identities, thus in particular the community's byzantine penetration is $\beta = 1/3$ and the sybil penetration is $\sigma = 1/9$. The fraction of internal attack edges to the volume of the honest part of the community graph, defined below, is $\gamma_e = 1/8$.

C. High-Level Approach

After defining the three types of population in the community, it is clear that the corrupt identities are the adversary to the goal of growing a community without sybils. Without corrupt identities, if the first identity in the community is not a sybil (therefore it is honest), and given that, by definition, honest identities have no trust edges with sybils, then sybils cannot join the community. To gain intuition regarding the two assumptions on the power of the adversaries, consider an extreme case, as shown in Figure 2, where the power of the adversary is minimal. The graph on top represents the first assumption, that honest identities are averse to corrupt identities. The graph below represents the second assumption, that there are not too many corrupt identities. In this extreme example the graph is not constrained in any way, which shows that even a weak adversary can add as many sybils as it wants, without additional measures. Our approach will be to measure the connectivity of the graph and derive a bound on the number of byzantines based on this measurement. The example in Figure 2 shows that some simple measurements of connectivity are fruitless for the goal of sybil detection. One such measurement is how dense the graph is, or what is the lower bound on the number of edges within the community. Both graphs show a community where the lower bound on the number of edges is of order $n/2$, and yet the corrupt identities are able to introduce as many sybils as they wish. Another simple measurement is the diameter of the graph, which is also very low in these two communities - 3 at the top and 2 at the bottom.

Yet there is a clear bottleneck in these extreme examples between honest and byzantines. The measures that capture precisely this type of bottleneck are conductance, when the bottleneck is in the edges, and vertex expansion when the bottleneck is in the vertices. The ability to protect the graph

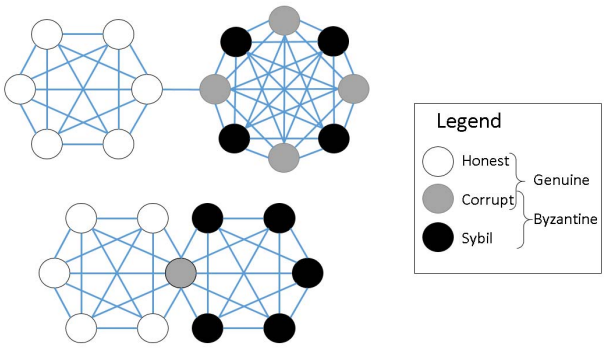

Fig. 2. Illustration of an extreme example: Both community graphs have one cluster of honest identities and one cluster of byzantine identities. In each cluster everyone trusts everyone (the sub graph of the cluster is a clique), yet there is almost no trust between the clusters. The graph at the top demonstrates the case where honest identities don't trust corrupt identities. The graph at the bottom demonstrates the case where there are almost no corrupt identities.

from byzantine penetration is based on the key assumption that, while there could be arbitrarily many byzantines wanting to join the growing community, they will have limited connectivity to the current community. Indeed, this observation was applied in the context of fending off sybils [10], [11], [17], [24].

In general, while the connectivity of the whole network is typically fairly low, a social network usually contains many clusters that reflect real life communities. The connectivity of the subgraphs restricted to each of these clusters may be high. In that sense, following Alvisi *et al.* [17], we adopt a local perspective and focus on the connectivity of the community, regardless of the connectivity of the entire network. In contrast to Alvisi *et al.* [17], however, we are interested in growing the community and not in whitelisting. Unlike the situation treated by Alvisi *et al.* [17], which can be viewed as whitelisting, initiated at a singleton community (that is, from a single non-sybil vertex), here we consider arbitrarily-large communities and aim to bound, but not detect or eliminate, the sybils in them.

Specifically, our framework makes use of a "target conductance" parameter Φ_e, or a "target vertex expansion" parameter Φ_v, and aims to grow, that is, admit new members, while retaining a conductance of at least Φ_e, or Φ_v respectively, at the larger community. Assuming that the initial community harbors a limited attack power and a bounded fraction of byzantines, this paper shows how to safely grow the community, indefinitely. The number of members that may join in each increment is a parameter of the algorithm and is related to the bound on byzantines the community maintains. The lower the bound the more members the community can add in each increment. The bound on byzantines, in turn, depends on the target conductance or vertex expansion that the community maintains. The higher the connectivity of the community, the better the bound on byzantines.

Remark 1: Note that our methods are deterministic. That is, they guarantee – deterministically – that, if the parameters have certain values and if the assumptions hold, then the conclusion – namely, that the growing community retains a low fraction of sybil penetration – holds.

D. Paper Structure

The paper begins with graph theory terminology and formal definition of graph conductance and vertex expansion in Section II. For simplicity, the framework describes undirected and unweighted graphs. Note, however, that it may easily be modified and applied to directed and weighted graphs as well. The model is formally described in Section III, by defining types of identities, communities and community history. Then, Section IV describes the first method, based on the assumption of little trust and the use of conductance, and showing sufficient conditions for safe community growth. Section V, shows that the framework is compatible with sparse trust graphs and provides some quantitative estimations of its guarantees. Section VI and Section VII introduce and analyze the second method, based on the assumption that there are not too many corrupt identities. Section VIII presents a simulation using synthetic data. The simulation shows even better results than the analytic estimation. Section IX analyzes data from a real social network, showing why an existing social network is not a good trust graph for the method in this paper, yet it also shows how such a social graph can be used efficiently as the underlying graph from which the desired trust graph can be generated. Section X concludes with intriguing open questions for future research.

II. Preliminaries

This section provides some needed definitions regarding graphs and graph connectivity. Refer to any graph theory textbook, like Diestel's Graph Theory [25] for additional background.

Let $G = (V, E)$ be an undirected graph. The *degree* of a vertex $x \in V$ is:

$$\deg(x) := |\{y \in V \mid (x, y) \in E\}|$$

G is *d-regular* if $\deg(x) = d$ holds for each $x \in V$. The *volume* of a given subset $A \subseteq V$ is the sum of degrees of its vertices:

$$vol(A) := \sum_{x \in A} \deg(x)$$

Additionally, denote the subgraph induced on the set of vertices A as $G|_A$, the degree of vertex $x \in A$ in $G|_A$ by $\deg_A(x)$, and the volume of a set $B \subseteq A$ in $G|_A$ by:

$$vol_A(B) := \sum_{x \in B} \deg_A(x)$$

Given two subsets $A, B \subseteq V$, the size of the cut between A and B is denoted by:

$$e(A, B) = |\{(x, y) \in E \mid x \in A, y \in B\}|$$

Definition 1 (Conductance): Let $G = (V, E)$ be a graph. The conductance *of G is defined by*:

$$\Phi_e(G) = \min_{\emptyset \neq A \subset V} \frac{e(A, A^c)}{\min\{vol(A), vol(A^c)\}}$$

where $A^c := V \setminus A$ is the complement of A.

Remark 2: Generally speaking, graph conductance aims to measure the connectivity of the graph by quantifying the

minimal cut normalized by the volume of its smaller subset. Conductance should be thought of as the weighted and irregular analogue of edge expansion [26], where both notions are essentially equivalent for regular graphs. To get a quantitative grip of this measure, notice that for all graphs, $\Phi_e \in [0, \frac{1}{2}]$. *Intuitively, the conductance of a highly connected graph approaches* $\frac{1}{2}$. *For example, cliques and complete bipartite graphs satisfy* $\Phi_e = \frac{1}{2}$, *while in a poorly connected graph this measure may be arbitrarily small; for example, a disconnected graph satisfies* $\Phi_e = 0$.

The next sections provide theoretical guarantees on sybil safety, given that one can compute conductance. However, determining the exact conductance of a given graph is known to be coNP-hard [27]. Luckily, the Cheeger inequality [28] provides a direct relation between conductance of a graph and the second eigenvalue of its random walk matrix, which can be calculated in polynomial time, and approximated in nearly linear time. Refer to [26], [29] and [30] for comprehensive surveys regarding efficient algorithms for measuring conductance.

Definition 2 (Inner Boundary Vertex Expansion): Let $G = (V, E)$ *be a graph. Given two subsets* $A, B \subseteq V$, *define the* inner boundary of A w.r.t. B *by*

$$\partial_v(A, B) := \#\{x \in A | \exists y \in B \ s.t. \ (x, y) \in E\}$$

The inner boundary vertex expansion is then defined by:

$$\Phi_v(G) := \min_{0<|A|\leq \frac{|V|}{2}} \frac{\partial_v(A, A^c)}{|A|}$$

Like conductance, vertex expansion also aims to measure the connectivity of the graph, this time by quantifying the minimal vertex cut, rather than the minimal edge cut.

To get a quantitative grip of this measure, note that for all graphs $\Phi_v \in [0, 1]$. Intuitively, the vertex expansion of a highly connected graph approaches 1. For example, a clique satisfies $\Phi_v = 1$, while in a poorly connected graph this measure may be arbitrarily small and a disconnected graph satisfies $\Phi_v = 0$. Also note the relation between conductance and vertex expansion, given by $\Phi_v/d \leq \Phi_e \leq \Phi_v$ for d-regular graphs.

III. Formal Model

A. Community Trust Graphs

The relation between people and their identities is rich and multifaceted. For the purpose of this paper, assume that some identities are *genuine* and others are not, in which case they are called *sybils*. We represent trust relations among identities via a trust graph, in which vertices represent identities and edges represent trust among identities.

Definition 3: A trust graph $G = (V, E)$ *is an undirected graph with vertices that represent identities and edges that represent trust among them.*

The concept of a community trust graph follows, which depicts the community that grows within such a trust graph.

Definition 4: A community trust graph $G = (A, V, E)$ *is a trust graph with vertices* V, *edges* E, *and a* community $A \subseteq V$.

B. Community Histories and Transitions

The aim of this paper is to find conditions under which a community may grow safely. A graph of identities represents the community. Once establishing some conditions on a given community, we want to verify that these conditions hold under the operation of adding additional identities to the community graph. As the newly added identities threaten these conditions (for example, assume that the community has a bound on the ratio of corrupt identities, and then the added identities may be corrupt and the new community will cross this bound), the model breaks the growth of the community into steps of incremental growth.

Definition 5 (Community History): A community history $\mathcal{G}_V$ *over a set of vertices*[1] V, *is a sequence of community trust graphs* $\mathcal{G}_V = G_1, G_2, \ldots$, *where* $G_i = (A_i, V, E)$, *such that* $\forall i \ A_i \subset A_{i+1}$.

C. Types of Identities

There are two types of identities: *genuine* and *sybil*. Next, community histories distinguish between two types of genuine identities – honest and corrupt: An identity is *corrupt* in a community history if it shares an edge with a sybil, and *honest* if it does not. Lumping together sybils and corrupt identities, they form the group of *byzantines*.

Below and in the rest of the paper we use disjoint union $A = B \uplus C$ as a shorthand for $A = B \cup C$, $B \cap C = \emptyset$.

Definition 6 (Types of Identities, Attack Edges, Sybil Penetration): Let V *be a set of vertices that consist of two disjoint subsets* $V = T \uplus S$ *of genuine* T *and sybil* S *vertices, and let* $\mathcal{G}_V$ *be a community history over* V. *Then, a genuine vertex* $t \in T$ *is* corrupt *in* $\mathcal{G}_V$ *if it trusts a sybil at anytime in* $\mathcal{G}_V$, *namely, there is some* $(t, s) \in E$, *with* $t \in T$, $s \in S$, *for some* $G = (A, V, E) \in \mathcal{G}_V$. *A genuine vertex that is not corrupt is said to be* honest. *Thus,* $\mathcal{G}_V$ *partitions the genuine identities* $T = H \uplus C$ *into honest* H *and corrupt* C *identities. An edge* $(h, c) \in E$ *is an* attack edge *if* $h \in H$ *and* $c \in C$. *The* sybil penetration $\sigma(G)$ *of a community trust graph* $G = (A, V, E) \in \mathcal{G}_V$ *is*

$$\sigma(G) = \frac{|A \cap S|}{|A|}$$

In the worst case, sybils and their corrupt perpetrators would cooperate; thus, to allow for incremental community growth, it must bound their combined presence in the community, as defined next:

Definition 7 (Byzantines and Their Penetration): Let $\mathcal{G}_V$ *be a community history over* $V = T \uplus S$ *that partitions* $T = H \uplus C$ *into honest* H *and corrupt* C *identities. Then, a vertex* $v \in V$ *is* byzantine *if it is a sybil or corrupt and the* byzantines $B = S \uplus C$ *are the union of the sybil and corrupt vertices. The* byzantine penetration $\beta(G)$ *of a community trust graph* $G = (A, V, E) \in \mathcal{G}_V$ *is*

$$\beta(G) = \frac{|A \cap B|}{|A|}$$

[1]As the set of vertices V is fixed in a community history, it does not explicitly model the birth and death of people; modeling this aspect is the subject of future work.

As $A = (A \cap H) \uplus (A \cap B)$, it would occasionally be convenient to use the equivalence between byzantine penetration to the community A and the fraction of byzantines w.r.t. genuine identities in A. Formally,

$$\frac{|A \cap B|}{|A|} \leq \beta \quad \text{iff} \quad \frac{|A \cap B|}{|A \cap H|} \leq \frac{\beta}{1-\beta} \tag{1}$$

IV. Conductance-Based Approach

The goal of this section is to find the conditions under which a community can grow while bounding the penetration of byzantines and sybils. The reader may read the following remedy as high level instructions to achieve this goal:

1) Start with an initial community.
2) Choose the desired bound on byzantine penetration.
3) Measure the fraction of edges within the community, out of all edges stemming out of the community.
4) Estimate a bound on the connectivity between honest and sybil/byzantine identities.
5) Admit new candidates to the community only if the connectivity within the target community is sufficiently large.

The following provides sufficient conditions for byzantine-resilient community growth, under the assumption that honest people tend to trust honest people and distrust corrupt people.

Theorem 1: Let $\mathcal{G}_V$ be a community history. Set parameters $\alpha \in [0,1], \beta \leq \frac{1}{2} - \frac{1}{|A_1|}, \gamma_e \in [0, \frac{1}{2}], \delta = 1 - 2\beta$. Assume:

1) *All communities have a bounded degree, both above and below*:
$$\alpha \cdot d \leq deg_{A_i}(v) \leq d \text{ for all } v \in A_i \, i \in \mathbb{N}$$
2) *Byzantine penetration to the initial community is bounded*:
$$\beta(G_1) \leq \beta$$
3) *The edges between honest and byzantine identities are relatively scarce*:
$$\frac{e(A_i \cap H, A_i \cap B)}{vol_{A_i}(A_i \cap H)} \leq \gamma_e$$
4) *Community growth is bounded*:
$$|A_i \setminus A_{i-1}| \leq \delta |A_{i-1}|$$
5) *The conductance within A_i is sufficiently high*:
$$\Phi_e(G|_{A_i}) > \frac{\gamma_e}{\alpha} \cdot \left(\frac{1-\beta}{\beta}\right)$$

Then, every community $G_i \in \mathcal{G}_V$ has Byzantine penetration $\beta(G_i) \leq \beta$.

Roughly speaking, Theorem 1 suggests that whenever: (1) Each graph $G_i|_{A_i}$ has a bounded degree, both above and below; (2) Byzantine penetration to A_1 is bounded; (3) Edges between honest and byzantine identities are scarce; (4) Community growth in each step is bounded; (5) The conductance within $G_i|_{A_i}$ is sufficiently high; Then, the community may grow indefinitely with bounded byzantine penetration.

Theorem 1 follows by induction from the following Lemma:

Lemma 1: Let $G = (A, V, E)$ and $G' = (A', V, E)$ be two community trust graphs, where $A \subset A'$. Set parameters $\alpha \in [0,1]$ and $\beta, \gamma, \delta \in [0, \frac{1}{2}]$. Assume:

1) *Each vertex in A' has a bounded degree, both above and below*:
$$\alpha \cdot d \leq deg_{A'}(v) \leq d \quad \forall v \in A'$$
2) *Byzantine penetration to the initial community is bounded*:
$$\beta(G) + \frac{\delta}{2} \leq \frac{1}{2}$$
3) *The edges between honest and byzantine identities are relatively scarce*:
$$\frac{e(A' \cap H, A' \cap B)}{vol_{A'}(A' \cap H)} \leq \gamma_e$$
4) *Community growth is bounded*:
$$|A' \setminus A| \leq \delta |A|$$
5) *The conductance within A' is sufficiently high*:
$$\Phi_e(G|_{A'}) > \frac{\gamma_e}{\alpha} \cdot \left(\frac{1-\beta}{\beta}\right)$$

Then, $\beta(G') \leq \beta$.

Proof: First note that even if all the added identities from A to A' are byzantines, it still follows that

$$|A' \cap B| \leq |A \cap B| + |A' \setminus A| = \beta(G) \cdot |A| + |A'| - |A|$$

Applying assumption (2):

$$\begin{aligned} |A' \cap B| &\leq \frac{(1-\delta)|A|}{2} + |A'| - |A| \\ &= \frac{|A'|}{2} - \frac{\delta|A|}{2} + \frac{|A'|}{2} - \frac{|A|}{2} \end{aligned}$$

Applying assumption (4):

$$|A' \cap B| \leq \frac{|A'|}{2} - \frac{\delta|A|}{2} + \frac{\delta|A|}{2} = \frac{|A'|}{2}$$

As $V = B \uplus H$, it follows that:

$$|A' \cap B| \leq |A' \cap H| \tag{2}$$

Now utilizing assumption (1):

$$\begin{aligned} vol_{A'}(A' \cap B) &:= \sum_{a \in A' \cap B} |\{x \in A' \mid (a,x) \in E\}| \\ &\geq \sum_{a \in A' \cap B} \alpha d = \alpha d |A' \cap B|. \end{aligned} \tag{3}$$

Similarly, the following holds:

$$vol_{A'}(A' \cap H) \geq \alpha d |A' \cap H| \tag{4}$$

Inequalities 2 and 4 imply that:

$$vol_{A'}(A' \cap H) \geq \alpha d |A' \cap B|$$

and together with Inequality 3:

$$\min\{vol(A' \cap H), vol(A' \cap B)\} \geq \alpha d |A' \cap B| \tag{5}$$

Now, Inequality 5 and assumption (5) imply that:

$$\frac{e(A' \cap H, A' \cap B)}{\alpha d |A' \cap B|} \geq \frac{e(A' \cap H, A' \cap B)}{\min\{vol(A' \cap H), vol(A' \cap B)\}} > \frac{\gamma_e}{\alpha} \cdot \left(\frac{1-\beta}{\beta}\right)$$

or equivalently:

$$\frac{e(A' \cap H, A' \cap B)}{d\gamma_e |A' \cap B|} \geq \frac{1-\beta}{\beta} \tag{6}$$

Assumptions (1) and (3) imply:

$$\frac{e(A' \cap H, A' \cap B)}{d|A' \cap H|} \leq \frac{e(A' \cap H, A' \cap B)}{vol_{A'}(A' \cap H)} \leq \gamma_e$$

or equivalently:

$$|A' \cap H| \geq \frac{e(A' \cap H, A' \cap B)}{d\gamma_e} \tag{7}$$

Combining Inequalities 6 and 7:

$$\frac{|A'|}{|A' \cap B|} = \frac{|A' \cap H| + |A' \cap B|}{|A' \cap B|} \geq \frac{e(A' \cap H, A' \cap B)}{d\gamma_e |A' \cap B|} + 1 > \left(\frac{1-\beta}{\beta}\right) + 1 = \frac{1}{\beta}$$

where the first equality holds as $A = (A \cap H) \uplus (A \cap B)$, the second inequality stems from Equation 7 and the third inequality stems from Equation 6. Flipping the nominator and the denominator then gives $\beta(A') := \frac{|A' \cap B|}{|A'|} < \beta$. □

Remark 3: A potential application of lemma 1 is a byzantine-resilient union of two communities. Let $A, A' \subseteq V$ denote two communities that have some overlap (non-empty intersection) and wish to unite into $A_2 := A \cup A'$. Then, if lemma 1 holds for (A_1, A_2) in case $A_1 := A$ and also in case $A_1 := A'$, this would provide both A and A' the necessary guarantee that the union would not result in an increase of the sybil penetration rate for either community.

V. Analysis of the Conductance-Based Approach

Our results show the conditions under which a community can grow and maintain sybil safety. It is still not clear however if such conditions are practical. This section takes a closer look at graphs, graph conductance and the interplay between the parameters. We show that under the range of possible parameters in the model and the required conductance derived from these parameters there are indeed many such graphs that meet the requirements. Theoretically, a fully connected graph easily holds these requirements, but trust graphs are rather sparse graphs, so specifically the question is whether sparse graphs can hold these requirements.

A. Sparse Graphs

Recall that the safety of the community growth, more specifically the required level of conductance for the community to grow safely, relies upon the parameters α, β, and γ_e. While a given community may evolve wrt. any choice of parameters, some choices will inevitably yield degenerate outcomes; one case is as the model requires $\Phi_e(G|_{A'}) > \frac{\gamma_e}{\alpha} \cdot \left(\frac{1-\beta}{\beta}\right)$, while the conductance of any graph is upper bounded by $\frac{1}{2}$. Specifically, whenever $\gamma_e \left(\frac{1-\beta}{\beta}\right) > \frac{1}{2}$, the community cannot possibly grow, regardless of the choice of α. While complete graphs and complete bipartite graphs are the classic examples of graphs which satisfy $\Phi_e(G|_{A'}) = \frac{1}{2}$, the fact that their degree is of order $d = \Theta(n)$ makes them unrealistic in our setting, where agents may potentially trust only a uniformly-bounded number of identities. In this context, the main question seems to be the following: *Could a given community safely grow while retaining a given maximal degree d?* Surprisingly, not only that the answer is affirmative, it also holds for a plethora of trust graphs. We utilize Friedman's classical result:

Theorem 2 (Friedman [31], Rephrased): Let G be a random d-regular graph on n vertices. Then, for any $0 < \epsilon$, $\lambda(G) \leq \frac{2\sqrt{d-1}}{d} + \epsilon$ holds with probability $1 - o_n(1)$.

Thus, almost all d-regular graphs on n vertices satisfy $\lambda_2 \leq \frac{2}{\sqrt{d}}$. Applying this term in Cheeger's inequality yields that such graphs satisfy

$$\frac{1}{2} - \frac{1}{\sqrt{d}} \leq \Phi_e \tag{8}$$

meaning that the choice of d affects the level of conductance one hopes to achieve.

B. Parameter Interplay

The following subsection considers numerical examples to better appreciate the analysis above. First, consider the realistic assumption where each identity is assumed to trust up to $d = 100$ identities (notice that this can be enforced by the system). Equation 8 now suggests that a random graph of degree d on n vertices (where d may be constant wrt. n) satisfies $\Phi_e > \frac{2}{5}$. For simplicity, we take this quantity as a benchmark. It follows that whenever $\frac{\gamma_e}{\alpha} \cdot \left(\frac{1-\beta}{\beta}\right) < \frac{2}{5}$, there exist a plethora of potential community histories for which a given community may potentially grow to be arbitrarily large.

Some further examples:

1) If $\gamma_e = 0$, then any community history that begins with a connected byzantine-free community would retain 0-byzantine penetration;
2) The choice $\beta = 0$ is not attainable, corresponding to the intuition that one can never guarantee a completely byzantine-free community growth.

Figure 3 illustrates the parameter interplay further. Notice that the key assumption, stating that honest people tend to trust honest people more than they tend to trust corrupt people, implies that $\gamma_e < \beta$ (as $\gamma_e > \beta$ implies that honest people trust corrupt people more than their relative share in the community).[2]

C. Parameter Estimation

While α and Φ_e can be decided by the community (either by the foremothers of the community or by a global, decentralized

[2]In a separate line of research (in preparation) we consider processes and mechanisms that help lowering γ_e even further.

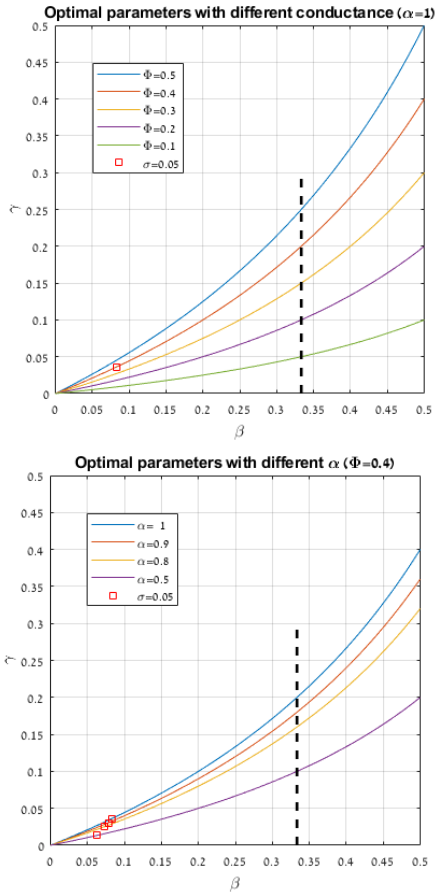

Fig. 3. Parameter Interplay. The top plot shows γ_e as a function of β, for $\alpha = 1$, where each line represents a different conductance Φ_e value. It shows, for example, that if the community fixes $\alpha = 1$ and sets $\Phi_e = 0.4$, then to achieve $\beta = 0.2$ it can tolerate $\gamma_e = 0.1$. The bottom plot shows the effect of α, for $\Phi_e = 0.4$. In both plots, the red rectangles show respective β and γ_e values ensuring $\sigma = 0.05$.

democratic decision making process), $\beta(G)$ and γ_e rely on the dynamics of the community history. To incrementally grow the community at a given time, one may settle for estimating the current state of affairs, as follows. Specifically, assuming that a thorough examination of a given identity could determine whether it is genuine or sybil, one may apply random checks to empirically estimate $\beta(G)$ and γ_e. This could be carried out in the following manner:[3]

1) Examination of an identity $x \in V$ determines whether it is genuine or sybil
2) Examination of the neighbors of a genuine identity $x \in V$ (the ball of radius 1 around it) determines whether it is explicitly (but not latently) corrupt
3) Examination of the ball of radius 2 around an honest identity x determines whether its neighbors are explicitly byzantine

VI. Vertex Expansion Approach

The next section presents our second assumption, which focuses on the corrupt identities themselves, rather then the trust between honest identities and corrupt identities. Thus, we simply assume that there is a bound on how many identities in a community are corrupt. In a trust graph this results in a limited number of vertices on the boundary between honest identities and sybil identities. The following provides sufficient conditions for byzantine-resilient community growth, under the assumption that the population of corrupt identities in the community is bounded. This time we use vertex expansion to derive a bound on the number of byzantine identities.

Theorem 3: Let $\mathcal{G}_V = G_1, G_2, \ldots$ be a community history over V. Let $\beta \leq \frac{1}{2} - \frac{1}{2|A_1|}$, $\gamma_v \in [0, \frac{1}{2}]$, and $\delta = 1 - 2\beta$. Assume:

1) *Byzantine penetration to the initial community is bounded*:
$$\beta(G_1) \leq \beta$$
2) *The population of corrupt identities is bounded*:
$$\frac{|A_i \cap C|}{|A_i|} \leq \gamma_v$$
3) *Community growth is bounded*:
$$|A_i \setminus A_{i-1}| \leq \delta |A_{i-1}|$$
4) *The vertex expansion within A_i is sufficiently high*:
$$\Phi_v(G|_{A_i}) > \frac{\gamma_v}{\beta}$$

Then, every community $G_i \in \mathcal{G}_V$ has Byzantine penetration $\beta(G_i) \leq \beta$.

Notice that there is one less parameter α in the vertex based version of the model. While it was required in the edge based version, to establish a lower bound on the volume of H, and although it has a strong intuition for our goal (the more honest identities trust each other, the harder it is for the untrusted to penetrate their community), the theorem for the vertex based version will hold without it. This makes this version slightly simpler, as there is one less parameter that the community needs to decide upon.

As before, theorem 3 follows by induction from the following Lemma:

Lemma 2: Let $G = (A, V, E)$ and $G' = (A', V, E)$ be two community trust graphs, where $A \subset A'$. Set parameters $\beta, \gamma, \delta \in [0, \frac{1}{2}]$. Assume:

1) *Byzantine penetration to the initial community is bounded*:
$$\beta(G) + \frac{\delta}{2} \leq \frac{1}{2}$$
2) *The population of corrupt identities is bounded in A':*
$$\frac{|A' \cap C|}{|A'|} \leq \gamma_v$$
3) *Community growth is bounded*:
$$|A' \setminus A| \leq \delta |A|$$
4) *The vertex expansion within A' is sufficiently high*:
$$\Phi_v(G|_{A'}) > \frac{\gamma_v}{\beta}$$

Then, $\beta(G') \leq \beta$.

[3]A related sampling-based approach to estimate the number of sybils is briefly discussed by Shahaf *et al.* [3, Remark 2].

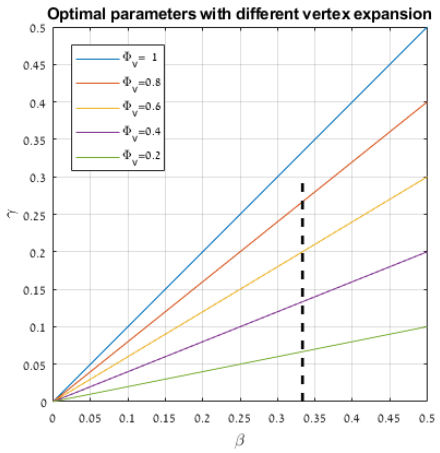

Fig. 4. Parameter Interplay. The plot shows γ_v as a function of β, where each line represents a different vertex expansion Φ_v value. It shows, for example, that if the community sets $\Phi_v = 0.6$, then to achieve $\beta = \frac{1}{3}$ it can tolerate $\gamma_v = 0.2$. There are only three parameters in the vertex expansion approach, as it does not require α-solidarity, hence there is just one graph to show in this section.

Proof: Similarly to the proof of lemma 1, assumptions (1) and (3) imply that:

$$|A' \cap B| \leq |A' \cap H| \tag{9}$$

Inequality 9 and assumption (4) imply that:

$$\frac{\gamma_v}{\beta} \leq \Phi_v(G'|_{A'}) \leq \frac{\partial_v(A' \cap B, A' \cap H)}{|A' \cap B|} \leq \frac{|A' \cap C|}{|A' \cap B|}$$

where the last inequality stems from definition 6 (there are no edges between H and S, therefore the boundary between B and H is a subset of C). Applying assumption (2) it follows that:

$$\frac{\gamma_v}{\beta} \leq \frac{\gamma_v |A'|}{|A' \cap B|}$$

which leads to

$$\frac{|A' \cap B|}{|A'|} \leq \beta$$

That is, G' has byzantine penetration $\beta(G') \leq \beta$. □

Remark 4: Our two results for community growth, one based on conductance and the other based on vertex expansion, are very similar. The main difference between them lies in the premises of the two corollaries. The first assumes that honest people tend to trust honest people more than they tend to trust corrupt people. The second, which may be more naïve, directly assumes that there are not too many corrupted people in a given community to begin with. Again, the conditions under which we assume either of these bounds to be low is the subject of a separate line of work.

VII. Analysis of the Vertex Expansion Approach

Given a d-regular graph it can be shown that the inner boundary vertex expansion of the graph is at least as high as the graph conductance. Assume w.l.o.g that $|A| \leq |A^c|$, since $\partial_v(A, A^c) \cdot d \geq e(A, A^c)$ it follows that:

$$\frac{\partial_v(A, A^c)}{|A|} = \frac{d \cdot \partial_v(A, A^c)}{d \cdot |A|} \geq \frac{e(A, A^c)}{vol(A)}$$

Going back to the numeric example in subsection V-B, now setting $\Phi_v = \frac{2}{5}$ then it follows that whenever $\frac{\gamma_v}{\beta} < \frac{2}{5}$, there exist a plethora of potential community histories for which a given community may potentially grow to be arbitrarily large. As an example, if the community wishes to achieve $\beta = 0.2$ then it can tolerate $\gamma_v = 0.08$. Figure 4 illustrates the parameter interplay further. The line $\Phi_v = 1$ shows a theoretical example where for each subset $A \subset V$, for every $x \in A$ there exist $y \in A^c$ such that $(x, y) \in E$. Assuming there is at least one honest identity in the community, and remembering that there cannot be an edge between an honest identity and a sybil identity, it follows that there are no sybils in any such community in V. The line $\Phi_v = 1$ expresses this result as it shows that $\gamma_v = \beta$, which leads to $S = \emptyset$.

Maintaining $\Phi_v = 0.5$ leads to $\beta = 2\gamma_v$ which means that the number of sybils in any such community is at most the number of corrupted identities that are willing to share an edge with a sybil identity. Unfortunately, the down side of using vertex expansion over conductance is that, as far as we know, there is no known way to measure or approximate vertex expansion better than the relation between vertex expansion and conductance shown above. We are also unaware of any method to construct a graph with vertex expansion 0.5 or higher with a constant degree d.

VIII. Simulation

The mathematical analysis described above shows some bound on the byzantine penetration. In particular, the example given in the previous section shows that a random d-regular graph with $d = 100$, where the community maintains a bound of no more than 8% corrupt identities, will maintain with high probability a vertex expansion measure that will guarantee no more than 20% of byzantines in the community. In this section we report on simulations we ran to investigate whether better results are achievable in practice. As we are not aware of existing randomized d-regular social networks we turn to simulations on synthetic data. There are several parameters to play with in the algorithms presented here and a comprehensive simulation is a topic for further research. Thus, here we concentrate on a single simulation run that exemplifies the usage of the above algorithms. As it is easier to simulate the vertex expansion approach, we start with this direction.

The pseudocode presented in Method 1 describes the specific simulation we performed. It accepts the following parameters:

- n - the number of identities in the final community.
- d - the degree of the community graph.
- $corruptLimit$ - the bound on the number of corrupt identities allowed in the community.
- $corruptProb$ - the probability of a newcomer to be corrupt.
- $sybilProb$ - the probability of a newcomer to be a sybil.

Method 1 Community Growth Simulation

function VERTEXSIMULATION(n, d, $corruptLimit$, $corruptProb$, $sybilProb$, $lambda$)
  init A with a $d+1$ clique
  init $ident$ and $failed$ with zeros
  $ident(1:d+1) \leftarrow 1$
  $index \leftarrow d+1$

  while $index < n$ **do**
    $B \leftarrow A$
    $nc \leftarrow$ draw a newcomer with given probs
    Avoid corrupt if limit reached
    loop $\frac{d}{2}$ times
      $edge \leftarrow$ draw a random edge
      Avoid edges between honest and sybils
      Connect newcomer between both vertices
      $v \leftarrow$ eigenvalues of B
    end loop
    if max($v(2)$, abs($v(last)$))< $lambda$ **then**
      $A \leftarrow B$
      $index \leftarrow index + 1$
      $ident(index) \leftarrow nc$
    else
      $failed(nc, index) \leftarrow failed(nc, index) + 1$
    end if
  end while
  return A, $ident$, $failed$
end function

- $lambda$ - the bound on the second eigenvalue of the random walk matrix.

The simulation starts with a clique of $d+1$ honest nodes and loops until the size of the community reaches n. At each iteration a newcomer is trying to join the community. Its type (honest, corrupt or sybil) is drawn according to the given parameters, while avoiding corrupt newcomers, if the bound of corrupt identities in the community was reached. The simulation then draws $\frac{d}{2}$ edges from the graph and connects the newcomer in the middle of these edges (connects the newcomer with edges to both sides of the existing edge, and removes it). This growing method guarantees that the graph remains d-regular at all times. The simulation then calculates the eigenvalues of the random walk matrix of the resulted graph. If the second eigenvalue is lower than $lambda$, then the graph is updated accordingly (the newcomer joins the community).

We ran the simulation with the following parameters: $n = 1000$, $d = 10$, $corruptLimit = 0.2$, $corruptProb = 0.1$, $sybilProb = 0.8$ and $lambda = 0.6$. The results of the simulation are shown in figure 5. It shows that, although 80% of newcomers were sybils, most of them were rejected by the community. Overall, 754 honest identities, 199 corrupt, and only 47 sybils joined the community. This is significantly better than the analytical analysis that guarantees 30% of sybils with the above parameters ($\Phi_v > 1 - 0.6 = 0.4$, $\gamma_v = 0.2$, $\beta = 0.5$, $\sigma = \beta - \gamma_v = 0.3$).

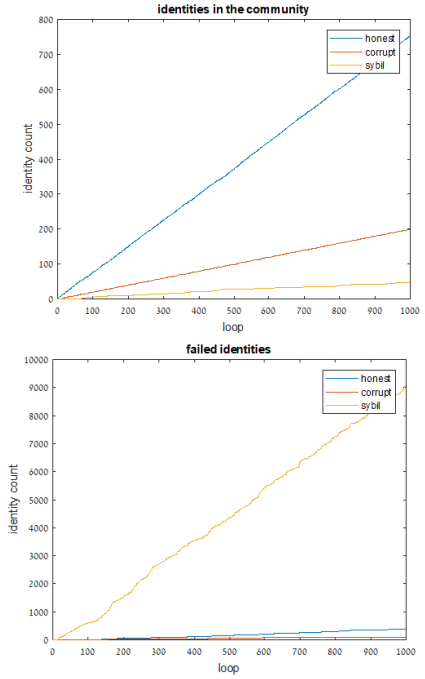

Fig. 5. Community growth simulation. The top plot shows the identities that were accepted into the community. It shows that, although many sybils attempted to join, only a bounded number succeeded. The bottom plot shows the identities that were rejected, as the expansion of the graph was not high enough.

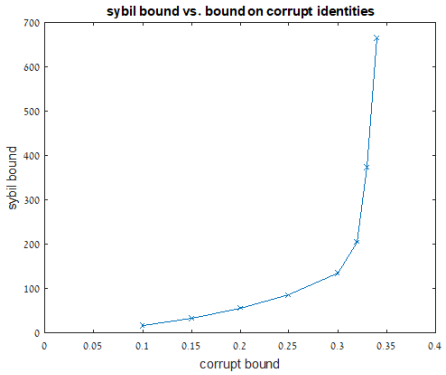

Fig. 6. Actual bound on sybils as a function of the limit on corrupt identities.

We ran the simulation for several other values of $corruptLimit$. The results are shown in figure 6. It shows that there is a tipping point around $corruptLimit = 0.32$, at which point the number of byzantines (corrupt and sybils together) may exceed one half, after which sybils can infiltrate the graph freely, turning the honest minority to be the under connected part of the graph.

IX. REAL DATA

So far we have shown that, if a community can maintain a d-regular trust graph with a good enough second eigenvalue of the random walk matrix, then it can maintain a good bound on the penetration of sybils. We also showed that a random d-regular graph maintains a good enough second eigenvalue with high enough probability. In this section we consider

the efficiency of our method with respect to existing social networks. Friendship graphs of such networks usually are not d-regular, and they are not chaotic as random graphs. This makes it harder to separate between sybils and non-sybils, as the graph structure of these two communities may look similar to the structure of any two communities of genuine identities, with few connections between the two.

In the previous section we presented method 1, which constructs a simulated random graph, classifies its nodes into the three types of populations, and grows a community based on the conditions presented in section VI. The first experiment in this section uses the same community growth method, but rather than generating a simulated random graph, it uses a sample from Facebook [32] as the underlying graph G. This sample, taken from the work of Leskovec and Mcauley, is a collection of 10 identities from facebook and their friends, a total of 4039 connected nodes.

A. A Small Sample From Facebook

The sample contains all genuine identities, which Leskovec and Mcauley manually analyzed to construct their social circles. Section V shows analytically that a d-regular graph of genuine identities with a good second eigenvalue has a good bound of sybils. This experiment focuses on testing whether genuine identities can form such a graph, based on their personal acquaintances. It therefore regards all identities in the sample as genuine. Figure 7 Shows the results. Since method 1 uses random decisions on whom to add next to the community, we ran the first test 1000 times, in which we checked how big a community the algorithm generates on each attempt. The graph at the top shows the histogram of the size of the final community. The largest community the method managed to grow had 294 members out of 4039 nodes in the underlying graph.

Next, we relax the algorithm to allow friends to introduce new friends to each other. When a newcomer does not have enough edges within the growing community, we create new edges over the underlying graph G, first by allowing friends of friends to connect with each other, then allowing also for identities that are further away from each other to connect. The purpose of this test is to check whether the assumption that trust is transitive to some degree can help the algorithm to generate bigger communities. We first measure how big can a community grow, then how much transitivity of trust is required.

The graph at the bottom (Figure 7) shows the results of a single run. It shows the distribution of distances as the community grows. Distance 1 shows the ratio of edges in the community A which are also edges in the underlying friendship graph G. Distance 2 shows the ratio of edges in A that are friends of friends in G, and so on. The results show that the relaxed algorithm managed to grow a community to a size of 2306 members, out of 4039 nodes in G. The first 134 members in the community were only connected to direct friends in G. As the community reaches its maximal size, almost half of the edges in A, are either edges in G (31%) or friends of friends in G (18%). The conclusion is therefore that, by itself, social networks are not a good enough trust graphs

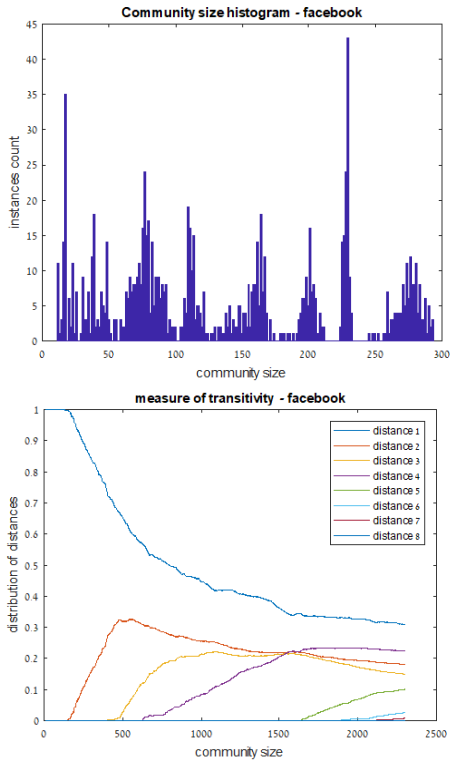

Fig. 7. Results of running method 1 on a sample from Facebook. The graph at the top shows the histogram of the maximal size of the community within 1000 runs. The largest community has 294 members out of 4039 identities in the sample. The graph at the bottom shows the distribution of distances between neighbours in the community (edges in A), when friends introduce new friends to create more trust edges.

to grow sybil resilient communities, but by allowing friends to introduce new friends to each other, bigger sybil resilient communities can be formed.

B. A Bigger Network From DBLP

The sample from Facebook is quite small; thus, to check whether method 1 is also suitable for larger social networks, next we examine its performance on a graph from the DBLP computer science bibliography database. The data is taken from the work of Yang and Leskovec [33], which evaluates different structural definitions of network communities based on a ground truth. They used the DBLP network, taking joint publication venues as an indication of joint scientific communities. This network consists of scientific publications in computer science, with scientists as nodes in the graph and edges connecting scientists that share at least one joint publication. The data in this database is the main connected component of this network, as used by Yang and Leskovec, which includes 317,080 nodes (computer scientists) and 1,049,866 edges (joint academic papers authorship).

Figure 8 shows the results of running method 1 on the DBLP network. As before, the graph at the top shows the histogram of running the method as is, for 1000 times. The results here are worse than before: The largest community found had only 129 members, and most runs ended with less than 20 members in the community. This is expected, as the Facebook sample consists of friends of a small group of identities (10), with

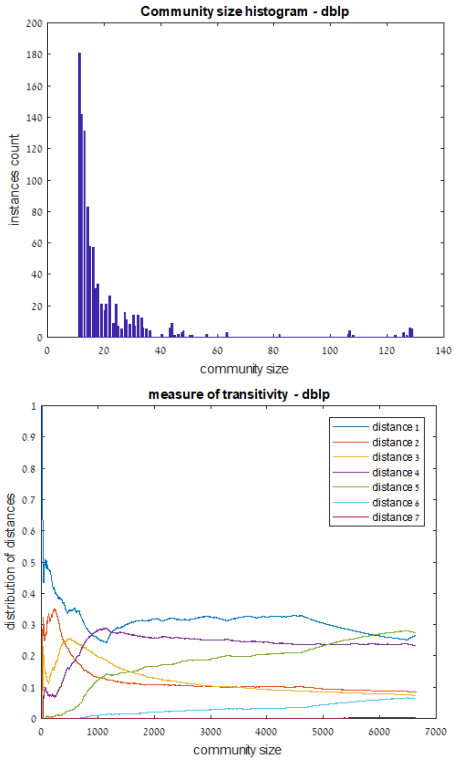

Fig. 8. Results of running method 1 on a the DBLP network. The graph at the top shows the histogram of the maximal size of the community within 1000 runs. The largest community has 129 members out of 317,080 identities in the underlying social graph. The graph at the bottom shows the distribution of distances between neighbours in the community (edges in A), when friends introduce new friends to create more trust edges. It shows that about 30% of the edges in A are friends in the underlying social graph G (distance 1), about 10% are friends of friends (distance 2), and so on.

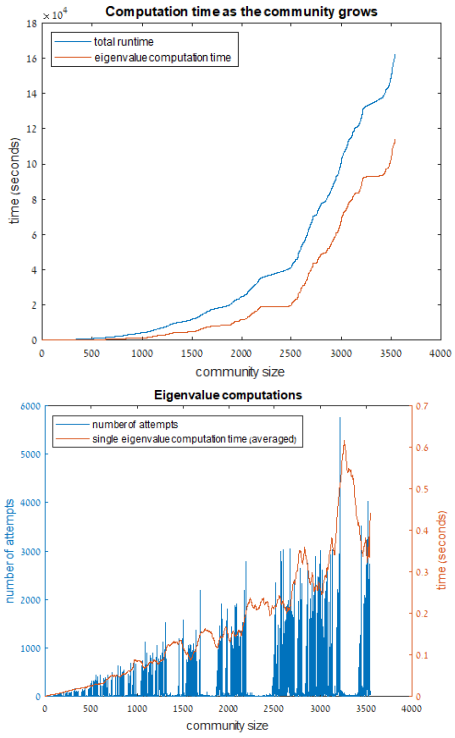

Fig. 9. Time measurements. The graph at the top shows the overall running time of method 1 and the overall eigenvalue computation time. Eigenvalue computation time is about 70% of the overall running time for a community of 3500 identities. The graph at the bottom shows the computation time of a single call to the eigenvalue calculation function (right y axis, averaged over 100 cycles), together with the number of attempts to add a new member to the community (left y axis).

many friends; while the DBLP network is larger and more dispersed. Therefore most attempts to randomly find highly connected communities within this network come short.

The second test exposes the processing time limitations of method 1. It took about two weeks to grow a community of about 6600 members, at which point we paused the run to take the data in the bottom graph of figure 8. The program continued to run for about six weeks without reaching an end. The growing community crossed 10,000 members, and it is unclear how far can it grow further, given more time. The graph shows that the DBLP network requires even more transitivity of trust than the Facebook sample, though still about half the edges in the growing community are of distance 3 or less in the underlying graph.

We ran the DBLP database once more, to measure processing time. All tests where done with Matlab R2017b, on an Intel i7-7700 CPU with 4 cores of 3.6 GHz and 16 GB of RAM. The measurements show that the main bottleneck is the computation of the second eigenvalue of the random walk matrix. The graph at the top (figure 9) shows the overall running time and the overall eigenvalue computation time. It shows that the eigenvalue computation time takes about 70% of the total running time. The graph at the bottom splits this bottleneck into the cumulative computation time of the second eigenvalue of a single matrix (right y axis) and the number of attempts (left y axis) to find a suitable new member who satisfies the condition in line 17 of method 1.

To conclude, this paper shows that the connectivity of the graph (either conductance or vertex expansion) can bound the penetration of sybils and that the second eigenvalue of the random walk matrix can be used to bound the connectivity of the graph. The conclusion of this subsection is that method 1 is effective for building sybil resilient communities of up to a few thousands of members. Bigger communities will require more efficient ways to bound the connectivity.

X. Outlook

We proposed two methods which allow a digital community to grow in a sybil-safe way. We analyzed them mathematically and showed that they are not only safe, but also feasible. Simulations show even better results than the mathematical analysis. Existing social networks are not a good candidate for these methods, but we showed a possible direction how to leverage the trust in existing social networks to construct efficient trust-graphs. A naïve imlementation of our method works well for a community of few thousand members. Bigger communities require more efficient ways to compute the connectivity of the graph. Future research also includes mechanisms for penalizing the creation of attack edges while rewarding sybil hunting, modeling the possibility of honest identities abandoning the community, and expanding with more simulations and real data analysis to better understand the dynamics of safe growth.

ACKNOWLEDGMENT

The authors would like to thank the Braginsky Center for the Interface between Science and the Humanities for their generous support.

REFERENCES

[1] O. Poupko, G. Shahaf, E. Shapiro, and N. Talmon, "Sybil-resilient conductance-based community growth," in *Computer Science—Theory and Applications*, R. van Bevern and G. Kucherov, Eds. Novosibirsk, Russia: Springer, Jul. 2019, pp. 359–371.

[2] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Trans. Program. Lang. Syst.*, vol. 4, no. 3, pp. 382–401, Jul. 1982.

[3] G. Shahaf, E. Shapiro, and N. Talmon, "Sybil-resilient reality-aware social choice," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Macao, China, Aug. 2019, pp. 572–579. [Online]. Available: https://www.ijcai19.org/

[4] G. Shahaf, E. Shapiro, and N. Talmon, "Genuine personal identifiers and mutual sureties for Sybil-resilient community formation," 2019, *arXiv:1904.09630*. [Online]. Available: http://arxiv.org/abs/1904.09630

[5] E. Shapiro, "Global cryptodemocracy is possible and desirable," in *Cloud Communities: The Dawn of Global Citizenship?* Florence, Italy: Globalcit, 2018. [Online]. Available: http://globalcit.eu/cloud-communities-the-dawn-of-global-citizenship/15/

[6] E. Shapiro, "Point: Foundations of e-democracy," *Commun. ACM*, vol. 61, no. 8, pp. 31–34, 2018.

[7] J. R. Douceur, "The Sybil attack," in *Proc. Int. Workshop Peer-to-Peer Syst.*, 2002, pp. 251–260.

[8] J. Newsome, E. Shi, D. Song, and A. Perrig, "The Sybil attack in sensor networks: Analysis & defenses," in *Proc. IPSN*, 2004, pp. 259–268.

[9] B. N. Levine, C. Shields, and N. B. Margolin, "A survey of solutions to the Sybil attack," Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep. 2006-052, 2006, p. 224, vol. 7.

[10] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, "SybilGuard: Defending against Sybil attacks via social networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 267–278, Aug. 2006.

[11] H. Yu, "Sybil defenses via social networks: A tutorial and survey," *ACM SIGACT News*, vol. 42, no. 3, pp. 80–101, Oct. 2011.

[12] G. Danezis and P. Mittal, "SybilInfer: Detecting Sybil nodes using social networks," in *Proc. NDSS*, 2009, pp. 1–15.

[13] D. N. Tran, B. Min, J. Li, and L. Subramanian, "Sybil-resilient online content voting," in *Proc. NSDI*, 2009, pp. 15–28.

[14] N. Tran, J. Li, L. Subramanian, and S. S. M. Chow, "Optimal Sybil-resilient node admission control," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 3218–3226.

[15] W. Wei, F. Xu, C. C. Tan, and Q. Li, "SybilDefender: Defend against Sybil attacks in large social networks," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1951–1959.

[16] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in *Proc. NSDI*, 2012, p. 15.

[17] L. Alvisi, A. Clement, A. Epasto, S. Lattanzi, and A. Panconesi, "SoK: The evolution of Sybil defense via social networks," in *Proc. SP*, May 2013, pp. 382–396.

[18] Y. Boshmaf *et al.*, "Ìntegro: Leveraging victim prediction for robust fake account detection in large scale OSNs," *Comput. Secur.*, vol. 61, pp. 142–168, Aug. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404816300633

[19] C. Liu, P. Gao, M. Wright, and P. Mittal, "Exploiting temporal dynamics in Sybil defenses," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*. New York, NY, USA: Association for Computing Machinery, Oct. 2015, pp. 805–816, doi: 10.1145/2810103.2813693.

[20] S. Friebe, P. Martinat, and M. Zitterbart, "Detasyr: Decentralized ticket-based authorization with Sybil resistance," in *Proc. IEEE 44th Conf. Local Comput. Netw. (LCN)*, Oct. 2019, pp. 60–68.

[21] N. Alon, E. Mossel, and R. Pemantle, "Distributed corruption detection in networks," 2015, *arXiv:1505.05637*. [Online]. Available: http://arxiv.org/abs/1505.05637

[22] Y. Jin, E. Mossel, and G. Ramnarayan, "Being corrupt requires being clever, but detecting corruption Doesn't," 2018, *arXiv:1809.10325*. [Online]. Available: http://arxiv.org/abs/1809.10325

[23] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[24] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, "SybilLimit: A near-optimal social network defense against Sybil attacks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 885–898, Jun. 2010.

[25] R. Diestel, *Graph Theory* (Graduate Texts in Mathematics). Heidelberg, Germany: Springer, 2017. [Online]. Available: http://diestel-graph-theory.com/

[26] S. Hoory, N. Linial, and A. Wigderson, "Expander graphs and their applications," *Bull. Amer. Math. Soc.*, vol. 43, pp. 439–561, Aug. 2006.

[27] M. Blum, R. M. Karp, O. Vornberger, C. H. Papadimitriu, and M. Yannakakis, "The complexity of testing whether a graph is a superconcentrator," *Inf. Process. Lett.*, vol. 13, nos. 4–5, pp. 164–167, Jan. 1981.

[28] J. Cheeger, "A lower bound for the smallest eigenvalue of the Laplacian," in *Proc. Princeton Conf. Honor Professor S. Bochner*, 1969, pp. 195–199.

[29] M. Jerrum and A. Sinclair, "The Markov chain Monte Carlo method: An approach to approximate counting and integration," in *Approximation Algorithms for NP-Hard Problems*. Boston, MA, USA: PWS Publishing, 1996, pp. 482–520.

[30] F. R. Chung, *Spectral Graph Theory* (CBMS Regional Conference Series in Mathematics), no. 92. Providence, RI, USA: American Mathematical Society, 1997.

[31] J. Friedman, "A proof of Alon's second eigenvalue conjecture," in *Proc. STOC*, 2003, pp. 720–724.

[32] J. Leskovec and J. Mcauley, "Learning to discover social circles in ego networks," in *Advances in Neural Information Processing Systems*, vol. 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 539–547. [Online]. Available: https://proceedings.neurips.cc/paper/2012/file/7a614fd06c325499f1680b9896beedeb-Paper.pdf

[33] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowl. Inf. Syst.*, vol. 42, no. 1, pp. 181–213, Jan. 2015, doi: 10.1007/s10115-013-0693-z.