Growing a Web of Trust

Benjamin Schiller and Thorsten Strufe
Privacy and Data Security
TU Dresden, Germany
benjamin.schiller1@tu-dresden.de
thorsten.strufe@tu-dresden.de

Dirk Kohlweyer
Computational Economics
Bielefeld University, Germany
dkohlweyer@uni-bielefeld.de

Jan Seedorf
NEC Laboratories Europe
Heidelberg, Germany
jan.seedorf@neclab.eu

***Abstract*—**
Web of Trust (WoT) graphs represent trust relations between people. They are used for research and analysis in various domains. Real-world instances are only available in sizes of up to 55k vertices. This renders the analysis of larger systems based on realistic input graphs impossible. To close this gap, we develop a growth model to generate WoT graphs of arbitrary size. New edges are formed based on realistic assumptions about trust establishment. We analyze the growth of a real-world WoT and perform a parameter study of our model. We compare both with many existing models and show that ours is the only one that matches the properties of the real-world WoT.

I. Introduction

Trust between users of a system is often represented by a trust graph called a *Web of Trust* (WoT). Users are represented as vertices and their trust relations are indicated by edges between them. Using such a WoT, users can decide who they trust based on various metrics like, e.g., their distance in the WoT, the number of disjunct paths between them, or the degree of the respective vertex. Hence, WoT instances are required for the development and evaluation of such trust metrics [5], [23]. They are also used to model trust in many different fields including ad-hoc networks [21], [24], the semantic web [7], [8], and opportunistic networks [25]. Furthermore, WoTs have been used as the topology for evaluating routing algorithms in social networks [16] and darknets [9] and as the foundation for self-certifying names in fragmented mobile networks [23].

Real-world WoTs are only available in small sizes with less than 60,000 vertices. Hence, the evaluation of the aforementioned systems based on a WoT can only be done for a strictly limited number of users. Since it is not feasible to create larger trust networks from surveys or other systems, a model is required to generate WoT graphs of arbitrary sizes. Such a model enables the research in areas like trust-metrics, opportunistic networks, and darknets to scale to arbitrary size without resorting to the use of unrealistic trust graphs.

When creating such a WoT model, we need to consider the properties it must possess. First, a WoT model should create graphs with degree distributions similar to those of real-world WoTs. Since people appear more trustworthy in case many others trust them, it is crucial to correctly map this property. In addition, a model should correctly reflect the extent to which users trust each other in contrast to one-way trust relations. Second, the shortest path length distribution of real-world WoTs indicates the fractions of trusted users, depending on the distance threshold. Therefore, it is crucial for a model to correctly reflect the distances between vertices, especially for the development of meaningful trust metrics and routing applications. Third, the clustering coefficient of generated graphs should be close to the number of connected neighbors in a WoT. It reflects the fraction of triangular trust relationships between neighboring users. Thereby, mapping this property correctly ensures that realistic connections between trusted users are created. In addition, other graph-theoretic properties like rich-club connectivity, community structure, and motif frequencies can be analyzed. We consider these properties but focus on the development of a realistic WoT model that achieves degree distribution, shortest path length distribution, and clustering coefficients close to real-world trust graphs.

A well-known WoT is the certificate graph of *Pretty Good Privacy* (PGP). *PGP* is a popular public-key cryptography system for the encryption and authentication of email communication [29], [1]. Instead of relying on a central authority to sign a user's certificate, users sign each other's certificates to gain and express trust. This concept can be modeled as a graph, the *PGP WoT*. Each user is modeled as a vertex and each signature is represented as an edge from singer to signee.

The remainder of this paper is structured as follows: We introduce terminology and definitions in Section II, present an analysis of a small trust graph in Section III, and derive generative principles for a WoT model. We discuss related work in Section IV and introduce a new WoT model in Section V. In Section VI, we analyze existing graph models and evaluate the capabilities of our model to generate larger WoT graphs. We summarize our work in Section VII and describe future work.

II. Terminology and Definitions

In this Section, we introduce our terminology for graphs and define graph-theoretic metrics used for analysis and evaluation.

A. Graphs

A simple directed graph $G = (V, E)$ consists of a set of vertices $V = \{v_1, v_2, \dots v_{|V|}\}$ and a set of edges $E \subseteq (V \times V) \backslash \{(v, v), v \in V\}$. In a WoT, users are represented by vertices. A directed edge $e = (v, w)$ expresses that v trusts w. In case two users v and w trust each other, an edge in both directions exists, i.e., $(v, w) \in E \wedge (w, v) \in E$. We refer to

$e^{-1} = (v, w)$ as the inverse edge of $e = (w, v)$. We call e bidirectional in case e^{-1} exists and unidirectional otherwise.

We denote the outgoing connections of a vertex v as $out(v) := \{w \in V : (v, w) \in E\}$ and its incoming connections as $in(v) := \{w \in V : (w, v) \in E\}$. Vertices connected to v over a bidirectional edge are denoted as $n(v) := in(v) \cap out(v)$. The (in-/out-)degree of a vertex v is then defined as $d_{in}(v) := |in(v)|$, $d_{out}(v) := |out(v)|$, $d(v) := d_{in}(v) + d_{out}(v)$. We define the set of all edges between elements of V' as $E(V') := E \cap (V' \times V')$.

B. Graph-theoretic metrics

The *degree distribution* of a graph is defined as $P_d(X = x) := \frac{|\{v \in V : d(v) = x\}|}{|V|}$. We denote the average, minimum, and maximum vertex degree as d_{avg}, d_{min}, and d_{max}. We define the *bidirectionality* of a graph as the fraction of bidirectional edges in E: $d_{bid} := |\{e \in E : e^{-1} \in E\}| \cdot |E|^{-1}$. We refer to the fraction of vertices with degree 2 that have a bidirectional instead of two unidirectional edges as $d_{bid,2}$.

The shortest path length between two vertices v and w is denoted as $spl(v, w)$. The *shortest path length distribution* is then defined as $P_{spl}(X = x) := \frac{|\{(v,w), v \neq w : spl(v,w) = x\}|}{|V| \cdot (|V|-1)}$. We denote the average and median shortest path length as spl_{avg} and spl_{med}. We denote the diameter as spl_{max} and the effective diameter as $spl_{90\%}$.

The *local clustering coefficient* of a vertex v measures how many connections exist between its neighbors and is defined as $lcc(v) := \frac{|E(n(v))|}{|n(v)| \cdot (|n(v)|-1)}$ for $|n(v)| > 1$ and $lcc(v) := 0$ otherwise. The *clustering coefficient* of a graph is defined as the average: $cc := \frac{\sum_{v \in V} lcc(v)}{|V|}$. The *transitivity* of a graph is defined as the fraction of existing edges between all 2-hop neighbors, i.e., $t := \frac{\sum_{v \in V} |E(n(v))|}{\sum_{v \in V} |n(v)| \cdot (|n(v)|-1)}$. It is a more global measure and not biased by low-degree vertices.

A community $C_i \subseteq V$ is a set of vertices that are closely interconnected but have only a few connections to nodes in other communities. When assinging each node to a single community, the set of all communities $C = \{C_1, C_2, \dots C_{|C|}\}$ defines a partition of the vertex set V. We define the *community size distribution* of a graph for a community partition C as $P_C(X = x) := \frac{|\{C_i \in C : |C_i| = x\}|}{|C|}$.

The *rich-club connectivity* measures how interconnected the top-k vertices, ranked by their degree, are. The rich-club connectivity of the k highest-degree vertices, denoted as RC_k is then defined as $rcc(k) := \frac{E(RC_k)}{|RC_k| \cdot (|RC_k|-1)}$ for $2 \leq k \leq |V|$.

There exist 13 distinct graph structures of 3 connected vertices called the directed 3-vertex motifs $(m_1, m_2, \dots m_{13})$ [17]. They differ in the number and pattern of edges between the 3 vertices. We refer to the *motif signature* as the fraction of subgraphs isomorph to each motif m_k contained in a graph: $P_m(X = x) := \frac{\text{occurrences of } m_x}{\text{total number of motifs}}$.

III. Analysis of a real-world WoT

In this Section, we analyze a snapshot of the *PGP WoT* taken in February 2005[1], in the following referred to as *WoT*$_{25k}$. Based on the properties of this largest strongly connected component of the *PGP WoT*, we formulate basic insights and derive generative properties of a WoT.

A. Properties of the PGP WoT

In February 2005, the *PGP WoT* consisted of 25,487 vertices connected by 230,455 edges, i.e., $d_{avg} \approx 18.08$. The degree distribution follows a power-law: 20% of vertices have a degree of 2 while the maximum degree is 1,368 (cf. Figure 1a). We performed an ordinary least squares estimation on the logarithmized frequencies and obtained a power-law exponent of 1.69. $d_{bid} = 52\%$ of all edges are bidirectional and $d_{bid,2} = 86\%$ of the vertices with the minimum degree of 2 have a single bidirectional instead of two unidirectional edges. Because of the graph's strong connectivity, each vertex has at least one incoming and one outgoing connection.

Transitivity $t = 0.39$ and clustering coefficient $cc = 0.37$ are high. Nearly 40% of all possible connections between neighbors of a vertex exist. This means that neighborhoods in the *PGP WoT* are even more densely interconnected than in many other social networks.

Since the *PGP WoT* is strongly connected, there exists a path between any two vertices. The graph has short average path lengths of 5.99 and a median path length of 6. While the graph has a high diameter of 25, 90% of all shortest paths have a length of 8 or less (cf. Figure 1b). This indicates that the *PGP WoT*, like many social networks, exhibits the small-world phenomenon, often explained by the observed power-law degree distribution and a set of well-connected, central vertices with a high degree.

This high interconnection of vertices with a high degree is well documented by the rich-club connectivity of the WoT shown in Figure 1c. The top-10 vertices with highest degree are highly interconnected: 72% of all possible edges exist. Even for the top-100 vertices, 41% of all possible edges exist.

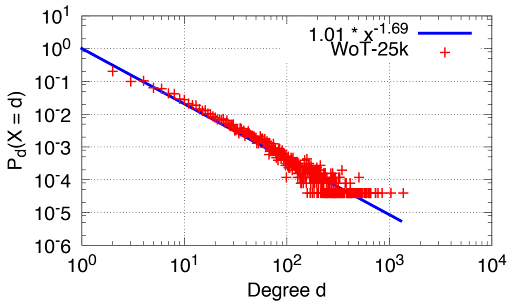
The size distribution of communities also appears to follow a power-law (cf. Figure 1d). Most of the 4,238 communities, found by the fast unfolding community detection algorithm [4], contain less than 3 vertices. While the average community size is 6.01, the largest community contains 173 vertices.

B. Basic insights from the PGP WoT

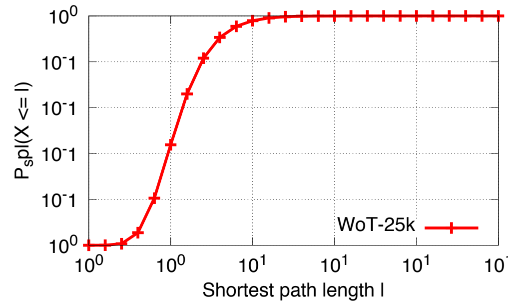
For evolving graphs, power-law degree distributions are often explained and produced by preferential attachment: High-degree vertices have a higher chance of getting further connections. In the context of the *PGP WoT*, it is reasonable to expect that users who already signed many certificates are more likely to sign further ones. Therefore, we assume that new vertices connect to a WoT with preferential attachment.

The high values of transitivity and average clustering coefficient imply that neighborhoods of vertices are often densely interconnected. This means that connections are often formed between vertices that have a neighbor in common [14] and are similar to each other [20]. In the *PGP WoT* this means that users seem more likely to sign the certificate of a user already

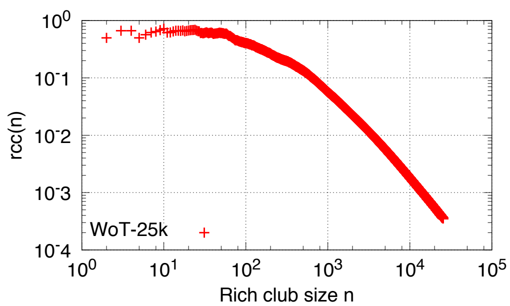
[1] Snapshot obtained from http://www.lysator.liu.se/∼jc/wotsap/wots2/



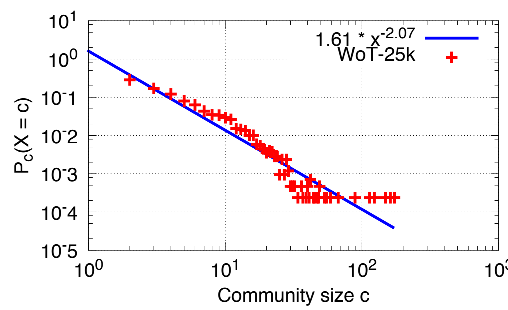
(a) Degree disribution: P_d



(b) Shortest paths: P_{spl} (cdf)



(c) Rich-club connectivity: rcc



(d) Community sizes: P_C

Fig. 1. Graph-theoretic properties of WoT_{25k}

connected to a neighbor and therefore deemed trustworthy. This interpretation is based on the idea that trust in a WoT is transitive, i.e., a user trusts and signs the users her contacts trust. Hence, we assume that new connections in a WoT are created between vertices that already share a neighbor.

The observed power-law distribution of community sizes indicates that there is a small number of large communities that are well interconnected. Smaller communities must be connected to these larger ones. Otherwise, the existence of many small communities would increase the average path lengths noticeably. In the *PGP WoT*, this observation can be explained by the existence of real-world communities, e.g., at a university, in a city, or a country. While they are well interconnected, central users have connections to other communities and thereby create short overall path lengths. Therefore, we assume a community structure of well interconnected vertex subsets in a WoT. While they are connected to many other communities, we assume that most of them maintain connections to a small set of large communities whose high-degree vertices act as bridges between the smaller ones.

C. Generative concepts for WoT graphs

Based on our insights into the *PGP WoT*, we assume the following concepts to evolve a WoT over time:

1) New vertices connect by preferential attachment
2) New edges are created between 2-hop neighbors
3) Community sizes follow a power-law

IV. Related Work

The existing literature provides many different models for generating artificial graphs with desirable properties.

The Erdős-Rényi model [6] (*ER*) generates a random graph with a specified number of edges and a Gaussian degree distribution which is unrealistic for most real-world graphs. The Power-law model [19] (*PL*) generates graphs that follow a parametrized power-law degree distribution.

Many real-world graphs are considered to be so-called *small-world* graphs [2]. They are characterized by small shortest paths and high clustering as observed in many social networks [18]. The Watts-Strogatz model [27] (*WS*) generates graphs with this property. It starts with a regular ring and creates shortcuts by rewiring a fraction of the edges.

The Barabási-Albert model [3] (*BA*) simulates the growth of a graph by adding vertices one at a time to a random graph. Based on the idea of preferential attachment, new vertices favor connections to high-degree vertices which results in a power-law degree distribution. Newer models achieve a high rich-club connectivity [28] or enrich the preferential selection with vertex properties like popularity or similarity [20].

The copying model [10] (*CP*) also starts with a small random graph. A new vertex connects to at least one bootstrap vertex and copies a subset of its connections. This generates graphs with high clustering. The basic idea has been used in many other models which change the selection of bootstrap vertex and connections to copy [11], [12].

Capkun et al. proposed a model explicitly designed to match the properties of the *PGP WoT* [26] (*PGP*). Similar to *WS*, *PGP* first generates a regular graph and then creates shortcuts by rewiring edges. The main difference to *WS* is that *PGP* requires a degree distribution as input which makes graph generation very artificial. Also, it neither allows to grow a given graph nor describes how the graph evolves over time.

The Forest-Fire model [13] (FF) combines multiple principles: preferential attachment, copying, and a community structure. The model connects new vertices to the network similar to a spreading forest fire. With probabilities for connecting to outgoing and incoming connections of visited vertices, the resulting graphs densify over time while their diameter decreases. Both properties have been shown to reflect growth in many real-world networks.

Some of the discussed models (*ER, PL, WS, PGP*) do not grow over time but generate a static graph with certain properties. Other models (*BA, CP, FF*) describe the evolution of a graph over time by defining the connection of new vertices to an existing graph. New edges are always connected to new nodes. Hence, there is no model yet that actually evolves the existing graph by adding edges to it. We close this gap by developing a model that generates graphs with desired properties, adds vertices over time, and also evolves the existing graph by adding connections between its vertices.

V. Model

In this Section, we present two models for generating WoT graphs. They use the generative principles we identified as a result of our analysis of the *PGP WoT*.

A. WoT$_{gr}$ - *a WoT growth model*

WoT graphs grow over time as new vertices are added and new connections between existing ones are formed. Therefore, we developed a model to reproduce this growth from any existing graph, e.g., an instance of an actual WoT. New vertices and edges are added based on the generative principles we identified during our analysis of *WoT*$_{25k}$ (cf. Section III).

WoT$_{gr}$ adds a vertex and edges to an existing graph G. As parameters, it takes the number of new edges d, the target overall fraction of bidirectional edges d_{bid}, and the probability of a bidirectional bootstrap $d_{bid,2}$. An overview of *WoT*$_{gr}$ is given in Algorithm 1.

First, a new vertex v is created to reflect the join of a new user to the WoT. It is connected to bootstrapping vertices $w, u \in V$ by edges (w, v) and (v, u) to maintain strong connectivity. We assume that this process is guided by preferential attachment, i.e., $w, u = pref(V)$ with $P(pref(V') = v) := \frac{d_{out}(v)}{\sum_{w \in V'} d_{out}(w)}$. The high fraction of $d_{bid,2}(WoT_{25k}) = 0.86$ implies that bootstrapping often results in a bidirectional edge, i.e., $w = u$. Therefore, a bidirectional bootstrapping is performed with probability $d_{bid,2}$ and w, u selected independently otherwise.

Second, new edges are created to reflect the establishment of trust relations between existing users. For each of the $d - 2$ remaining edges, a source v and a destination w must be determined. We assume that every user has the same probability of establishing a new trust relation. Hence,

Algorithm 1: Adding a new vertex and edges with *WoT*$_{gr}$

```
Data: G, d, d_bid, d_bid,2
begin
    v = new vertex; V.add(v);                 // add new vertex
    w = pref(V);                              // 1. bootstrap
    if U[0,1] ≤ d_bid,2 then  u = w;          // bid. bootstrap
    else  u = pref(V);                        // 2. bootstrap
    E.add((w,v), (v,u)); e = 2;               // add bootstrap edges
    while e < d do
        v = rand(V);                          // sel. src
        u = pref(out(v)); w = rand(out(u));   // sel. dst
        E.add((v,w)); e++;                    // add new edge
        if d_bid(G) < d_bid then
            E.add((w,v)); e++                 // bidirectional edge
```

we select v uniformly at random, i.e., $v = rand(V)$ with $P(rand(V') = v) := |V'|^{-1}$. The destination w is then selected from the 2-hop neighbors of v to achieve the high connectivity of neighbors observed for *WoT*$_{25k}$ with $cc = 0.37$. First, we select an outgoing connection u of v preferentially, i.e., $u = pref(out(v))$. Then, the destination w is selected uniformly at random from u's outgoing connections, i.e., $w = rand(out(u))$. With this selection, all outgoing 2-hop connection of v have the same probability to be chosen. The creation of edge (v, w) means that v now trusts w, a user already trusted by one of v's trusted connections u. In case the current fraction of bidirectional edges in G is below the target of d_{bid}, the inverse edge (w, v) is added as well.

This model can be used to grow an existing graph to any size by executing *WoT*$_{gr}$ once for each new vertex. We can also use *WoT*$_{gr}$ to generate a graph completely by, e.g., starting with the largest strongly connected component of a random graph $R(N, E)$ with N vertices and E edges. We denote such a generated graph as *WoT*$_{gr,r}(N, E, d_{bid}, d_{bid,2})$.

B. WoT$_{com}$ - *a community-based WoT model*

For *WoT*$_{25k}$, we observed a community structure with a power-law size distribution. Networks that are grown using preferential attachment, copying, or hierarchical growth do not create such communities [15]. We assume that the absence of community structures in graphs grown using *WoT*$_{gr}$ or generated with *WoT*$_{gr,r}$ leads to shorter paths and lower clustering compared to real-world instances. Therefore, we developed a model for generating WoT graphs from scratch based on the idea of separate interconnected communities. This *WoT*$_{com}$ model consists of four steps: (1) generate a list of community sizes, (2) generate each community separately with *WoT*$_{gr,r}$, (3) connect the largest community to all others, and (4) further interconnect all communities.

As parameters, *WoT*$_{com}$ takes the target network size N, the fraction of vertices in the largest community c_c, the power-law exponent c_{exp} for the community size distributoon, cutoff values c_{min} and c_{max} for the community size distribution, and the number of bidirectional edges between communities c_d.

In addition, *WoT*$_{com}$ takes the same parameters as *WoT*$_{gr}$ for generating each community: d, d_{bid}, and $d_{bid,2}$. An overview of *WoT*$_{com}$ is given in Algorithm 2.

Algorithm 2: Generating a complete graph with *WoT*$_{com}$

```
Data: N, c_c, c_exp, c_min, c_max, c_d, d, d_bid, d_bid,2
begin
    // Compute communities sizes C'
    C' = {c_c · N}
    while Σ_{c∈C'} c < N do
        c ~ C_size(c_exp, c_min, c_max);
        if N − (|V| + c) < c_min then  c = N − |V|;
        C'.add(c);
    // Generate all separate communities
    V = ∅; E = ∅; G = (V, E); C = ∅;
    for c ∈ C' do
        G' = WoT_gr,r(c, c · d, d_bid, d_bid,2);
        V.add(v ∈ V'); E.add(e ∈ E'); C.add(V');
    // Connect central community C_c to others
    for C_i ∈ C, C_i ≠ C_c do
        v = pref(C_c); w = pref(C_i);
        E.add((v, w), (w, v));
    // Interconnect communities
    for i = 0; i < c_d · |C|; i++ do
        C_j = pref(C); C_k = pref(C); v = pref(C_j);
        w = pref(C_k); E.add((v, w), (w, v));
```

First, we determine the sizes C' for all communities. We start with the largest or central community C_c with a size of $N \cdot c_c$. The sizes of all remaining communities are drawn from a power-law distribution C_{size} with exponent (c_{exp} and values limited to $c_{min} \leq x \leq c_{max}$).

Second, we generate a separate community graph for each size $c_i \in C'$. We start with the largest connected component of a random graph with $d-2$ vertices and d edges. Using *WoT*$_{gr}$, we grow this initial graph to the target size of c_i vertices.

Third, we connect each community $C_i \neq C_c$ to the central community with a bidirectional edge. The vertices creating this connection are chosen preferentially as we assume it is more likely that well connected users know and trust users from different communities. Thereby, the complete graph becomes strongly connected.

Fourth and finally, we create additional $c_d \cdot |C|$ bidirectional edges between communities. For each edge, we select two communities C_j and C_k preferentially, i.e., with probabilities proportional to their size. From each community, we select a vertex preferentially assuming that well connected users are more likely to establish new trust relationships. By creating a bidirectional edge between vertices in C_j and C_k, both communities are directly connected.

VI. Evaluation

In this Section, we perform a parameter study for *WoT*$_{com}$ based on *WoT*$_{25k}$ and evaluate to which extent existing graph models are capable of reproducing the properties of a WoT. Then, we evaluate the capabilities of our models against the Forest Fire model to grow a WoT over time.

A. Implementation of models and analysis

We implemented all models in *GTNA*, a framework for the graph-theoretic analysis of network snapshots [22][2]. We generated 20 instances for each model, parameter set, and size and averaged their graph-theoretic properties.

B. Parameter study for WoT$_{com}$

The parameters used for *WoT*$_{gr}$ follow directly from the properties of *WoT*$_{25k}$ (cf. Section III): We use $d = 9$ to produce an average degree of ≈ 18 and bidirectionality parameters of $d_{bid} = 0.5$ and $d_{bid,2} = 0.85$.

To determine the community-related parameters, we performed a parameter study to observe the impact of values on relevant properties like d_{max}, spl_{avg}, cc, and t. We selected each parameter from a reasonable range and determined the impact that an increase of the parameter in the respective range has on each property (cf. Table I). For each parameter and property, we determined if an increase does highly increase (↑↑) or decrease (↓↓) the property, slightly increase (↑) or decrease (↓) it, or if it has no noticeable effect (−).

P	Range	d_{max}	spl_{avg}	cc	t	choice
c_c	[0.1, 1]	↑↑	↓↓	↓↓	↓↓	0.45
c_{exp}	[2, 4]	−	−	↑	↑	3.0
c_{min}	[25, 200]	−	↑	↓	↓	100
c_{max}	[250, 2k]	−	−	−	−	1,000
c_d	[0, 5]	−	↓	−	−	3

TABLE I
Impact of parameter values on properties of *WoT*$_{com}$

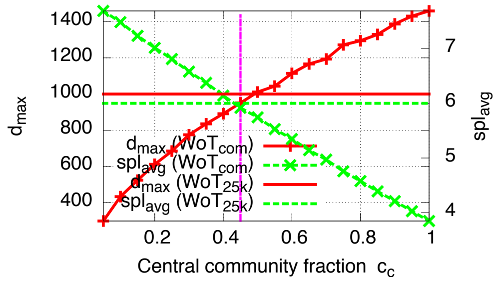
The impact of the parameters that influence community sizes (c_{exp}, c_{min}, c_{max}) and number of interconnecting edges (c_d) is small. Therefore, we selected the values for these parameters arbitrarily from the investigated parameter ranges.

In contrast, the choice of c_c has a high impact on all properties. While increasing c_c, d_{max} changes from 220 to 1,480 and spl_{avg} is decreased from 7.8 to 3.8 (cf. Figure 2b). Clustering coefficient and transitivity are also highly influenced and range from 0.43 to 0.31 and from 0.37 to 0.07 respectively (cf. Figure 2a). Based on these results, we chose $c_c = 0.45$ as it generates graphs with properties closest to *WoT*$_{25k}$.

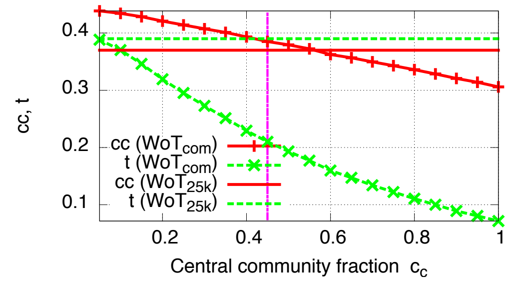
C. Comparison with existing graph models

We evaluate the capabilities of all models introduced in Section IV to generate graphs with properties close to *WoT*$_{25k}$. We selected the parameters for *WoT*$_{gr}$ and *WoT*$_{com}$ according to our parameter study. For all other models, we chose parameters so that the resulting graphs have an average degree close to the observed one of *WoT*$_{25k}$: *ER*: $d_{avg} = 18$, *PL*: $out_{exp} = 1.8, out_{min} = 1, out_{max} = 1103, in_{exp} = 1.78, in_{min} = 1, in_{max} = 1642$, *BA*: $edgesPerNode = 5$,

[2]open-source implementations of all models: http://bit.ly/1KvTTdi



(a) Maximum degree and average shortest path length



(b) Clustering coefficient and transitivity

Fig. 2. Parameter study of central community fraction $c_c \in [0, 1]$ regarding impact on certain properties for *WoT*$_{com}$ ($c_c = 0.45$ highlighted)

WS: $succ = 5, \beta = 0.5$, *CP*: $k = 10$, *PGP*: $\phi = 0.2$, and *FF*: $p = 0.409, p_b = 0.32$. In addition, we implemented a modified version of Forest Fire (*FF*$_d$) where new vertices establish a bidirectional connection during bootstrap and all other connections are formed bidirectionally with probability $d_{bid} = 0.5$.

Only *PL*, *PGP*, *FF*, and *FF*$_d$ produce graphs with degree distributions similar to *WoT*$_{25k}$ (cf. Figure 3a). Even though all models produce a similar average degree, d_{min} and d_{max} differ significantly (cf. Table II). For *BA* and *WS*, d_{min} is too high (10) while d_{max} of *ER* and *WS* is to low (< 40). For *BA* and *WS*, all edges are bidirectional while the bidirectionality of *ER* and *WS* is close to 0. *PGP* stands out with a fraction of at least 0.08 bidirectional edges. The only model that is capable to reproduce the high bidirectionality of *WoT*$_{25k}$ is *FF*$_d$. This is because we specifically modified the model to achieve this. While $d_{bid,2}$ is 0.85 for *WoT*$_{25k}$, it is 0 for all models except for *FF*$_d$ which has a value of 1.

CP and *FF*$_d$ are the only models that generate graphs with spl_{avg} close to *WoT*$_{25k}$. They exhibit values of 5.46 and 5.37 respectively (cf. Table II). *FF* stands out with the lowest value of only 3.88 but only less than 10% of all vertex pairs are connected (cf. Figure 3b). While spl_{avg} of graphs generated by *BA*, *PGP*, *ER*, and *WS* are 18% to 34% too low, the average paths length in *PL* graphs is 75% higher that the target value of *WoT*$_{25k}$ (cf. Table II).

Clustering coefficient $cc = 0.37$ and transitivity $t = 0.39$ of *WoT*$_{25k}$ are not matched by any model (cf. Table II). *ER*, *BA*, and *PL* produce graphs with values close to 0, hence, neighborhoods are basically never interconnected. *WS* and *PGP* generate graphs with higher values. Overall, the closest match can be observed for *CP* with cc =0.38 and t =0.25. *FF* has very low values of only $cc = 0.17$ and $t = 0.09$. With cc =0.35, *FF*$_d$ is very close as well but exhibits low transitivity of only 0.08.

The rich-club connectivity of *WoT*$_{25k}$ is not matched by any model. Only *CP*, *PGP*, *FF*$_d$, and *BA* exhibit close values with $rcc(10) \geq 0.4$ and $rcc(100) \leq 0.14$.

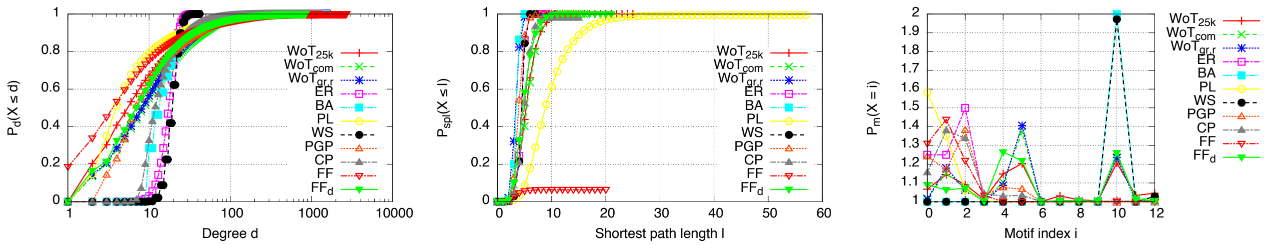
All models except *FF*$_d$ exhibit a different motif signature than *WoT*$_{25k}$ (cf. Figure 3c). In *WoT*$_{25k}$, no motif makes up for more than 21%. In clear contrast, some models produce graphs in which single motifs are present in up to 99% of all vertex triplets.

Our community-based model *WoT*$_{com}$ is able to accurately reproduce most of the properties of *WoT*$_{25k}$. Unsurprisingly, the bidirectionality is exactly as specified by the parameters of the model and hence very close to the target values. Average path length, median path length, and effective diameter are closely matched. The clustering coefficient is very close to the target value of *WoT*$_{25k}$. With a value of 0.21, the transitivity is not as close but still closer than all models except *CP*. These results are not surprising since the parameters of our model have been tuned to reproduce the properties of *WoT*$_{25k}$. Even though not targeted during the parameter study, the rich-club connectivity is not perfectly matched but with values of $rcc(10) = 0.98$ and $rcc(100) = 0.25$ closer than most models. Most surprisingly, the motif signature of *WoT*$_{com}$ is very close to *WoT*$_{25k}$. Even though we did not consider the motif signature during the design and parameter study of our model, *WoT*$_{com}$ is actually able to match the motif signature of *WoT*$_{25k}$ closely. Since the motif signature of a graph is believed to be characteristic to the evolving mechanisms behind the modeled system [17], this similarity to the original graph indicates that the generative principles we used in our model could be close to reality.

D. Evaluation of growing larger WoT graphs

So far, we have seen that *WoT*$_{com}$, *WoT*$_{gr,r}$, and *FF*$_d$ are the only models capable of producing graphs with many properties close to *WoT*$_{25k}$. Now, we evaluate to which extent they are capable of growing or generating larger WoT graphs. We also include *FF* to showcase the benefits introduced by our modified *FF*$_d$ version. As a baseline, we selected 31 snapshots from the *PGP WoT* with sizes between 25k and 55k vertices[3]. We refer to those snapshots as *WoT*. For each graph size, we generated *FF*, *FF*$_d$ and three instances of our model *WoT*$_{com}$, *WoT*$_{gr}$, and *WoT*$_{gr,r}$. *WoT*$_{com}$ is the community-based model where communities are grown separately and interconnected afterwards. *WoT*$_{gr}$ is the growth model applied to the *WoT*$_{25k}$. For *WoT*$_{gr,r}$, the growth model is applied but instead of

[3]http://www.lysator.liu.se/~jc/wotsap/wots2/ and https://wot.siccegge.de/



(a) Degree distribution P_d (cdf) (b) Shortest path length distribution P_{spl} (cdf) (c) Motif signature P_m

Fig. 3. Comparison of relevant graph properties from our models and existing ones to *WoT*$_{25k}$

	d_{avg}	d_{max}	$d_{bid,2}$	spl_{avg}	spl_{med}	spl_{max}	$spl_{90\%}$	cc	t	sc	$rcc(10)$	$rcc(100)$	$rcc(1000)$
WoT	**18.08**	**1368**	**0.85**	**5.99**	**6**	**25**	**8**	**0.37**	**0.39**	**1.00**	**0.72**	**0.40**	**0.05**
WoT_{com}	19.25	955	0.85	5.94	6	14	8	0.38	0.21	1.00	0.96	0.26	0.01
$WoT_{gr,r}$	19.49	1460	0.84	3.84	4	9	5	0.30	0.07	1.00	0.95	0.28	0.02
ER	18.00	37	0.00	4.86	5	8	6	0.00	0.00	0.99	0.00	0.00	0.00
BA	19.99	1169	0.00	3.92	4	6	5	0.00	0.00	1.00	0.45	0.08	0.00
PL	13.61	569	0.00	10.16	9	39	15	0.00	0.00	0.99	0.15	0.14	0.03
WS	19.99	39	0.00	4.90	5	7	6	0.08	0.08	1.00	0.00	0.00	0.00
PGP	19.04	592	0.00	4.45	4	12	6	0.10	0.17	1.00	0.45	0.26	0.04
CP	17.67	931	0.00	5.43	5	13	7	0.38	0.26	0.97	0.52	0.07	0.00
FF	17.19	2175	0.00	3.92	4	17	5	0.17	0.09	0.00	0.22	0.25	0.06
FF_d	18.11	693	1.00	5.36	5	19	7	0.35	0.08	1.00	0.39	0.22	0.04

TABLE II

PROPERTIES OF GRAPHS WITH 25,487 VERTICES GENERATED WITH ALL MODELS (**BASELINE**, CLOSEST VALUES)

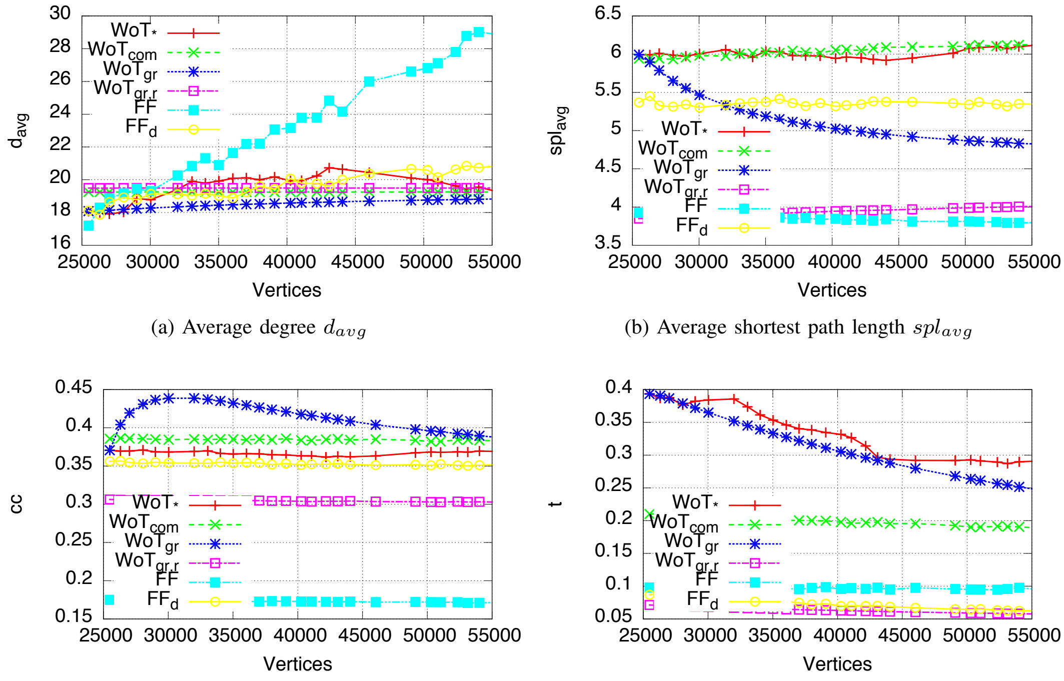
starting with an original WoT graph, the graph is initialized with a small random graph, i.e., the community-based model is executed with a single community (c_c =1.0).

We expect WoT_{com} to perform best regarding shortest paths since WoT_{gr} and $WoT_{gr,r}$ do not incorporate the generation of new communities over time. As new vertices are connected to a single component this should lead to shorter paths in both cases. Also, we expect WoT_{gr} to achieve a rich-club connectivity and motif signature closer to the original WoT since the WoT_{25k} still makes up for a large part of the overall graph. Furthermore, we expect *FF* and FF_d to increase the average degree over time and decrease shortest paths since these are the desired properties of the original Forest Fire model.

In our model, we assume a constant average degree over time. While WoT_{com} produces graphs with an average degree around 19.25 for all sizes, it increases from 18 to 20.7 and then falls again to 19.5 for *WoT*. The average degree of WoT_{gr} starts with WoT_{25k} at 18 and is then slightly increased towards WoT_{com}'s average degree as more vertices are added. While the additional connections between communities seem to increase the average degree noticeably from the target 18, it seems to actually bring the WoT_{com} closer to the original average degree as it develops over time. As expected, *FF* and FF_d lead to an increasing average degree. While this increase is not as steep for the modified version FF_d, the densification is a property that does not appear in our observations of a WoT. In case even larger graphs would be generated using *FF* of FF_d, this densification would progress even further and not resemble real-world properties any more.

As we expected, the average path length of WoT_{com} is very close to *WoT* whose values vary around 6 for all graph size (cf. Figure 4b). When growing the WoT_{25k} with the growth model WoT_{gr}, spl_{avg} decreases as the graph grows. We assume that this is due to the absence of new communities that are attached to the central community which would create longer paths. Instead, WoT_{gr} and $WoT_{gr,r}$ further interconnect a single component and therefore decrease shortest paths despite the addition of further vertices. For $WoT_{gr,r}$, the average path length starts very low at 3.85 and slightly increases as the graph is grown further. Here, we also assume that these short paths are caused by the absence of small communities attached to a densely connected central component. *FF* produces graph with low spl_{avg} that even decrease as the graphs grow, a property that is desired by the model but does not match the development of real-world WoT graphs. For FF_d, spl_{avg} only shrinks slightly but is still far lower than *WoT* with a value around 5.4.

As the original graph grows, its transitivity decreases noticeably from 0.39 to 0.28, a trend well matches by WoT_{gr} (cf. Figure 4d). In contrast, the transitivity of WoT_{com} is rather low with values around 0.2 but also decreases as the graphs get larger. The same trend can be observed for $WoT_{gr,r}$ even though it already starts with a very low transitivity of 0.07. The clustering coefficient of *WoT* is always close to its initial value of 0.37 (cf. Figure 4c). Similarly, the clustering coefficient of WoT_{com} stays close to a slightly higher value of 0.38. In clear contrast, WoT_{gr}'s clustering coefficient increases



(a) Average degree d_{avg}

(b) Average shortest path length spl_{avg}

(c) Clustering coefficient cc

(d) Transitivity t

Fig. 4. Comparison of our growth models with Forest Fire and the original WoT_* for increasing vertex count

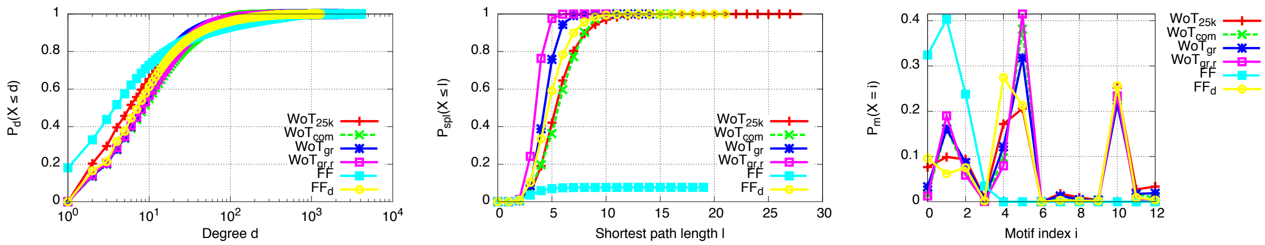
to a maximum of 0.44 for 31k vertices before decreasing again as the graph size grows. Again, the value produced by $WoT_{gr,r}$ is lower but does not change as the graph grows. The clustering coefficient of *FF* and FF_d are also close to constant. While our modified version produces values around 0.36 close to the baseline, the original *FF* model exhibits very low values around 0.17. With transitivities of $t < 0.1$, both model produce very low values that indicate an overall very different interconnection of neighborhood than *WoT*. This shows that the rather general generative principles of a forest fire are not applicable for the generation of trust graphs as they do not generate the observed high interconnection of neighborhoods.

Despite the different evolution of clustering coefficient, transitivity, and average shortest paths for the different model instances, WoT_{com} and WoT_{gr} produce graphs whose properties are very similar to *WoT* even when investigating the largest instances with 55k vertices. Even though WoT_{gr} is grown to more than twice its initial size, its degree distribution and motif signature are very close to those of *WoT* (cf. Figures 5a and 5c). The properties of the community-based model are also very close to the original even though slightly worse that WoT_{gr}. As expected, the shortest paths are shorter for WoT_{gr} and $WoT_{gr,r}$, most probably because of the absence of explicitly created separate communities that naturally increase the overall path lengths. It is remarkable to see how closely WoT_{com} matches the original graph's average path length even though its parameters were selected based on a graph of half the size. Together with the closely matches motifs signature, these results show that the generative principles we use in our models could very well resemble realistic user behavior. The original Forest Fire model does not match many of the desired and investigated properties. Our modified version FF_d matches most properties better but still not as good as our models. We believe that the generative principles of a forest fire do not resemble the behavior of users that lead to the growth of WoT and therefore do not represent a viable alternative to our model. Also, our model is the only one that creates further connections between existing vertices as it happens in reality.

In summary, we have shown that WoT_{com} and WoT_{gr} are capable of generating realistic WoT graphs with properties close to the original *WoT*. While a modified version of the Forest Fire model produces close results as well, the original model does not reproduce the growth of a WoT. The growth of an initial WoT or random graph results in rather short paths between all vertices while maintaining rich-club connectivity and degree distribution. The community-based model is capable of generating graphs with accurately replicated path lengths and a similar clustering coefficient. In all three instances of the model, the motif signature is very close to the original network which implies that the strategies for connecting new vertices to the network as well as interconnecting existing ones to grow the network reproduce realistic user behavior.

VII. Summary, Conclusion, and Outlook

In this work, we developed and evaluated graph models to generate realistic WoT graphs. Based on an analysis of the graph-theoretic properties of a small PGP WoT of only 25k



(a) Degree distribution P_d (cdf) (b) Shortest path length distribution P_{spl} (cdf) (c) Motif signature P_m

Fig. 5. Comparison of relevant graph properties of our growth models, Forest Fire, and WoT_{55k}

vertices, we determined key properties of a WoT and identified generative principles that explain how these properties can evolve in a graph that is grown over time. Using these principles as basic guidelines for the development of a graph model, we created two models. The growth model WoT_{gr} allows us to grow any input graph with the characteristic principles identified in the analysis by adding vertices and edges one after the other. The community-based model WoT_{com} enables us to create WoT graphs without an initial graph by creating a set of separate communities first and interconnecting them afterwards.

We investigated various existing graph models to determine the extent to which they are able to produce graphs with the desired properties. Even the most promising model, a modification of the Forest Fire model, did not match the properties of a real-world WoT as good as our models. The generative principles of our model seem accurate as our model is able to grow a WoT to more than twice the size of the smallest one based on which we determined its parameters. Therefore, we believe that WoT_{gr} is the first model to actually reflect the user behavior that contributes to the growth of a WoT.

While growth and community-based models are both able to reproduce the properties of a WoT graph at arbitrary sizes, the properties of a graph grown from an original WoT still deviate. We assume that this is because of the absence of new communities created during this process Therefore, we will perform a more detailed evaluation of existing WoT graphs to better understand the creation of communities over time. Based on these insights, we will investigate how to incorporate the generation of new communities during the addition of new vertices in the specific growth process of trust graphs.

Acknowledgment

This work has been partially supported by the GreenICN project (EC contract no. 608518, NICT contract no. 167).

References

[1] A. Abdul-Rahman. The pgp trust model. In *EDI-Forum: the Journal of Electronic Commerce*, volume 10, 1997.

[2] Amaral et al. Classes of small-world networks. *National Academy of Sciences*, 97(21), 2000.

[3] Barabasi and Albert. Emergence of scaling in random networks. *Science*, Oct. 1999.

[4] Blondel et al. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, 2008(10), 2008.

[5] Caronni. Walking the web of trust. In *WET ICE*, 2000.

[6] Erds and Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 1959.

[7] Golbeck et al. Trust networks on the semantic web. In *LNCS 2782*, 2003.

[8] Golbeck and Hendler. Accuracy of metrics for inferring trust and reputation in semantic web-based social networks. In *Engineering knowledge in the age of the semantic web*. 2004.

[9] Hofer et al. Greedy embedding, routing and content addressing for darknets. In *Networked Systems*, 2013.

[10] Kleinberg et al. The web as a graph: Measurements, models, and methods. In *Computing and combinatorics*. 1999.

[11] Krapivsky and Redner. Organization of growing random networks. *Physical Review E*, 63(6), 2001.

[12] Krapivsky and Redner. Network growth by copying. *Physical Review E*, 71(3), 2005.

[13] Leskovec et al. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 2007.

[14] Leskovec et al. Microscopic evolution of social networks. In *SIGKDD*, 2008.

[15] Leskovec et al. Statistical properties of community structure in large social and information networks. In *WWW*, 2008.

[16] Marti et al. Sprout: P2p routing with social networks. In *Current Trends in Database Technology*, 2005.

[17] Milo et al. Network motifs: simple building blocks of complex networks. *Science*, 2002.

[18] Newman et al. Random graph models of social networks. *National Academy of Sciences*, 99(suppl 1), 2002.

[19] Palmer and Steffan. Generating network topologies that obey power laws. In *GLOBECOM*, 2000.

[20] Papadopoulos et al. Popularity versus similarity in growing networks. *Nature*, 489(7417), 2012.

[21] Pirzada and McDonald. Establishing trust in pure ad-hoc networks. In *Australasian conference on Computer science*, 2004.

[22] Schiller et al. GTNA: a Framework for the graph-theoretic Network Analysis. In *Spring Simulation Multiconference*, 2010.

[23] Seedorf et al. Decentralised binding of self-certifying names to real-world identities for assessment of third-party messages in fragmented mobile networks. In *INFOCOM WKSHPS*, 2014.

[24] Sun et al. Information theoretic framework of trust modeling and evaluation for ad hoc networks. *Selected Areas in Communications*, 2006.

[25] Trifunovic et al. Social trust in opportunistic networks. In *INFOCOM*, 2010.

[26] Čapkun et al. Small worlds in security systems: an analysis of the PGP certificate graph. In *Workshop on New security paradigms*, 2002.

[27] Watts and Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684), 1998.

[28] Zhou and Mondragón. Accurately modeling the internet topology. *Physical Review E*, 70(6), 2004.

[29] Zimmermann. *The official PGP user's guide*. 1995.