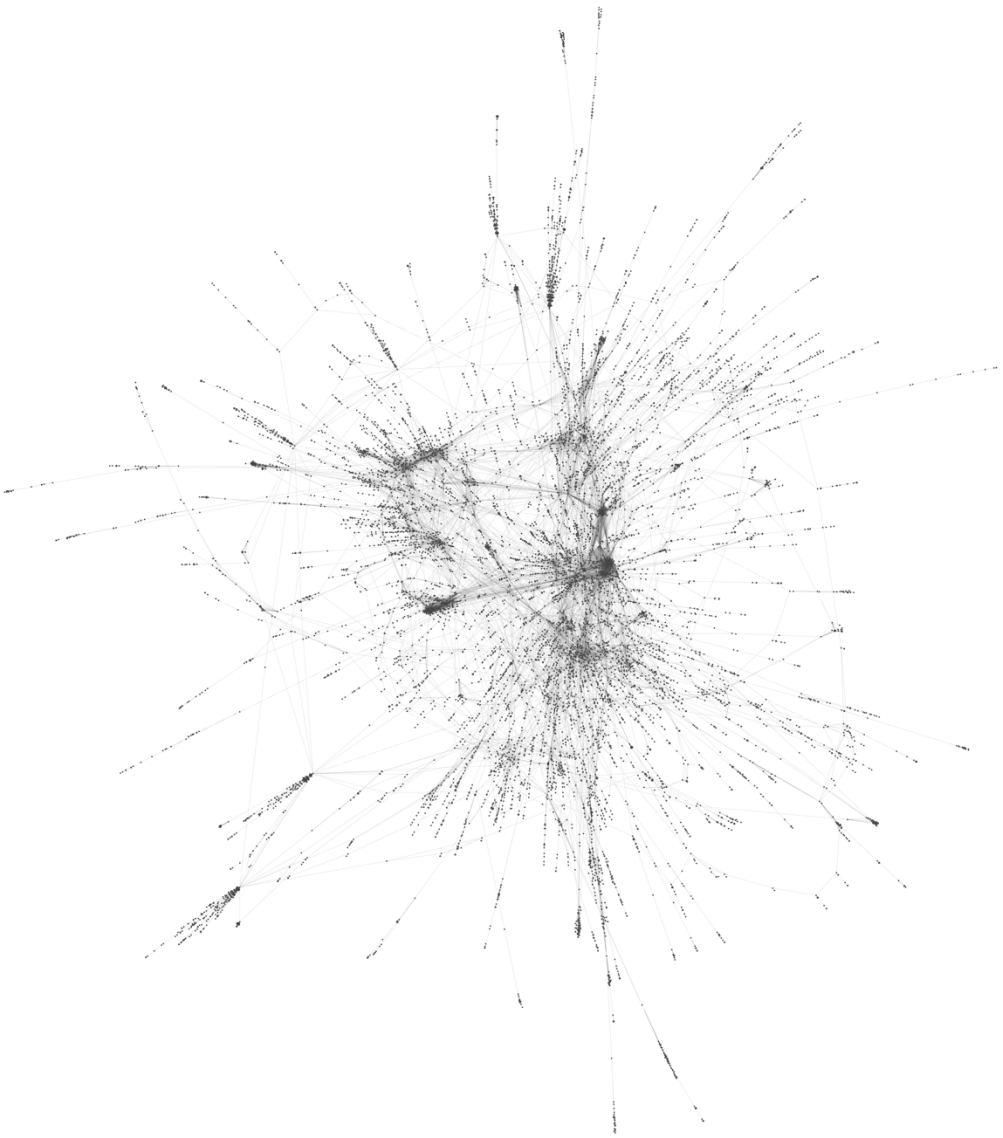
FFHS
Fernfachhochschule Schweiz
Zürich | Basel | Bern | Brig



PGP - THE WEB OF TRUST

AUTOR
MARIUS GIGER
PIBS 2015
SWISSCOM AG
marius.giger@swisscom.com

BETREUT DURCH
URSULA DERIU
DOZENTIN NETWORK ANALYSIS
FFHS
ursula.deriu@ffhs.ch

ZÜRICH, 3. JANUAR 2019

ABSTRACT

Soziale Netzwerke haben in den letzten Jahren durch das Aufkommen der sozialen Medien und dem Internet vermehrt an Aufmerksamkeit gewonnen. Die Wissenschaft steht jedoch immer noch am Anfang und so gibt es diverse Bereiche deren volles Potenzial noch nicht ausgelotet wurde. In dieser Arbeit geben wir eine Einführung in das soziale Netzwerk von Pretty Good Privacy (PGP) - einer dezentralisierten Publik-Key-Infrastruktur und ergänzen dadurch das Bild von Vertrauensnetzwerken mit einem Praxisbeispiel. Wir zeigen die Kerneigenschaften des Web of Trusts von PGP auf und legen dar, dass die Gradverteilung des PGP Netzwerks lognormal-verteilt ist, dass das Netzwerk Clustering aufweist, eine interessante Community-Struktur vorliegt sowie, dass es sich bei dem Netzwerk um ein assortatives Netzwerk handelt und bestätigen und verfeinern dadurch die Erkenntnisse aus vorangegangenen Arbeiten.

INHALT

I. EINFÜHRUNG

Pretty Good Privacy (PGP) [1] ist ein Verschlüsslungsprogramm für den sicheren Informationsaustausch, das auf Public-Key-Kryptographie basiert. Das zu Grunde liegende Protokoll sowie eine entsprechende Software wurde im Jahr 1991 von Phil Zimmermann vorgeschlagen und entwickelt [2]. Im Jahr 1996 wurde ein dazugehörender RFC Standard für das Protokoll eingeführt [3]. Der Standard wurde seither zweimal überarbeitet und fand seine momentane Version im «OpenPGP Message Format» [4]. PGP hat die Prämisse ein «Web of Trust» aufzubauen, d.h. ein dezentralisiertes und fehlertolerantes Vertrauensnetzwerk für die Zugehörigkeit eines Public Keys zu erzeugen. Das Protokoll standardisiert dabei das Format für den Nachrichtenaustausch, die Signaturen sowie die Zertifikate. OpenPGP ist in den vergangenen zwanzig Jahren zu einem weitverbreiteten Standard für die Verschlüsselung von Emails geworden. In dieser Arbeit zeigen wir die Eigenschaften des «Web of Trust» von PGP auf und diskutieren deren Sicherheit.

Die Verteilung von Zertifikaten ist Teil einer Public Key Infrastruktur[1] (PKI) und wurde vor allem aufgrund der SSL-Zertifikate[2] ein existentieller Bestandteil des Internets. Eine PKI vereinfacht die Verwendung von asymmetrischer Kryptographie im Alltag, indem der Besitzer des öffentlichen Schlüssels durch eine Signatur einer öffentlichen Stelle - genannt Certificate Authority - verifiziert werden kann. Eine herkömmliche Certificate Authority (CA) signiert Zertifikate mit ihrem privaten Schlüssel, und garantiert dadurch den Link zwischen einem Public Key und der Identität eines Besitzers (z.B. einer Website). Die Trust-Struktur ist hierarchisch gegliedert und nimmt die Form eines Baums an. Es gibt dementsprechend nur eine Wurzel - die sogenannte Root Certificate Authority. In der Realität sind mehrere Root CAs vorhanden, die jeweils einen separaten Baum bilden. Das Vertrauen zieht sich entlang den Pfaden im Baum. Je näher ein Knoten an der Wurzel ist, desto stärker ist das Vertrauen eines Benutzers in dieses Zertifikat. Im Gegensatz zu einer klassischen hierarchischen Certificate Authority, die Zertifikate an Subentitäten ausstellt, ist in PGP jeder Teilnehmer eine Certificate Authority. Das bedeutet, dass jeder Benutzer die Schlüssel von beliebigen anderen Nutzern signieren kann. Es ergibt sich dadurch eine netzwerkartige Trust-Struktur, die die Form eines stark verknüpften generischen Graphen annimmt, mit zahlreichen Zyklen und mehreren Pfaden zwischen zwei Knoten. Diese Struktur wird *Web of Trust* genannt. Im Verlauf der Zeit akkumuliert ein Benutzer die Schlüssel von anderen Benutzern, denen er vertraut. Er signiert diese Schlüssel, damit andere die ihm vertrauen dies sehen. Somit bauen sich die Benutzer graduell eine Sammlung von Signaturen auf. Dies führt zur Entstehung von einem dezentralisierten und fehlertoleranten Vertrauensnetz.

Das PGP Netzwerk kann als gerichteten Graphen angeschaut werden, wobei die Knoten die PGP Schlüssel sind und die Kanten die Signaturen (ich vertraue dir). Gibt es einen Pfad zwischen Schlüssel A und B, so vertraut A der Identität von B. Die Distanz, also die Länge des kürzesten Pfads, zwischen A und B kann dabei als Indikator für die Stärke des Vertrauens verwendet werden. Das sogenannte *strong set* - die grösste zusammenhängende Komponente - ist das grösste Set von Schlüsseln, in dem für jedes Schlüsselpaar ein Pfad zueinander gefunden werden kann. Diese Arbeit beleuchtet nur das strong set.

In Abschnitt II werden soziale Netzwerke kurz eingeführt und die grundsätzlichen Merkmale hervorgehoben. Abschnitt III stellt verwandte Forschungsarbeiten vor. Nachfolgend werden in Abschnitt IV die verwendeten Methoden erklärt und die dazu erhobenen Resultate präsentiert. Schliesslich werden in Abschnitt V die Findings reflektiert und es wird ein Ausblick gegeben.

[1] https://www.bsi.bund.de/DE/Themen/DigitaleGesellschaft/ElektronischeIdentitaeten/sicherPKI/sicherheitsmechanismenPKI.html [Letzter Zugriff: 20.12.2018]

[2] SSL- Zertifikate werden verwendet für HTTPS

II. Soziale Netzwerke

Soziale Netzwerke widerspiegeln die Komplexität von menschlichen Interaktionen. Die Darstellung eines sozialen Netzwerks als Graph hilft soziale Strukturen auf einer generelleren Ebene zu untersuchen und dadurch mit Hilfe von mathematischen Mitteln mit ähnlichen Netzwerken zu vergleichen sowie solche Netze zu simulieren. Die Knoten stellen die Nutzer des Netzwerks dar und die Kanten bilden die Interaktionen zwischen den Nutzern ab. Auch wenn diese Abstraktion hilft die Topologie von echten Netzwerken zu verstehen, kann dadurch noch keine Aussage über deren Entstehung resp. Veränderung über die Zeit gemacht werden. Die Regeln der Entstehung eines sozialen Netzwerks zu untersuchen ist ein komplexes Unterfangen. So sind viele Faktoren für die Strukturen verantwortlich. Z.B. machen Individuen, die ähnliche Interessen teilen oder sich an denselben Orten aufhalten, schneller Bekanntschaften untereinander. Es gibt diverse psychologische Modelle, die versuchen diesem Umstand gerecht zu werden (z.B. [5]) und versuchen die Gründe für die Entstehung von Interaktionen abzubilden. Diese Modelle sind jedoch meist relativ schwer auf die Netzwerkanalyse anzuwenden, weshalb die Frage gestellt wurde, ob ein mathematisches Modell ein ähnliches Netzwerk erstellen kann, ohne die Umstände der Netzwerknutzer einzubeziehen. Ein solches Modell wird von Boguñá et al. in [6] vorgestellt. Dabei wird mit Hilfe der sozialen Distanz zwischen verschiedenen Netzwerkteilnehmer deren Abstand berechnet. Hierfür wird der Grad der Nähe eines Individuums oder einer Gruppe zu einem anderen Individuum oder Gruppe in einem sozialen Netzwerk quantifiziert. Individuen machen dabei Bekanntschaften mit einer Wahrscheinlichkeit die relativ zur sozialen Distanz abnimmt.

Die Beschaffenheit von sozialen Netzwerken hat sich als fundamental anders herausgestellt als viele Netzwerke, die in der Natur beobachtet wurden [7]. Boguñá et al. nennen in [6] folgende Eigenschaften: Transitivität der Beziehung zwischen Knoten (Clustering), Korrelation zwischen der Anzahl der Bekanntschaften eines Knotens (Grad eines Knotens) und die Präsenz einer stark gegliederten Community-Struktur. Die statistischen Eigenschaften von sozialen Netzwerken können gemäss Boguñá et al. mit folgenden Metriken zusammengefasst werden:

- **Grosser Clustering-Koeffizient:** Die Anzahl Verbindungen zwischen transitiven Peers ist merklich grösser, als dies durch ein rein zufälliges Modell erklärbar wäre. Dies ist auf den Umstand zurückzuführen, dass die Freunde von Freunden oftmals auch Freunde eines Netzwerkteilnehmers sind, was mit dem Clustering-Koeffizienten [8] quantifiziert werden kann.
- **Positive Gradkorrelation:** Soziale Netzwerke weisen eine positive Gradkorrelation auf [9], was bedeutet, dass sich Knoten mit hohem Grad vorzugsweise mit anderen Knoten mit hohem Grad verknüpfen. Dieser Umstand wird *assortatives Mixing* genannt. Nicht soziale Netzwerke weisen oftmals *disassortatives Mixing* auf, was so viel heisst wie, dass sich Knoten mit hohem Grad vorzugsweise mit Knoten mit tiefen Grad verknüpfen und umgekehrt.
- **Community Struktur:** In sozialen Netzwerken werden oft komplexe Community-Strukturen beobachtet [10], da Individuen typischerweise zu Gruppen mit einer hohen Anzahl an Verknüpfungen zwischen einander gehören, welche jedoch nur schwach untereinander verknüpft sind. Diese Struktur ist oftmals hierarchisch, was bedeutet, dass Gruppen in Gruppen existieren und sich die Gruppen oftmals auch überlappen.

III. VERWANDTE FORSCHUNGSARBEITEN

Boguñá et al. führen in [6] das Modell der sozialen Distanz ein. Diese quantifiziert den Grad der Nähe resp. der Akzeptanz, welche ein Individuum oder eine Gruppe gegenüber einem anderen Individuum oder Gruppe in einem sozialen Netzwerk verspürt. In diesem Modell bauen Individuen Verbindungen basierend auf der sozialen Distanz auf, die Wahrscheinlichkeit nimmt dabei mit der relativen sozialen Distanz ab. Insbesondere beleuchten Boguñá et al. das Web of Trust von PGP basierend auf dem in dieser Arbeit verwendeten Netzwerk. Sie zeigen, dass das Netzwerk die Schlüsseleigenschaften eines sozialen Netzwerks aufweist und evaluieren ihr Modell mit dem PGP Netz. Sie gehen jedoch nicht darauf ein, ob dieser Vergleich repräsentativ für andere soziale Netzwerke ist.

Cederlöf präsentiert auf seiner Website [11] eine Analyse der Vertrauensbeziehungen im *strong set* von PGP. Er zeigt, dass die visualisierte Vertrauensmatrix eine blattartige Struktur aufweist. Die Daten wurden durch das Level von Vertrauen eines Nutzers in eine Signatur erhoben. Darauf aufbauend analysiert Penning auf seiner Website [12] verschiedene Distanzmetriken im PGP Netzwerk. Er zeigt insbesondere wie sich das Netzwerk durch das Entfernen von Schlüssel verhält (z.B. durch Angriffe). Die zusammenhängende Komponente bricht dabei ab einem bestimmten Zeitpunkt auf. Dieser Zeitpunkt hängt von der Art der attackierten Schlüssel ab. Beim Entfernen von k *core keys* (Schlüssel, die eine kleine *mean shortest distance* (MSD)[3] haben) müssen 1/3 aller Schlüssel entfernt werden, bevor der Graph in Einzelteile aufbricht (vgl. Abb. 1). Dies legt nahe, dass die Top-1000 Schlüssel von PGP nicht besonders speziell sind. Werden k *fringe keys* entfernt (Schlüssel mit hoher MSD) bleibt der Graph bestehen (vgl. Abb. 2). Ebenso beim Entfernen von k zufälligen Schlüssel, jedoch bricht der Graph schlussendlich dann doch auf (vgl. Abb. 3). Diese Analysen zeigen, dass das PGP Netzwerk resistent gegen Angriffe ist und dass eine grosse Anzahl von Schlüssel entfernt werden müssen, um grossflächigen Schaden im Netzwerk anzurichten.

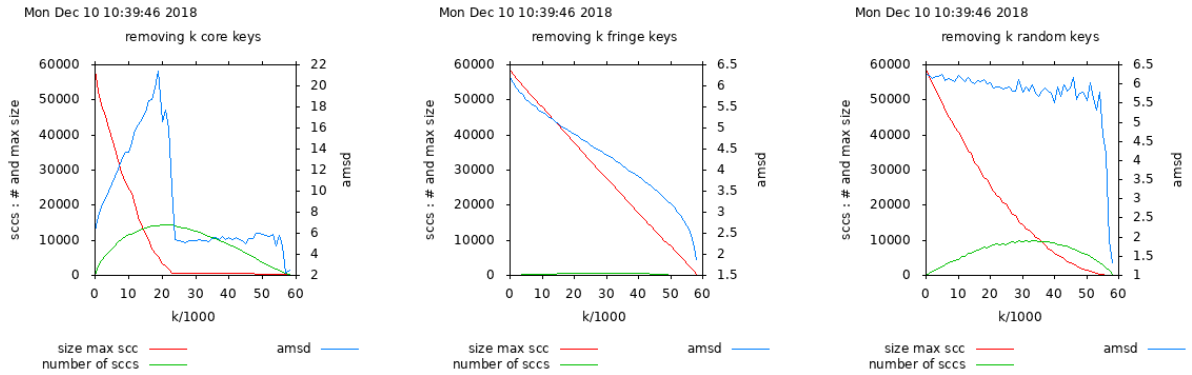


Abbildung 1 - Grösse des strongly connected components (SCC) beim Entfernen von k Schlüsseln. Rote Linie: Grösse des verbleibenden SCC, Grüne Linie: Anzahl von SCCs, Blaue Linie: Schätzung der durchschnittlichen Mean Shortest Distance im SCC

Quelle: https://pgp.cs.uu.nl/plot/ [Letzter Zugriff: 27.12.2018]

Die verwandten Forschungsarbeiten zeigen, dass das PGP Netzwerk bereits mehrmals Forschungsgegenstand einer Arbeit war. Dies ist auch nicht weiter verwunderlich, da PGP durch seine dezentralisierte Natur Eigenschaften aufweist, die so vorher noch nicht beobachtet werden konnten - zum Beispiel die Entstehung eines globalen Web of Trusts. Des Weiteren können durch die öffentliche Verfügbarkeit relativ einfach Daten erhoben werden.

[3] die *mean shortest distance (MSD)* wird berechnet, indem der Durchschnitt aller kürzesten Pfade von diesem Schlüssel zu allen anderen Schlüssel gebildet wird. Sie kann dadurch als Indikation verwendet werden, wie vertrauenswürdig ein PGP Schlüssel ist, wobei ein Schlüssel mit kleinerer Zahl vertrauenswürdiger ist.

IV. Analyse des PGP Netzwerks

Das vorliegende Netzwerk [13] ist eine Liste der Knoten des «strongly connected components» der Netzwerkbenutzer von PGP aus dem Jahr 2004. Dabei stellen die Knoten die Benutzer und die Kanten die Interaktionen zwischen den Benutzern dar. Eine Interaktion bedeutet dabei ein gegenseitiges Signieren der Schlüssel («ich vertraue dir, du vertraust mir») alle anderen Kanten (nur einseitiges Signieren) wurden weggelassen. Dadurch bildet der Graph das Web of Trust von PGP ab.

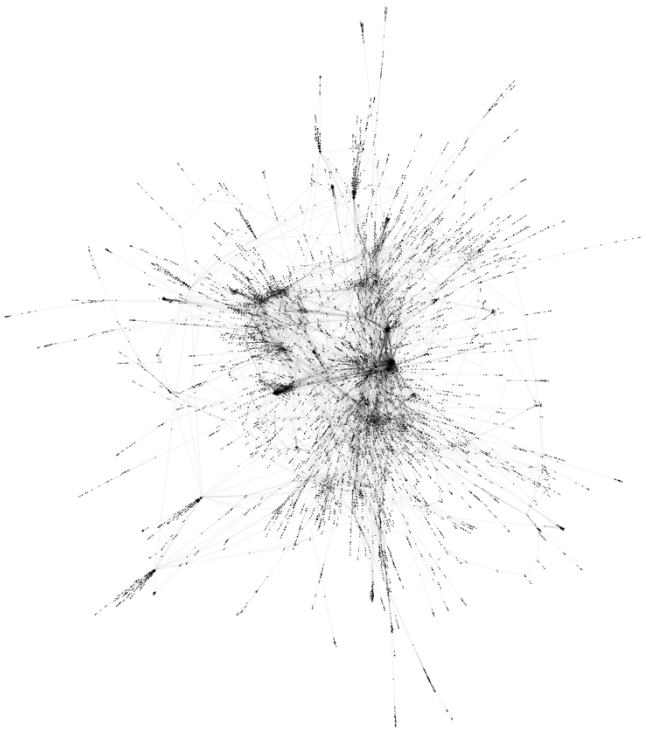


Abbildung 2 - Darstellung Netzwerk erstellt mit Gephi

Das Netzwerk weist folgende Eigenschaften auf[4].

Knoten	10’680
Kanten	24’316
durchschnittlicher Grad <k>	4.55
maximaler Grad	205
minimaler Grad	1
Grösse der grössten Komponente	10'680 (Netzwerk ist verbunden)
durchschnittlicher Clustering-Koeffizient <c>	0.266
durchschnittliche Pfadlänge	7.486
Diameter	24
Dichte	0.00043

Tabelle 1 - Netzwerkeigenschaften

[4] Diese Daten wurden mit NetworkX erhoben. Der Source-Code kann dem folgenden Git-Repository entnommen werden: https://github.com/mariusgiger/network-analysis-pgp

A. GRADVERTEILUNG

Nachfolgend wird die Gradverteilung des PGP Netzwerks analysiert. Diese wird verwendet, um das Netzwerk als skalenfrei oder nicht-skalenfrei zu kategorisieren. Es werden folgende Hypothesen aufgestellt.

> *H0: Die Gradverteilung des PGP Netzwerks kann nicht mit einer Powerlaw-Verteilung angenähert werden.*
> *H1: Die Gradverteilung des PGP Netzwerks folgt einer Powerlaw-Verteilung.*

Die Hypothesen werden getestet, indem mit dem Python-Module Powerlaw[5] die Gradverteilung berechnet und ein Powerlaw-fit erstellt wird. Abbildung 3 zeigt die Gradverteilung auf einer log-log-Skala. In Abbildung 4 wird ein Powerlaw-fit auf die komplementäre kumulative Verteilungsfunktion (ccdf) angewendet, da diese besonders für extreme Werte der Verteilung sinnvolle Resultate liefert [14].

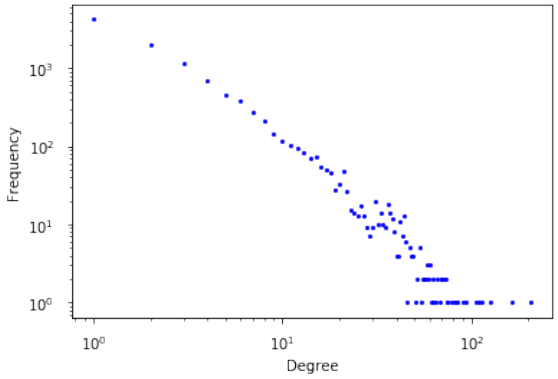


Abbildung 3 - Gradverteilung (log-log) erstellt mit NetworkX

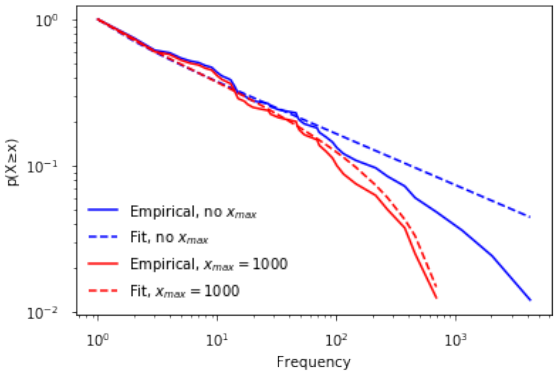


Abbildung 4 - Powerlaw-Fit erstellt mit Powerlaw
Alpha kein xmax: 1.35, Sigma kein xmax: 0.38
Alpha mit xmax=1000: 1.29, Sigma mit xmax=1000: 0.03

Für $x > 100$ kann dabei eine deutliche Abweichung vom Fit festgestellt werden. Dies legt nahe, dass nur eine begrenzte Powerlaw-Verteilung vorliegt. Aus diesem Grund wurde die Methode von Clauset et al. [14] zum Vergleich von Heavy-Tailed-Verteilungen verwendet, um die Gradverteilung von PGP mit ähnlichen Verteilungen zu vergleichen. Dabei wird die Anpassgüte (goodness of fit) basierend auf der Kolmogorov-Smirnov-Statistik von parametrisierten Verteilungen berechnet. Tabelle 1 zeigt die entsprechenden Resultate. Die R-loglikelihood ist positiv, falls die Verteilung der Daten mehr zur ersten Verteilung passt und negativ falls die zweite Verteilung besser geeignet ist. Die Signifikanz für diese Indikation wird dabei durch den p-Wert spezifiziert. Falls $p > .05$ ist keine der Verteilungen signifikant besser geeignet andernfalls kann davon ausgegangen werden, dass die Indikation gegeben durch die R-loglikelihood signifikant ist.

R-loglikelihood	**p-Wert**	**Verteilung 1**	**Verteilung 2**
140.5	0.000004	powerlaw	exponential
-3.77	0.006	powerlaw	truncated_powerlaw
-3.59	0.049	powerlaw	lognormal
0.179	0.86	truncated_power_law	lognormal

Tabelle 2 - Vergleich von Fitting-Modellen

Die Werte in Tabelle 2 legen nahe, dass ein truncated-powerlaw- oder ein lognormal-Fit signifikant besser geeignet sind als ein Powerlaw-Fit, um die Gradverteilung des PGP Netzwerks anzunähern. Diese Erkenntnis wird in Abbildung 5 reflektiert. Es zeigt sich, dass die Gradverteilung von PGP mit einem lognormal-Fit sehr gut angenähert werden kann.

[5] https://pypi.python.org/pypi/powerlaw

Folgende Gleichung stellt eine lognormal-Verteilung dar:

$$f(x) = \frac{1}{x}\exp\left[-\frac{(\ln x - \mu)^2}{\sigma^2}\right]$$

Für den Fit an die CCDF des PGP Netzwerks wurden dabei folgende Parameter berechnet: $\mu = -0.474$ und $\sigma = 3.424$.

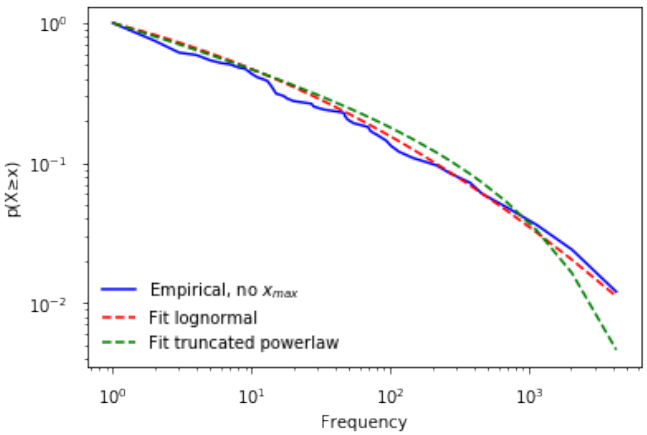


Abbildung 5 - Gradverteilung mit lognormal- und truncated-powerlaw-Fit

Aus den obigen Berechnungen geht hervor, dass das PGP Web of Trust nicht skalenfrei ist, sondern über eine Heavy-Tailed-Verteilung verfügt, die mit der lognormal-Verteilung am besten angenähert wird und gemäss Abbildung 4 nur für kleine x einer Powerlaw-Verteilung unterliegt. Diese Erkenntnis stimmt auch mit den Analysen von Boguña et al. [6] überein, die jedoch keine explizite alternative Verteilung erwähnen.

B. ASSORTATIVITÄT

Dieser Abschnitt untersucht die Assortativität von Knoten, sprich der Korrelation zwischen dem Grad eines Knotens und den Verbindungen mit anderen Knoten.

> *H0: Hubs im PGP Netzwerk verknüpfen sich zufällig mit anderen Knoten.*
> *H1: Hubs im PGP Netzwerk tendieren dazu sich mit anderen Hubs zu verknüpfen.*

Dafür wird basierend auf dem Algorithmus von Newman [15] der Assortativitäts-Koeffizient ausgerechnet. Das PGP Netzwerk verfügt über einen Assortativitäts-Koeffizienten von $r = 0.238$.

Es gilt:

- Assortatives Netzwerk $r > 0$
- Neutrales Netzwerk $r = 0$
- Disassortatives Netzwerk $r < 0$

für $-1 < r < 1$.

Da der Koeffizient r positiv ist, handelt es sich um ein assortatives Netzwerk. Dies bedeutet, dass Hubs dazu tendieren sich mit anderen Hubs zu verknüpfen. Je höher der Grad eines Knotens k ist, desto höher ist daher auch der Grad der umliegenden Knoten. Betrachtet man dies aus Perspektive des PGP Netzwerks ist diese Korrelation naheliegend, da Schlüssel mit vielen Signaturen tendenziell anderen Schlüsseln mit vielen Signaturen mehr vertrauen.

C. COMMUNITIES

In diesem Abschnitt wird das PGP Netz auf Communities analysiert. Insbesondere wird die Struktur einiger Communities genauer untersucht. Diese Analyse wird nahe gelegt durch das Statement von Boguñá et al. [6] wonach soziale Netzwerke eine starkgegliederte Community-Struktur aufweisen. Dabei sollen folgende Hypothesen überprüft werden:

H0: Im vorliegenden Netzwerk existieren keine Communities.
H1: Das PGP Netzwerk weist verschiedene Communities auf.

Der Clustering-Koeffizient C_i gibt den Bruchteil der Nachbarn eines Knotens i an, die verbunden sind [16, pp. 63-64]:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

wobei L_i die Anzahl Kanten zwischen den Nachbarn von Knoten i angibt und k_i dessen Grad ist. Der durchschnittliche Clustering-Koeffizient $\langle c \rangle$ wird folgendermassen berechnet:

$$\langle c \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

Der durchschnittliche Clustering-Koeffizient nimmt eine reale Zahl zwischen 0 und 1 ein, wobei für 0 kein Clustering existiert und für 1 das Netzwerk aus disjunkten Cliquen besteht. Der durchschnittliche Clustering-Koeffizient im Web of Trust ist $\langle c \rangle = 0.266$.

Betrachtet man den durchschnittlichen lokalen Clustering-Koeffizienten (siehe Abb. 6, berechnet als Durchschnitt von allen Knoten mit demselben Grad), ist zu sehen, dass Knoten mit tiefem Grad (~2-80) einen höheren Clustering-Koeffizienten haben als Knoten mit höherem Grad. Dies bedeutet, dass wenn

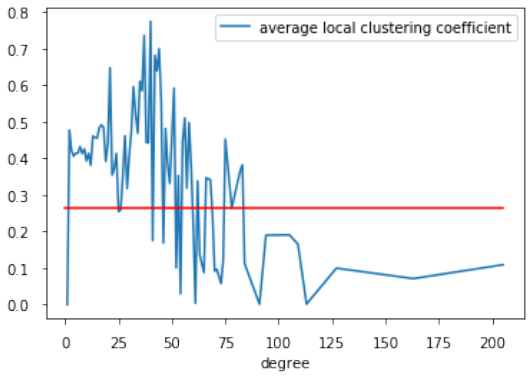


Abbildung 6 - durchschnittlicher lokaler Clustering-Koeffizient (rote Linie: durchschnittlicher globaler Clustering Koeffizient)

ein Schlüssel viele Signaturen hat, haben die signierenden Schlüssel tendenziell weniger Verbindungen untereinander als bei Schlüsseln mit weniger Signaturen. Grundsätzlich deckt sich diese Erkenntnis mit den Annahmen, da ein Schlüssel mit vielen Signaturen höchstwahrscheinlich mehrere Communities verbindet. Der Clustering-Koeffizient von $\langle c \rangle = 0.266$ gibt eine Indikation, dass im PGP Netzwerk Communities existieren, dies wird gestützt durch die Betrachtung des lokalen Clustering-Koeffizienten, der besagt, dass für Knoten mit kleinen Graden tendenziell mehr Clustering besteht.

Ein Mass zur Erhebung von Communities in einem Netzwerk ist die Modularität [16, pp. 339-340]. Zur Bestimmung der Modularität wird der Unterschied zwischen den im Netzwerk vorhandenen Verknüpfungen A_{ij} und der erwarteten Anzahl von Verknüpfungen zwischen den Knoten i und j, falls das Netzwerk zufällig entstanden wäre, gemessen:

$$M_C = \frac{1}{2L} \sum_{(i,j) \in Cc} A_{ij} - p_{ij}$$

Falls M_C positiv ist, dann hat der Subgraph C_C mehr Kanten als durch Zufall erwartet werden und repräsentiert deswegen eine mögliche Community. Für $M_C = 0$ können die Verbindungen zwischen den Knoten durch Zufall erklärt werden und für ein negatives M_C existieren keine Communities.

Das gesamte vorliegende PGP-Netzwerk weist eine Modularität von 0.879 auf. Für die Community-Analyse wurde der Algorithmus von Blondel et al. [17] verwendet, der basierend auf Optimierung der Modularität, in kurzer Zeit eine heuristische Annäherung für mögliche Communities macht.

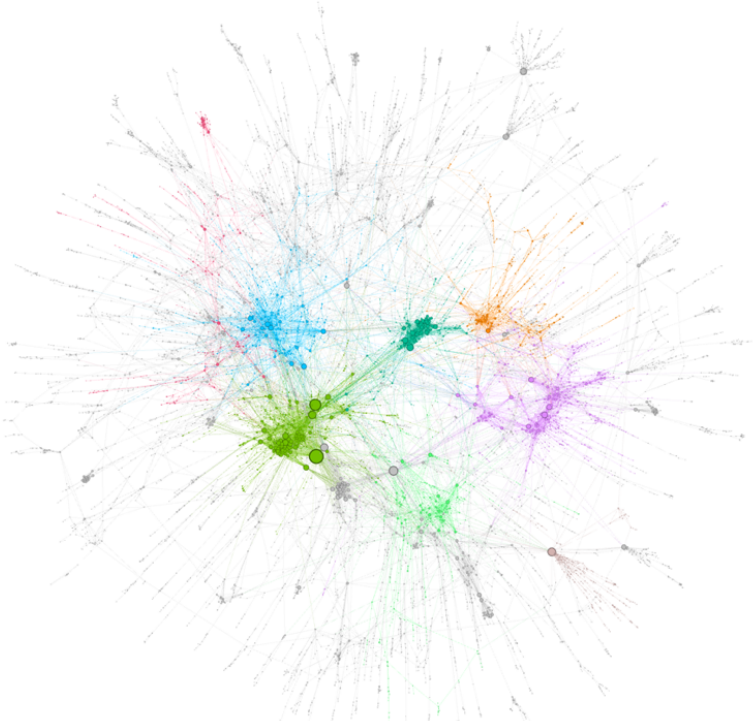


Abbildung 7 - Communityanalyse basierend auf der Modularität erstellt mit Gephi. Die Knoten wurden dabei anhand der Modularitäten eingefärbt. Die relative Grösse der Knoten gibt deren Grad an.

Es wurden durch den Algorithmus 94 Communities gefunden. Betrachtet man die Communities genauer, so sind einige interessante Eigenschaften hervorzuheben:

- Das PGP Netzwerk weisst Communities auf, die einer traditionellen Certificate Authority ähneln, in dem eine baumartige Struktur vorhanden ist (Siehe braune und graue Community in Abbildung 8).
- Das PGP Netzwerk weisst Communities auf, die einen dezentralisierten Charakter haben (siehe rote und blaue Community in Abbildung 9).

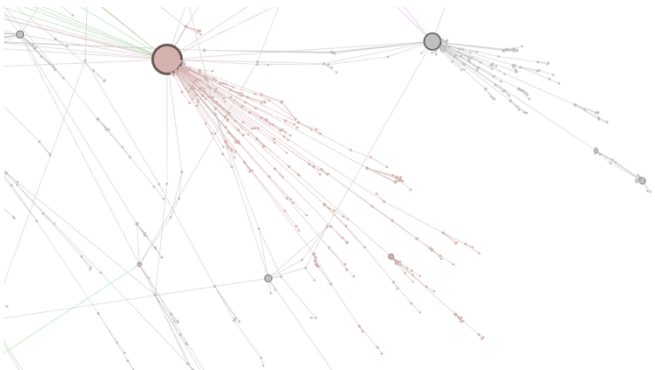


Abbildung 8 - baumartige Community

Abbildung 9 - Dezentralisierte Community

Die Community-Analyse hat gezeigt, dass das PGP-Netzwerk tatsächlich über Communities verfügt. Die obigen Beobachtungen legen nahe, dass das PGP Netzwerk einerseits keine uniforme Verteilung hat und andererseits Vergleiche mit anderen sozialen Netzen nur mit Vorsicht angestellt werden dürfen, da diese Strukturen mit grosser Wahrscheinlichkeit nicht durch Zufall entstanden sind, sondern von Menschenhand aufgrund einer bestimmten Art der Verwendung des Netzwerks geschaffen wurden. Das Modell von Boguñá et al. [6] beachtet diese Charakteristiken dabei nicht und geht von einer uniformen Verteilung aus.

D. WICHTIGKEIT VON SCHLÜSSELN

In diesem Abschnitt wird analysiert welche Schlüssel als besonders vertrauenswürdig gelten und wie sich diese von den anderen Schlüsseln abheben. Dafür werden die Masse *Betweenness Centrality* [18], *Mean Shortest Distance (MSD)* und *Degree Centrality* zur Einordnung eines Schlüssels verwendet.

> *H0: Aufgrund der Zentralitätsmasse eines Schlüssels kann keine Aussage über dessen Wichtigkeit im Netzwerk gemacht werden.*
> *H1: Die Zentralitätsmasse eines Schlüssels geben eine Indikation für dessen Wichtigkeit im Netzwerk.*

Die *Betweenness Centrality* wird folgendermassen berechnet [18]:

$$C_B = \frac{Anzahl\ der\ k\ddot{u}rzesten\ Pfade\ durch\ einen\ Knoten}{Anzahl\ m\ddot{o}glicher\ k\ddot{u}rzester\ Pfade\ durch\ einen\ Knoten} = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

wobei V ein Set von Knoten ist, $\sigma(s,t)$ die Anzahl möglicher kürzester Pfade durch v ist und $\sigma(s,t|v)$ die kürzesten Pfade durch v berechnet. Die Betweenness Centrality gibt an welche Knoten im Netzwerk als Brücke fungieren.

Die *Degree Centrality* wird wie folgt berechnet[6]:

$$C_D = \frac{Grad\ eines\ Knotens}{Anzahl\ aller\ Knoten} = \frac{1}{N-1} * \deg(v)$$

Die Degree Centrality weisst einem Knoten einen Score basierend auf dessen Grad zu und ist nützlich um sehr verknüpfte Knoten zu identifizieren.

Die *Mean Shortest Distance (MSD)* wird berechnet, indem der Durchschnitt aller kürzesten Pfade von einem Schlüssel zu allen anderen Schlüssel gebildet wird. Sie kann dadurch als Indikation verwendet werden, wie vertrauenswürdig ein PGP Schlüssel ist, wobei ein Schlüssel mit kleinerer Zahl vertrauenswürdiger ist (vgl. [12]).

Die durchschnittliche Pfadlänge im PGP Netzwerk beläuft sich auf: 7.486, wobei die Mean Shortest Distances gemäss Abbildung 10 verteilt sind. Der Durchschnitt der MSDs entspricht dabei der durchschnittlichen Pfadlänge. In Abbildung 11 ist zu erkennen, dass zum Teil auch Knoten, die nicht im Zentrum des Netzwerks liegen eine tiefe MSD haben.

[6] https://networkx.github.io/documentation/stable/_modules/networkx/algorithms/centrality/degree_alg.html#degree_centrality [Letzter Zugriff: 30.12.2018]

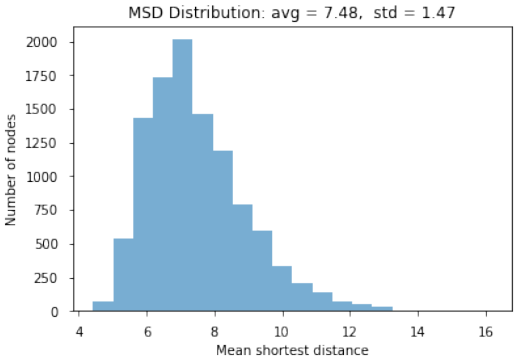


Abbildung 10 - Verteilung MSDs

Abbildung 11 - MSDs visualsiert mit Gephi
(grünere und grössere Knoten haben eine tiefere MSD)

Abbildung 12 und 13 lassen erkennen, dass nur für einige Schlüssel mit kleiner MSD sowohl die Degree Centrality wie auch die Betweenness Centrality höher ist als für andere Schlüssel mit kleiner MSD. Jedoch tendenziell falls eine hohe Betweenness Centrality oder eine hohe Degree Centrality vorliegt auch eine tiefe MSD vorhanden ist. Dies legt nahe, dass die MSD und die entsprechenden Centralities nur begrenzt austauschbar sind. Abbildung 14 zeigt, dass auch die Betweenness Centrality und die Degree Centrality nicht unbedingt miteinander verglichen werden können, obwohl wahrscheinlich schon eher von einer Korrelation ausgegangen werden kann. Der statistische Nachweis dafür wird dabei einer zukünftigen Arbeit überlassen. Die Annahme, dass die Zentralitätsmasse nur eine begrenzte Austauschbarkeit haben, bedeutet, dass je nach Mass jeweils verschiedene Schlüssel als zentral identifiziert werden. Die Betweeness Centrality identifiziert Schlüssel die als Brücken zwischen verschiedenen Schlüssel agieren. Die Degree Centrality identifiziert Schlüssel mit vielen Signaturen und

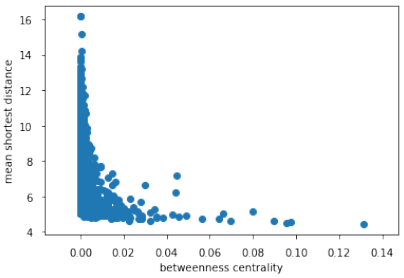


Abbildung 12 - Korrelation MSD und Betweenness Centrality

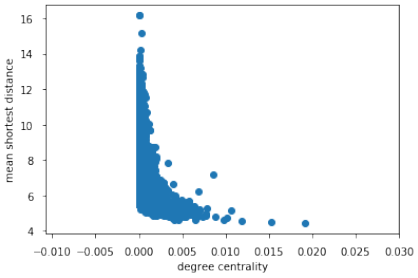


Abbildung 13 - Korrelation MSD und Degree Centrality

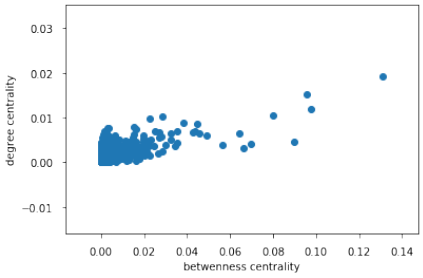


Abbildung 14 - Korrelation Degree Centrality und Betweenness Centrality

die Mean Shortest Distance legt Schlüssel offen, die im Durchschnitt von anderen Schlüssel schnell erreicht werden können.

Durch den grossen Clustering-Koeffizienten, die positive Gradkorrelation sowie der Anwesenheit von Communities kann darauf geschlossen werden, dass es sich bei PGP um ein soziales Netzwerk handelt. Dies ist keine neue Erkenntnis und wurde bereits durch Boguñá et al. [6] gezeigt, wird aber durch die in dieser Arbeit erhobenen Ergebnisse gestützt. Im nächsten Abschnitt werden die Resultate diskutiert und es wird ein Ausblick gegeben.

V. Diskussion

Die Analyse des Web of Trusts von PGP hebt zwei wichtige Erkenntnisse hervor, die zum besseren Verständnis von PGP beitragen können: einerseits das PGP Netzwerk weist verschiedene spezielle Charakteristiken auf und andererseits das PGP Netzwerk war zum Zeitpunkt der Erhebung der Netzwerkdaten sehr robust. Für die speziellen Charakteristiken kann zum einen die Gradverteilung verantwortlich gemacht werden, die besagt, dass das PGP nicht skalenfrei ist, sondern einer lognormal-Verteilung unterliegt und zum anderen die verschiedenen Communities, die zeigen, dass das Netzwerk verschiedenartig verwendet wird. So existieren Teile des Netzwerks, die wie eine herkömmliche Certificate Authority strukturiert sind und Teile, die den erwarteten dezentralisierten Charakter eines Web of Trusts verkörpern. Zukünftige Arbeiten sollten dabei das Modell von Boguñá et al. [6] anhand eines anderen sozialen Netzwerks ausser PGP validieren, um zu sehen ob diese Netzwerke wirklich verglichen werden können und dieses Modell für andere Netzwerke ebenfalls verwendet werden kann. Die Annäherung der Gradverteilung mit einer lognormal-Verteilung könnte diesbezüglich ebenfalls validiert werden, da möglicherweise mit dieser Verteilung ein Overfitting stattfindet, insbesondere könnte diese Annahme für andere Web of Trusts geprüft werden. Zu beachten ist hierbei, dass eine lognormal-Verteilung von einer kontinuierlichen Variable ausgeht und nur für $x \gg 1$ verwendet werden darf. Die Robustheit wird einerseits durch die Erkenntnisse von Penning [12] nahegelegt und andererseits durch die Analyse der Zentralitätsmasse, die zeigt, dass die MSDs mehr oder weniger normal verteilt sind, also viele Schlüssel existieren, die von anderen Schlüsseln schnell erreicht werden können.

Für ein besseres Verständnis des PGP Netzwerks als Gesamtes müssten noch mehr Daten erhoben werden, da es sich bei PGP lediglich um Infrastruktur handelt, weshalb es interessant wäre zu schauen, für was die verschiedenen Signaturen verwendet wurden. Somit könnten auch die verschiedenartigen Communities besser eingeteilt und interpretiert werden. Dennoch wurde durch diese Arbeit klar, dass ein diverses Netzwerk wie PGP verschiedene Formen aufweist, da auch die Benutzung unterschiedlich ist. Aus diesem Grund geht hervor, dass Subcommunities von Vertrauensnetzwerken separat betrachtet werden sollten und nur generelle Aussagen über das gesamte Netz gemacht werden können.

PGP wird höchstwahrscheinlich zeitnah durch eine Blockchain-Technologie abgelöst [19], da vor allem das Aufbauen eines Web of Trusts sehr mühselig ist und viel Zeit bedarf. So wird von den Protokollherstellern von PGP geraten an Key Signing Sessions teilzunehmen[7], dies ist jedoch überhaupt nicht praktikabel und alles andere als sicher. Dennoch sind die hier erhobenen Analysen wertvoll, um den Charakter von Web of Trusts besser zu verstehen. Insbesondere folgende Eigenschaften konnten dabei dank einer Analyse des PGP Netzwerks besser verstanden werden:

- eine **Visualisierung** hilft die Netzwerkstruktur besser zu verstehen und insbesondere spezielle Charakteristiken wie z.B. die verschiedenartigen Communties zu erahnen.
- die **Sicherheit** eines Vertrauensnetzwerks kann durch die Berechnung von Zentralitätsmassen wie der Mean Shortest Distance besser eingeschätzt werden.
- die Annahme, dass es sich bei einem Vertrauensnetzwerk um ein **soziales Netzwerk** handelt, hilft gewisse Methoden für dieses Modell anzuwenden (wie z.B. die positive Gradkorrelation), sollte jedoch nicht als gegeben hingenommen werden, da sich das Netzwerk je nach Verwendung anders entwickelt.

Die Analyse sozialer Netzwerke ist eine immer wichtiger werdende Disziplin, um die vielschichtigen Strukturen unserer täglichen Handlungen zu verstehen. Sie erlaubt es uns einen besseren Einblick in die Organisation eines sozialen Systems zu erhalten und aufgrund einer bestimmten Konstellation eine Vorhersage über die zukünftige Entwicklung zu machen. Die Analyse des PGP Netzwerks hat gezeigt, wie komplex die realen Bedingungen sind und wie schwierig es ist sinnvolle Erkenntnisse daraus zu

[7] https://www.gnupg.org/gph/en/manual/x547.html

gewinnen. Dies legt nahe, dass noch viel Potenzial vorhanden ist, um die Analysen von sozialen Netzwerken einfacher und resistenter gegenüber Fehlannahmen zu machen. Ein Beispiel dafür ist ein generelleres Verständnis von den Regeln in einem Vertrauensnetzwerk wie dem PGP Web of Trust und deren Anwendbarkeit auf andere ähnliche Netzwerke. Zusammenfassend kann gesagt werden, dass unsere Arbeit einen soliden Einblick in die Untersuchung des PGP Netzwerks gibt und durch eine explorative Analyse des Netzwerks neue Erkenntnisse über die Gradverteilung und Community-Struktur verschafft. Dadurch wurden weitere Fragestellungen aufgeworfen, deren Beantwortung wir uns von zukünftigen Analysen von Vertrauensnetzwerken erhoffen.

Anhang

Der Sourcecode zur Analyse kann unter folgendem Link abgerufen werden:
https://github.com/mariusgiger/network-analysis-pgp

Referenzen

[1] M. Kerwin, “Pretty Good Privacy,” 2006.

[2] P. Zimmermann, “Why I Wrote PGP,” 1999. [Online]. Available: https://www.philzimmermann.com/EN/essays/WhyIWrotePGP.html.

[3] P. Zimmermann, “PGP Message Exchange Formats,” 1996. [Online]. Available: https://tools.ietf.org/html/rfc1991.

[4] P. Corporation, “OpenPGP Message Format.” [Online]. Available: https://tools.ietf.org/html/rfc4880.

[5] J. M. Sakoda, “The checkerboard model of social interaction.,” *J. Math. Sociol.*, vol. 1, no. 1, pp. 119–132, 1971.

[6] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, “Models of social networks based on social distance attachment,” *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, vol. 70, no. 5, p. 8, 2004.

[7] S. N. Dorogovtsev and J. F. Mendes, “Evolution of networks: From biological nets to the Internet and WWW,” *OUP Oxford*, 2013.

[8] D. J. Watts and S. H. Strogatz, “Collective dynamics of ’small-world’-networks,” *Nature*, vol. 393, no. 6684, p. 440, 1998.

[9] R. Pastor-satorras, V. Alexei, and A. Vespignani, “Dynamical and correlation properties of the Internet,” *Phys. Rev. Lett.*, vol. 87, no. 25, p. 258701, 2001.

[10] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Phys. Rev. E*, vol. 69, no. 2, p. 026113, 2004.

[11] J. Cederlöf, “Dissecting the Leaf of Trust.” [Online]. Available: http://www.lysator.liu.se/~jc/wotsap/leafoftrust.html. [Accessed: 20-Dec-2018].

[12] H. P. Penning, “Analysis of the strong set in the PGP web of trust,” 2018. [Online]. Available: https://pgp.cs.uu.nl/plot/. [Accessed: 20-Dec-2018].

[13] KONECT, “Pretty good privacy network dataset,” 2017. [Online]. Available: http://konect.uni-koblenz.de/networks/arenas-pgp. [Accessed: 10-Oct-2018].

[14] A. Clauset, C. R. Shalizi, and M. E. J. Newman, “Power-law distributions in empirical data,” *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, 2009.

[15] M. E. J. Newman, “Mixing patterns in networks,” *Phys. Rev. E*, vol. 67, no. 2, p. 026126, 2003.

[16] A.-L. Barabási, *Network Science*, 4th ed. Cambridge: Cambridge University Press, 2017.

[17] V. D. Blondel, J. Guillaume, J. L. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *J. Stat. Mech. theory Exp.*, vol. 2008, no. 10, p. P10008, 2008.

[18] U. Brandes, “A Faster Algorithm for Betweenness Centrality ∗,” *J. Math. Sociol.*, vol. 25, no. 2, pp. 163–177, 2001.

[19] W. Duane and G. Ateniese, “From pretty good to great: Enhancing PGP using bitcoin and the blockchain,” in *International conference on network and system security*, 2015.