Analyse des OpenPGP Web of Trust

Alexander Ulrich
ulricha@informatik.uni-tuebingen.de

Wilhelm-Schickard-Institut für Informatik
Universität Tübingen

Abschlussvortrag Studienarbeit
23.06.2010

Motivation

- In den letzten Jahren steigende Verfügbarkeit von Daten über große, komplexe Netzwerke (z.B. technologische, soziale, biologische)
- Analyse der Struktur dieser Netzwerke kann interessante Einblicke in die Mechanismen der zugrundeliegenden Systeme liefern
- OpenPGP (PGP/GnuPG): praktikables und weit verbreitetes Paket für Verschlüsselung und Authentifizierung, insbesondere von E-Mail
- Web of Trust: dezentraler Authentifizierungsmechanismus für Schlüssel liefert komplexes Netzwerk
- verfügbare Datensätze über das Web of Trust sind veraltet oder fehlerhaft
- bisherige Arbeiten beruhen meist auf sehr alten Daten oder haben sich nur am Rand mit der Netzwerkstruktur beschäftigt

Web of Trust

- Public-Key-Kryptographie: Verteilung von (öffentlichen) Schlüsseln über unsichere Kanäle
- Notwendig: Überprüfung der *Authentizität* von Schlüsseln
- Digitale Signaturen auf Schlüsseln nach Überprüfung der Identität
- X.509: zentrale *Certificate Authorities* (CA)
- Web of Trust: Jeder Teilnehmer kann als CA fungieren, jeder Teilnehmer entscheidet selbst, welchen CAs er *vertraut*
- Signaturen aller Teilnehmer ergeben einen gerichteten Graphen
- Aufbau von Signaturketten: jedem Kettenglied muss vertraut werden
- Hypothese: Signaturen spiegeln *soziale Beziehungen* wieder $\Rightarrow$ soziales Netzwerk
- Keysigning-Parties

Fragestellung

- Wie ist das Netzwerk insgesamt aufgebaut?
- Verfügt das Netzwerk über eine (regelmäßige) *Struktur* oder ist es vollständig chaotisch?
- Inwiefern spiegelt die Graphenstruktur die zugrundeliegenden Mechanismen (Signierung von Schlüsseln) wieder?
- Wie stark wird das System Web of Trust überhaupt benutzt?
- Wie gut erfüllt es seinen Zweck?

Übersicht über die Arbeit

- Implementierung der Datenextraktion aus Schlüsseldatenbank
- Analyse der Struktur des Graphen auf verschiedenen Ebenen
- FIXME

Software

Extraktion

- Integriert in SKS-Keyserver (OCaml, $\approx$ 1200 LOC)
- Reduzierung von OpenPGP-Schlüsseln auf interessante Daten
- Daten in SQL-Datenbank abgelegt
- speichert komplette Geschichte: Zeitpunkte von Schlüssel- und Signaturerzeugung, Ablaufdatum, Widerrufsdatum
- Zusätzliche Daten: Public-Key-Algorithmus und Hashalgorithmus, Schlüssellängen, UserIDs

Auswertung

- Sammlung von Kommandozeilenwerkzeugen (OCaml, $\approx$ 3800 LOC)
- All-pair-shortest-path und Betweeness centrality auf MPI-Cluster

Datensatz

- Datenbank: 2700000 Schlüssel, 1100000 Signaturen
- 410000 Schlüssel abgelaufen, 100000 widerrufen, 50000 defekt
- Komplett unvernetzte Schlüssel entfernt
- Graph mit 325000 Knoten, 817000 Kanten
- Großteil der verfügbaren Schlüssel ist nicht verifizierbar, kann keine Signaturketten verwenden
- unbekannte Anzahl von *nicht-öffentlichen* Signaturen
- Keine Aussage über *aktuelle* Anzahl von PGP-Benutzern möglich

Zusammenhangskomponenten

Bekannt: Eine gigantische starke Zusammenhangskomponente (MSCC) mit ca. 45000 Knoten. Aber: Wie ist der Rest strukturiert?

- ca. 240000 Komponenten
- Deutlich über 100000 mit Größe 1, 10000-20000 Knotenpaare, wenige zwischen 10 und 100
- Einzelne Knoten oder sehr kleine Cluster, die einzelne Kanten zu anderen Komponenten haben
- Komponenten untereinander kaum vernetzt
- *bow tie*: in 18000 Knoten, out 92000 Knoten

$\Rightarrow$ Nennenswerte Signaturaktivität nur in der MSCC, für restliche Schlüssel ist das Web of Trust kaum benutzbar.

Rest der Arbeit konzentriert sich auf MSCC: 45000 Knoten, 443000 Kanten (Capkun 2001: 12000 Knoten).

Strukturelle Merkmale

Gegenseitigkeit

- ca. 50% aller Kanten haben eine Gegenkante
- Erklärung: Signierung normalerweise beidseitig
- positive Korrelation zwischen ein- und ausgehendem Grad

Small-World

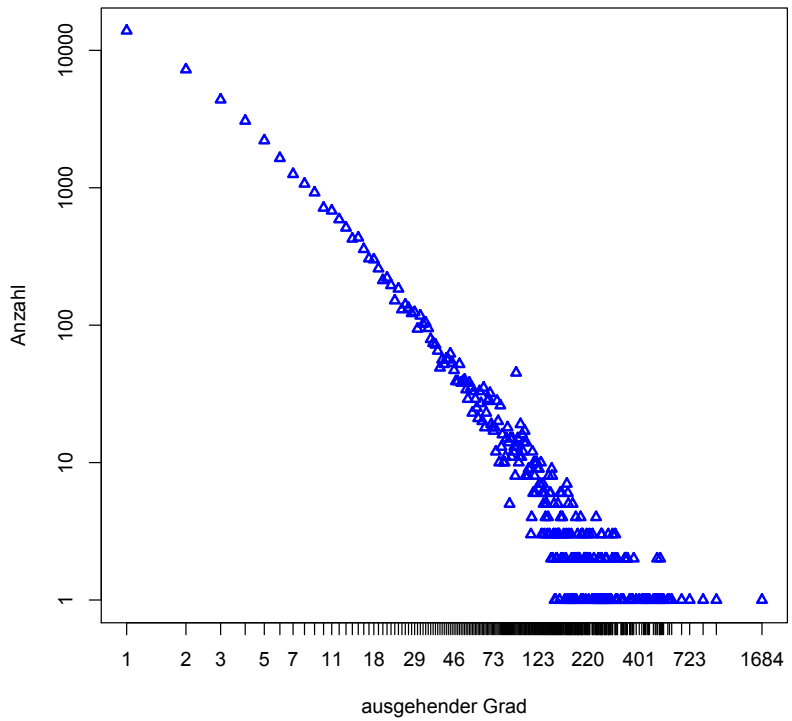
- Geringe charakt. Distanz $\approx 12 \Rightarrow$ Small-World-Effekt
- Interessant aufgrund weltweiter geographischer Verteilung der Teilnehmer
- Radius 16, Durchmesser 36

Clustering

- hoher Clustering coefficient $C = 0,460$ (configuration model $C = 0,013$)
- hohes Maß an Clustering charakteristisch für soziale Netzwerke (Newman 2003)
- Grund: Community-Struktur

Gradverteilung (1)

Durchschnittlicher ausgehender Grad 9,29

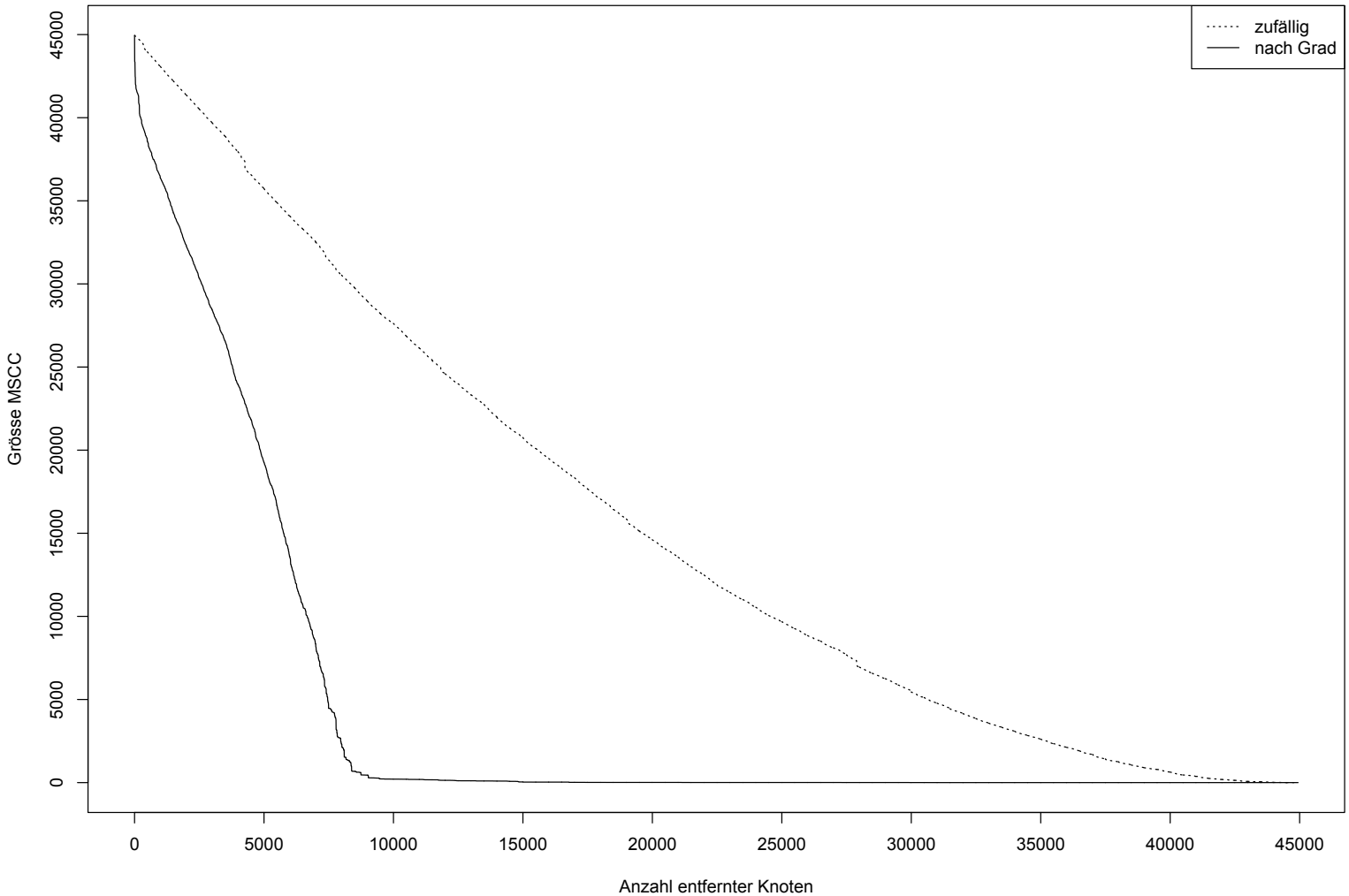


- Sehr inhomogene Verteilung, Durchschnitt ist wenig aussagekräftig.
- Annähernd gerade auf log-log-Plot $\Rightarrow$ Power-Law?

Gradverteilung (2)

- Bestimmung des Exponenten: keine lineare Regression
- Stattdessen Maximum-Likelihood-Methode (Clauset 2009): $x_{min} \approx 84$, $\alpha = 2,35$
- Überprüfung der Anpassungsgüte (Kolmogorov-Smirnov) schließt power law aus
- Zentrales Merkmal in Literatur zu skalenfreien Netzwerken: Hub-Struktur, "robust yet fragile" (Albert 2000)
- Power law weder hinreichend noch notwendig für Hub-Struktur (Li 2005)
- Verteilung mit hoher Variabilität, gut vernetzte Knoten primär mit anderen gut vernetzten Knoten verbunden
- Tatsächlich: (schwache) positive Korrelation zwischen Graden benachbarter Knoten

Robustheit (1)



Robustheit (2)

zufällige Schädigung Ablaufsdatum, Passphrase vergessen, neuer Schlüssel...

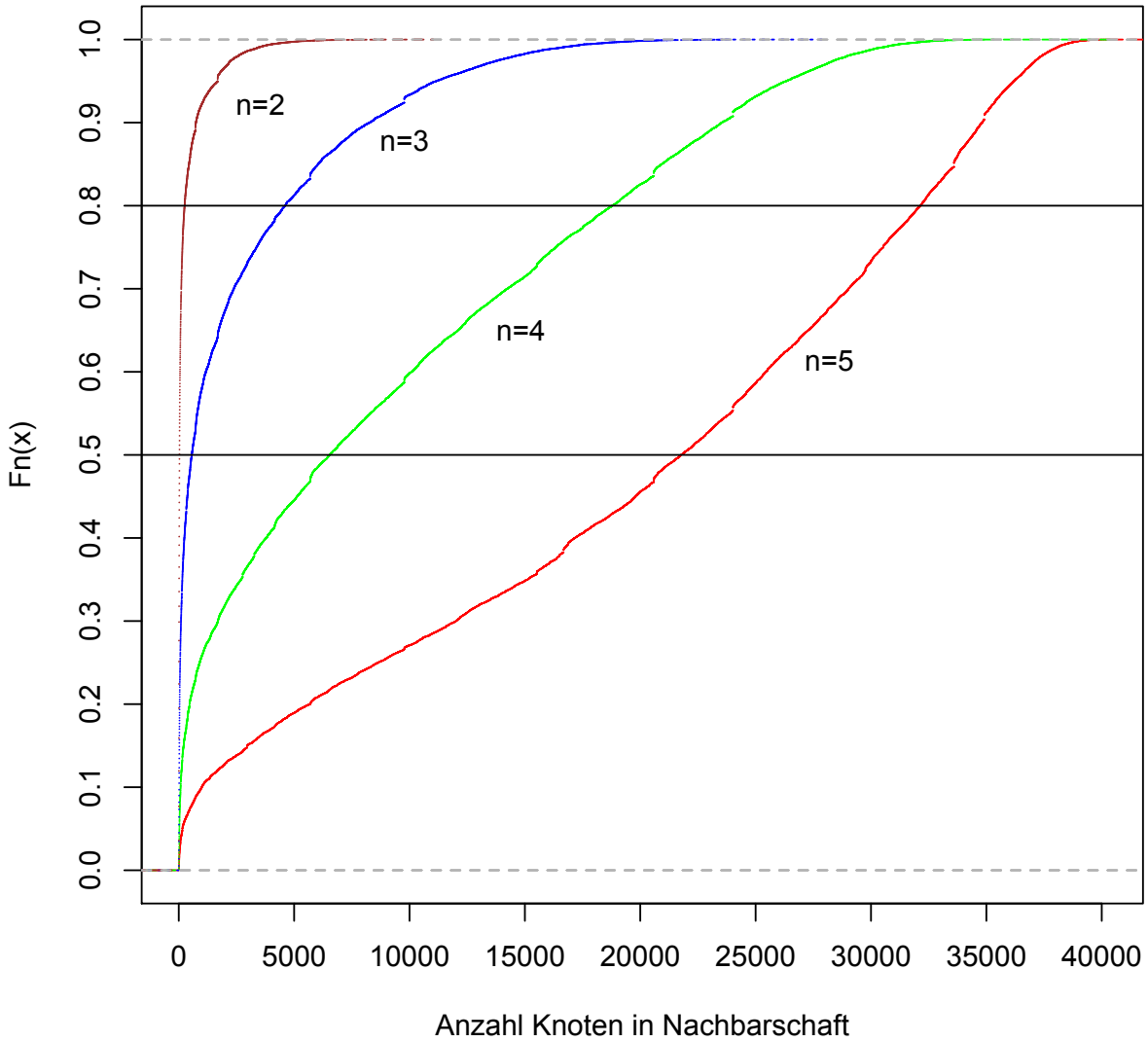
gezielter Angriff Kompromittierung

- Netzwerk ist sehr robust gegen Schädigung
- Netzwerk ist überraschend robust gegen gezielten Angriff, kein rapider Zerfall
- $\Rightarrow$ Zusammenhalt hängt nicht von wenigen gut vernetzten Knoten ab, keine ausgeprägte Hub-Struktur
- charakteristische Distanz/Durchmesser nicht betrachtet
- Knoten mit höchstem Grad sind nicht unbedingt die *zentralsten* Knoten

Nützlichkeit (1)

- Mindestvorraussetzung für Verifizierbarkeit: Signaturkette $\Rightarrow$ Pfad
- Wichtige Einschränkung: maximale Pfadlänge **5**
- Zusätzlich: *Vertrauen* in jedes Glied der Signaturkette
- Erhebliche Einschränkung für Knoten mit Grad 1 (ein-/ausgehend): keine Redundanz
- Radius 16, durchschnittliche Eccentricity ≈ 28, durchschnittliche Distanz 12
- Je länger die Kette desto mehr Vertrauen ist notwendig $\Rightarrow$ betrachte auch kürzere Ketten
- h-Nachbarschaften ($h = 1, \ldots, 5$)

Nützlichkeit (2)



Communities (1)

- Individuen in sozialen Netzwerken neigen zu Gruppenbildung: familiär, freundschaftlich, gemeinsame Interessen, professionell...
- Communities: Gruppen von Knoten, die untereinander über viele Kanten verfügen, nach ausserhalb nur wenige Kanten (Fortunato 2010)
- Hypothese: Signaturentstehung wird von sozialen Beziehungen und Keysigning-Parties bestimmt
- Frage: hat der Graph eine ausgeprägte Community-Struktur?
- Frage: Lässt sich für Communities entscheiden, wie sie entstanden sind?
- Kriterien: SLD aus UserIDs (Organisation, Land), zeitliche Korrelation der Signaturen (Keysigning-Party)

Communities (2)

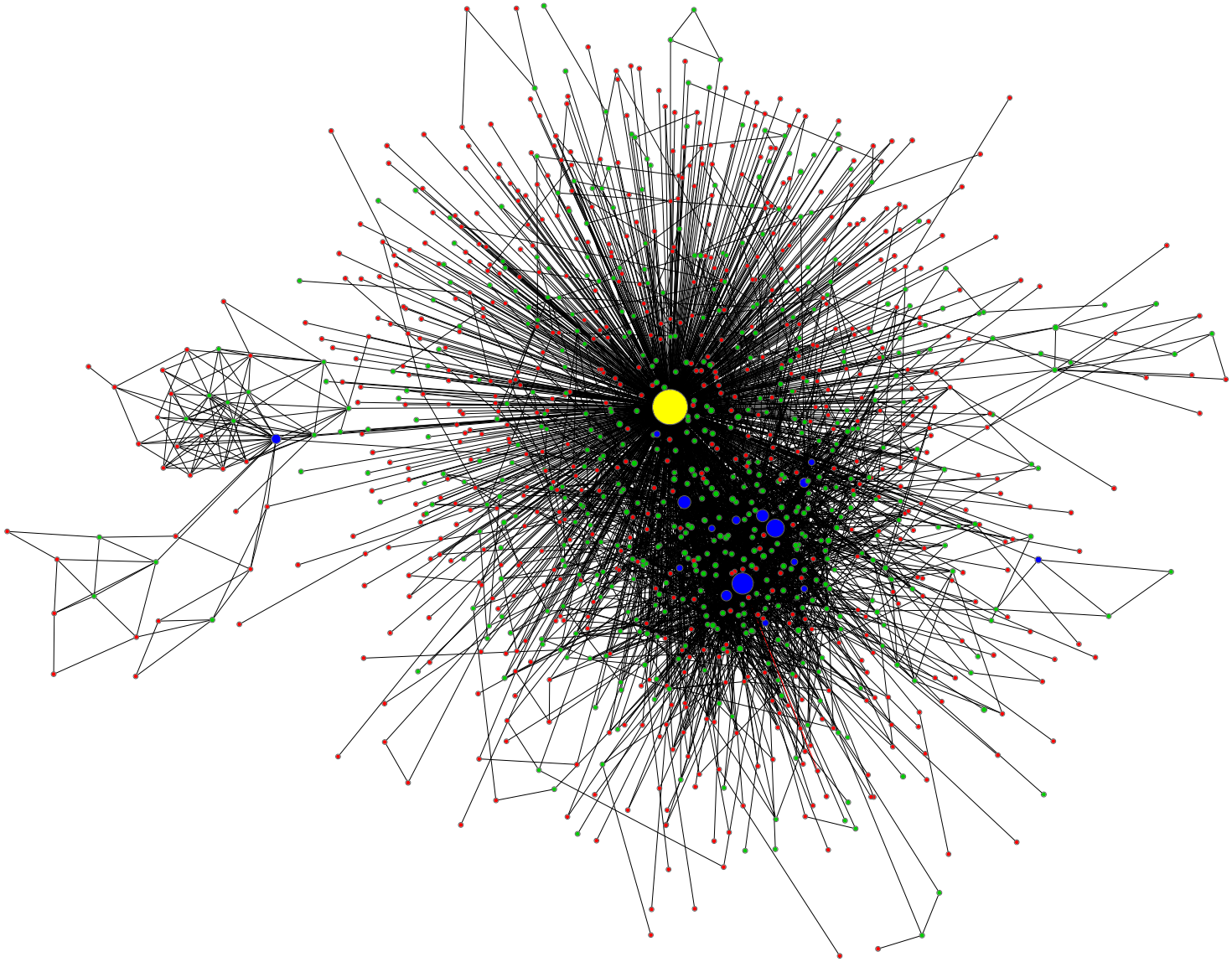


Abbildung: Struktur der Communities größer 5

Communities (3)

- 1421 Communities grösser 3, stark inhomogene Verteilung
- Aufteilung gibt die Struktur des Netzwerkes gut wieder (Modularity $Q = 0,780$) $\Rightarrow$ tatsächlich ausgeprägte Community-Struktur
- eine gigantische Community, Rest tendenziell sternförmig angeordnet
- Fast alle Communities sind einem Land zuordnenbar
- Zuordnung zu SLDs funktioniert nur bei kleineren Communities, nicht bei besonders vielen (Ausnahmen: apache.org (436, 48%), cert.org (97, 70%), ...)
- Zeitliche Korrelation bei 40% der Communities, hauptsächlich kleineren

Communities (4)

- Annahme nicht widerlegt
- Methoden zu primitiv, um Entstehungsmechanismus zu erklären
- Sehr beschränkte Daten über soziale Gruppenzugehörigkeit (nur UserIDs)
- Je aktiver die Teilnehmer, desto unschärfer wird das Bild $\Rightarrow$ Communities verschmelzen
- Beispiel: Grösste Community ist intern sehr dicht

Möglicherweise interessant:

- Anhand der Entstehungsgeschichte des Netzwerks Entwicklungsdynamik der Communities nachvollziehen
- Untersuchen, wie Communities untereinander in Bezug auf nationale/geographische Zuordnung vernetzt sind

Zusammenfassung

?