



Kommunikationsnetzwerke

Vortrag im Rahmen des Seminars Netzwerke

Michael Murawicki und Robert Richter

Agenda

- Instant Messaging Studie
 - Kommunikationscharakteristika
 - Netzwerkeigenschaften

- Small World Experimente
 - Milgram
 - Dodd

- IRC Kommunikationsgraphen
 - Piespy
 - Experiment (optional)

Weltweite instant-messaging (IM) Studie

-Autoren

- Jure Leskovec
 - Carnegie Mellon University, Pittsburgh, USA
- Eric Horvitz
 - Microsoft Research, Redmond, USA

-MS Instant Messenger Verbindungsdaten

- Kompletter Juni 2006
- Anonym, keine Inhalte
- Ziel: statistische Analyse

-Autoren behaupten

- Die Hälfte des weltweiten IM Verkehrs wurde erhoben
- Größte bisher analysierte IM Datenmenge

Instant Messaging

-Applikationen ermöglichen

- Textkommunikation
 - nahezu synchron
 - i.d.R. unicast (1-zu-1)
 - Aber auch multicast (1-zu-viele(nicht alle))
- Kontaktlisten
 - Zeigen Verbindungsstatus bekannter Kontakte

-Verschiedene Protokolle und Clients

- XMPP (Protokoll)
- Windows Live Messenger (ehemals MS Messenger; Protokoll und Client)
- ICQ (Protokoll und Client)
- ...

MS IM Studie: Datensatz

-Erfasst wurden

- Präsenz (zeit- und benutzermarkiert)
 - Login, logout
 - Hinzufügen, entfernen, blockieren eines „Buddies“; einladen eines neuen Benutzers
 - Statusänderung
- Kommunikation
 - Pro beteiligten Benutzer einer Kommunikationssitzung
 - Sitzungs-ID, Benutzer-ID, Eintritt, Austritt, nTx und nRx Nachrichten
 - Benutzer (freiwillige Selbstangabe im Profil)
 - Alter, Geschlecht, Ort, Sprache, IP-Adresse (für geographische Lokalisierung)

-Ungerichteter Graph mit 180 Mio. Knoten und 1,3 Mrd. Kanten

- 180 Mio. verschiedene kommunizierende Benutzer aus 240 Mio. Aktiven
- 1,3 Mrd. Konversation aus 30 Mrd. Nachrichten

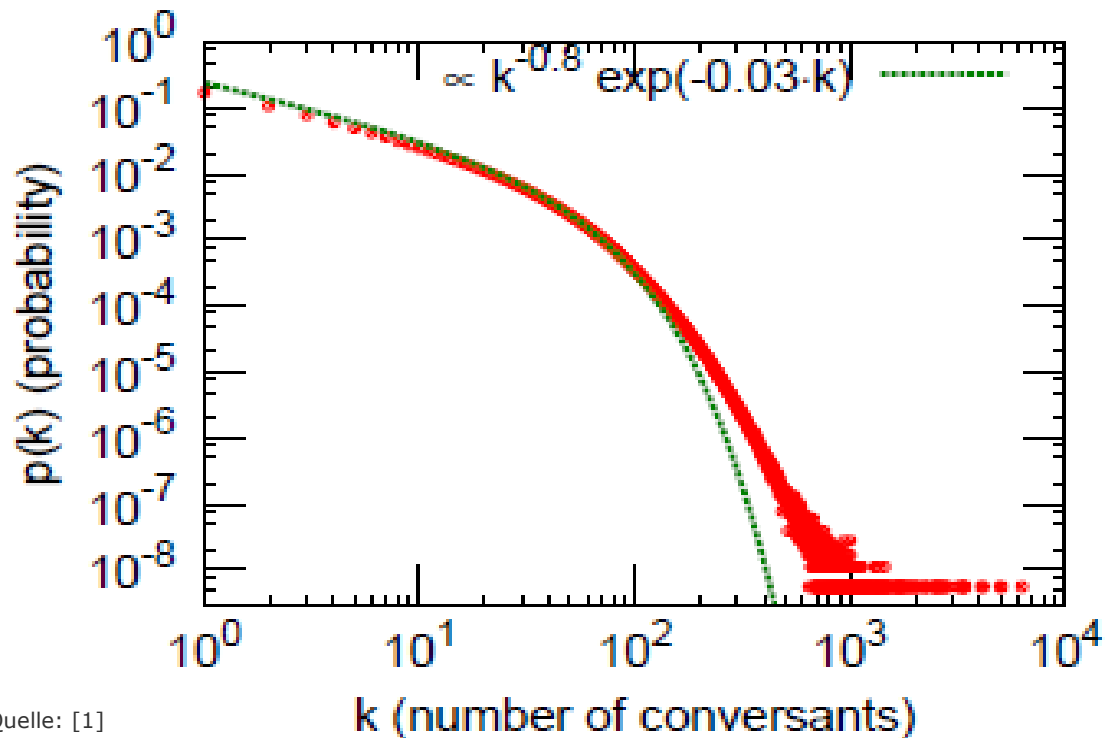
MS IM Studie: Analyse

-Analysiert wurden (Ausblick)

- Aktivität
 - Login, Logout, Sitzungsdauer etc.
- Demographie
 - Alter, Geschlecht, etc.
- Kommunikationscharakteristik
 - Anzahl Kommunikationen und Nachrichten
- Korrelationen zwischen
 - Kommunikationscharakteristik
 - Demographie
 - Geographie
- Homophilie der Kommunikation
- Das Netzwerk selbst: Eigenschaften des Kommunikationsgraphen
 - Cluster
 - Robustheit
 - Small world

MS IM Studie: Ergebnisse

- Knotengradverteilung k (Anzahl Kommunikationspartner)
 - Die Dichtefunktion der Wahrscheinlichkeitsverteilung der Knotengrade
 - folgt Potenzgesetz („power-law“) mit exponentiellem Abschluss
 - Potenz $a=-0.8$
 - Exponentkonstante $\text{Lambda}=-0.03$



Quelle: [1]

Wahrscheinlichkeitsverteilungen

-In Netzwerken kann man verschiedenste Verteilungen beobachten

-Kurze Wiederholung (oder auch Einführung)

- Normalverteilung
- „power-law“
- Logarithmische Normalverteilung
- Exponentialverteilung

-Generell

- Dichtefunktion (engl.: probability density function)
 - $p(x) = \Pr(X=x)$
 - Ableitung der Verteilungsfunktion
- Verteilungsfunktion (engl.: cumulative distribution function)
 - $P(x) = \Pr(X \leq x)$
 - Integration der Dichtefunktion

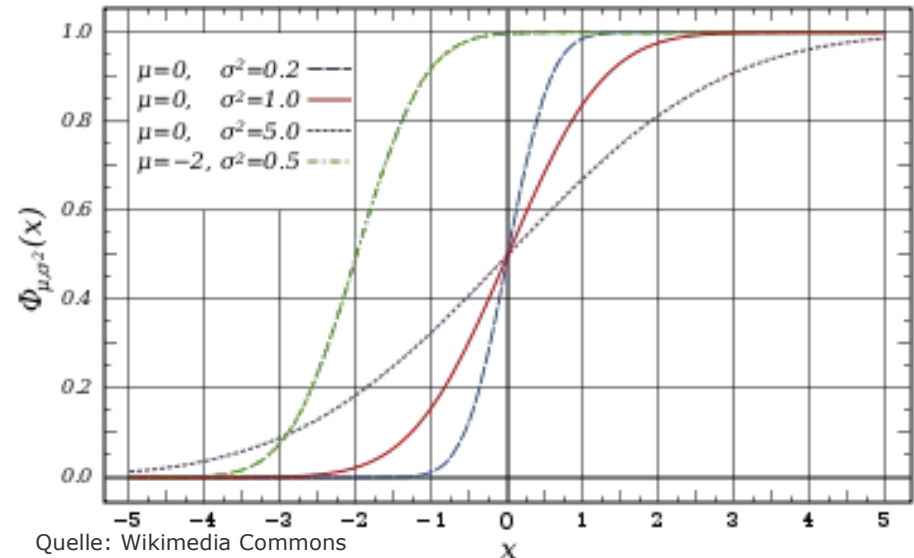
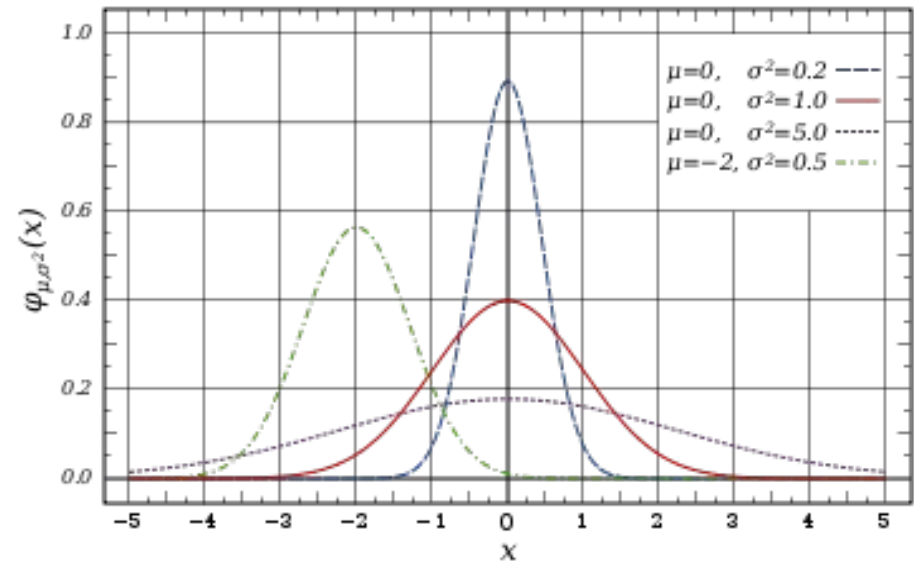


Normalverteilung

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}$$

-Zentraler Grenzwertsatz

- „Die Summe von n unabhängigen, identisch verteilten Zufallsvariablen ist im Grenzwert normal verteilt.“
- Deshalb wohl so häufig zu beobachten
- Vermutete man auch häufig in Netzwerken (Vergleiche [Film])
- Doch: häufiger in Netzwerken
 - „power-law“ →



Quelle: Wikimedia Commons

„power-law“

-Verteilung folgt dem „power-law“, wenn für ihre Dichtefunktion gilt

$$p(x) \propto x^{-a}$$

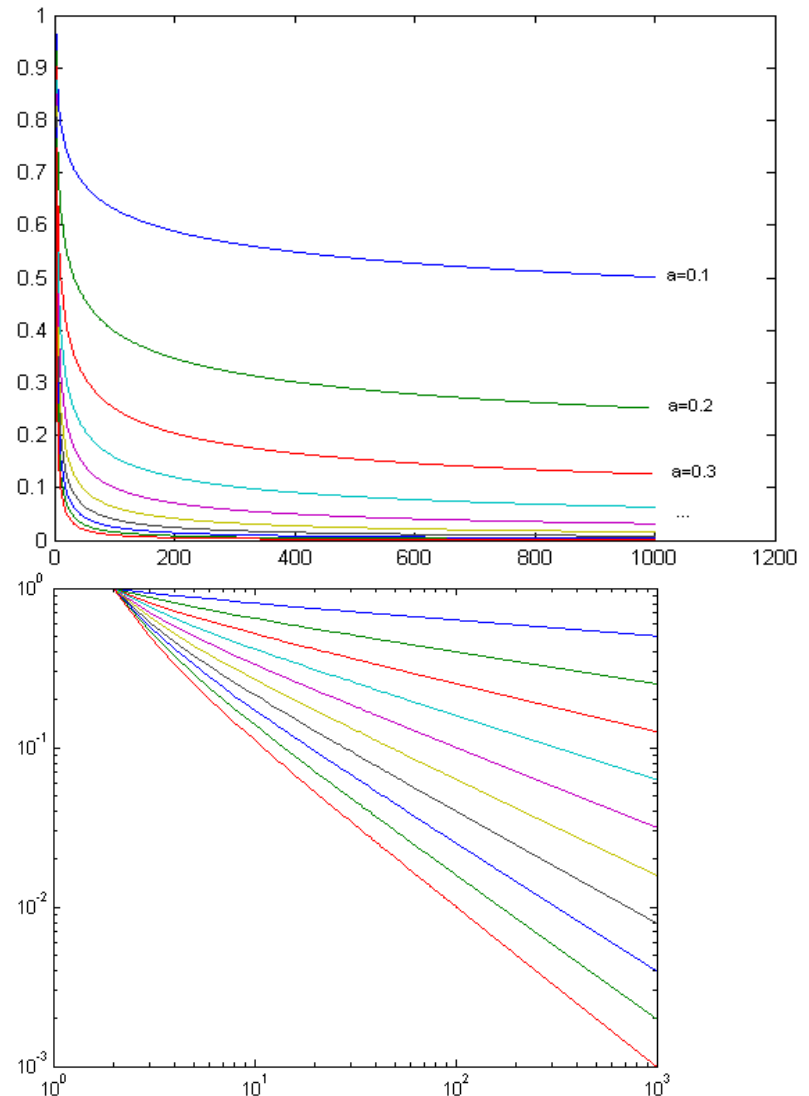
-„Große Ereignisse sind selten, kleine sind recht häufig“ [3]

-Klassisches Beispiel

- Es gibt wenige Städte mit vielen Einwohnern, aber viele mit wenigen.

-Grafische Darstellung

- Lineare Skala ist problematisch
- Deshalb logarithmische Skala



Logarithmische Normalverteilung

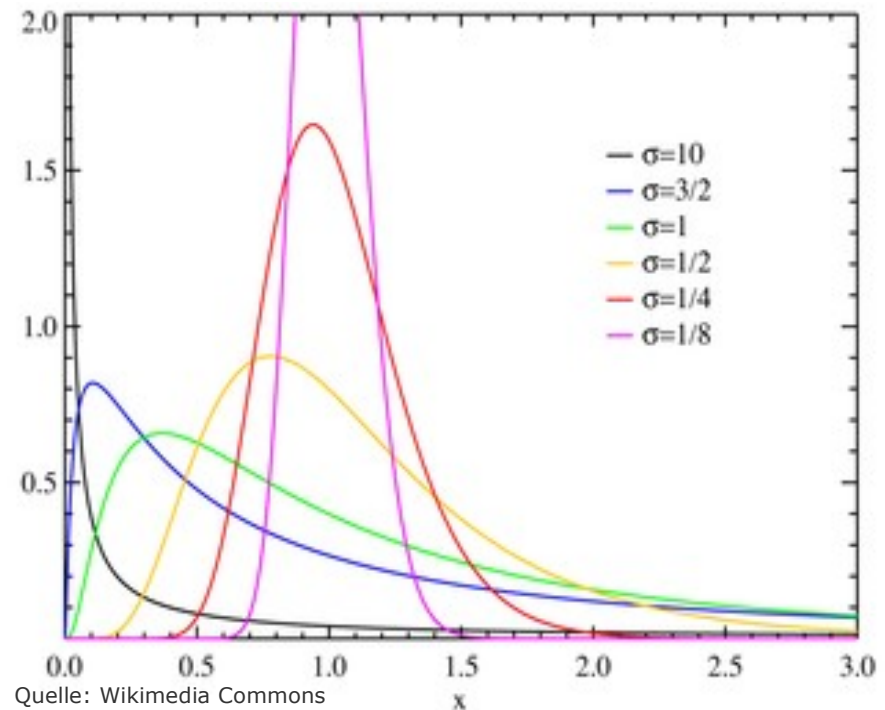
-Logarithmische Normalverteilung (LN) einer Zufallsvariablen X liegt vor, wenn

$$\ln(X) \sim N(\mu, \sigma^2) \Leftrightarrow X \sim LN(\mu, \sigma^2)$$

$$p(x) = \frac{1}{x \sigma \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

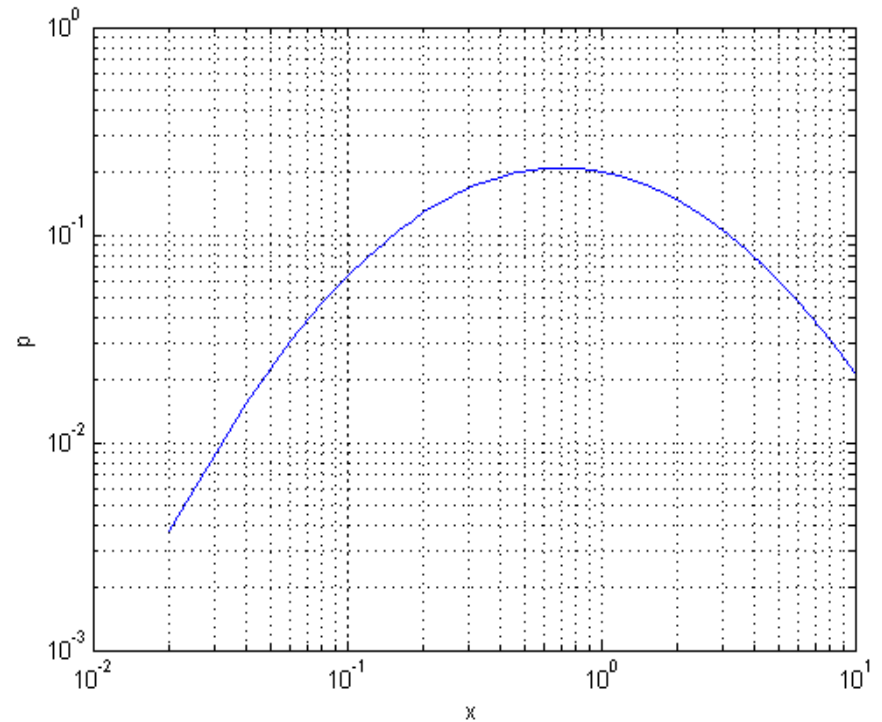
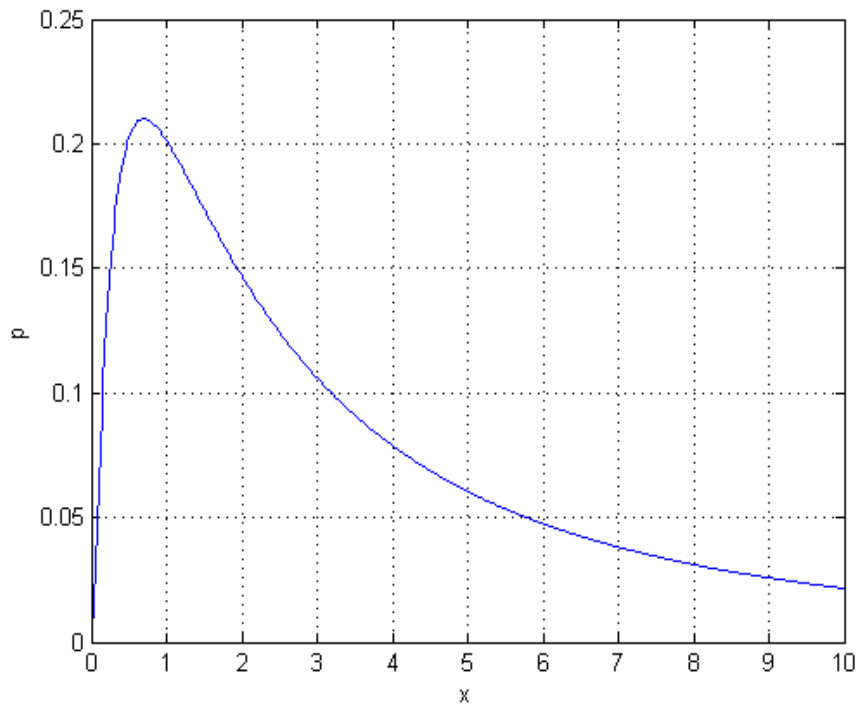
-Beispiel

- Einkommensverteilung in einem Unternehmen
 - Wenig hohe Gehälter
 - Viele vergleichsweise niedrigere Gehälter
 - Kaum sehr niedrige Gehälter

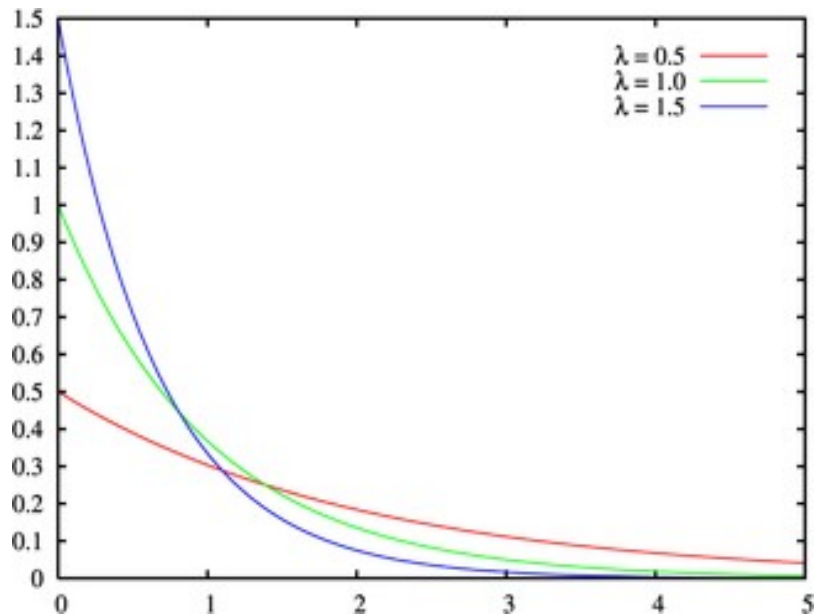


Logarithmische Normalverteilung

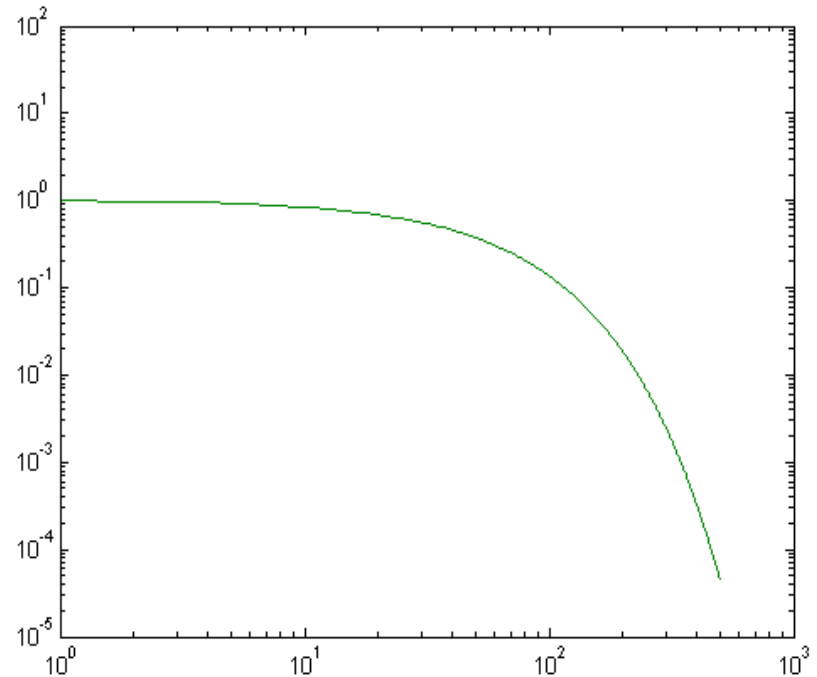
-LN auf voll logarithmischer Skala



Exponentialverteilung



Quelle: Wikimedia Commons



$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0 \\ 0 & , \text{sonst} \end{cases}$$

-Beispiel Ausfallwahrscheinlichkeit

- Lambda ist Ausfallrate
- Die Wahrscheinlichkeit, dass ein Bauteil am Tag x ausfällt ist exponentiell verteilt (Gegenwahrscheinlichkeit: $1-f$)

Exponentialverteilung

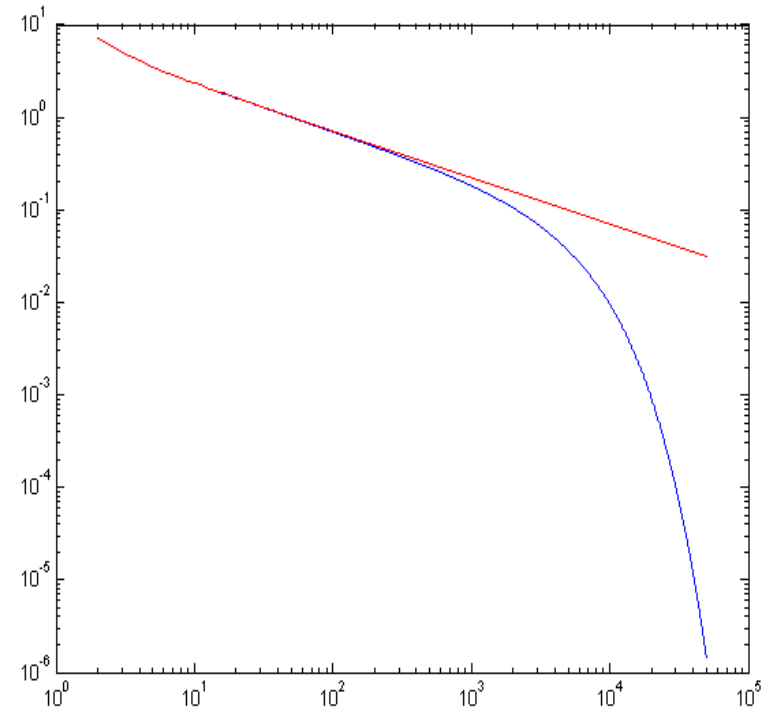
- „Power-law“ mit exponentiellem Abschluss

$$p(x) \propto x^{-a} e^{-\lambda x}$$

- Exponentieller Anteil dominiert mit steigendem x

- Beispiel: Gutenberg-Richter Gesetz

- Wahrscheinlichkeit von Erdbeben einer bestimmten Stärke
- Für weniger starke \rightarrow „power-law“
- Für sehr starke \rightarrow exponentiell
- Ursache
 - Begrenzte Energie in der Erdkruste



Literaturbeispiele

-Beispiele

- Komplementäre Verteilungsfunktion $P(X > x)$

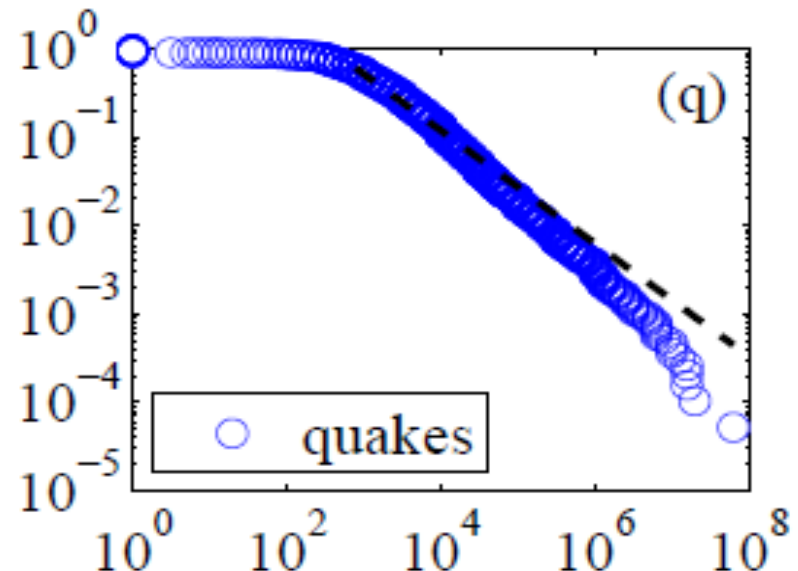
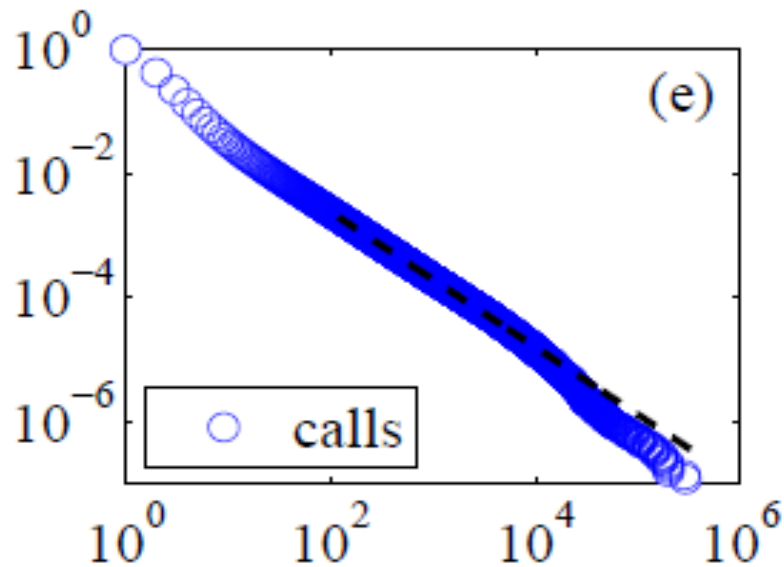
- (Erinnerung: Verteilungsfunktion $P(X \leq x)$)

- Links: erhaltene Anrufe pro AT&T Kunde

- Rechts: Intensität von Erdbeben

-Beobachtung: $P(x) \propto x^{-a}$

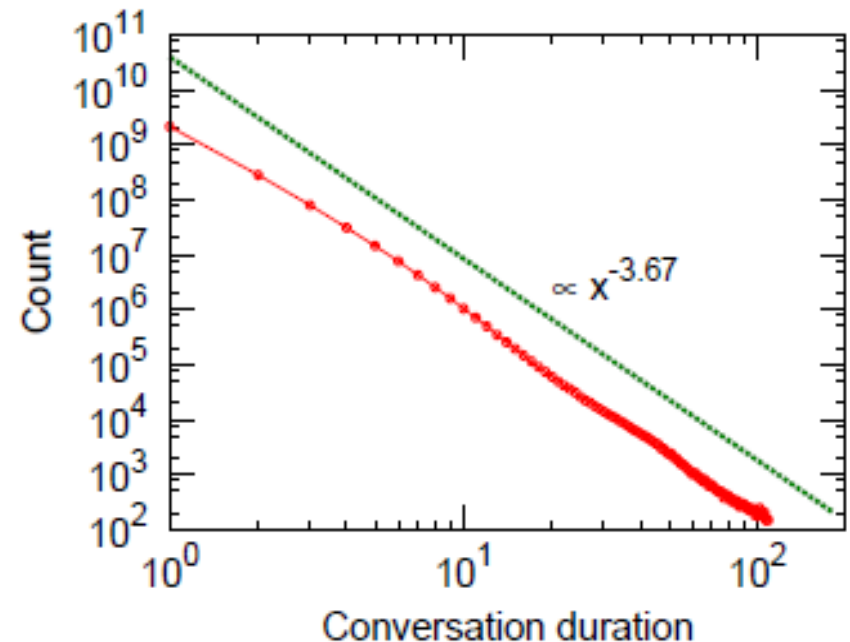
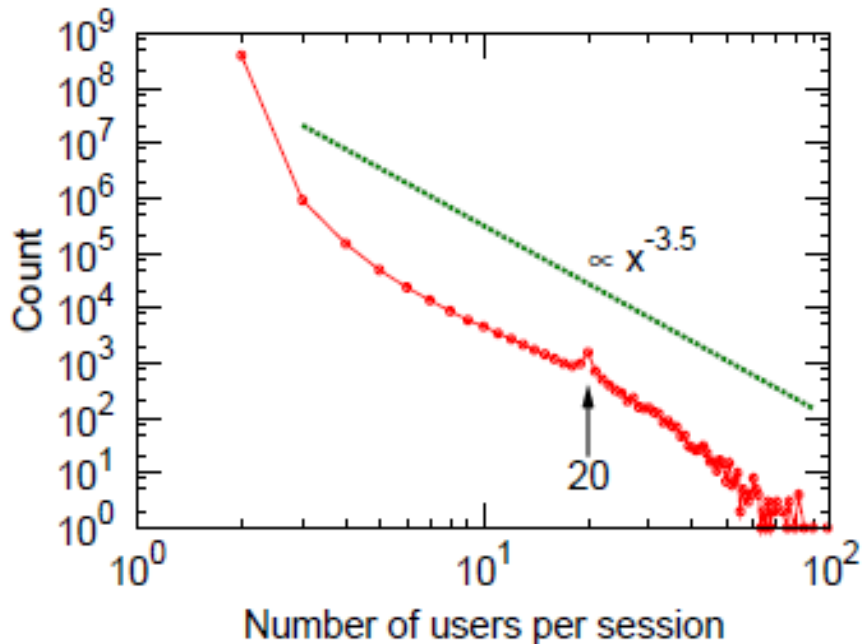
- Nach [2]: Dichtefunktion \sim PL gdw. Verteilungsfunktion \sim PL



Quelle [3]

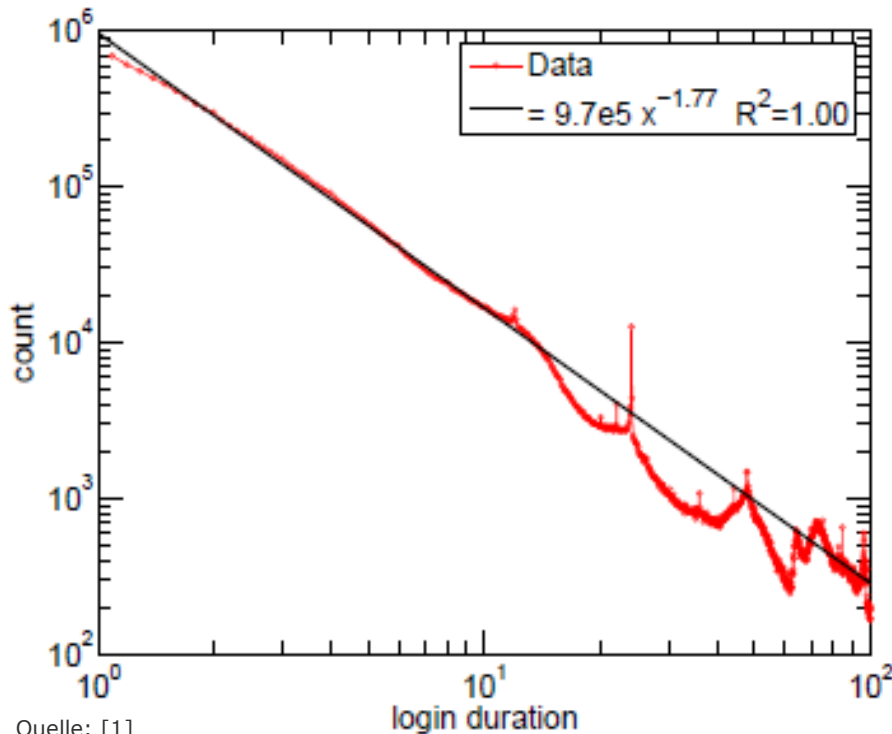
MS IM Studie: Nutzungsaktivität

- Links: Benutzer pro Sitzung („Chatraum“)
 - Benutzerlimit=20 (aber es zählen alle Nutzer, die einmal beteiligt waren)
 - 99% der Sitzungen waren zwischen 2 Benutzern (Ausreißer)
- Rechts: Sitzungsdauer (keine Quantität gegeben, vermutlich Stunden)

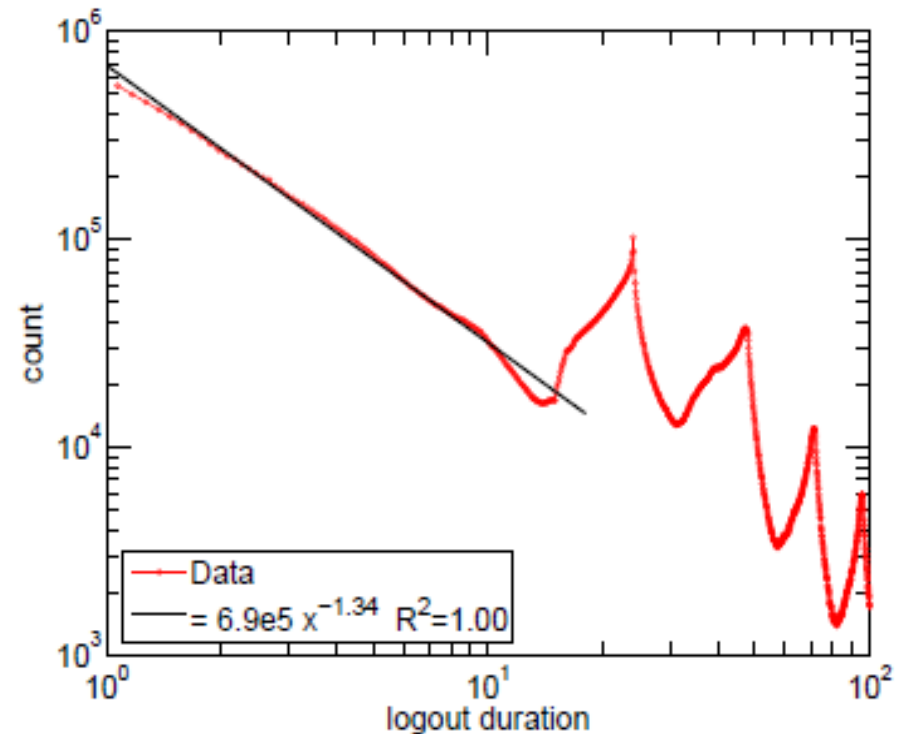


Nutzungsaktivität

- Links: Login-Dauer („online“)
 - Ausreißer bei 24 Std
- Rechts: Logout-dauer („offline“)
- Beobachtung: Login Zeiten fallen schneller
- „Interpretation: Es gibt noch ein Real Life“



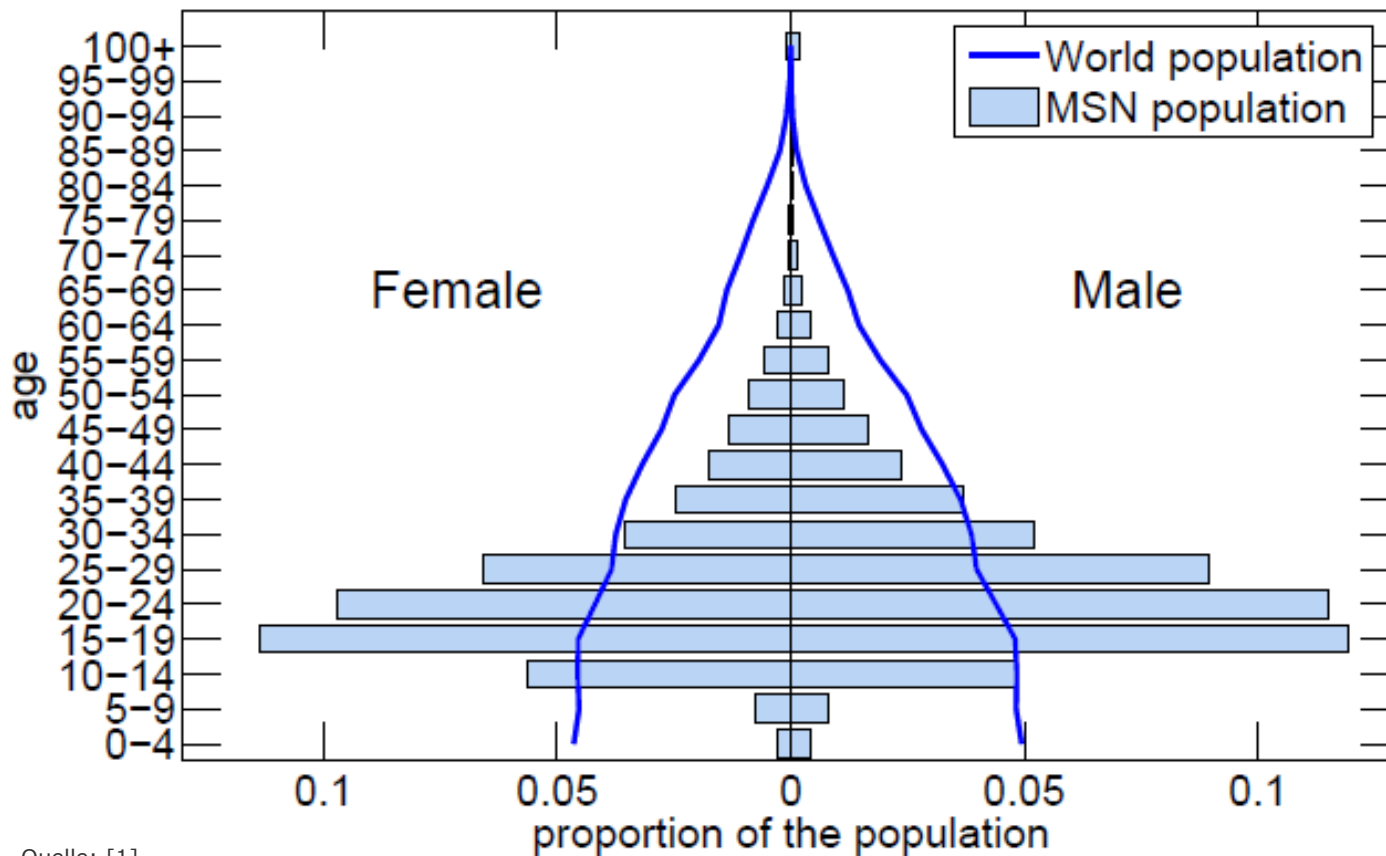
Quelle: [1]





-Benutzer

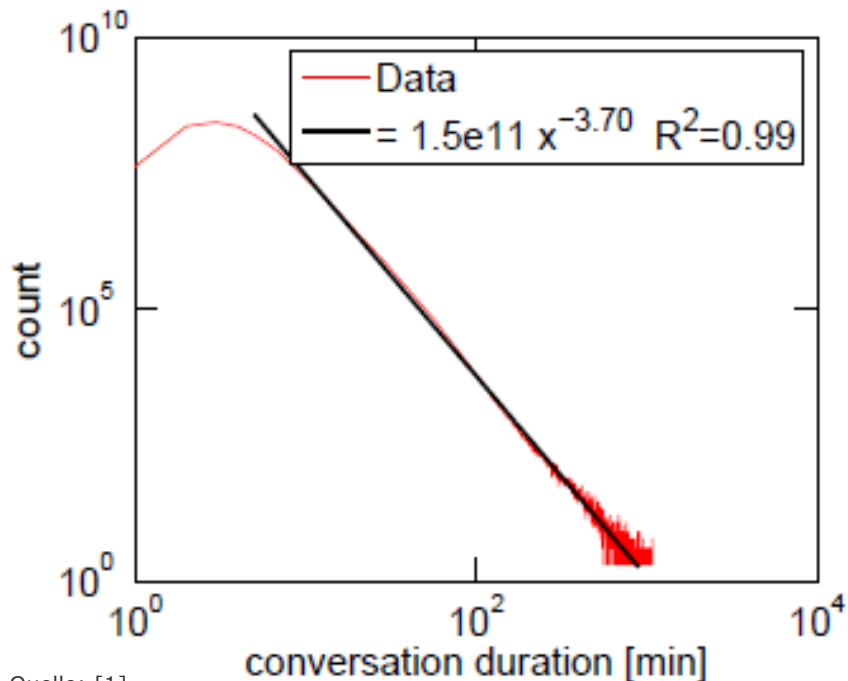
- Jung
- Eher männlich, aber Ausreißer bei weiblich 10-14



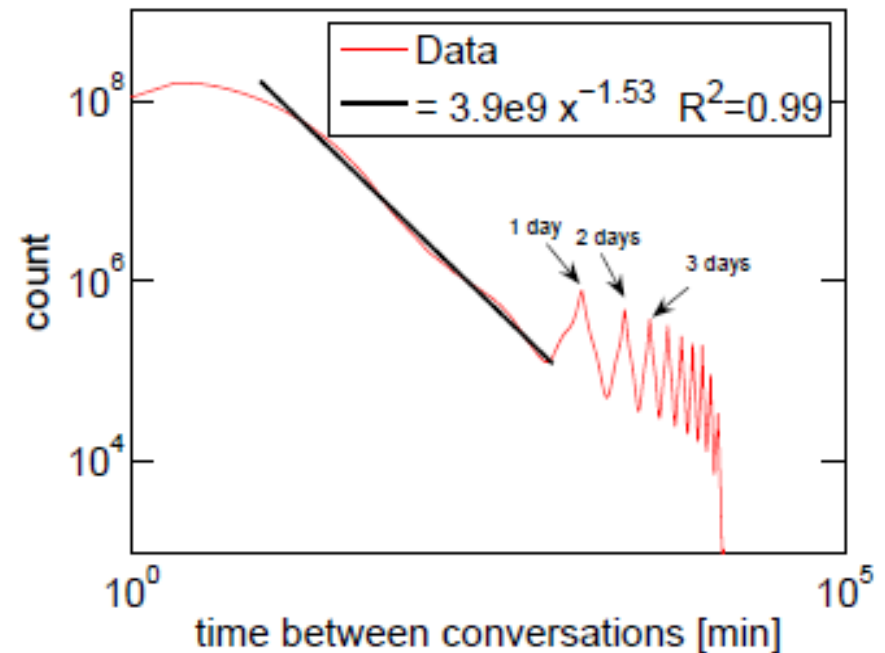
Quelle: [1]

Kommunikationscharakteristik

- Daten beschränkt auf zwei Benutzer Sitzungen (99%)
- Konversationsdauer nicht ganz „power-law“
 - Aber trotzdem „heavy-tailed“
- Ähnlich Zeit zwischen Konversationen

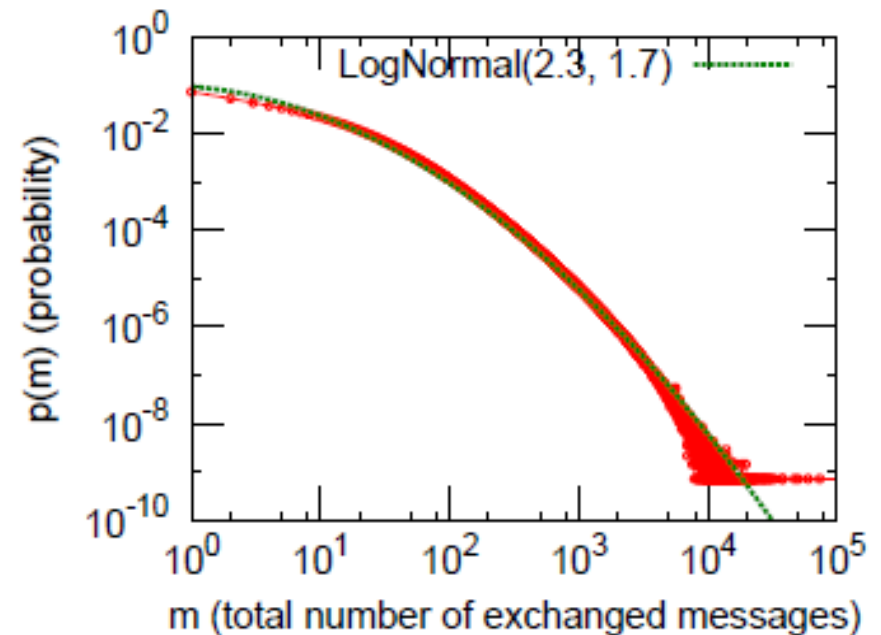
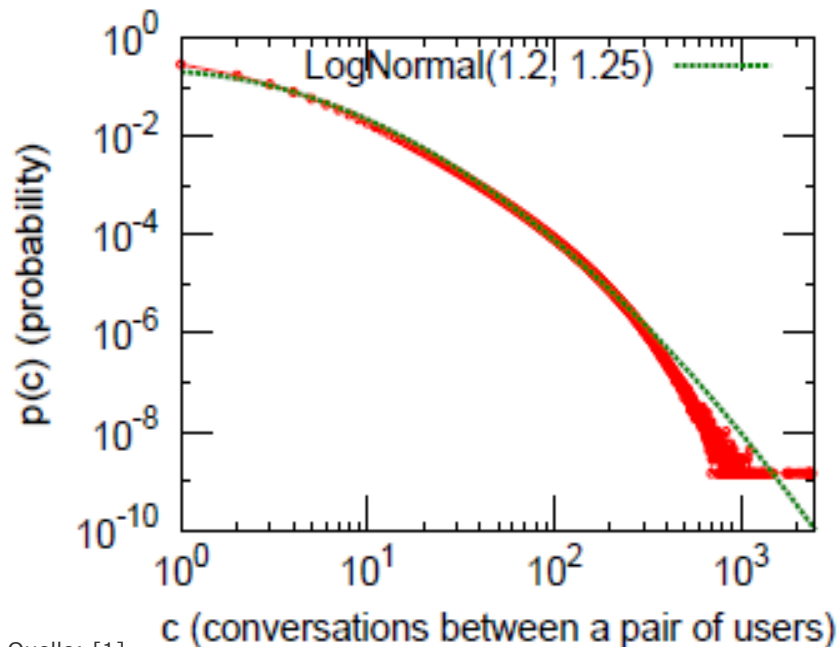


Quelle: [1]

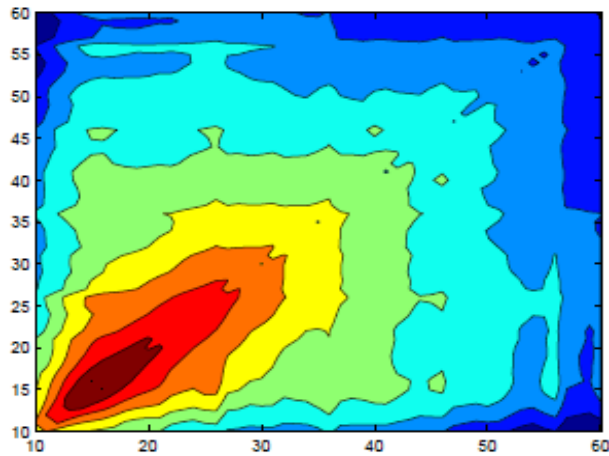


Kommunikationscharakteristik

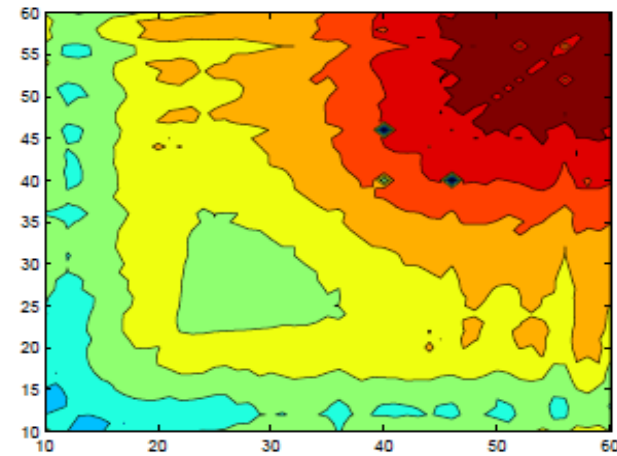
- Anzahl Konversationen zwischen zwei Benutzern und Anzahl ausgetauschter Nachrichten (Wahrscheinlichkeit)
 - Logarithmische Normalverteilung (auch „heavy-tailed“)



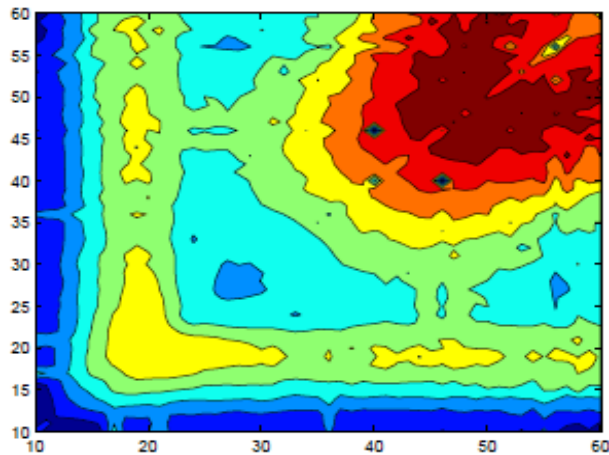
Quelle: [1]



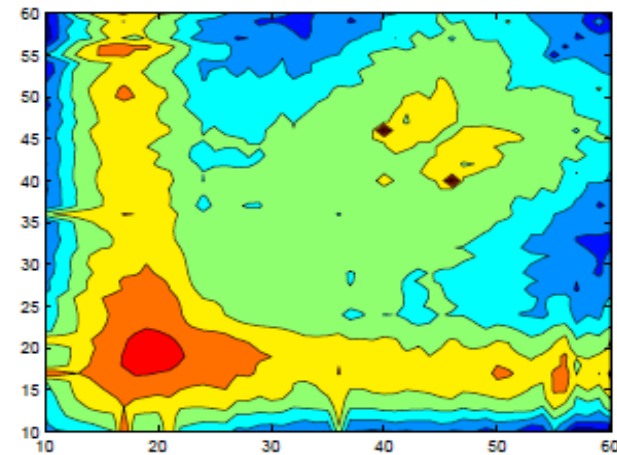
(a) Number of conversations



(b) Conversation duration



(c) Messages per conversation
Quelle: [1]



(d) Messages per unit time

Korrelation Geschlecht-Kommunikation

- Achtung: a) entspricht ungefähr Erwartungswert zufälliger Auswahl
 - Zeigt aber Tendenz Kreuzgeschlechtkonversation
- Aber: restliche Werte deuten hin auf
 - „sich etwas mehr Mühe geben“

Freiwillige Selbstangabe

	Unknown	Female	Male
Unknown	1.3	3.6	3.7
Female		21.3	49.9
Male			20.2

(a) Conversations

	Unknown	Female	Male
Unknown	277	301	277
Female		275	304
Male			252

(b) Conversation duration [sec]

	Unknown	Female	Male
Unknown	5.7	7.1	6.7
Female		6.6	7.6
Male			5.9

(c) Exchanged messages per conversation

Quelle: [1]

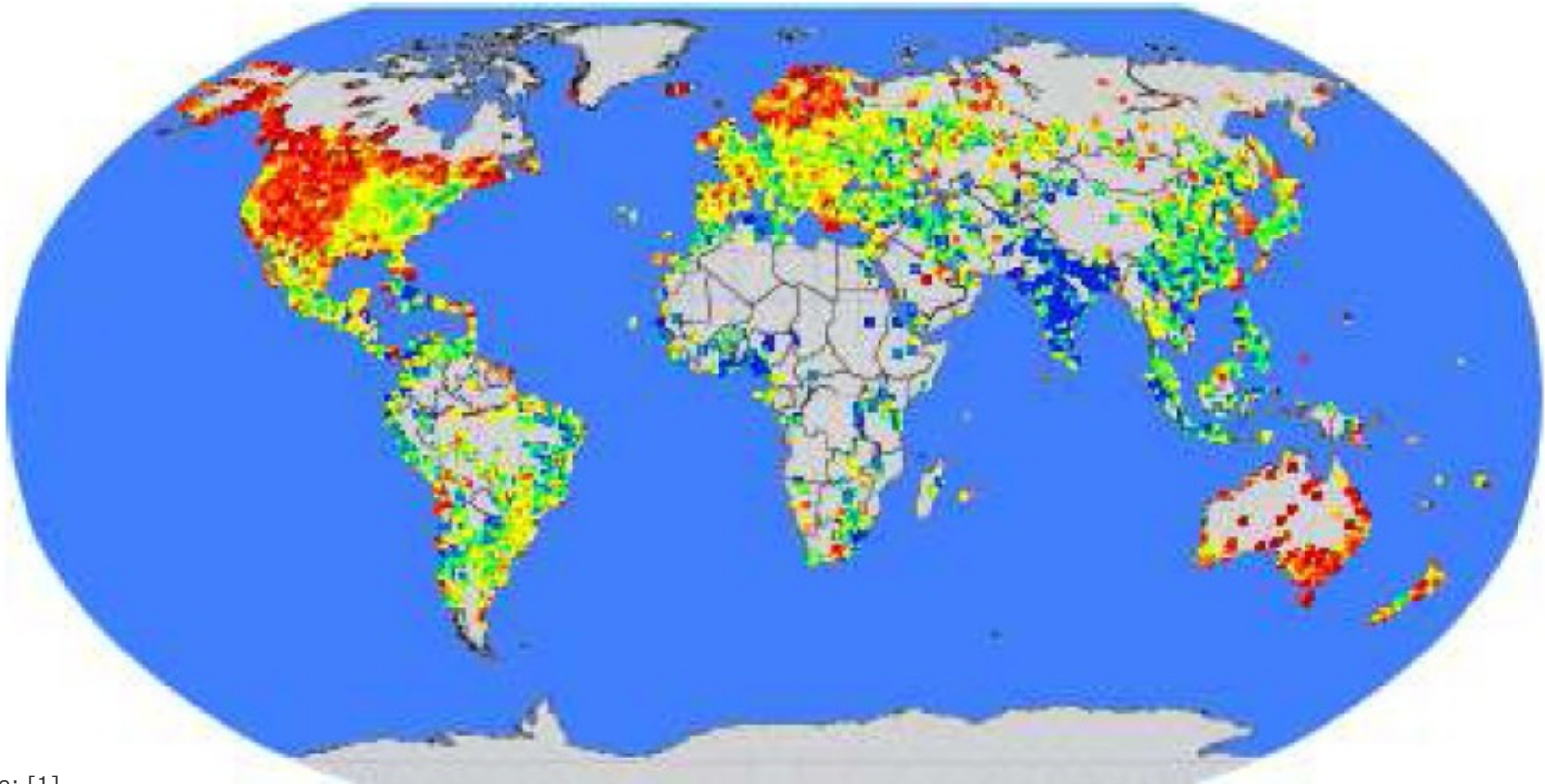
	Unknown	Female	Male
Unknown	1.25	1.42	1.38
Female		1.43	1.50
Male			1.42

(d) Conversation intensity

[Msg per min]

Korrelation: Ort-Kommunikation

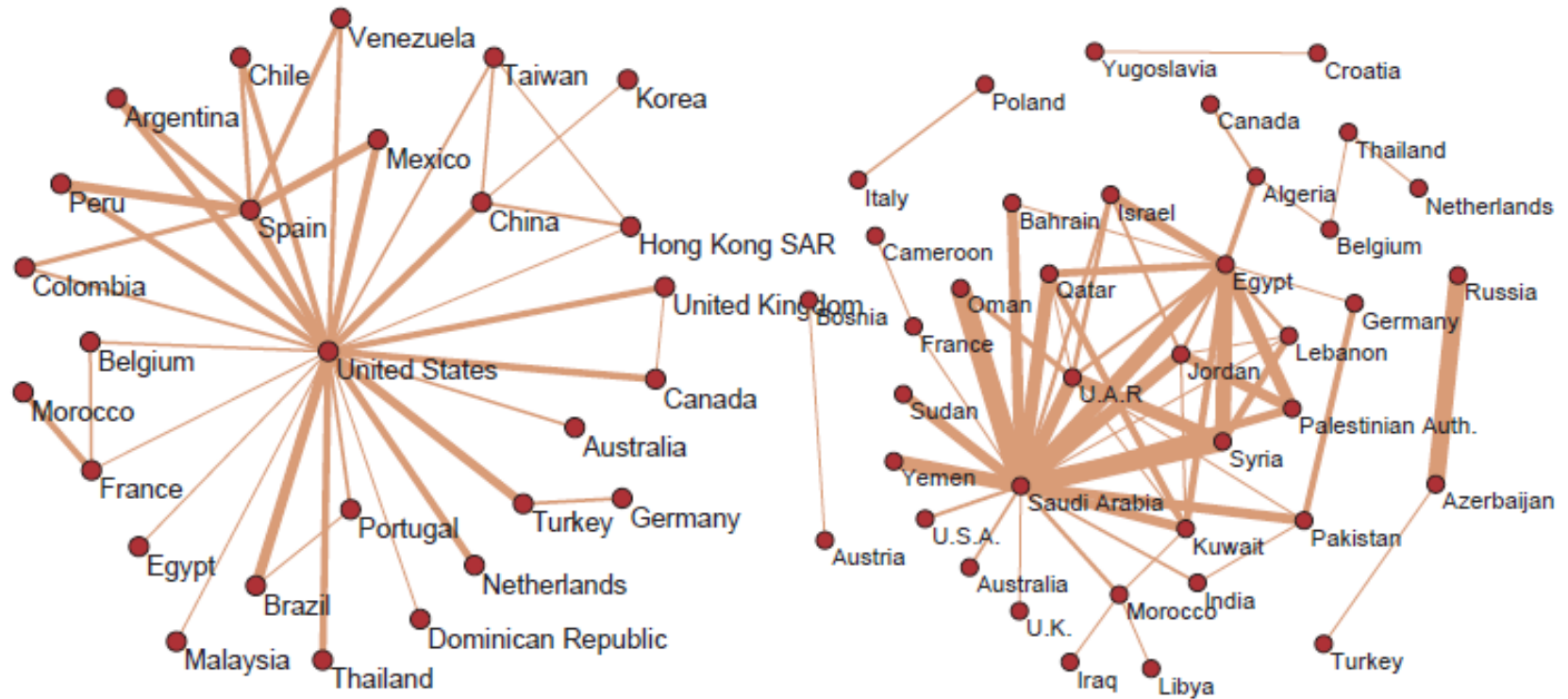
-Anzahl logins (log) pro Kopf pro Koordinate (IP basiert)



Quelle: [1]

Korrelation: Ort-Kommunikation

- Links: Anzahl Konversationen
- Rechts: Durchschnittliche Konversationslänge



Quelle: [1]

Korrelation: Ort-Kommunikation

- Links: Benutzerzahl zur Gesamtbevölkerung
- Rechts: Benutzeraktivität

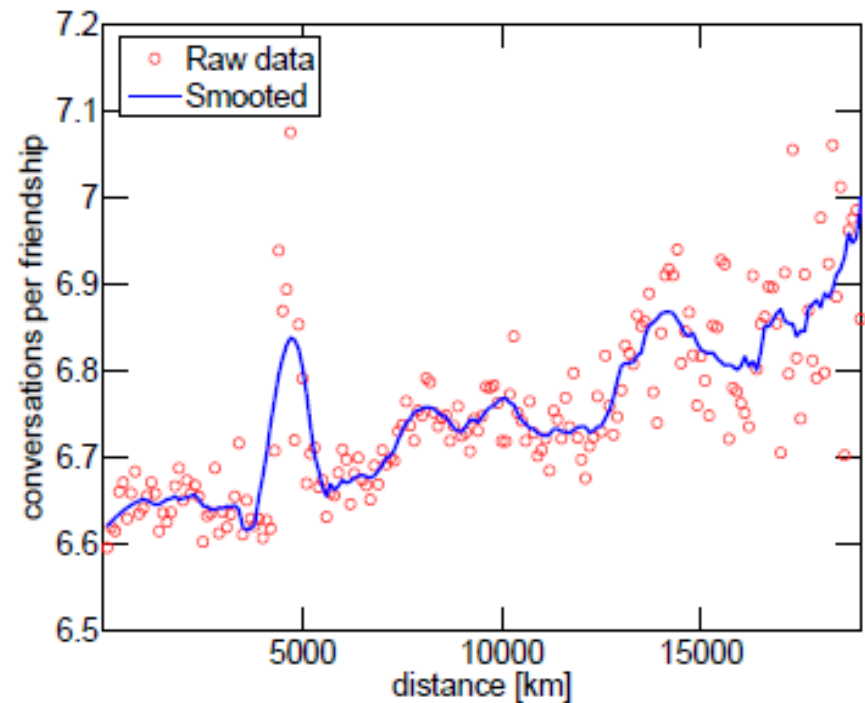
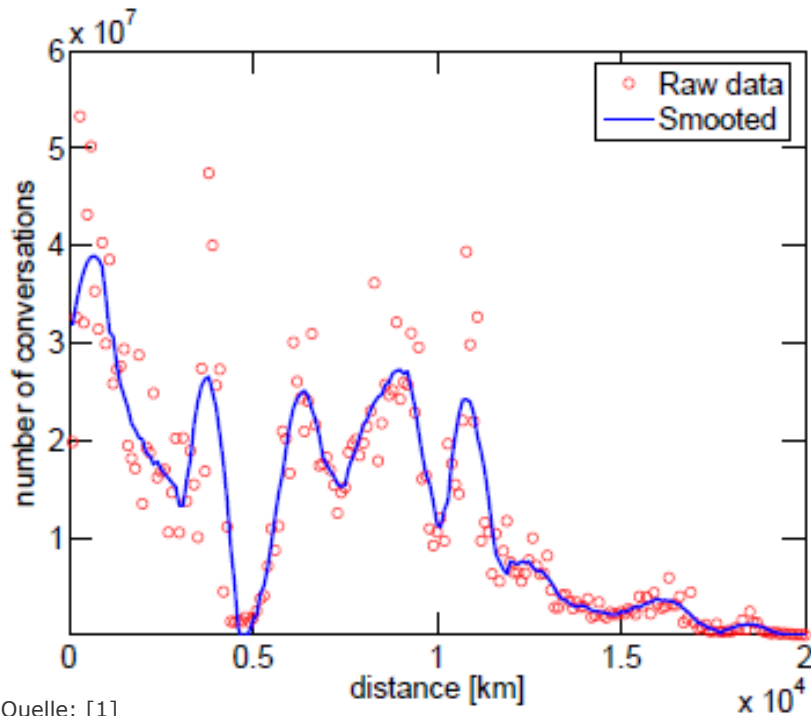
Country	Fraction of population
Iceland	0.35
Spain	0.28
Netherlands	0.27
Canada	0.26
Sweden	0.25
Norway	0.25
Bahamas, The	0.24
Netherlands Antilles	0.24
Belgium	0.23
France	0.18
United Kingdom	0.17
Brazil	0.08
United States	0.08

Country	Conversations per user per day
Afghanistan	4.37
Netherlands Antilles	3.79
Jamaica	2.63
Cyprus	2.33
Hong Kong	2.27
Tunisia	2.25
Serbia	2.15
Dominican Republic	2.06
Bulgaria	2.07

Quelle: [1]

Korrelation: Ort-Kommunikation

- Links: Anzahl Konversationen über Distanz absolut
 - Max bei 500km
 - Fällt mit Distanz
- Rechts: Konversationen pro Verbindung über Distanz
 - „Öfter mit einem weit entfernten Freund, als mit einem nahen.“



Quelle: [1]

Homophilie

-Wahrscheinlichkeit, dass zwei Knoten ein Attribut teilen

- Rnd: bei zufällig gewählten Knoten (gleichverteilt)
- Comm: bei Knoten, die kommuniziert haben

-Benutzer „lieben“:

- Gleiche Sprache (stark)
- Gleiches Land (stark)
- Gleicher Ort (mittel)
- Gleiches Alter (mittel)
- Anderes Geschlecht (leicht)

Attribute	Probability	
	Rnd	Comm
Age	0.030	0.162
Gender	0.434	0.426
ZIP	0.001	0.23
County	0.046	0.734
Language	0.030	0.798

Quelle: [1]

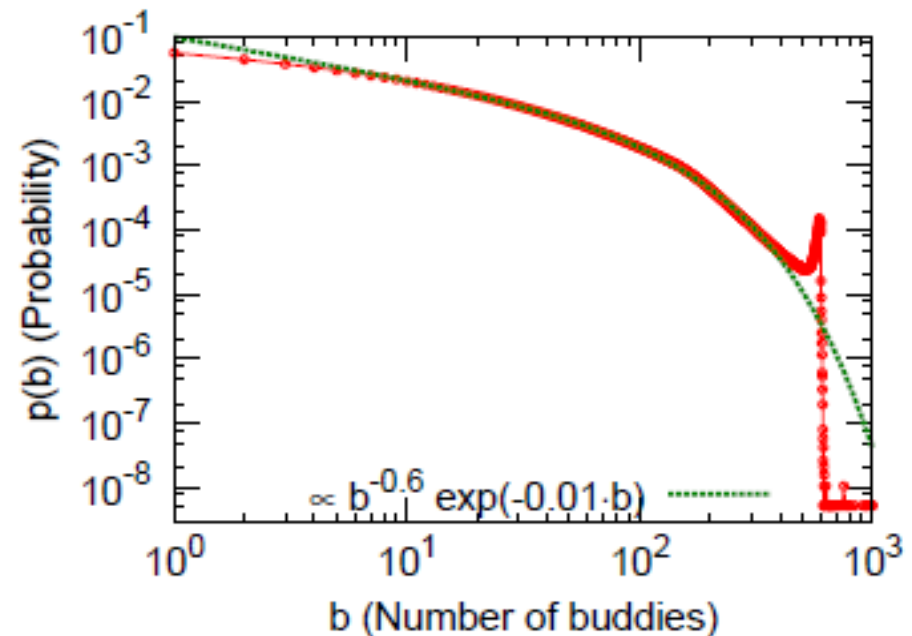
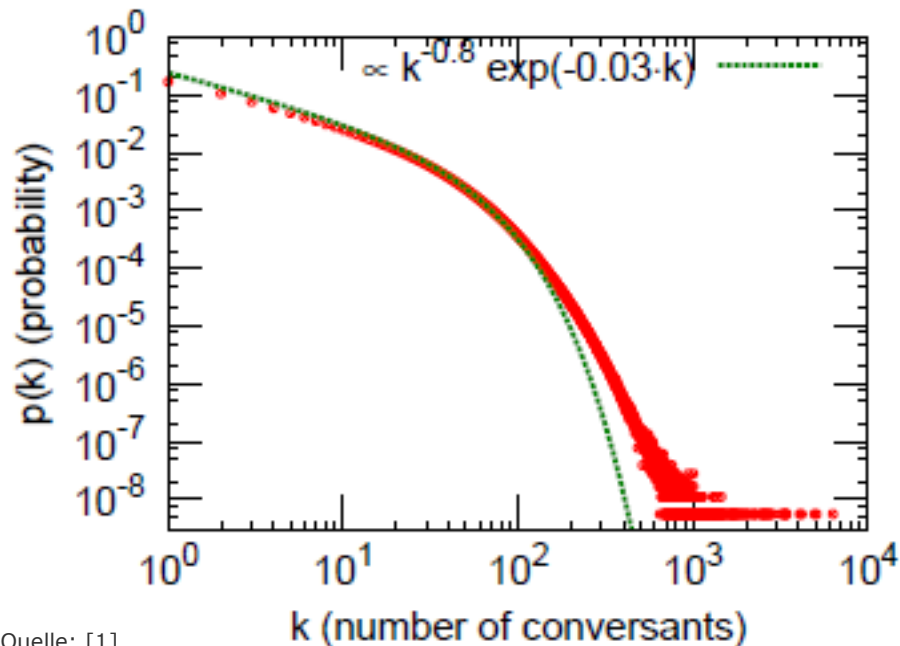
Das Netzwerk: Knotengrade

-Links: Kommunikationsgraph

- Dichtefunktion Knotengrade k (Anzahl Kommunikationspartner)
 - „power-law“ mit exponentiellem Abschluss (Erinnerung: Erdbeben)

-Rechts: Anzahl Freunde auf der Kontaktliste

- Lokales Maximum bei 600 \rightarrow maximale Listenlänge



Quelle: [1]

Das Netzwerk: Clustering

-Clustering-Koeffizient (c)

- Erinnerung: c_{Knoten}

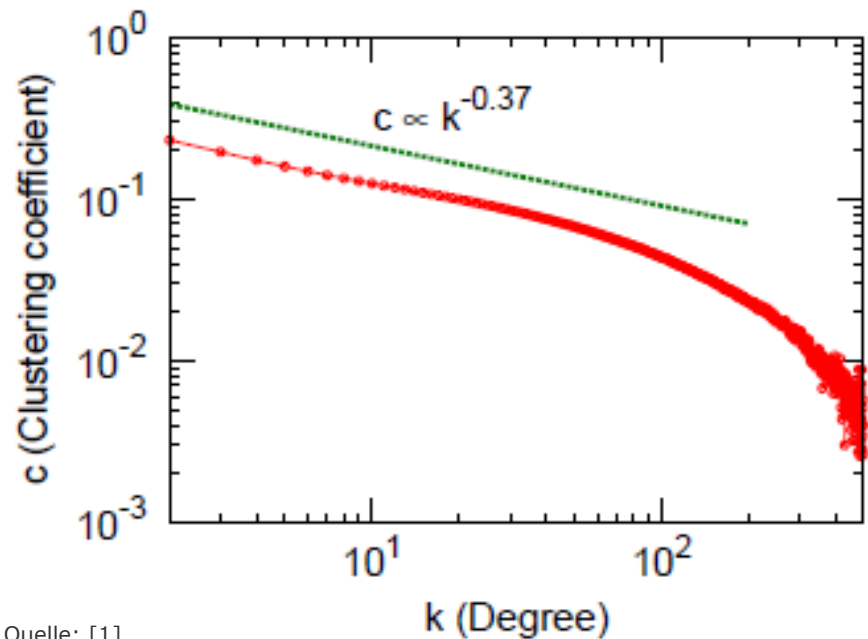
- Verbindungsgrad der Nachbarn im Verhältnis zum möglichen Verbindungsgrad (Clique hat $c=1$)

-Bild: c über Knotengrade k

- Man erwartet sinken mit steigendem Knotengrad (seltener Große Cluster)

- Aber: andere Arbeiten finden

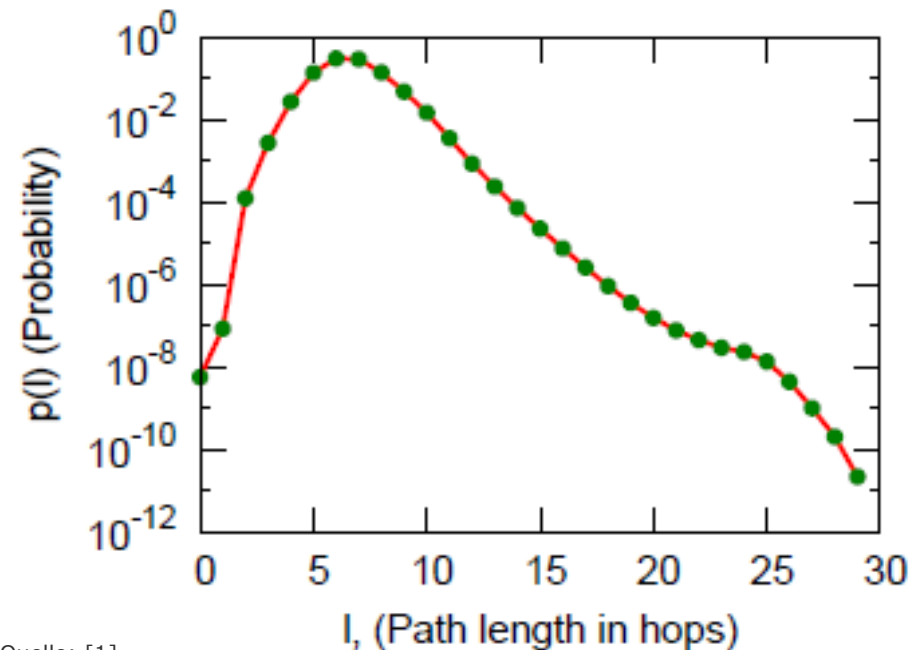
- $c \sim k^{-1}$
- Hier: Potenz relativ klein
 - Netz ist sehr stark Verbunden



Quelle: [1]

Das Netzwerk: Small World

- Durchmesser=29 (Längster kürzester Pfad)
- Modus=6 (maximale Wahrscheinlichkeit)
- Median=7 (Fläche links=Fläche rechts, Achtung: log-y)
- Mittelwert=6.6 (Geodätische Distanz)
- „7 degrees of separation“ (aber nur knapp)



Quelle: [1]

Das Netzwerk: Robustheit

-Größe Komponente bei entfernen von Knoten nach bestimmten Prioritäten

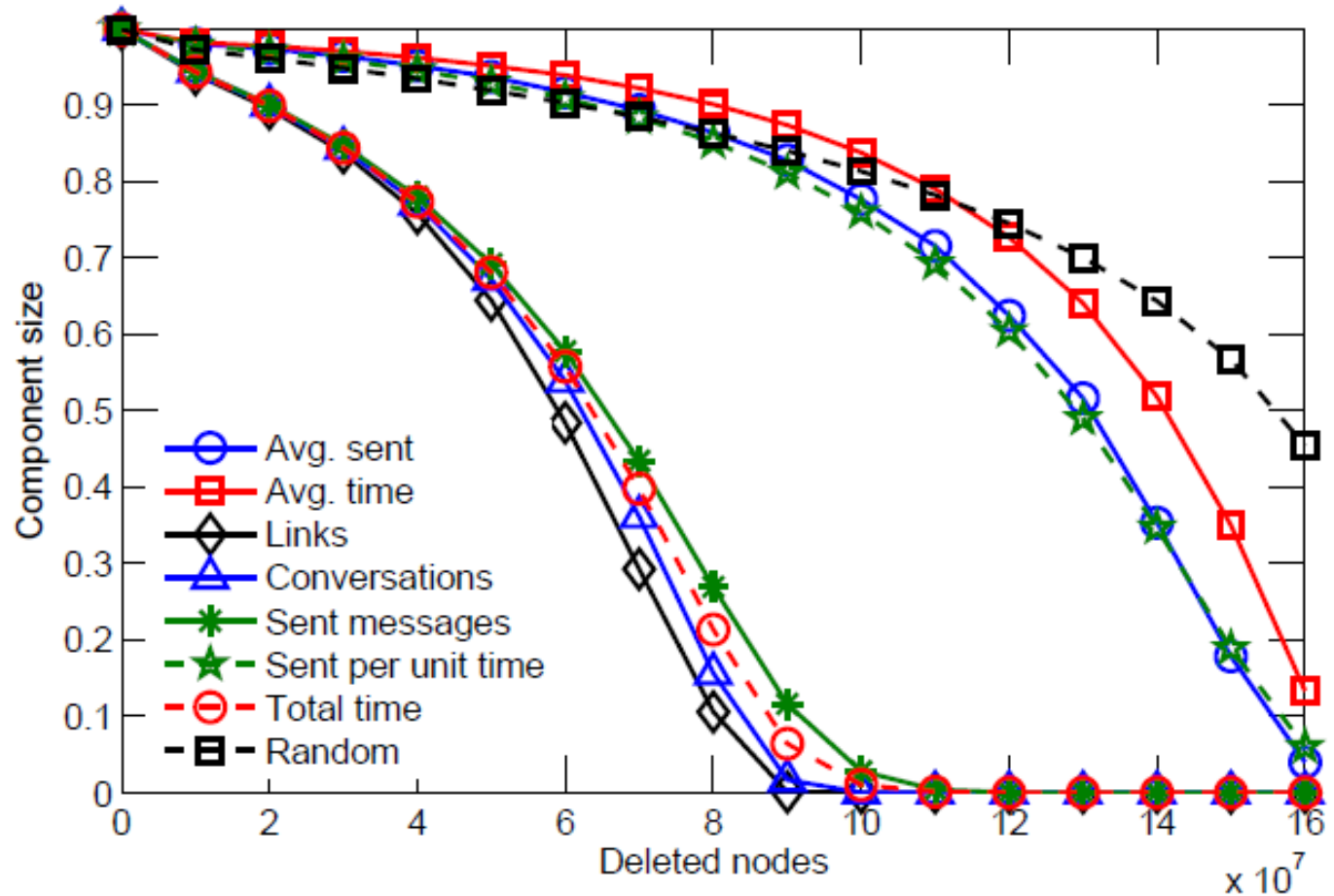
-Anfälligkeit

- Stark

- Grad

- Schwach

- Zufall



-Kommunikationsnetzwerk ist

- stark Verbunden und Robust (Zufallslöschung)
- Homophil außer beim Geschlecht
- „6 degrees of separation“ kann fast voll bestätigt werden

Milgram's Experimente (1960er)

Erste Experimente zur Untersuchung auf „Small World Phänomen“

1967:

- 60 zufällig ausgewählte Teilnehmer
- Eine festgelegte Zielperson in Boston
- 3 Pakete erreichten das Ziel
- Durchschnittliche Pfadlänge: 5,5

1969:

- 296 Sender in Nebraska und Boston, Zielperson in Massachusetts
- 64 Briefe haben ihr Ziel erreicht (20%)
- Durchschnittliche Pfadlänge: 6,5 → „*six degrees of separation*“

1970:

- Untersuchung auf mögliche Grenzen zwischen ethnischen Gruppen
- 270 gestartete Pakete
- 13% von 270 erreichten afro-amerikanische Zielperson
- 33% von 270 erreichten 'weiße' Zielperson

Milgram's Experiment - Kritik

- Wenige erfolgreiche Abschlüsse der Kette
(bei längeren Ketten ist die Abbruchwahrscheinlichkeit größer)
- Schwierigkeiten bei der link-Auswahl
- Isolierte Communities

Dodds' Experiment (2003)

(Peter Sheridan Dodds, Roby Muhamad, Duncan J. Watts)

- Globales E-Mail-Experiment at Columbia University
- **18** Zielpersonen aus **13** Länder
- Daten für mehr als **60.000** Teilnehmer aus **166** Länder gesammelt
- **25%** von 98.847 registrierten Personen, haben Ketten initiiert
- Partizipationsrate nach dem ersten Schritt: **37%**
- Insgesamt **24.163** Ketten
- **384** Ketten waren erfolgreich
- Teilnehmer mussten sich auf der Webseite des Experiments registrieren und haben zusätzliche Informationen angegeben, wie: geografische Lage, Geschlecht und optional: Alter, Einkünfte, Ausbildung, Beschäftigung, Religion

Dodds' Experiment - Zielpersonen (1)

Target	City	Country	Occupation	Gender	<i>N</i>	<i>N_c</i> (%)	<i>r</i> (<i>r₀</i>)	< <i>L</i> >
1	Novosibirsk	Russia	PhD student	F	8234	20(0.24)	64 (76)	4.05
2	New York	USA	Writer	F	6044	31 (0.51)	65 (73)	3.61
3	Bandung	Indonesia	Unemployed	M	8151	0	66 (76)	n/a
4	New York	USA	Journalist	F	5690	44 (0.77)	60 (72)	3.9
5	Ithaca	USA	Professor	M	5855	168 (2.87)	54 (71)	3.84
6	Melbourne	Australia	Travel Consultant	F	5597	20 (0.36)	60 (71)	5.2
7	Bardufoss	Norway	Army veterinarian	M	4343	16 (0.37)	63 (76)	4.25
8	Perth	Australia	Police Officer	M	4485	4 (0.09)	64 (75)	4.5
9	Omaha	USA	Life Insurance Agent	F	4562	2 (0.04)	66 (79)	4.5

Quelle: [4]

N – Anzahl Personen, die der entsprechenden Zielperson zugeordnet wurden

N_c – Anzahl erfolgreicher Ketten

r₀ – Bruchteil von Personen, die sich registriert haben, aber keine Nachrichten weitergeleitet haben

r – mittlerer Bruchteil von abgebrochenen Ketten (keine Weiterleitung nach dem ersten Schritt)

<L> - mittlere Pfadlänge von erfolgreichen Ketten

Dodds' Experiment - Zielpersonen (2)

10	Welwyn Garden City	UK	Retired	M	6593	1 (0.02)	68 (74)	4
11	Paris	France	Librarian	F	4198	3 (0.07)	65 (75)	5
12	Tallinn	Estonia	Archival Inspector	M	4530	8 (0.18)	63 (79)	4
13	Munich	Germany	Journalist	M	4350	32 (0.74)	62 (74)	4.66
14	Split	Croatia	Student	M	6629	0	63 (77)	n/a
15	Gurgaon	India	Technology Consultant	M	4510	12 (0.27)	67 (78)	3.67
16	Managua	Nicaragua	Computer analyst	M	6547	2 (0.03)	68 (78)	5.5
17	Katikati	New Zealand	Potter	M	4091	12 (0.3)	62 (74)	4.33
18	Elderton	USA	Lutheran Pastor	M	4438	9 (0.21)	68 (76)	4.33
Totals					98,847	384 (0.4)	63 (75)	4.05

Quelle: [4]

- Mittlere Pfadlänge: 4,05

Dodds' Experiment - Beobachtungen

- Männer haben Nachrichten öfter an andere Männer geschickt (57%), analog bei Frauen (61%)

	Nature of relationship	N_i	N_c	f_i	f_c	Δ
<ul style="list-style-type: none"> • Teilnehmer haben ihre Freunde bevorzugt, dann Business- und Familienbeziehungen • Erfolgreiche Ketten: Business- vor Freundschaftsbeziehungen 	Friend	22358	700	64.7	50.7	-13.9
	Relatives	3457	64	10.0	4.6	-5.4
	Sibling	1774	28	5.1	2.0	-3.1
	Spouse/Significant other	1238	33	3.6	2.4	-1.2
	Customer	79	8	0.2	0.6	+0.4
	Service provider	145	12	0.4	0.9	+0.5
	Business partner	234	20	0.7	1.4	+0.8
	Client	137	17	0.4	1.2	+0.8
	Junior	336	26	1.0	1.9	+0.9
	Other	1179	87	3.4	6.3	+2.9
	Senior	543	86	1.6	6.2	+4.7
	Co-worker	3103	299	9.0	21.7	+12.7

i - nicht-komplette Ketten

c - komplette Ketten

N - Frequenz der Kategorien

f - relative Frequenz

$$\Delta = f_{c,x} - f_{i,x}$$

Quelle: [4]



Dodds' Experiment - Weiterleitung von Nachrichten

Wahl des nächsten Empfängers:

<i>L</i>	<i>N</i>	Location	Travel	Family	Work	Education	Friends	Cooperative	Other
1	19,718	33	16	11	16	3	9	9	3
2	7,414	40	11	11	19	4	6	7	2
3	2,834	37	8	10	26	6	6	4	3
4	1,014	33	6	7	31	8	5	5	5
5	349	27	3	6	38	12	6	3	5
6	117	21	3	5	42	15	4	5	5
7	37	16	3	3	46	19	8	5	0

Quelle: [4]

- Strategie: geografische Nähe zur Zielperson, Berufsähnlichkeit
- Geografische Nähe weniger zitiert nach dem 3. Schritt; Beruf wird relevanter
- Stark verbundene Teilnehmer (Hubs) haben wenig Relevanz



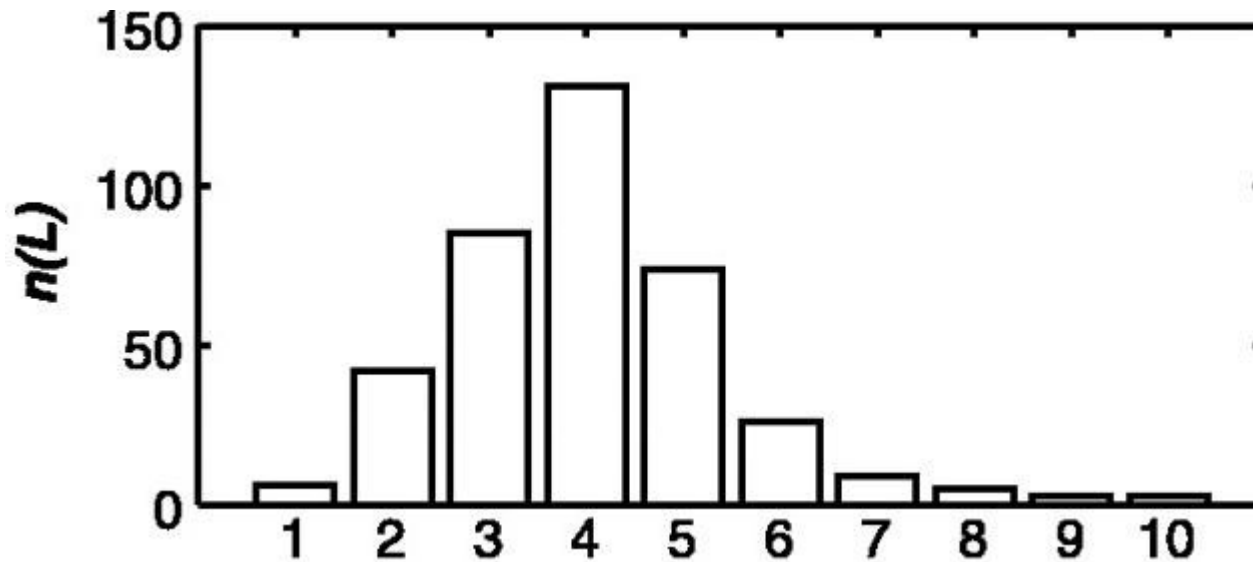
Dodds' Experiment - Weiterleitung von Nachrichten

How initially met acquaintance	N_i	N_e	Δ	Reason for choosing link	N_i	N_e	Δ
Immediate Family	4358	80	-6.8	Geographic	10825	183	-14.3
Internet	2189	44	-3.1	Travelled to target's location	4257	38	-9.5
Extended Family	2043	41	-2.9	Continue the chain	2477	6	-7.4
Grew up together	1269	13	-2.7	Lots of friends	2515	14	-6.6
School	2077	48	-2.5	Family origin	3331	58	-4.2
Friend of Family	1593	42	-1.6	Other	839	51	+3.1
Live(d) in same Neighborhood/Roommate	994	22	-1.3	Similar education	1147	65	+3.7
Hobby/Club	1197	32	-1.1	Work	2791	129	+5.7
Travel	645	11	-1.1	Similar profession	2449	325	+29.4
Mutual Friend	3173	113	-1.0	Strength	N_i	N_e	Δ
Other	542	14	-0.6	Extremely close	6628	123	-10.3
Place of worship	559	15	-0.5	Very close	7844	177	-9.9
Sport	245	7	-0.2	Fairly close	11366	433	-1.5
University/College	5320	321	+7.9	Casually	7507	516	+15.7
Work	8381	577	+17.6	Not close	1239	131	+5.9

Quelle: [4]



Verteilung von Kettenlängen:

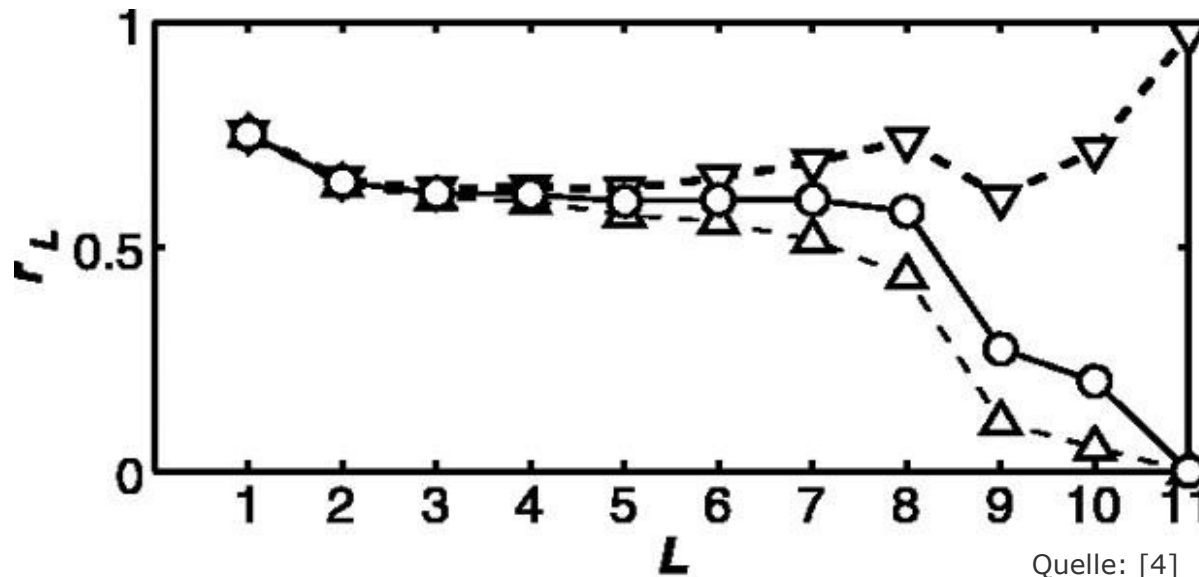


Quelle: [4]

$n(L)$ – Anzahl der Ketten, die in L Schritten erfolgreich abgeschlossen wurden

Attrition rate (Abwanderungsquote)

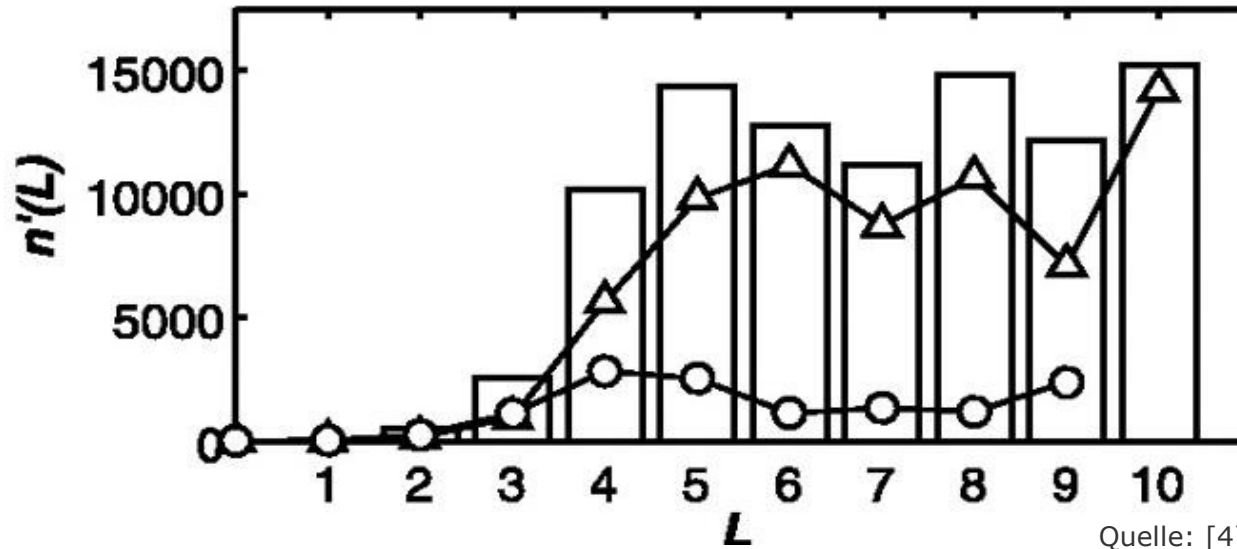
= Anzahl Teilnehmer, die nicht mehr mitmachen /
Anzahl aller Teilnehmer
(über eine gegebene Zeitspanne)



r_L – mittlere Attrition rate pro Schritt (Kreise)

Dreiecke: Konfidenzintervall (Vertrauensbereich)

Dodds' Experiment Ermittlung der mittleren Pfadlänge (3)



$$\mathbf{n'(L)} = n(L) / \prod_{i=0..L-1}(1-r_i) \quad (\text{ideale Verteilung von Pfadlängen})$$

Mittlere Pfadlänge:

$L^* = 5$ (Start und Ziel im gleichen Land)

$L^* = 7$ (Start und Ziel in unterschiedlichen Ländern)

$$\mathbf{5 \leq L^* \leq 7}$$

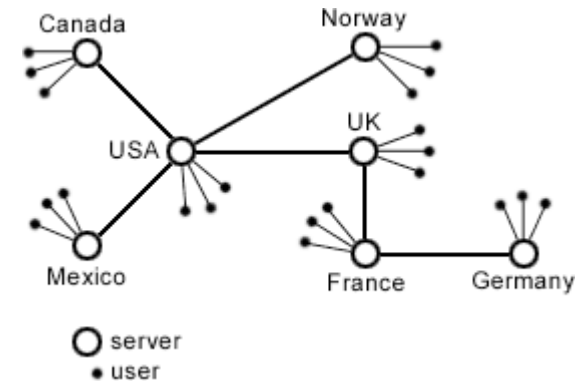
Dodds' Experiment - Zusammenfassung

Erfolgreiche Suche:

- Mittel-starke bis schwache Beziehungen
- Professionelle Beziehungen
- Hubs nicht notwendig
- Mittlere Pfadlänge: 5 bis 7 Schritte
- Motivation der Teilnehmer ist entscheidend
 - Der Rückgang von Attrition rate um 15% kann die Erfolgsquote um 800% erhöhen!

IRC (Internet Relay Chat)

- Textbasiertes Chatsystem
- Entwickelt in 1988, ab 1993 weitere IRC-Netzwerke
- Gruppenchat in Kanälen
+ private Gespräche
- IRC-Netzwerke bestehen aus mehreren IRC-Servern
 - **IRCNet, EFNet, QuakeNet, UnderNet, DALNet, Freenode..**
 - große Netzwerke haben gleichzeitig mehr als 100.000 verbundene Clients
- Viele IRC-Clients: **mIRC, ChatZilla, Miranda, BitchX, XChat...**
- Kommunikation mit einem Server:
`/join #channel, /whois Nickname, /msg Nickname <message>`



Beispiel für ein IRC-Netzwerk, Quelle: [6]

Wie kann man einen Kommunikationsgraphen eines IRC-Kanals erstellen?

IRC - Ableitung des Graphen eines IRC-Kanals

Bestimmung von Knoten: trivial
(einfach alle User eines IRC-Kanals nehmen)

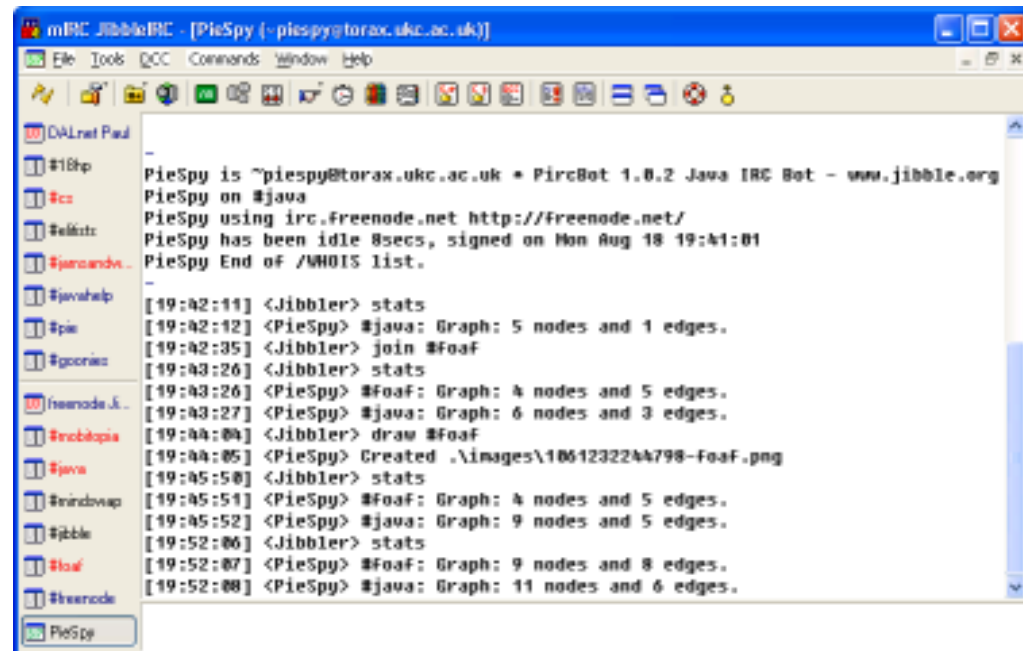
Bestimmung von Kanten (Beziehungen):

- direkte Addressierung

```
<Dave> Can someone ping me?
<Phil> Dave: Okay.
```
- Zeitbasierte Analyse (Temporal Proximity)
- Analyse der Konversationsdichte (Temporal Density)
- Überwachung von privaten Gesprächen (schwierig)

PieSpy - <http://www.jibble.org/piespy/>

- In Java geschriebenes IRC-Bot zur Überwachung mehrerer IRC-Kanäle (basiert auf PircBot – IRC-Bot Framework)
- GNU Lizenz / kommerzielle Lizenz
- Erstellung von Netzwerkgraphen (PNG-Grafiken) anhand Heuristiken aus der vorherigen Folie



```
mIRC JibbleIRC - [PieSpy (-piespy@torax.ukc.ac.uk)]
File Tools QCC Commands Window Help
DALnet Paul
#18hp
#cc
#elitz
#jancandv...
#javahelp
#pie
#ponic
#ircnode.j...
#ircbopia
#java
#windwap
#jibble
#foaf
#ircnode
PieSpy
PieSpy is ~piespy@torax.ukc.ac.uk * PircBot 1.0.2 Java IRC Bot - www.jibble.org
PieSpy on #java
PieSpy using irc.freenode.net http://freenode.net/
PieSpy has been idle 8secs, signed on Mon Aug 18 19:41:01
PieSpy End of /WHOIS list.
[19:42:11] <Jibbler> stats
[19:42:12] <PieSpy> #java: Graph: 5 nodes and 1 edges.
[19:42:35] <Jibbler> join #foaf
[19:43:26] <Jibbler> stats
[19:43:26] <PieSpy> #foaf: Graph: 4 nodes and 5 edges.
[19:43:27] <PieSpy> #java: Graph: 6 nodes and 3 edges.
[19:44:04] <Jibbler> draw #foaf
[19:44:05] <PieSpy> Created .\images\1061232244798-Foaf.png
[19:45:50] <Jibbler> stats
[19:45:51] <PieSpy> #foaf: Graph: 4 nodes and 5 edges.
[19:45:52] <PieSpy> #java: Graph: 9 nodes and 5 edges.
[19:52:06] <Jibbler> stats
[19:52:07] <PieSpy> #foaf: Graph: 9 nodes and 8 edges.
[19:52:08] <PieSpy> #java: Graph: 11 nodes and 6 edges.
```

PieSpy - Kommandos

Benutzung von PieSpy durch das Senden von privaten Nachrichten an das Bot:

```
/msg PieSpy [password] join #channel
```

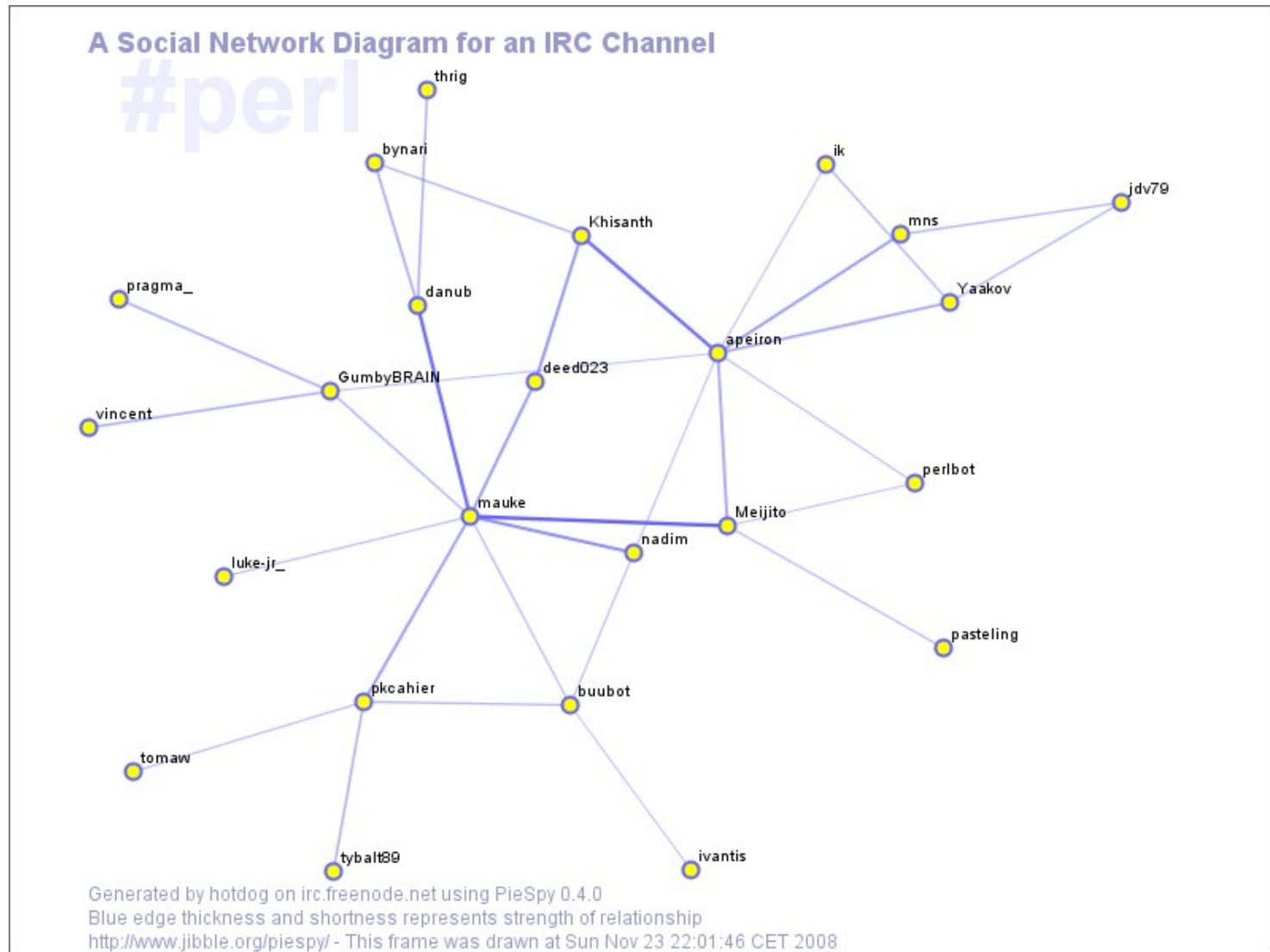
```
/msg PieSpy [password] part #channel
```

```
/msg PieSpy [password] stats
```

```
/msg PieSpy [password] draw #channel
```

```
/msg PieSpy [password] ignore nick
```

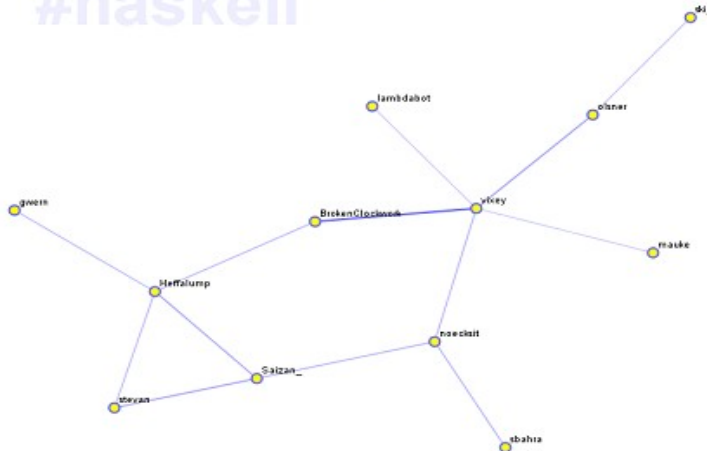
```
/msg PieSpy [password] raw [command]
```



PieSpy - #haskell

A Social Network Diagram for an IRC Channel

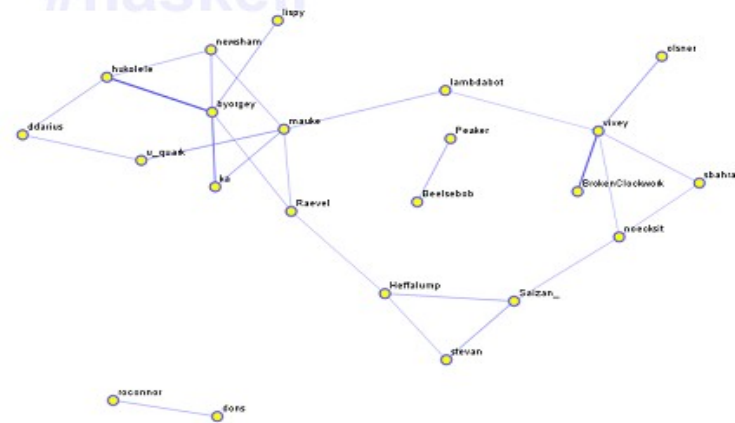
#haskell



Generated by hofdog on irc.freenode.net using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Sun Nov 23 19:26:32 CET 2008

A Social Network Diagram for an IRC Channel

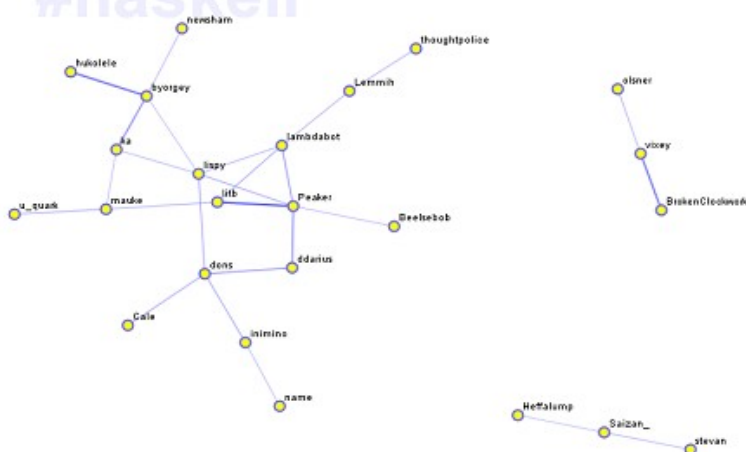
#haskell



Generated by hofdog on irc.freenode.net using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Sun Nov 23 20:05:58 CET 2008

A Social Network Diagram for an IRC Channel

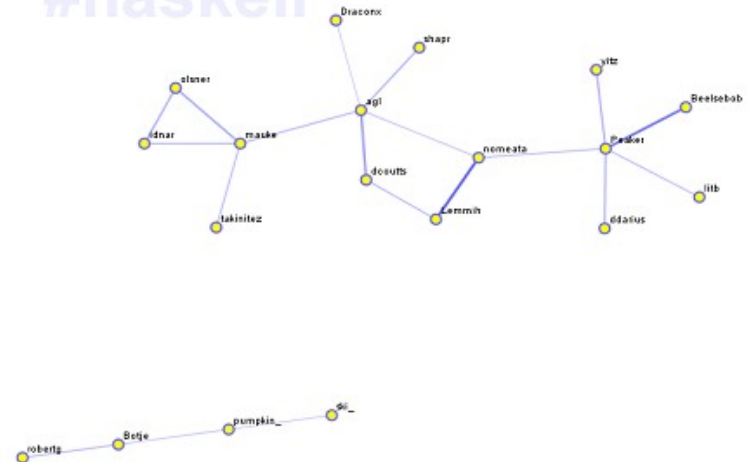
#haskell



Generated by hofdog on irc.freenode.net using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Sun Nov 23 20:28:13 CET 2008

A Social Network Diagram for an IRC Channel

#haskell

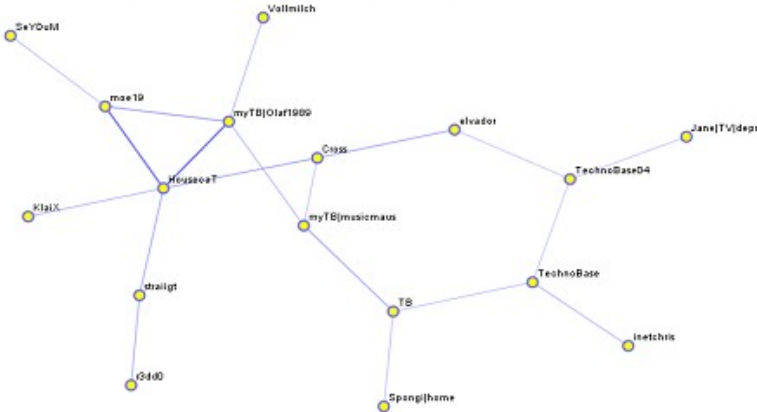


Generated by hofdog on irc.freenode.net using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Sun Nov 23 23:26:55 CET 2008

PieSpy - #TechnoBase

A Social Network Diagram for an IRC Channel

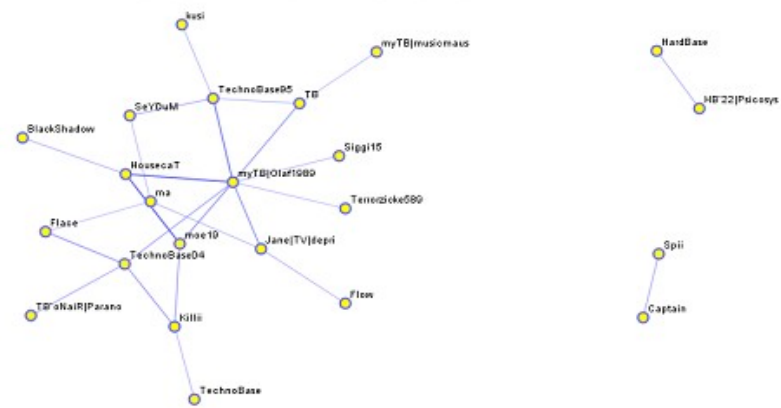
#TechnoBase



Generated by h0dd0g on irc.quakenet.org using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Tue Nov 19 21:08:49 CET 2008

A Social Network Diagram for an IRC Channel

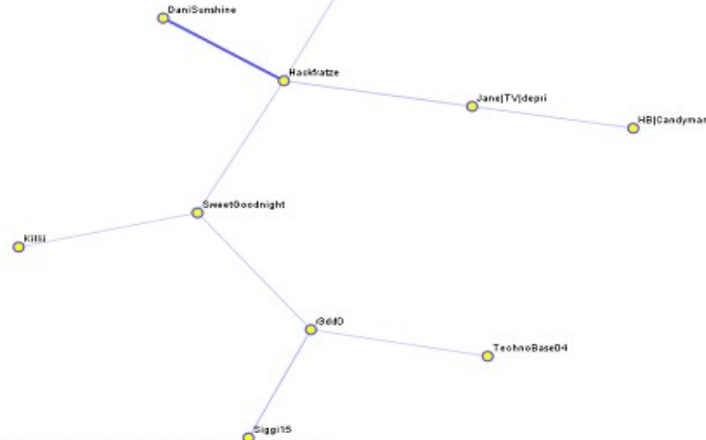
#TechnoBase



Generated by h0dd0g on irc.quakenet.org using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Tue Nov 18 21:31:45 CET 2008

A Social Network Diagram for an IRC Channel

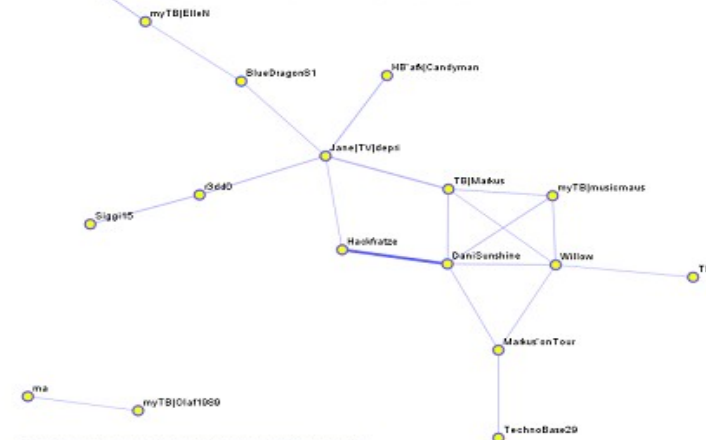
#TechnoBase



Generated by h0dd0g on irc.quakenet.org using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Tue Nov 18 22:58:54 CET 2008

A Social Network Diagram for an IRC Channel

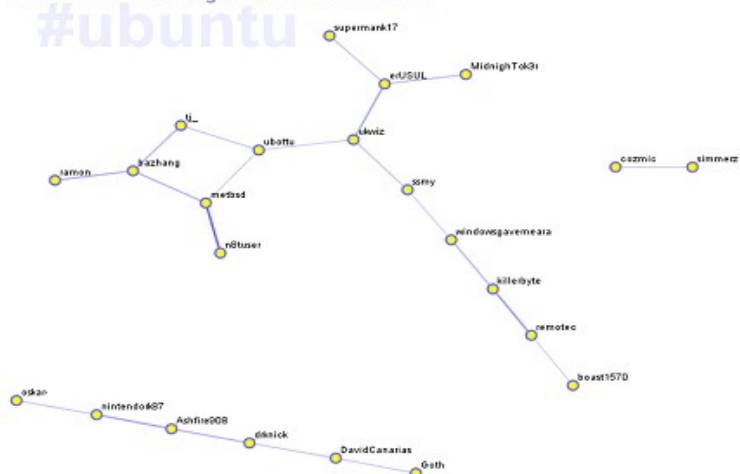
#TechnoBase



Generated by h0dd0g on irc.quakenet.org using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Tue Nov 18 23:18:04 CET 2008

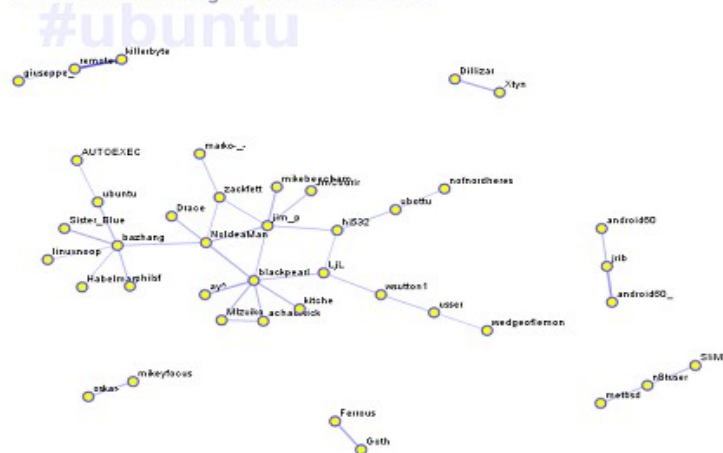
PieSpy - #ubuntu

A Social Network Diagram for an IRC Channel



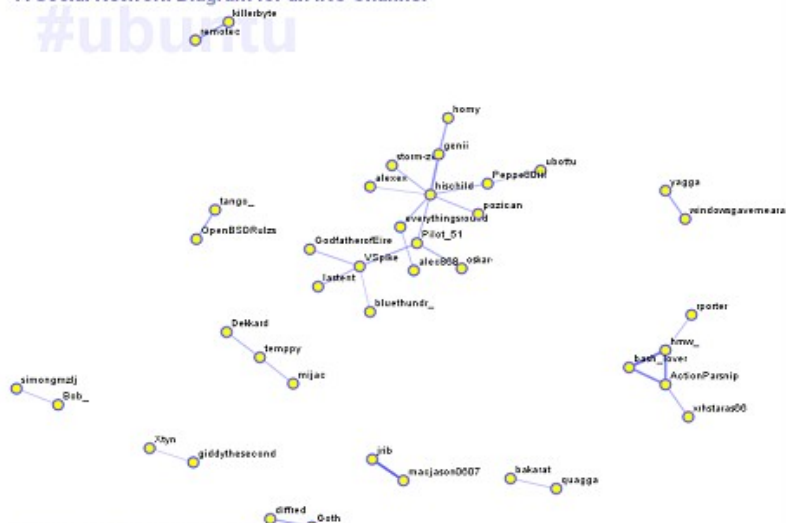
Generated by h0dd0g on irc.freenode.net using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Sun Nov 23 18:35:19 CET 2008

A Social Network Diagram for an IRC Channel



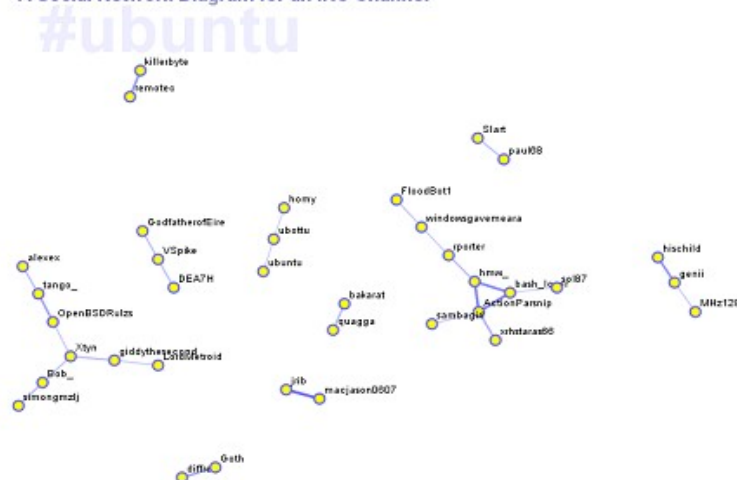
Generated by h0dd0g on irc.freenode.net using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Sun Nov 23 19:47:07 CET 2008

A Social Network Diagram for an IRC Channel



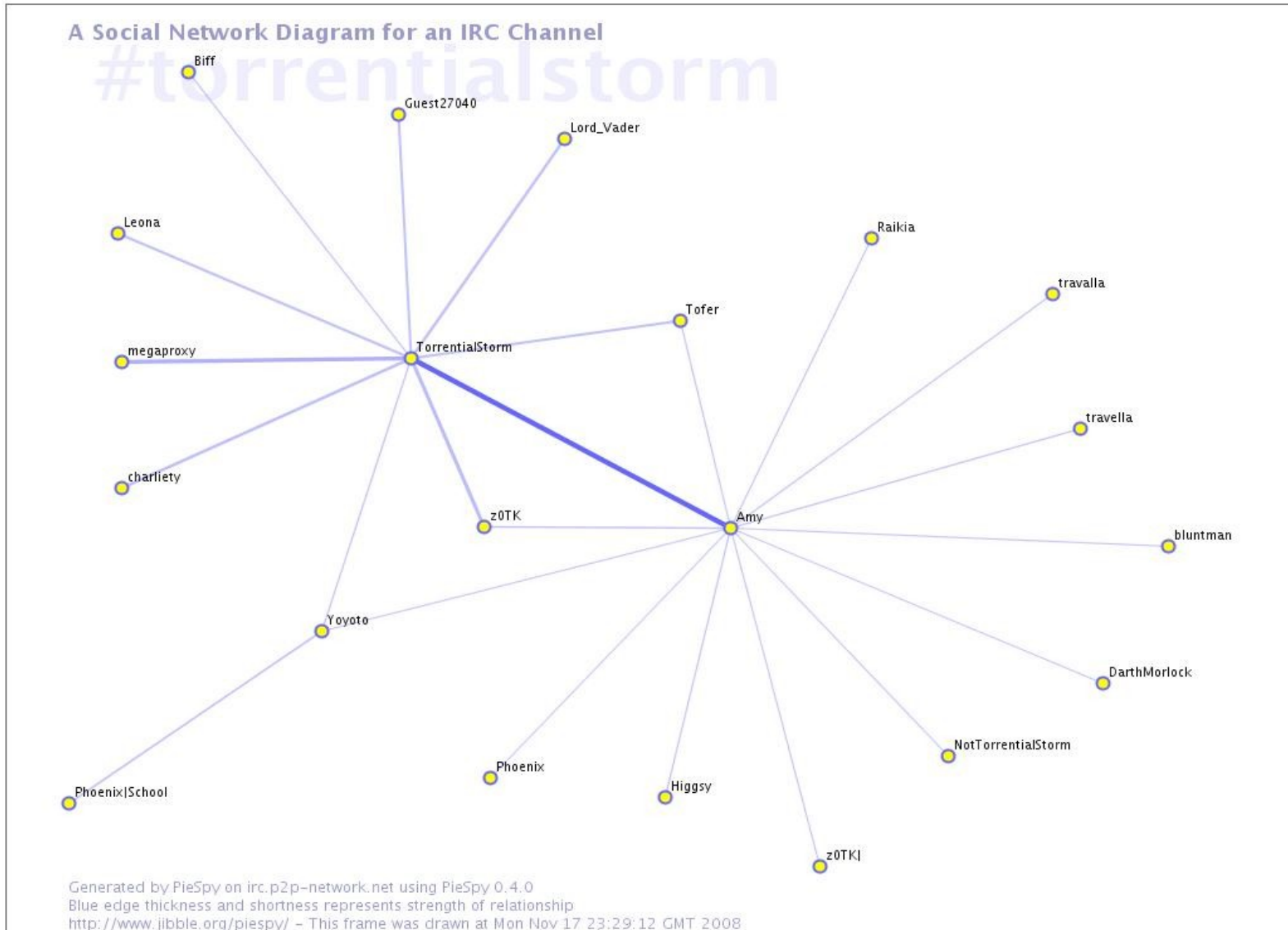
Generated by h0dd0g on irc.freenode.net using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Sun Nov 23 22:30:04 CET 2008

A Social Network Diagram for an IRC Channel



Generated by h0dd0g on irc.freenode.net using PieSpy 0.4.0
 Blue edge thickness and shortness represents strength of relationship
<http://www.jibble.org/piespy/> - This frame was drawn at Sun Nov 23 22:06:34 CET 2008

PieSpy - #torrentialstorm

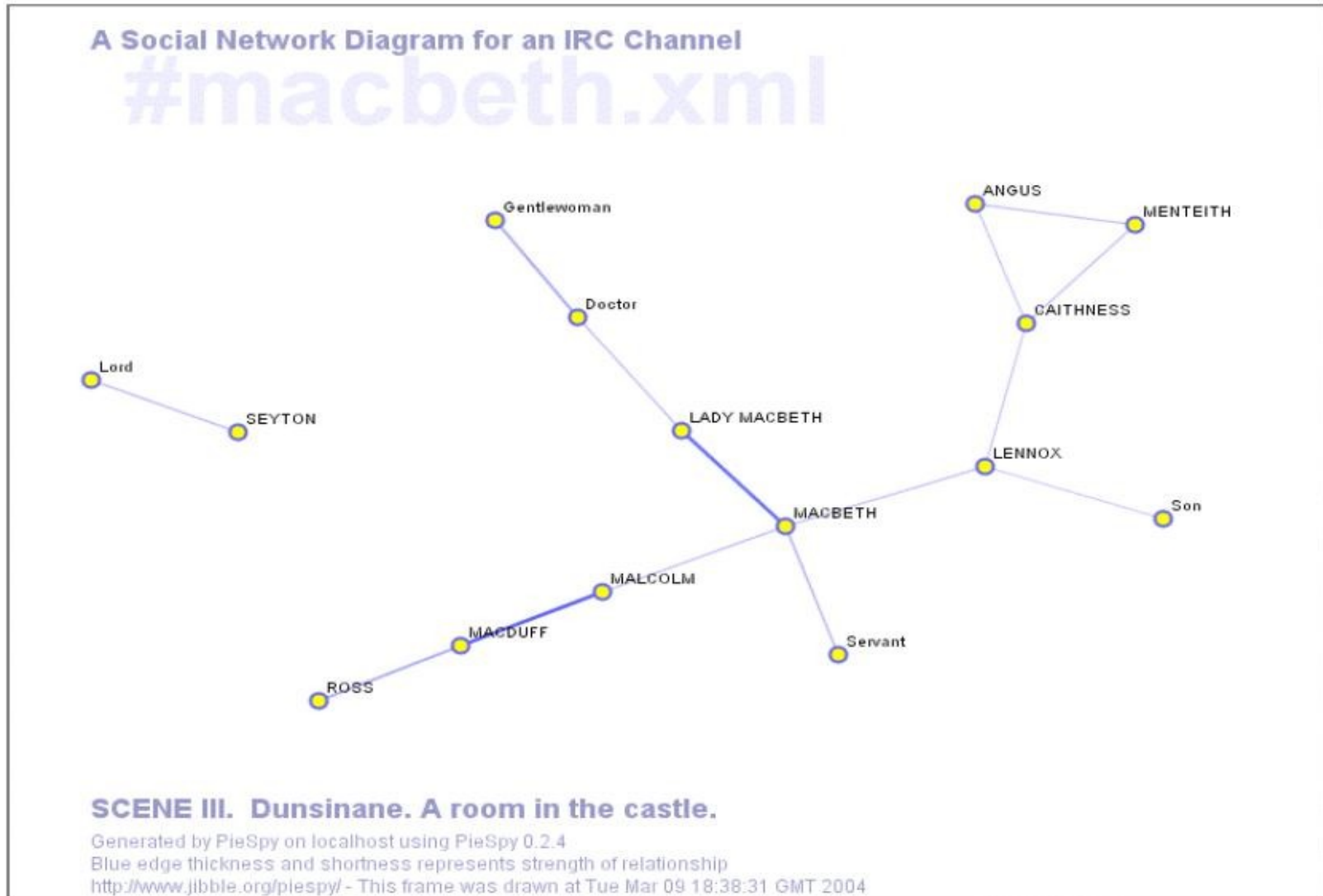


PieSpy - Anwendungen

- Integration mit IRC-Klienten
- Spam Filterung
- Channel Similarity Graphs
- Modellierung von sozialen Netzwerken in dramatischen Werken
- ...



PieSpy - Macbeth (W. Shakespeare)



Quelle: [5]

Referenzen

- [1] Planetary-Scale Views on an Instant-Messaging Network, Jure Leskovec, Eric Horvitz
 - arXiv:0803.0939v1 [physics.soc-ph]
- [2] Power laws, Pareto distributions and Zipf's law, M. E. J. Newman
 - arXiv:cond-mat/0412004v3 [cond-mat.stat-mech]
- [3] Zipf, Power-laws, and Pareto - a ranking tutorial
 - Lada A. Adamic, Information Dynamics Lab, Information Dynamics Lab, HP Labs, Palo Alto, CA 94304
 - Link: <http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html>
- [4] „An Experimental Study of Search in Global Social Networks“, Peter Sheridan Dodds, Roby Muhamad, Duncan J. Watts
 - Science 301, 827 (2003)
 - DOI: 10.1126/science.1081058
 - <http://www.sciencemag.org/cgi/content/full/301/5634/827>
- [5] „Inferring and Visualizing Social Networks on Internet Relay Chat“, Paul Mutton
 - <http://www.jibble.org/piespy/>
- [6] Guide to IRC: <http://gc.stonewallcs.com/IRC/guide.php>