

Charakterisierung der freigegebenen Dateien im modernen Gnutella

Yu, Xiao Tai
25.02.2009



Agenda

- Einführung
- Messung
- Analyse
- Ergebnisse



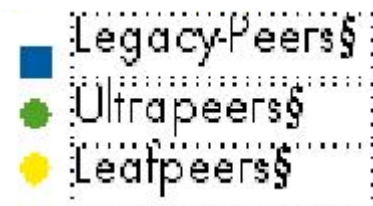
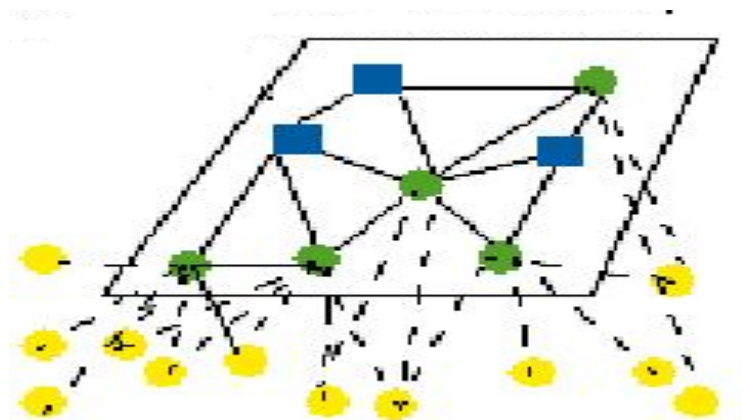
Gnutella

- P2P-Filesharing
- Populares Overlay-Netzwerk
- *Browse-Host Feature*



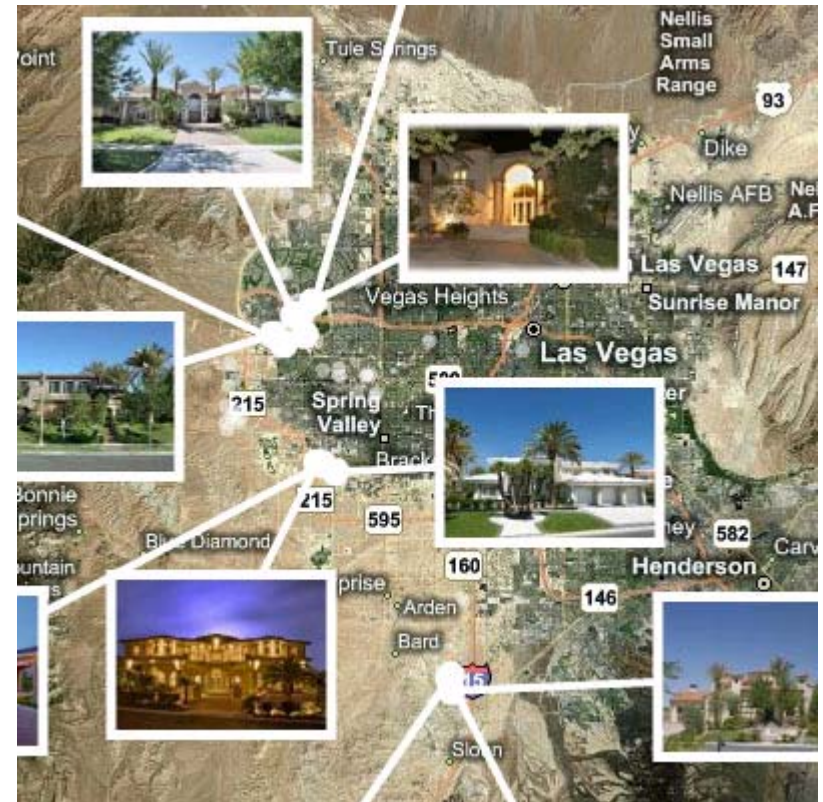
Zweischichtige Overlay-Topologie

- Obere Schicht
 - Ultrapeers
 - Legacy-Peers
- Untere Schicht
 - Leaf-Peers
 - Verbunden mit Ultrapeers



Messung

- Crawler
 - Snapshots
 - Sharing-Listen



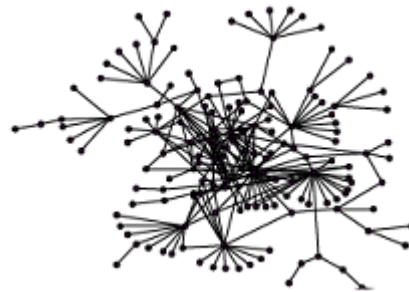
Snapshots

| Crawl Date | Type | Number | TCP Refused | Timeout | Conn. Lost | App. Refused |
|------------|-------------|-----------|-------------|---------|------------|--------------|
| Jun. 13 | Ultrapeers | 281,472 | 34.90% | 3.48% | 2.97% | 2.18% |
| | Leaf Peers | 1,932,944 | 85.39% | 0.90% | 0.88% | 0.55% |
| | Total Peers | 2,218,883 | 78.93% | 1.23% | 1.15% | 0.76% |
| Aug. 31 | Ultrapeers | 347,168 | 35.70% | 3.76% | 6.35% | 2.03% |
| | Leaf Peers | 2,320,088 | 86.98% | 0.86% | 1.08% | 0.46% |
| | Total | 2,674,452 | 80.27% | 1.24% | 1.82% | 0.67% |
| Oct. 13 | Ultrapeers | 320,063 | 36.05% | 3.47% | 5.89% | 2.25% |
| | Leaf Peers | 2,790,171 | 87.27% | 0.84% | 1.04% | 0.37% |
| | Total | 3,118,986 | 81.96% | 1.11% | 1.59% | 0.56% |



Das Tool Cruiser

- Topologie-Crawl
 - Für Topologie
 - 15 min
- Inhalt-Crawl
 - Für Sharing-Listen
 - 5,5 Stunden
- Präzise Snapshots
 - Zwei mal Topologie-Crawl
 - Langfristige vs. Kurzfristige Peers



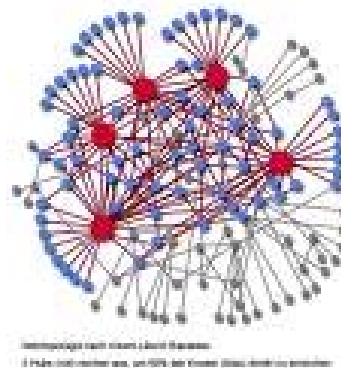
200,000 Peers

- Master/Slave-Architektur
- 6 Off-The-Shelf 1 GHz GNU/Linux Geräten



Analyse

- Statische Analyse
- Topologische Analyse
- Dynamische Analyse

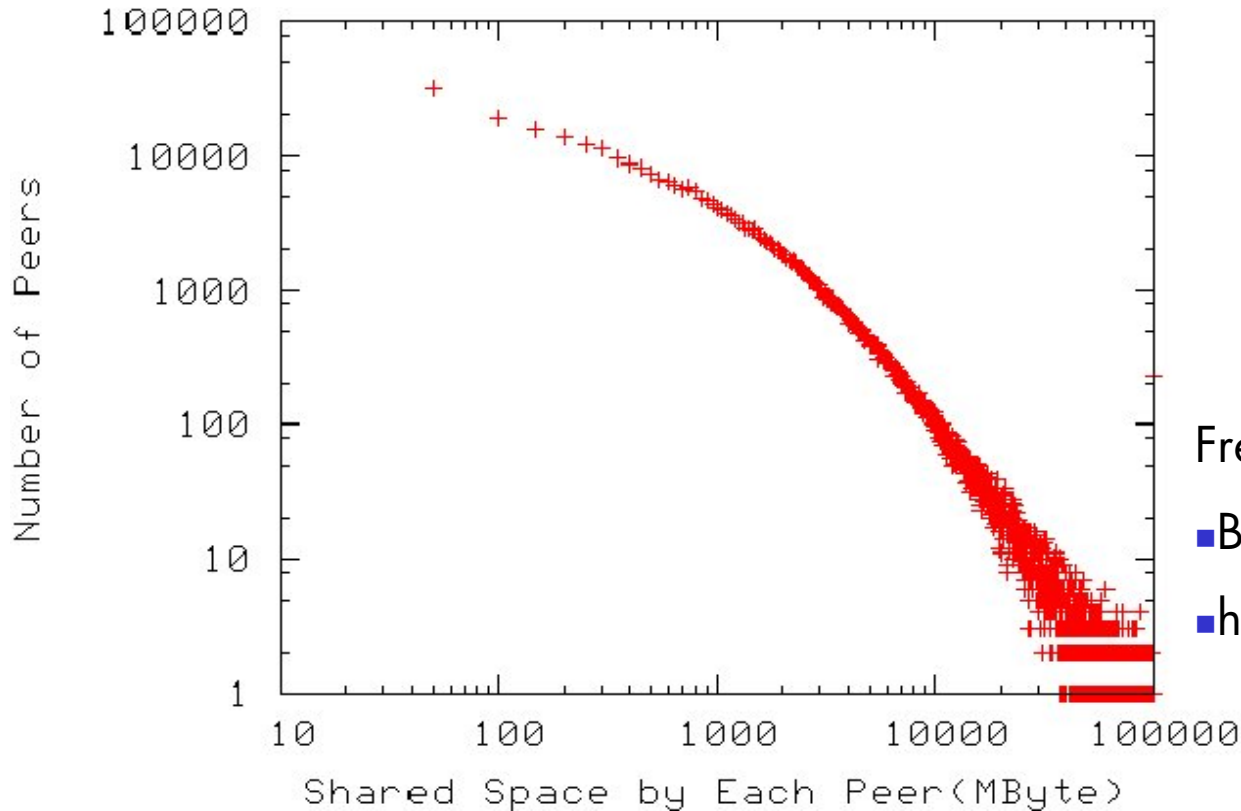


Distribution-Funktionen

- Power-Law-Distribution
 - $F(x) = ax^k + o(x^k)$
 - Kleiner Anteil häufig, großer Anteil selten
- Zipf-Distribution
 - Ranked
 - Kleiner Anteil häufig, großer Anteil selten



Power-Law-Distribution



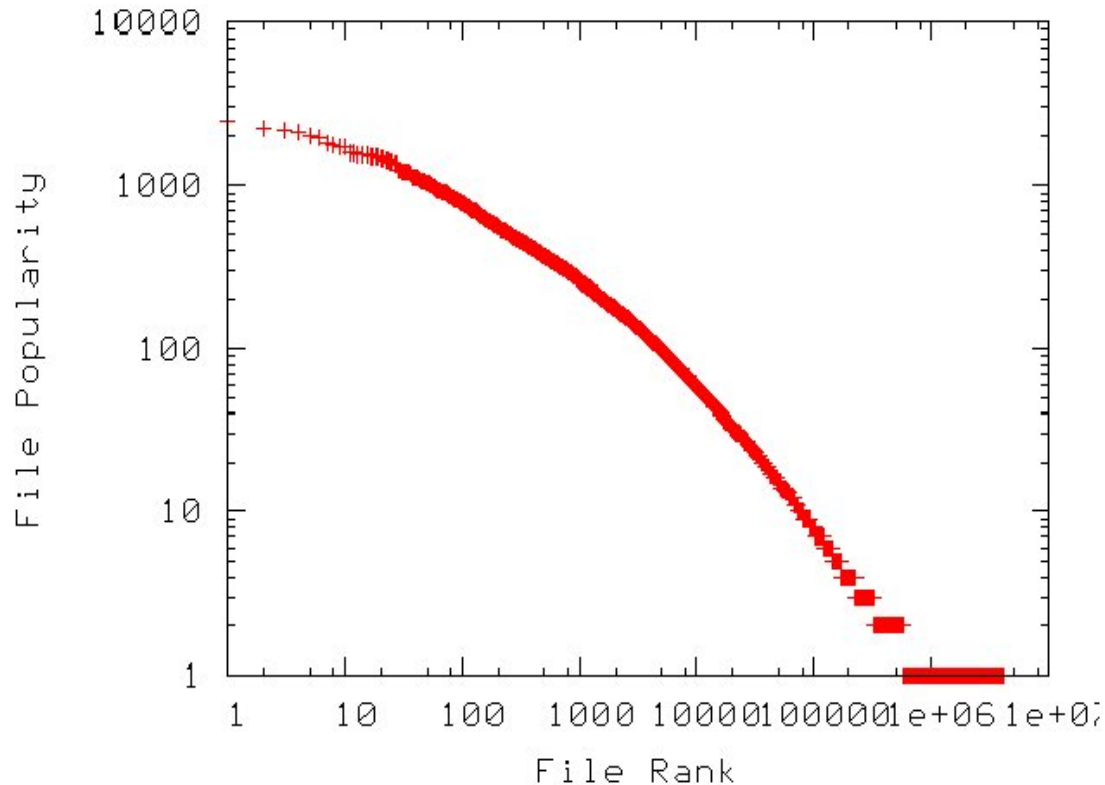
Freigegebene Speichergröße

- Bis zu 100GB, aber selten
- häufig 100MB



$$y = C x^a \text{ bedeutet } \log(y) = \log(C) - a \log(x)$$

Zipf-Distribution



Popularität

■ 10 Millionen Dateien

■ stabil innerhalb 4 Monaten



Popularität der freigegebenen Dateitypen

- Multimedia-Dateien (93%)
 - populärste MP3
 - 1/3 Speicherplatz
 - Videodateien:
 - 1/10 weniger als Audiodateien,
 - aber 25% Bytes mehr
 - Popularität und Speichergröße verdreifacht

| Major Audio Types | | |
|--------------------------|------------------|------------------|
| Type | Files (%) | Bytes (%) |
| mp3 | 61.06-61.54 | 36.96-38.40 |
| wma | 2.69-2.76 | 1.28-1.36 |
| wav | 1.83-1.98 | 0.69-0.73 |
| m4a | 1.33-1.47 | 0.71-0.78 |
| Total | 67.14-67.58 | 39.68-41.21 |



Ergebnisse der statischen Analyse

- Free-Riders
 - Anteil 13%
 - Keine Relation zwischen Sharing und Peertyp
 - Ultrapeers vs. Leaf-Peers
- Anzahl und Speichergröße
 - Power-Law-Distribution
- Popularität
 - Zipf-Distribution

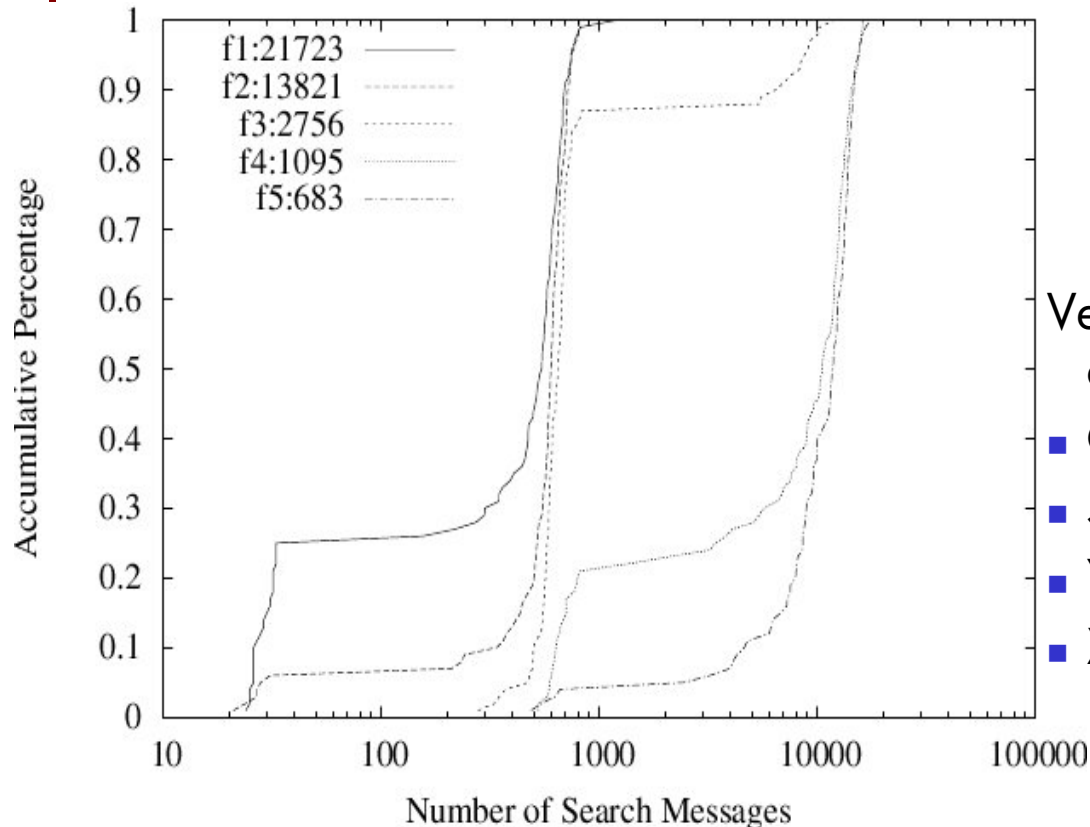


Topologische Analyse

- Verteilung der freigegebenen Dateien auf die Overlay-Topologie
- Gemeinsamkeit zwischen Peers



Cumulative Distribution Function (CDF)

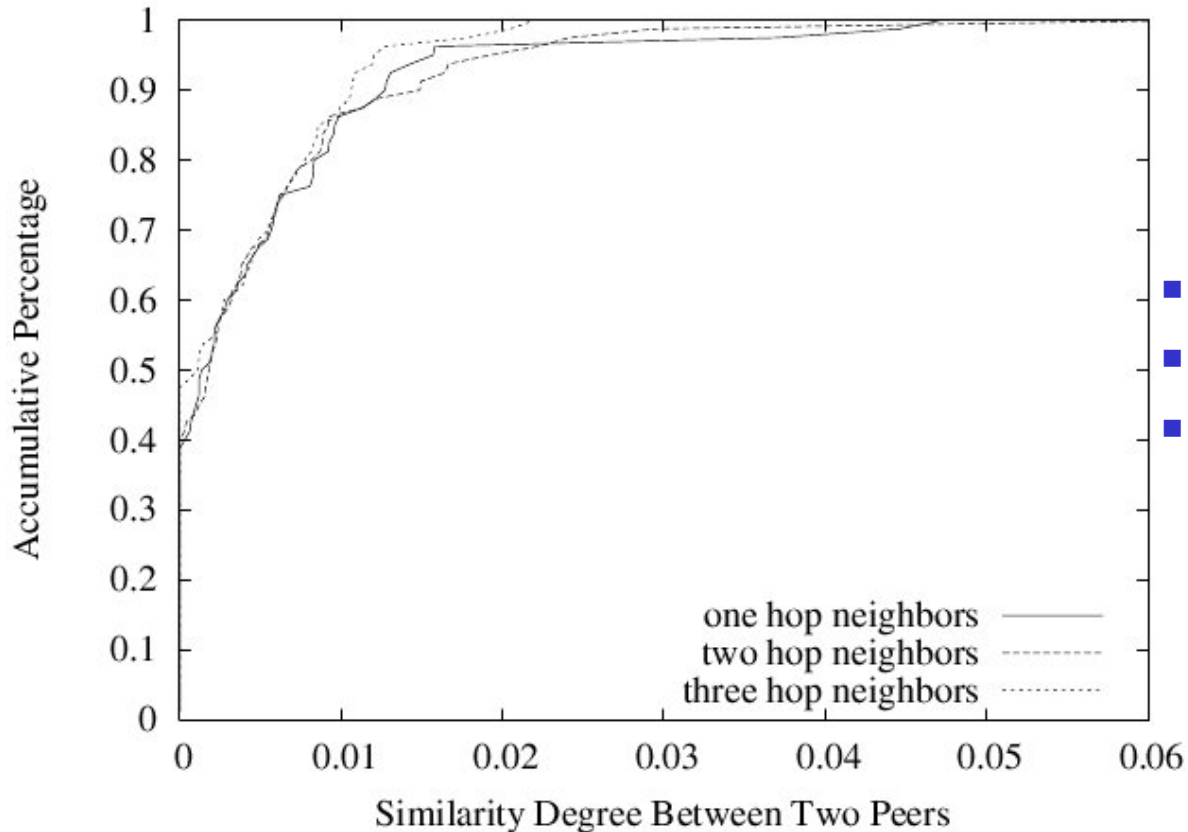


Verteilung der freigegebenen Dateien auf die Overlay-Topologie

- CDF: Suche 5 Kopien einer Datei
- 5 CDF: 5 Exemplare
- Y: 100 Peers
- X: minimale Suchanfragen



Gemeinsamkeit zwischen Peers



- 3 CDF: 1, 2 und 3-Hops
- Y: 100 Peers
- X: Gemeinsamkeit



Ergebnisse der topologischen Analyse

- Churn ist dominant
 - Sharing-Listen zufällig auszuwählen
 - Dateien zur Simulation für P2P-Filesharing zufällig zu verteilen

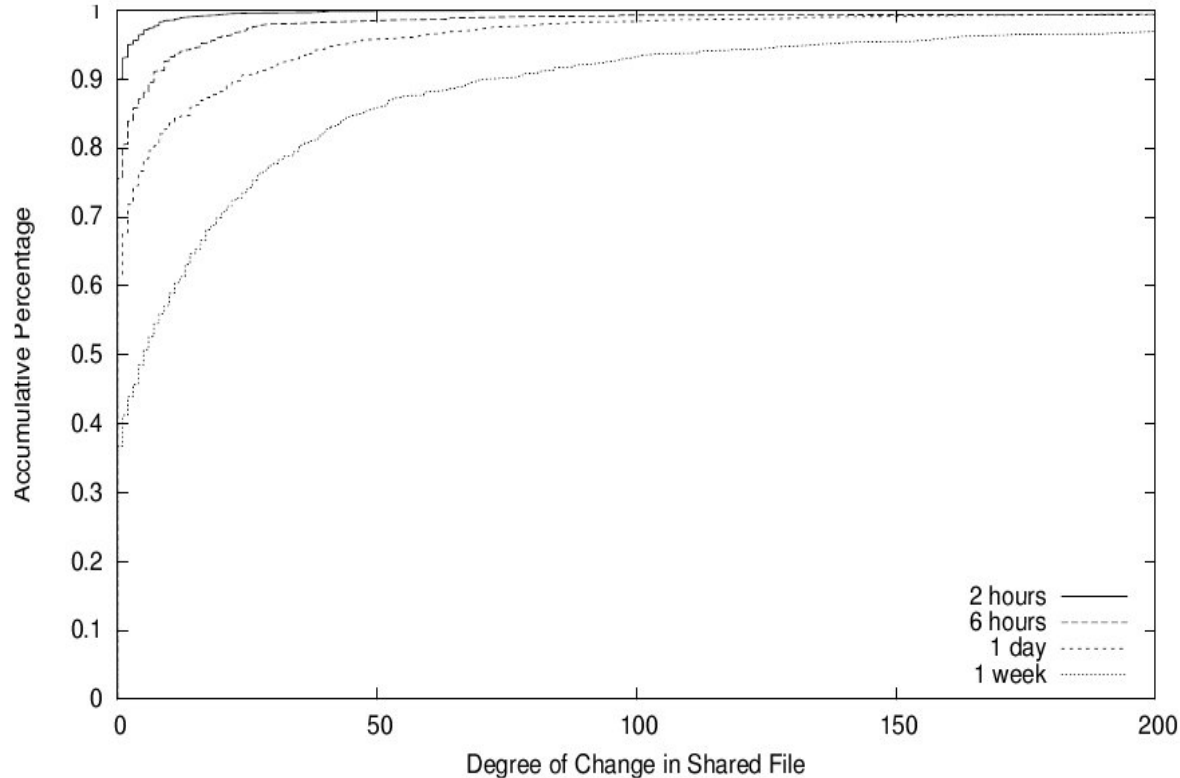


Dynamische Analyse

- Änderungen der freigegebenen Dateien
- Popularitätsänderungen
- Entwicklung der Popularitätsänderungen



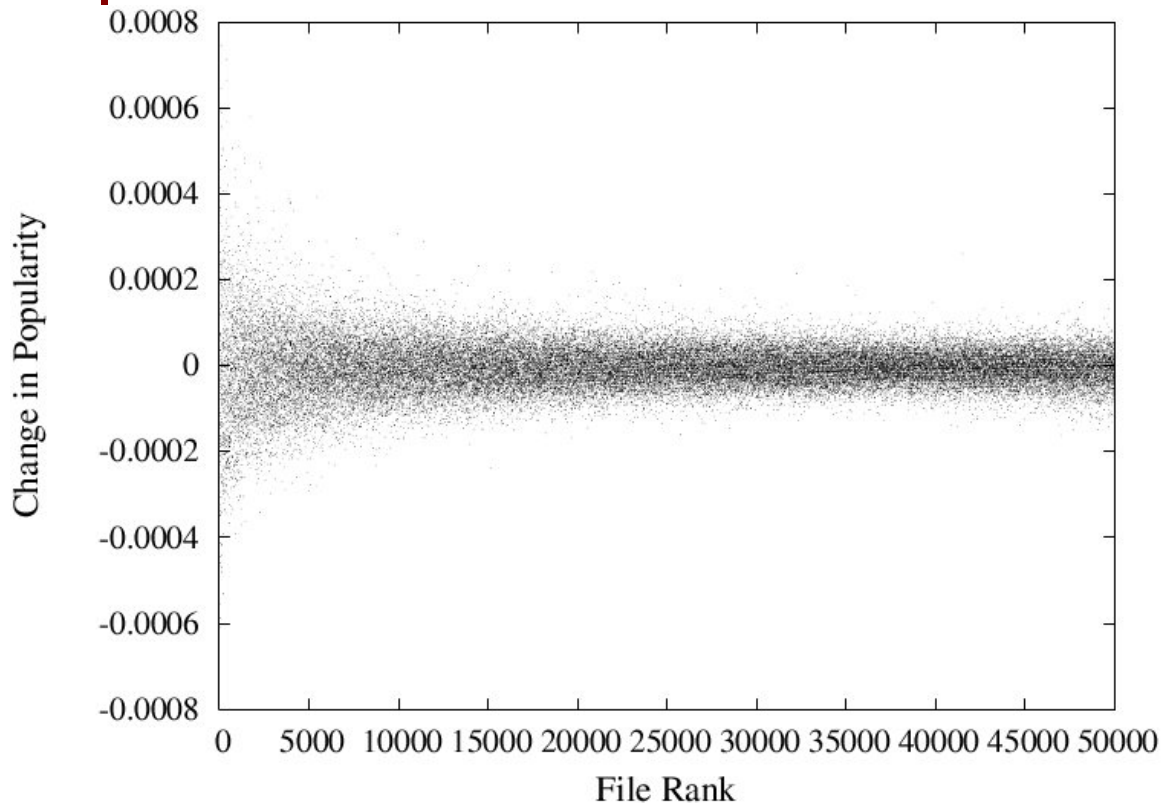
Änderungen der freigegebenen Dateien



- 4 CDF:
 - 2 Std., 6 Std., 1 T. und 1 W.
- Y: 3000 Peers
- X: Änderungen der Dateien



Popularitätsänderungen – ein Tag

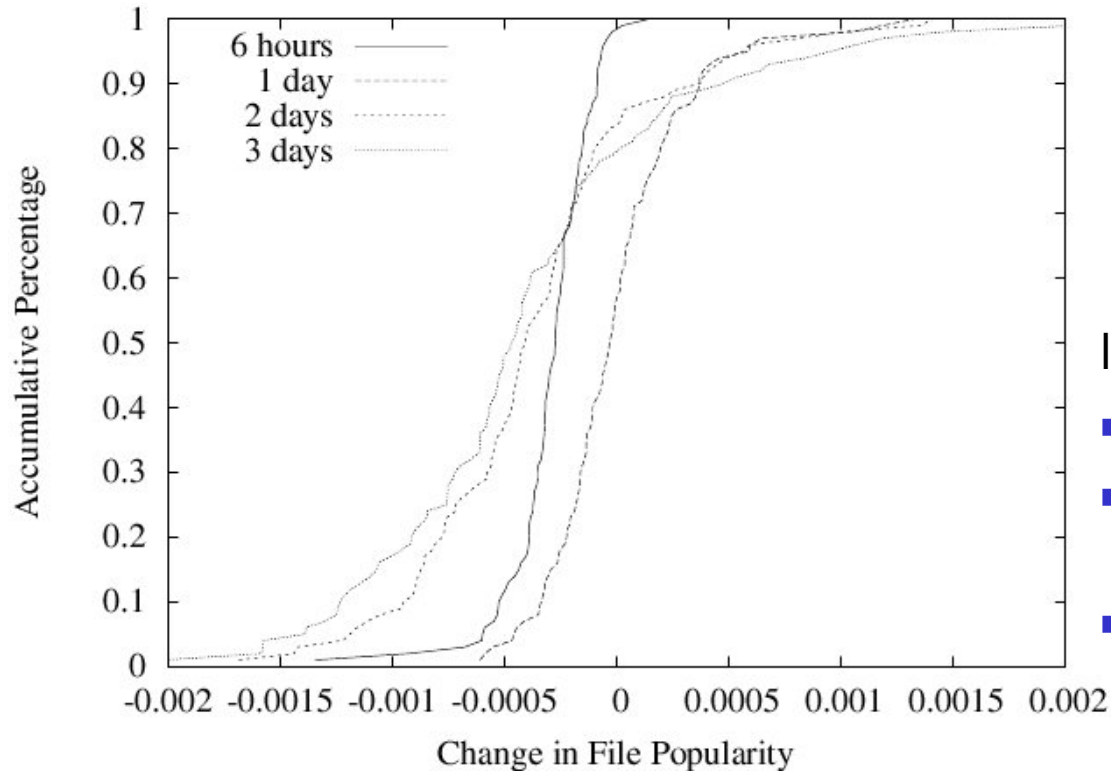


Änderungen an einem Tag

- X: 50,000 Rank
- Y: Popularitätsänderung



Popularitätsänderungen – Tage

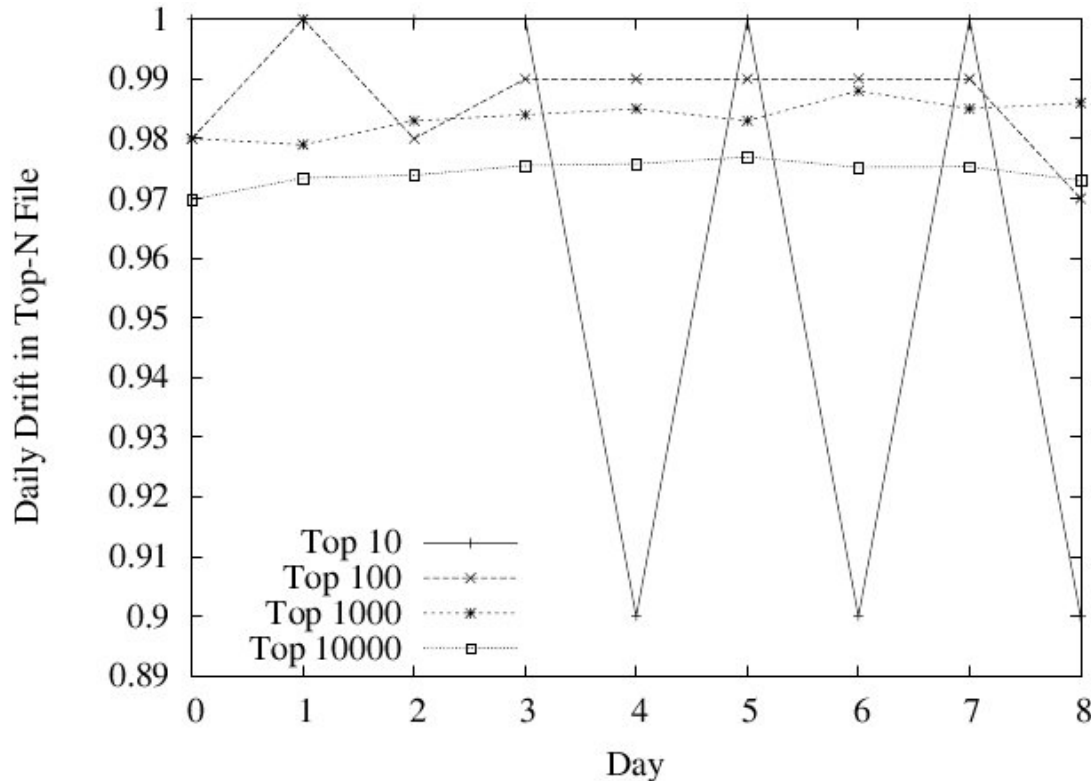


Im kurzen Zeitraum

- CDF: 6 Std. und 1, 2, 3 Tage
- Y: Top-1000 populärsten Dateien
- X: Popularitätsänderung



Popularitätsänderungen – 1 Woche

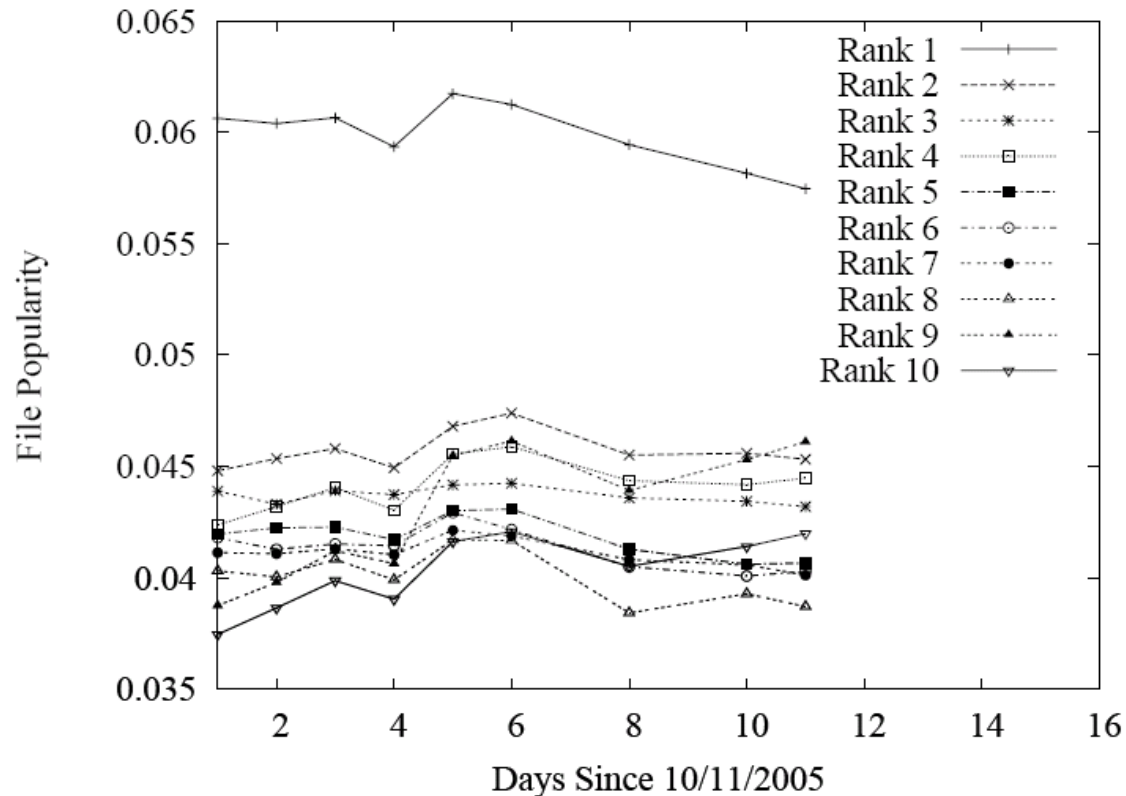


Top 10-10,000 populärsten Dateien

- X: Tage
- Y: Popularitätsänderungen



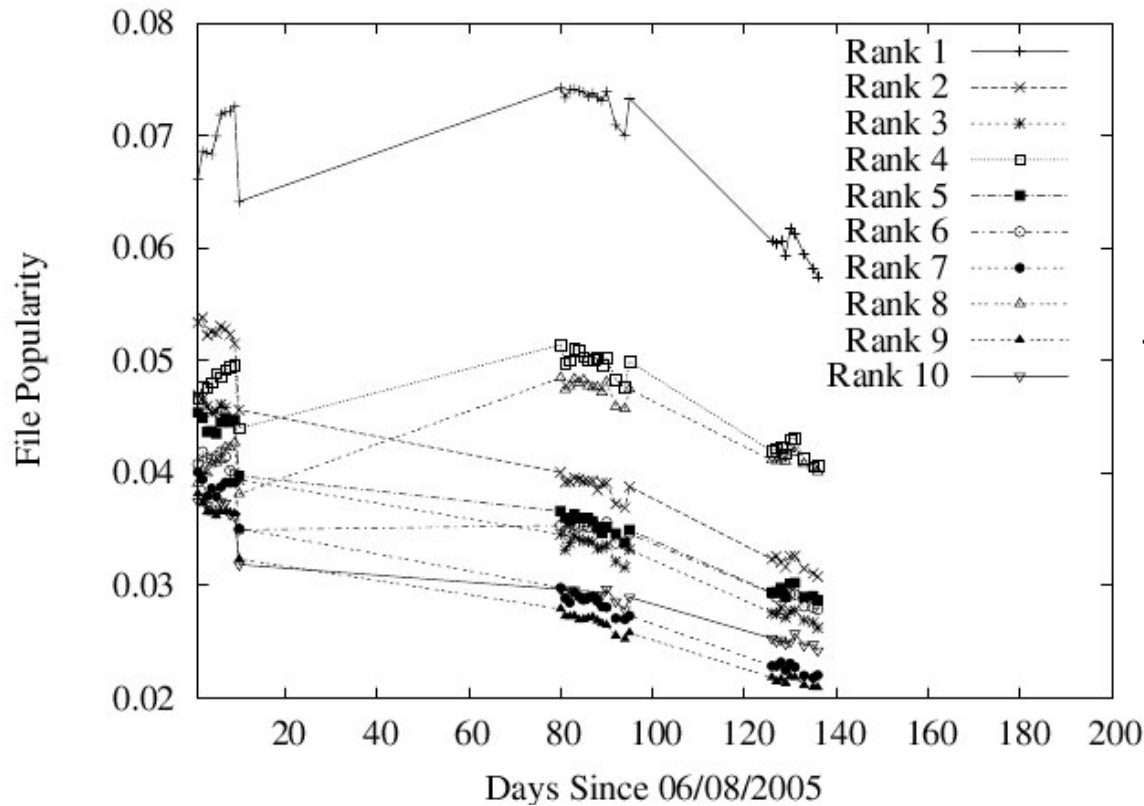
Entwicklung der Popularitätsänderungen – Tage



Top-10 populärsten Dateien
■ X: Tage
■ Y: Popularitätsänderungen



Entwicklung der Popularitätsänderungen – Monate



- Top-10 populärsten Dateien
- X: Tage
- Y: Popularitätsänderungen



Ergebnisse der dynamischen Analyse

- Änderungen
 - populärsten Dateien größer
 - schneller mit der Zeit
 - groß über Monate
 - klein innerhalb einer Woche
 - Caching bandbreitesparende Strategie



Zusammenfassung

- Das moderne Gnutella
- Messungsmethode
- Analyse
 - Anzahl, Speichergröße, Dateitypen und Popularität
 - Topologische Analyse
 - Änderungen der freigegebenen Dateien



Literatur Verzeichnis

- D.Stutzbach and R.Rejaie: "Characterizing the Two-Tier Gnutella Topology",
- S. Zhao, D. Stutzbach & R. Rejaie: "Characterizing Files in the Modern Gnutella Network: A Measurement Study"
- Verteilungsfunktionen:
<http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html>
- Ein ultimatives P2P-Buch ist gewünscht
 - Überblick des weiten Bereichs von P2P
 - Stand der Techniken
 - Architektur
 - Konzepte des Designs



Vielen Dank

Charakterisierung der freigegebenen Dateien



im modernen Gnutella



Technische Universität Berlin
Fakultät IV – Elektrotechnik und Informatik
Seminar: Internet Measurement
Intelligent Networks / Intelligente Netze (INET)
Research Group of Prof. Anja Feldmann, Ph.D.

Charakterisierung der freigegebenen Dateien
im modernen Gnutella - 28