



KTH Electrical Engineering



UNIVERSITY OF
CALGARY

Power-law revisited: A large scale measurement study of P2P content popularity

György Dán

School of Electrical Engineering
KTH, Royal Institute of Technology
Stockholm, Sweden

Niklas Carlsson

Department of Computer Science
University of Calgary
Calgary, Canada

27 April 2010, IPTPS
San Jose, CA

P2P Content Popularity

- Instantaneous popularity
 - Concurrent number of peers
 - Effectiveness of locality awareness
 - Little data available
 - Power-law ?
- Download popularity
 - Number of peers that downloaded content
 - Effectiveness of caching
 - Several measurements
 - Power-law but flattened head (Mandelbrot-Zipf)
- Measurements limited in time and coverage
 - How accurate are they?
 - How accurate can they be?



KTH Electrical Engineering



L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang, Measurement, Analysis, and Modeling of BitTorrent-like Systems, in Proc. ACM IMC, Oct. 2005.

M. Hefeeda and O. Saleh, Traffic modeling and proportional partial caching for peer-to-peer systems, IEEE/ACM Trans. on Networking, vol. 16, no. 6, pp. 14471460, 2008.

Measuring P2P Content Popularity

- Overlay structure → Measurement methodology
 - Tracker based (BitTorrent)
 - Peer harvesting
 - Tracker query - scrape
 - Deep packet inspection
 - Unstructured (Gnutella, Ares, FastTrack)
 - Monitoring queries and replies
 - Deep packet inspection
- Measurement = Sample of population wide popularity
 - Probability sampling - difficult
 - Opportunity sampling
 - Inference can be misleading



KTH Electrical Engineering



Measurement Methodology

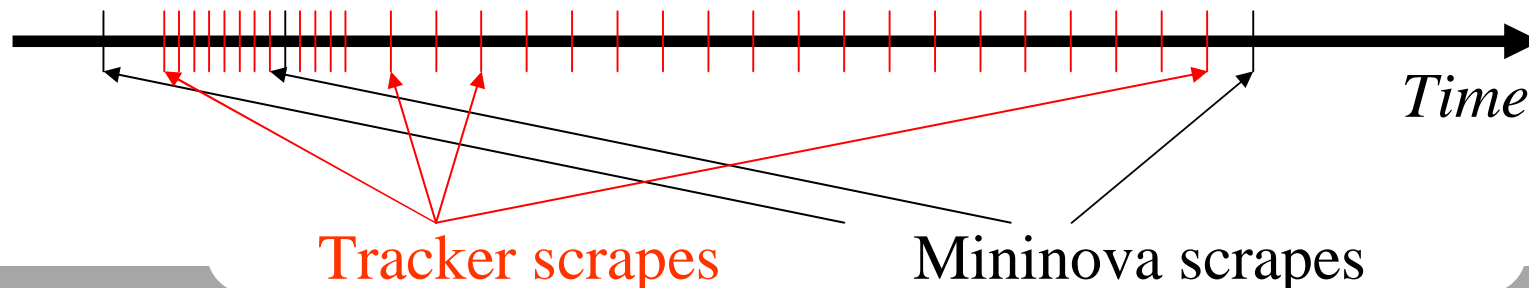
- Screen scrape of *Mininova.org*
 - Largest torrent search engine
 - 31 Aug. 2008, 15 Oct. 2008, 31 Aug. 2009
 - Scrape URL of 1690 BitTorrent trackers
- Scrape of 721 BitTorrent trackers (S,L,D)
 - 15 Sept. 2008 to 17 Aug. 2009
 - weekly, daily at 8pm GMT
 - Almost instantaneous (<30mins)



KTH Electrical Engineering



UNIVERSITY OF CALGARY



Zipf's Law and Beyond

- Zipf's Law

$$f_{\text{Zipf}(f_1, \theta)}(r) = \frac{f_1}{r^\theta}$$

- Heavy tail

$$\lim_{r \rightarrow \infty} e^{ar} f_{\text{Zipf}(f_1, \theta)}(r) = \infty \quad \forall a > 0, \lambda \geq 0$$

- Mandelbrot-Zipf Law

$$f_{\text{MZipf}(f_1, \lambda, \theta)}(r) = \frac{f_1}{(\lambda + r)^\theta}$$

- Flattened head

- Generalized Zipf Law

$$f_{\text{GZipf}(f_1, \lambda, \mu, \theta)}(r) = \frac{f_1}{(1 - \lambda / \mu + (\lambda / \mu)e^{(1/\theta)\mu r})^\theta}$$

- Light tail

$$\lim_{r \rightarrow \infty} e^{ar} f_{\text{GZipf}(f_1, \lambda, \mu, \theta)}(r) < \infty \quad \exists a > 0, \lambda \geq 0$$

- Flattened head

- Power-law trunk

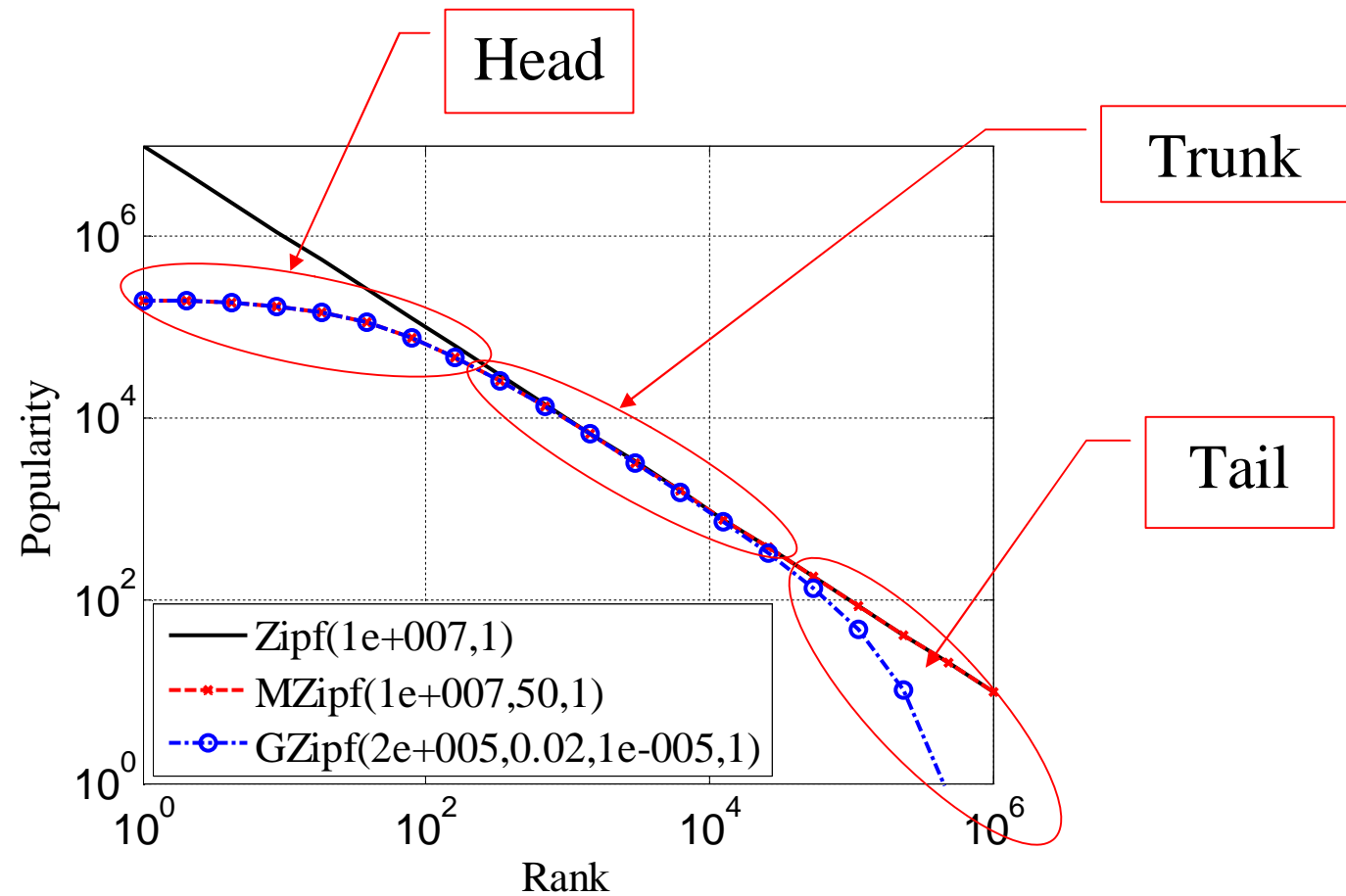


KTH Electrical Engineering



UNIVERSITY OF CALGARY

Zipf's Law and Beyond - Example



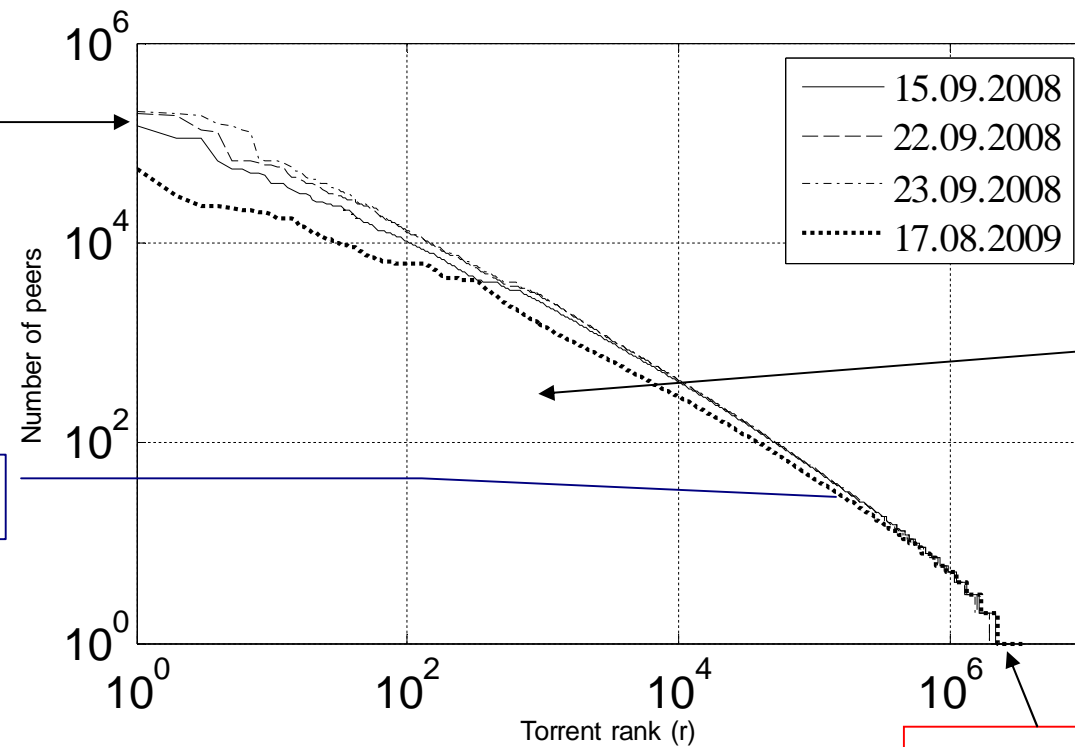
What we measured (I)

- Instantaneous popularity



Max peers

Power-law?

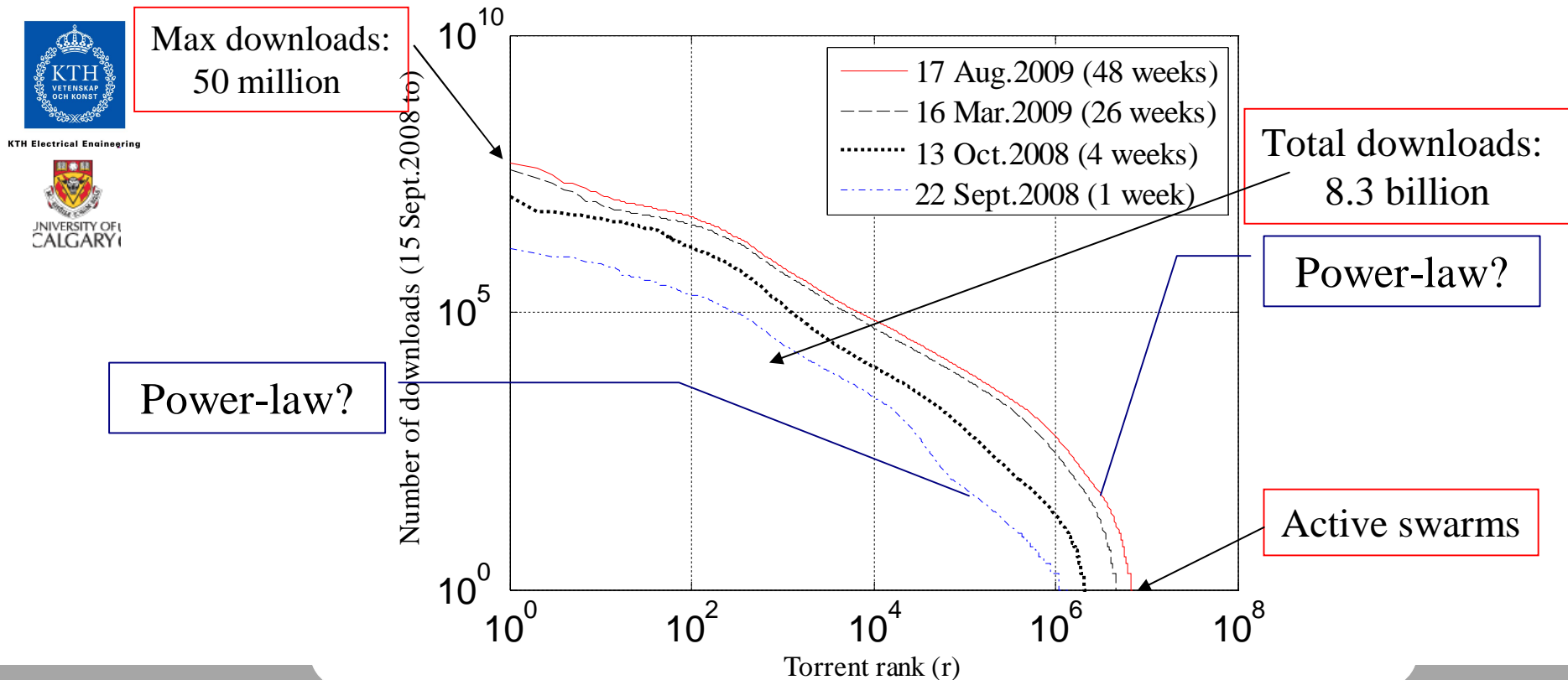


Total peers:
42 million

Active swarms

What we measured (II)

- Download popularity



Instantaneous Popularity

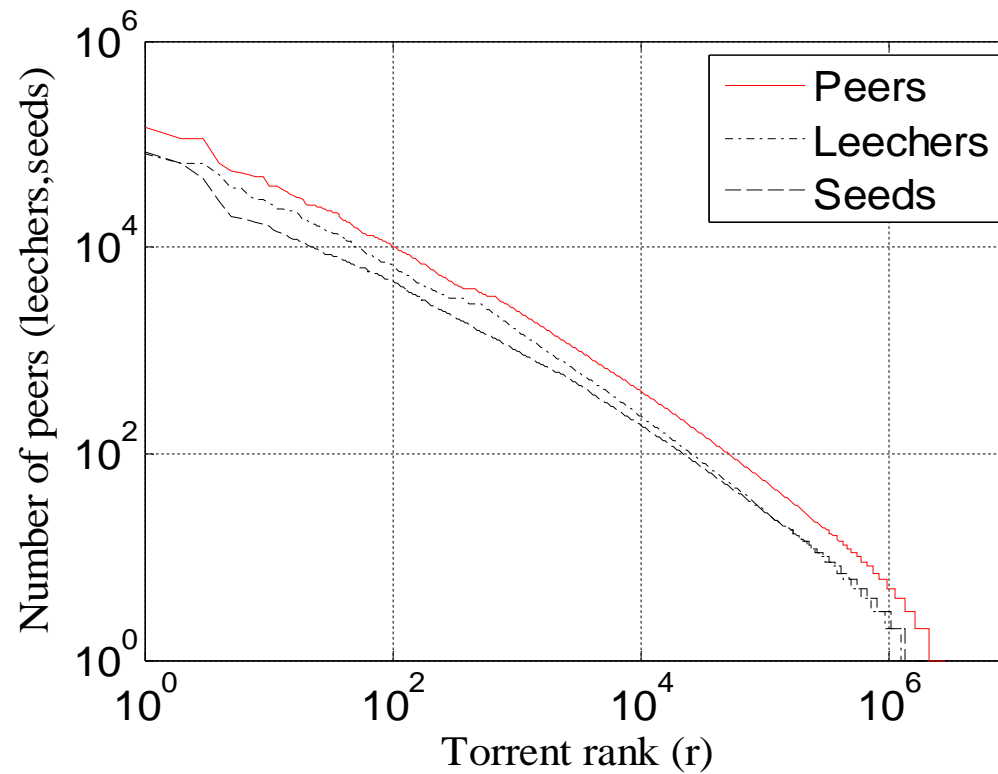
- Instantaneous popularity 15 Sept 2008, 8pm GMT
 - Max: 1.6×10^5 , Total: 4.23×10^7 , Active: 2.93×10^6



KTH Electrical Engineering

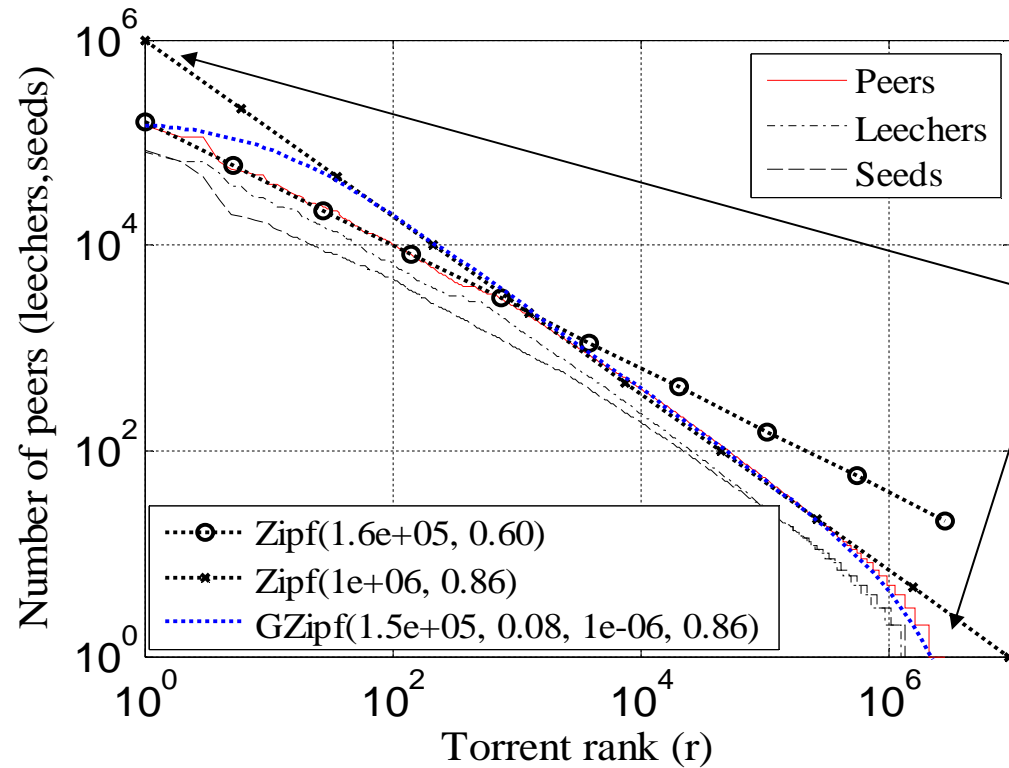


UNIVERSITY OF CALGARY



Power-law or Double-power-law?

- Instantaneous popularity 15 Sept 2008, 8pm GMT
 - Max: 1.6×10^5 , Total: 4.23×10^7 , Active: 2.93×10^6



Power-law trunk hypothesis:
-Max: 10^6
-Total: 6.1×10^7
-Active: 9.5×10^6

Sampling artifact?



Sampling and Exponential cutoff

- Instantaneous popularity 15 Sept 2008, 8pm GMT
 - 2.93×10^6 samples from Double-Zipf in two ways
 - PropTor (discover torrent proportional to its popularity)
 - UnifTor (discover torrent uniform at random)

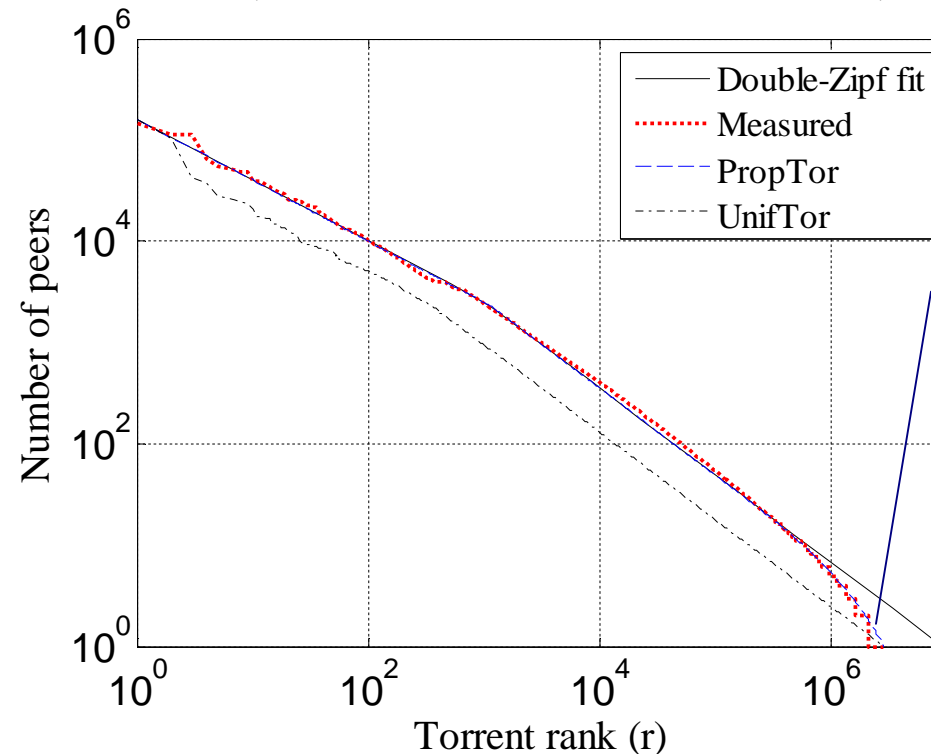


KTH Electrical Engineering



UNIVERSITY OF CALGARY

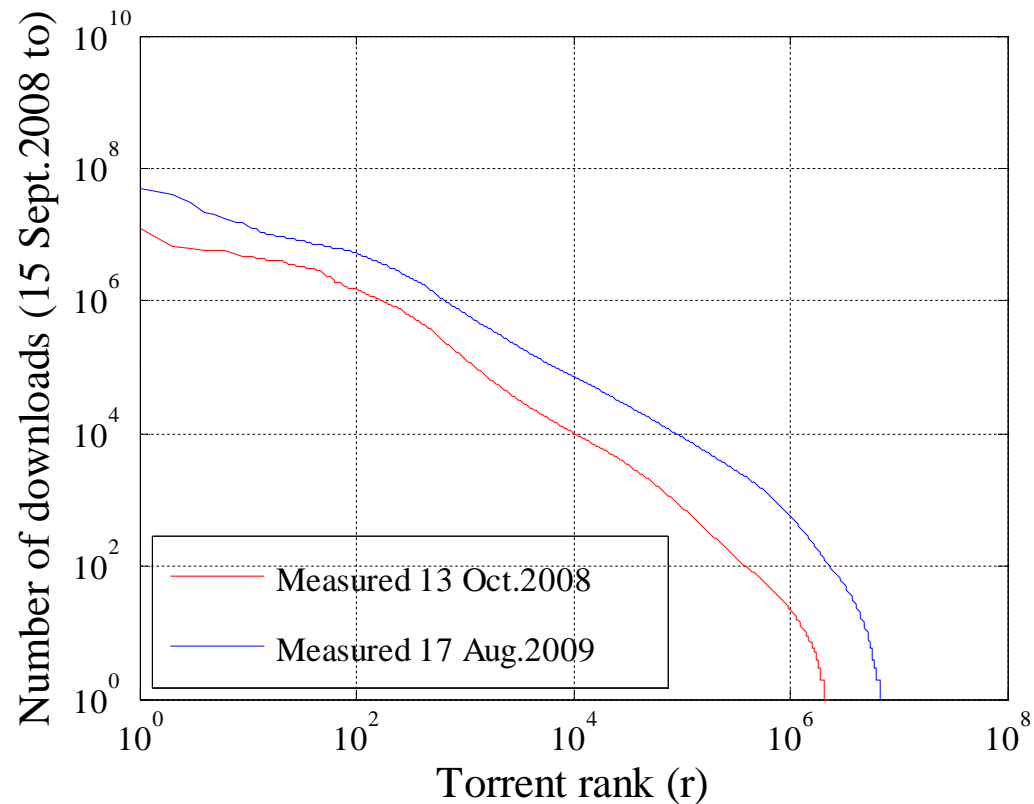
Total: 4.23×10^7
Total: 4.02×10^7
PMCC=0.99



PropTor
sampling
introduces
exponential
cutoff

Download Popularity

- Download popularity over 4 and 48 weeks
 - Active: 2.29×10^6 and 7.17×10^6 torrents



KTH Electrical Engineering



UNIVERSITY OF CALGARY

Power-law vs. Exponential cutoff

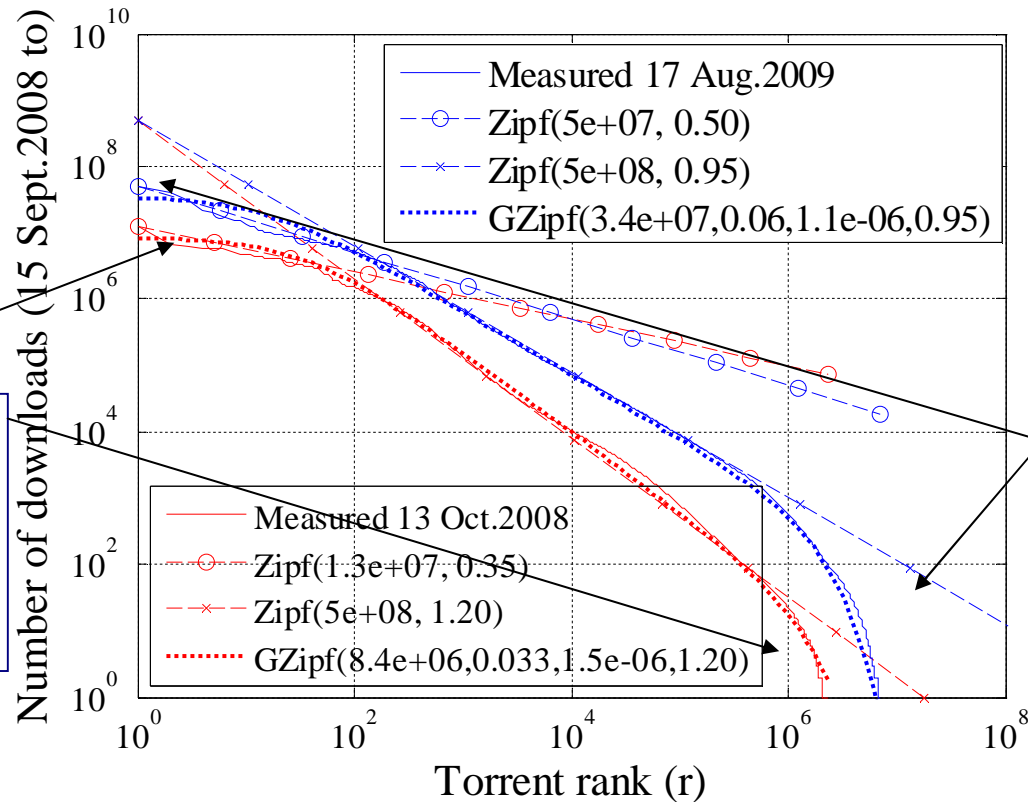
- Download popularity over 4 and 48 weeks
 - Active: 2.29×10^6 and 7.17×10^6 torrents



KTH Electrical Engineering



UNIVERSITY OF CALGARY



4 weeks
Double power-law hypothesis:
-Active: 1.77×10^7

48 weeks
Double power-law hypothesis:
-Active: 1.43×10^9

Sampling and Exponential cutoff

- Download popularity over 4 weeks (15 Sept.2008-13 Oct.2008)
 - 2.29×10^6 samples from Double-Zipf in two ways
 - PropTor (discover torrent proportional to its popularity)
 - UnifTor (discover torrent uniform at random)

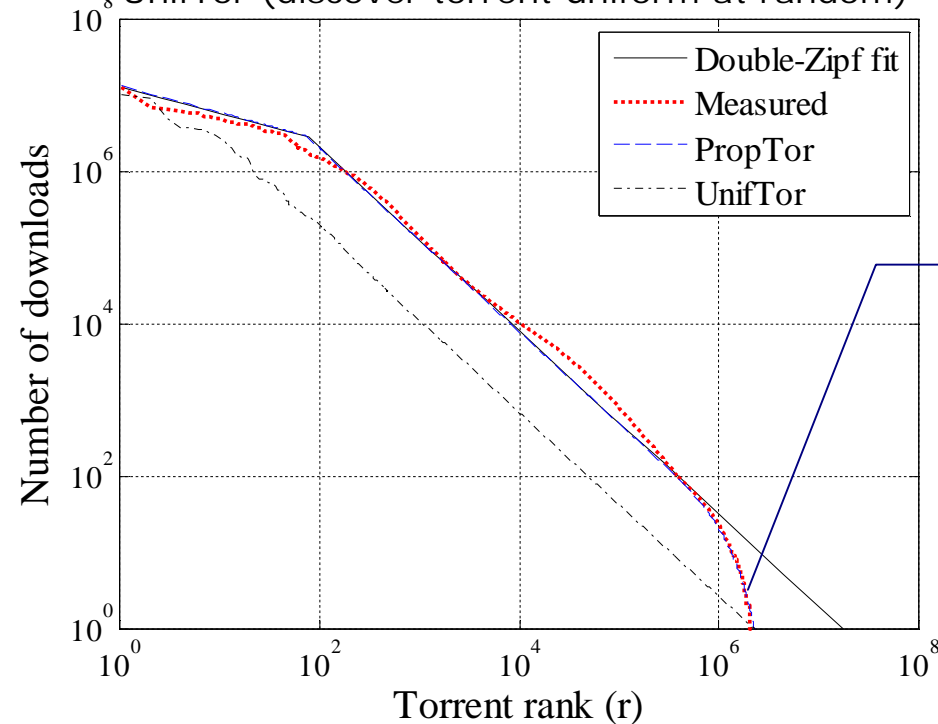


KTH Electrical Engineering



UNIVERSITY OF CALGARY

Total: 1.31×10^9
Total: 1.21×10^9
PMCC=0.99



PropTor
sampling
introduces
exponential
cutoff

Impact of Sampling

Large torrents overrepresented

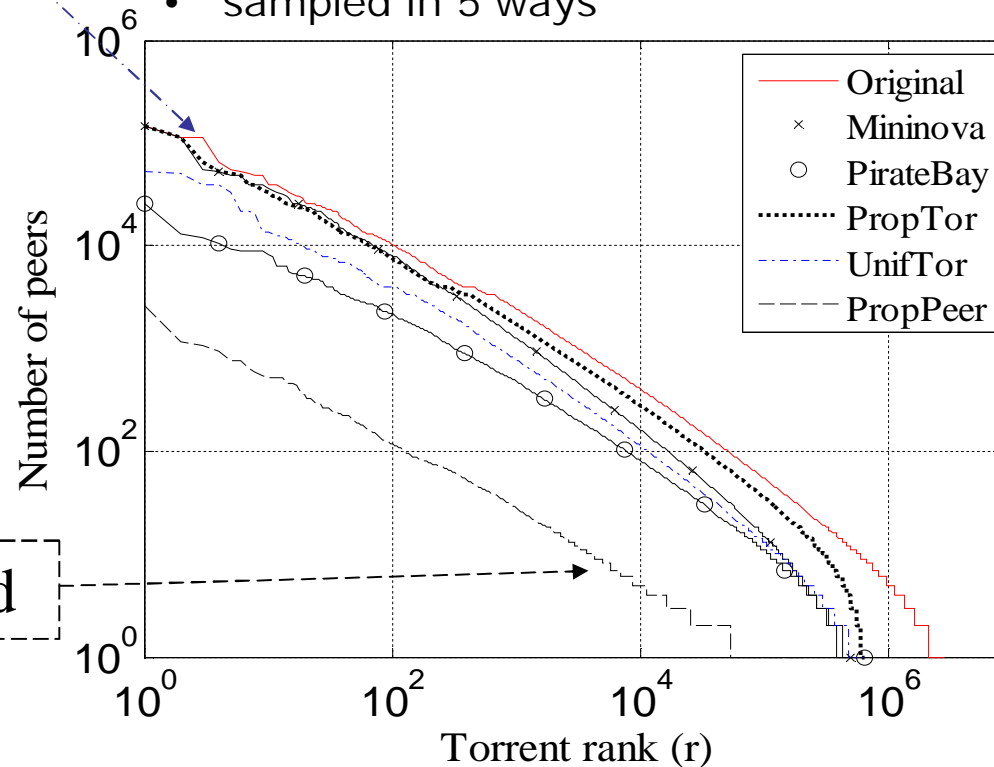
- Instantaneous popularity 15 Sept 2008, 8pm GMT
 - 2.93×10^6 active torrents, 4.23×10^7 total peers
 - sampled in 5 ways



KTH Electrical Engineering



UNIVERSITY OF CALGARY



Heavy-tailed

PirateBay, PropTor, UnifTor:
 6.55×10^5 torrents

PropPeer: 4.23×10^5 peers (1% of total)

Mininova: 9.7×10^5 torrents

Impact of Sampling

Large torrents
overrepresented

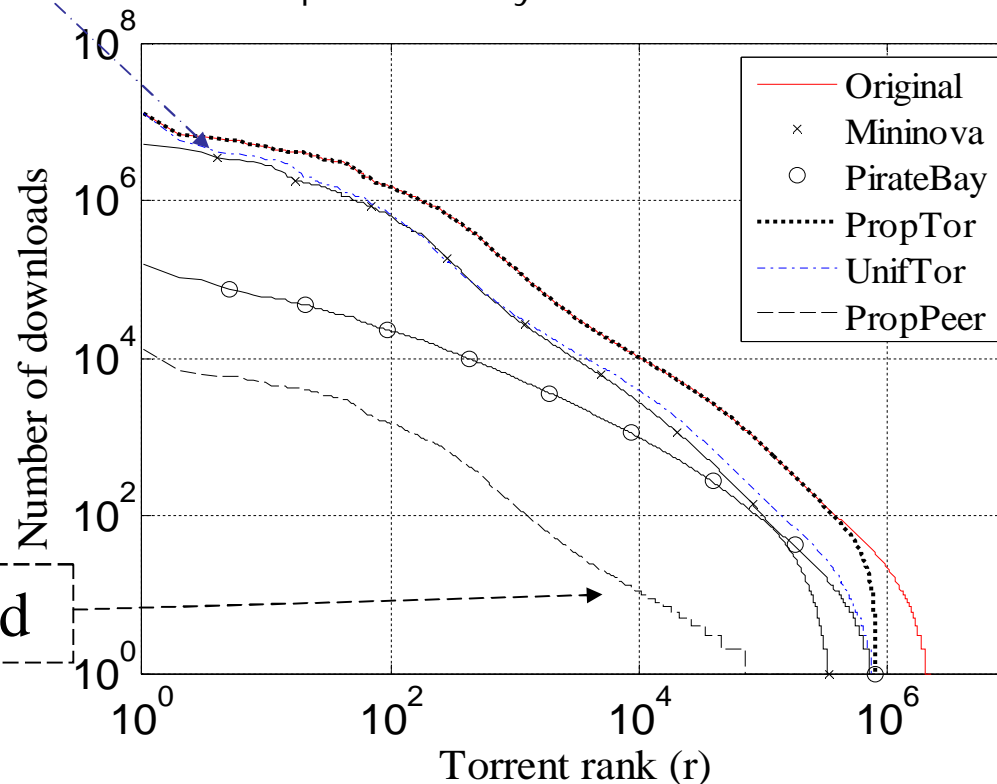
- Download popularity over 4 weeks
 - 2.29×10^6 active torrents, 1.31×10^9 total downloads
 - Sampled in 5 ways



KTH Electrical Engineering



UNIVERSITY OF CALGARY



PirateBay, PropTor,
UnifTor:
 1.69×10^6 torrents

Mininova: 4.95×10^5
active torrents

PropPeer: 1.31×10^6
peers (0.1% of total)

Heavy-tailed

Summary

- Large measurement study of P2P content popularity
 - Instantaneous popularity
 - Download popularity
- Instantaneous popularity
 - Power-law head?, power-law trunk
 - Tail *may be* power-law
- Download popularity
 - Flat head, power-law trunk
 - Tail *may be* power-law for short periods
 - Not power-law for long periods
- Sampling and measured characteristics
 - Infer with care



KTH Electrical Engineering





KTH Electrical Engineering



UNIVERSITY OF
CALGARY

Power-law revisited: A large scale measurement study of P2P content popularity

György Dán

School of Electrical Engineering
KTH, Royal Institute of Technology
Stockholm, Sweden

Niklas Carlsson

Department of Computer Science
University of Calgary
Calgary, Canada

27 April 2010, IPTPS
San Jose, CA