

A Longitudinal Characterization of Local and Global BitTorrent Workload Dynamics

Niklas Carlsson
György Dan
Anirban Mahanti
Martin Arlitt

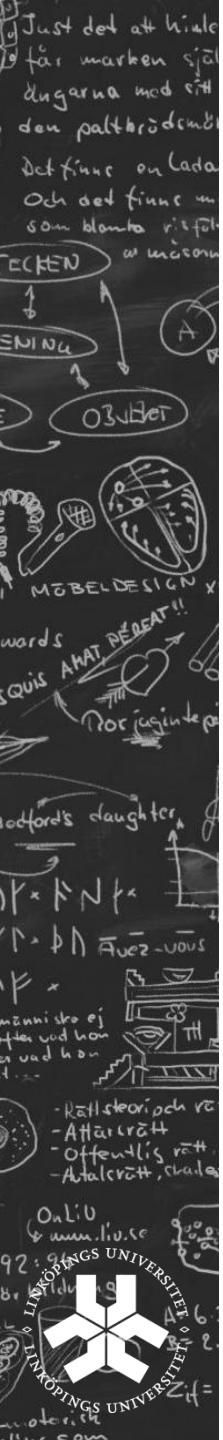
Linköping University
KTH Royal Institute of Technology
NICTA
HP Labs and University of Calgary



March 14, 2012

Motivation

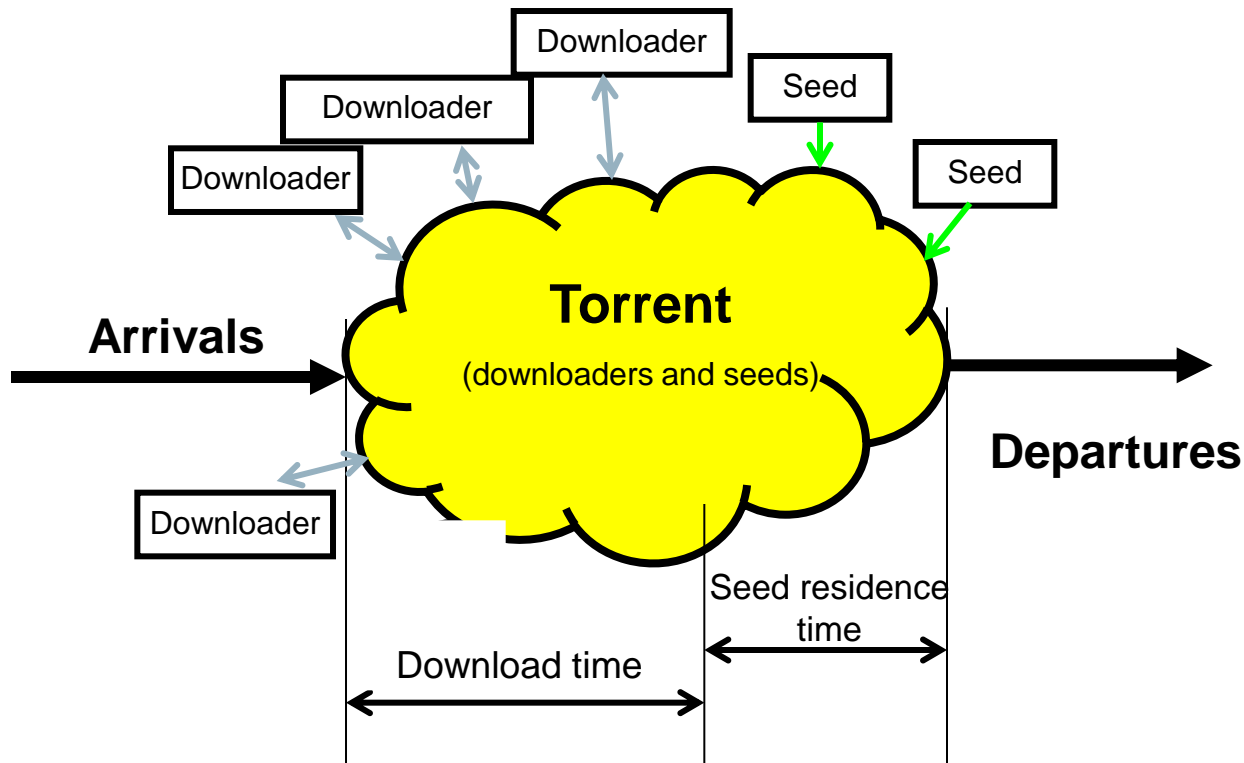
- Use of Internet for content delivery is massive ... and becoming more so
 - How to make scalable and efficient?
 - Server-based and peer-to-peer
- Chunk-based approach proven scalable
 - Files split into smaller chunks
 - Clients can download from both servers and other clients (peers)
- How to best manage large-scale content replication systems
 - E.g., where to place chunks?
 - Must first understand workload dynamics ...



Background: BitTorrent

Single file download

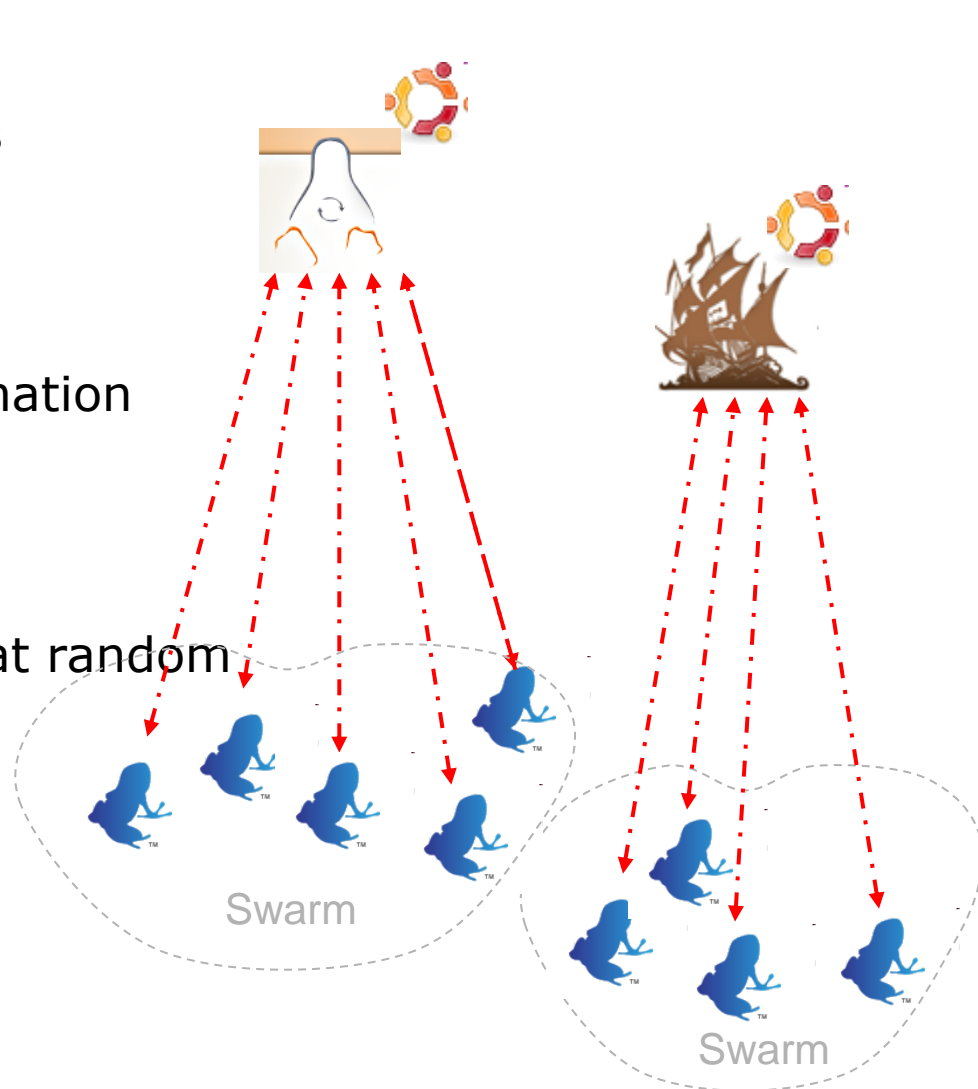
- File split into many smaller chunks
- Downloaded from both seeds and downloaders
- Distribution paths are dynamically determined
 - Based on data availability



Background: BitTorrent

Multi-tracked torrents

- Torrent file
 - "announce-list" URLs
- Trackers
 - Register torrent file
 - Maintain state information
- Peers
 - Obtain torrent file
 - Choose one tracker at random
 - Announce
 - Report status
 - Peer exchange (PEX)






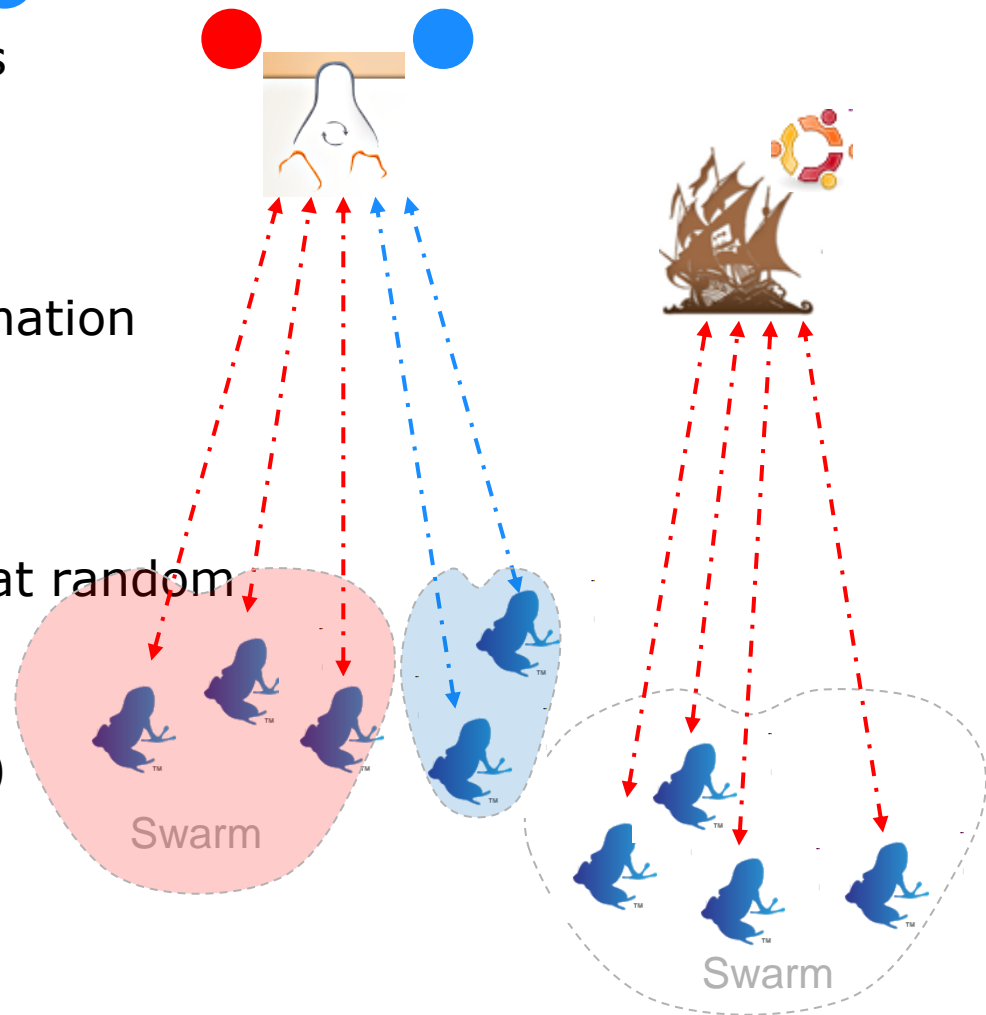
Just det att kiale
får marken själ
dugarna med sig
den paltbröden
det finns en lada
och det finns en
som blanda riktat
i mörk
ECKEN
ENINU
O3Ubet
MÖBELDESIGN
WARDS
SQUIS AKAT DÉBÉAT!!
Por jag inte p
mofford's daughter
Åvez-vous
- Kell teorin och va
- Affär rätt
- Offentlig rätt
- Atalcrätt, skade
OnLiU
www.liu.se
92
LÄNKÖPINGS UNIVERSITET



Background: BitTorrent

Multi-tracked torrents

- Torrent file  ● ●
 - "announce-list" URLs
- Trackers  
 - Register torrent file
 - Maintain state information
- Peers
 - Obtain torrent file
 - Choose one tracker at random
 - Announce
 - Report status
 - Peer exchange (PEX)



Just det att kiale
får marken själ
dugarna med sig
den paltbröden
det finns en lada
och det finns en
som blanda riktat
u masonu

ECKEN
ENINU
OZUbet

MÖBELDESIGN
WARDS
SQUIS AKAT DÉBÉAT!!
Por jaginte p

odford's daughter
BY x K N F x
V x P N Avez-vous
F x

männi sko ej
fater vad hon
en vad hon

- Rell teorin och va
- AHAU crätt
- Offentlig rätt
- Ahtalcrätt, skade

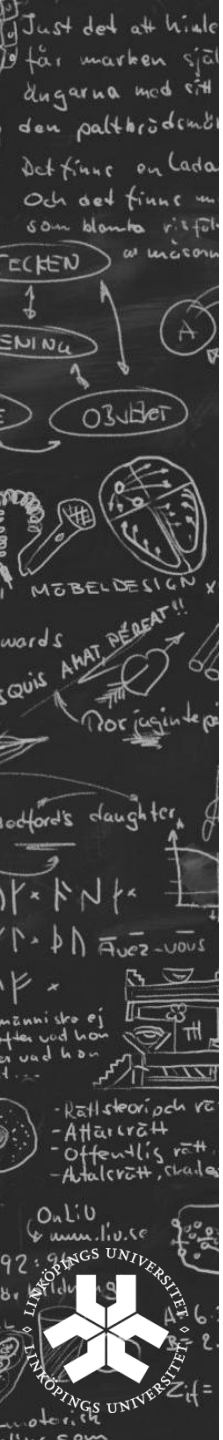
OnLiU
www.liu.se

92: g
LINKÖPINGS UNIVERSITET



Contributions

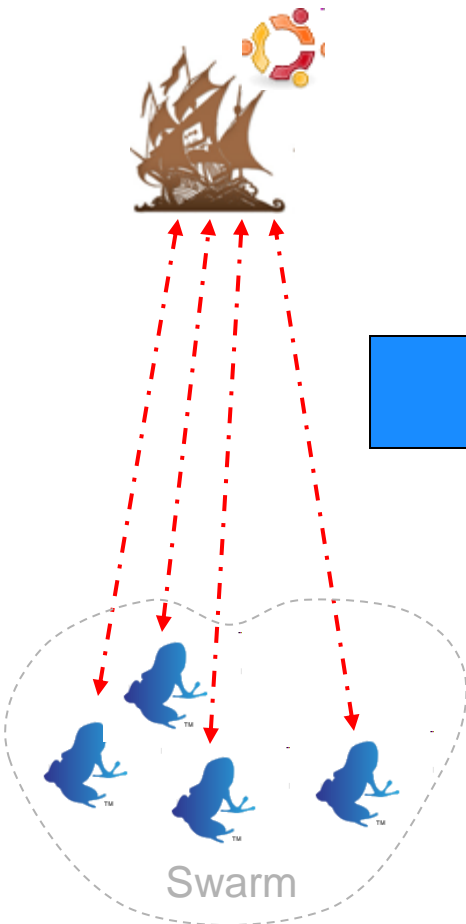
- Longitudinal multi-torrent analysis
 - 48 weeks from two vantage points
- Capturing differences in dynamics observed locally and globally
 - University campus vs. global tracker-based
- Example observations
 - Campus users download larger files
 - Campus users early adopters (except music)
 - High popularity churn
 - Most popular content peak later



Measurement overview

Active + passive measurements

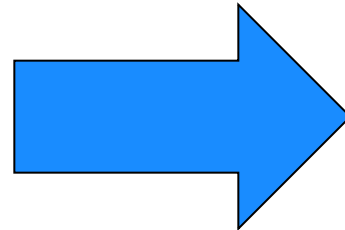
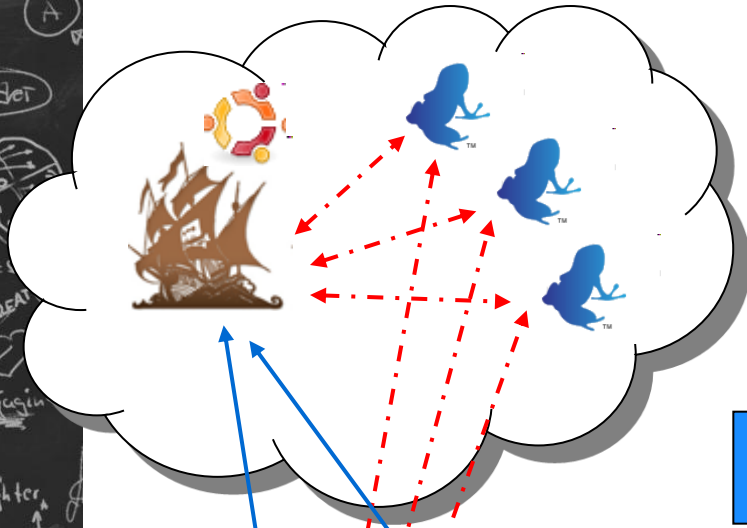
Popularity dynamics



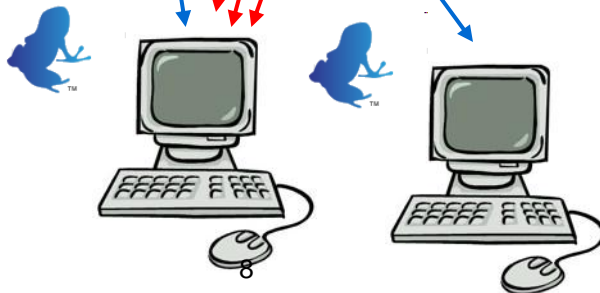
- Longitudinal data
- Two vantage points
 - University campus (ingress/egress)
 - Global trackers

University: tracker communication

Passive measurements



Extract HTTP peer-to-tracker traffic at campus ingress/egress



Just det att hiale
får marcken sjä
dugarna med sig
den paltbröden
det finns en lada
och det finns m
som blanda riktat
w mason

EKEN
ENINU
O3Ubet

MÖBELDE
wards
SQUIS AKAT DÉREAT
Por iugin

odford's daughter
P * K N P x
P * Avez-vous
P *

männi sko ej
fter vad hon
en vad hon

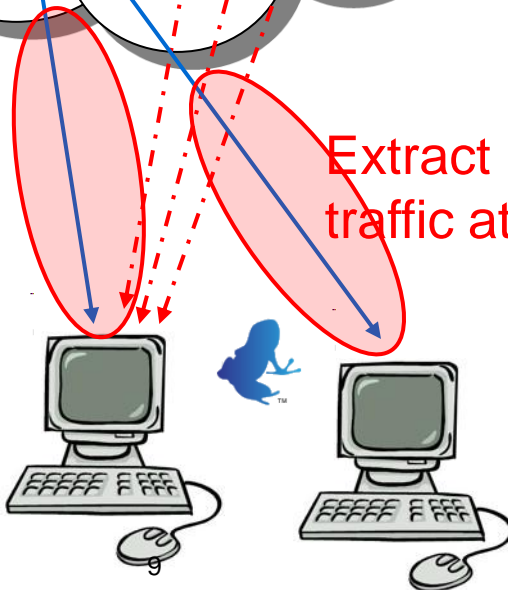
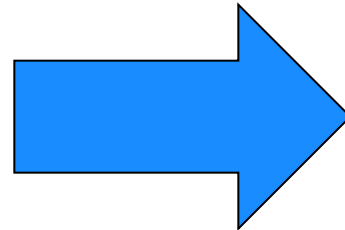
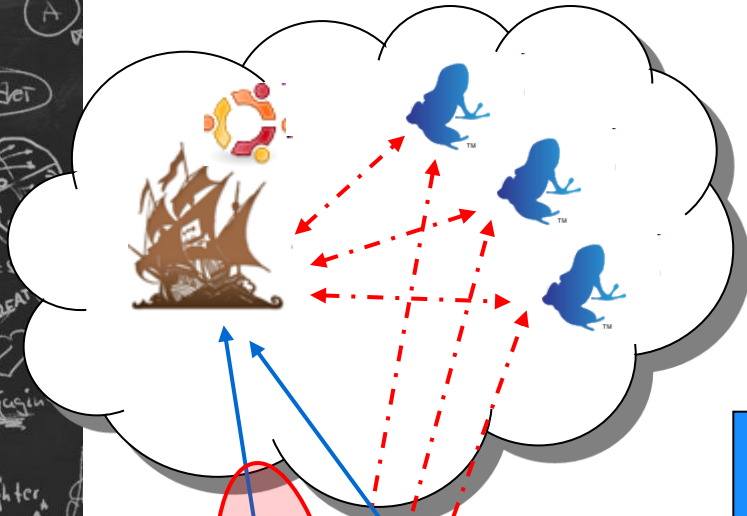
- Rätt teori och r
- Affär rätt
- Offentlig rätt
- Arterätt, skade

OnLiU
www.liu.se

92: g
B
LINKÖPINGS UNIVERSITET

University: tracker communication

Passive measurements



Extract HTTP peer-to-tracker traffic at campus ingress/egress

Just det att kiale
får marke sjö
dugarna med s
den paltbröden
det frunc on lada
Och det frunc m
som blanta riefat
w mason

EKEN
ENINU
O3Ubet

MÖBELDE
wards
SQUIS AHAT DEBEAT
Por iugiu

odford's daughter
Åvez-vous
minni sko ej
fter vad hon
en vad hon

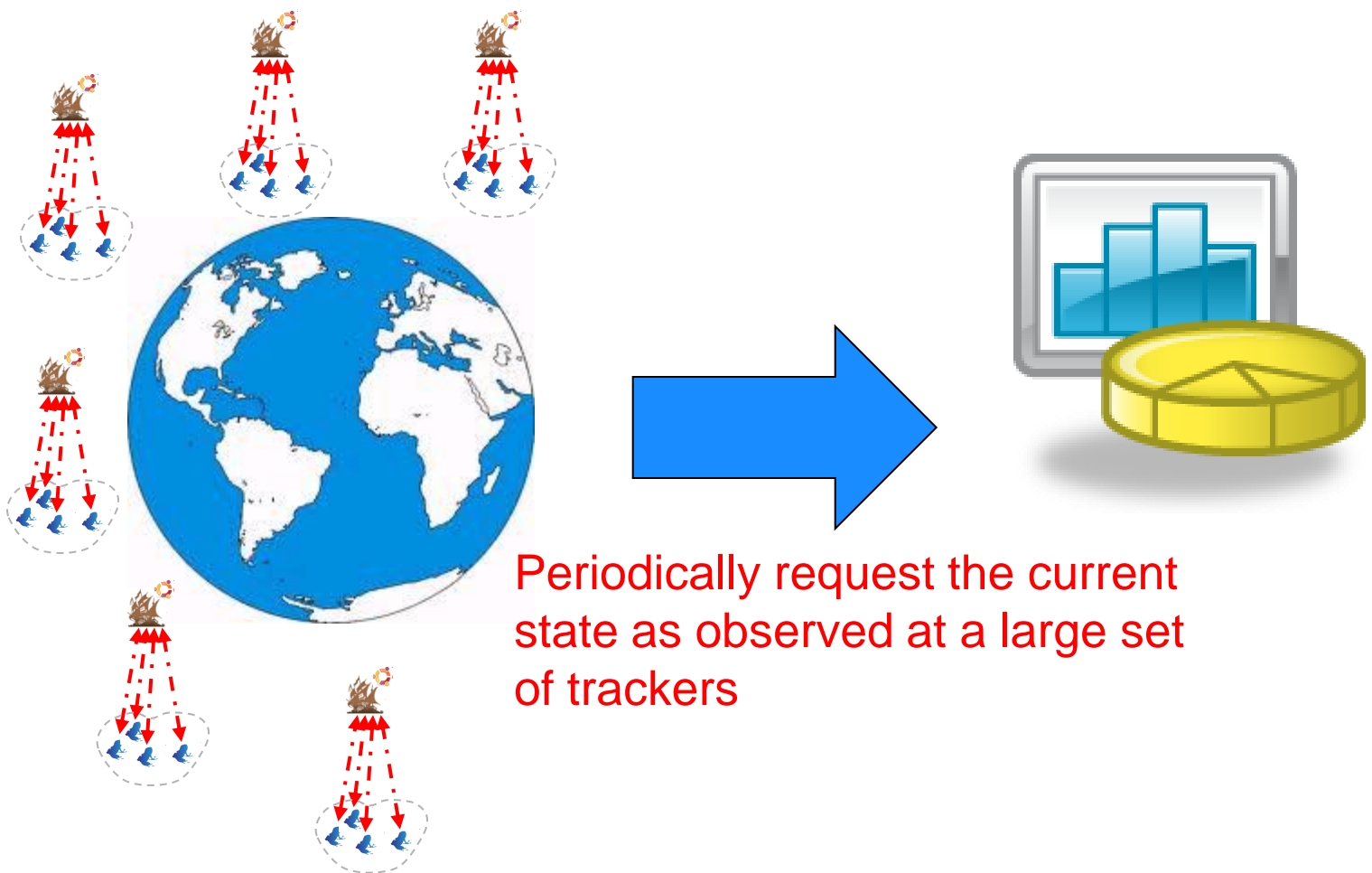
- Rätt teori och r
- Affär rätt
- Offentlig rätt
- Arterätt, skade

OnLiU
www.liu.se

LINKÖPINGS UNIVERSITET

Global: Tracker scrapes

Active measurements



Periodically request the current state as observed at a large set of trackers

Just det att kiale
fär märken sjä
dugarna med sja
den paltbröden
det finns en lada
och det finns m
som blanda riefat
u masonu

ECKEN
ENINU
OBJET

MÖBELDESIGN

words
SQUIS AKAT DÉBÉAT!!
Por jicinte p

ford's daughter

Åvez-vous

minni sko ej
fiter vad hon
en vad hon

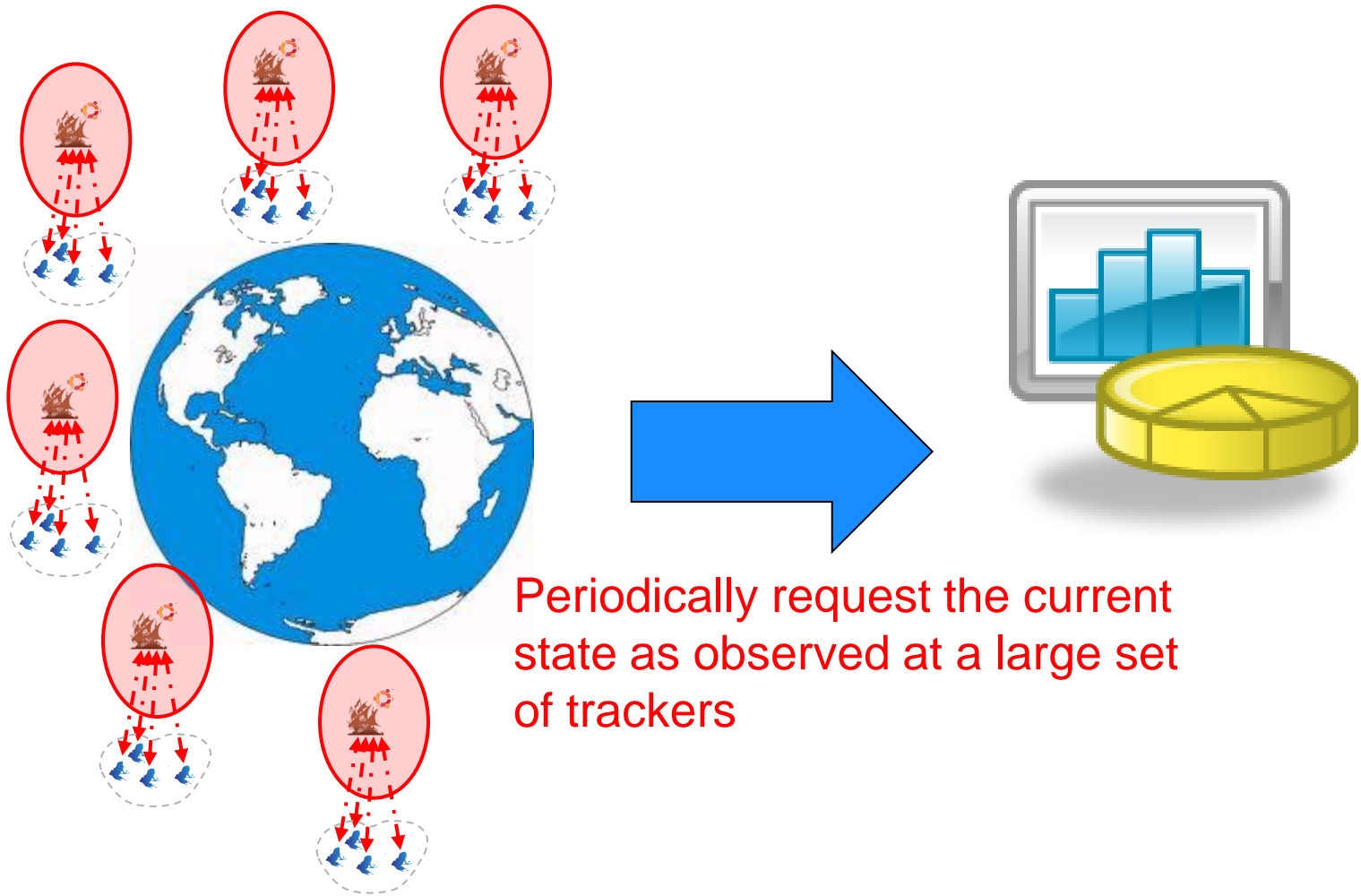
- Rätt teori och va
- Affär rätt
- Offentlig rätt
- Affär rätt, skola

OnLiU
www.liu.se

LINKÖPINGS UNIVERSITET

Global: Tracker scrapes

Active measurements



Periodically request the current state as observed at a large set of trackers

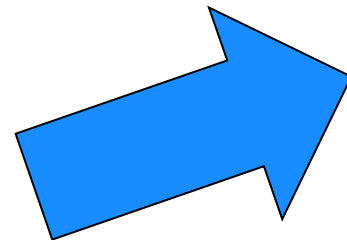
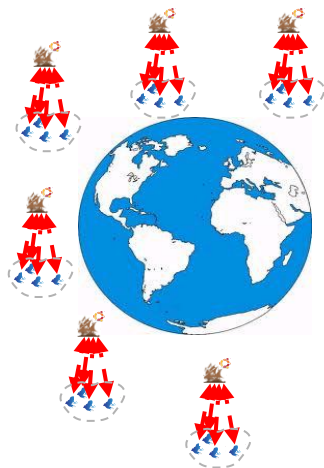
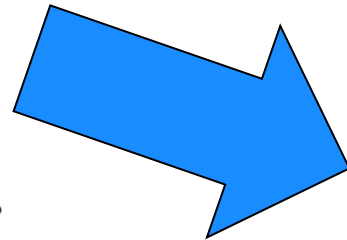
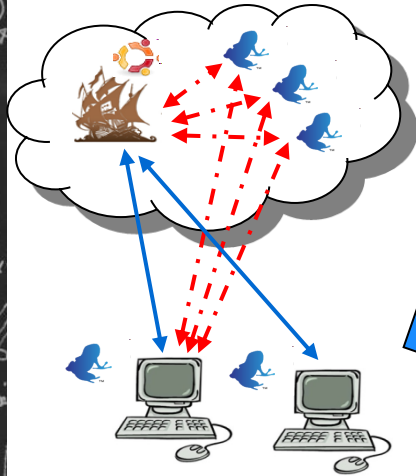
Just det att kiale
får marken själ
dugarna med sig
den paltbröden
det finns en lada
och det finns m
som blanda riefat
u masonu
ECKEN
ENINU
O3Ubet
MÖBELDESIGN
wards
SQUIS AKAT DÉBÉAT!!
Por jaginte p
mofford's daughter
Åvez-vous
- Rell teori och va
- Aftär rätt
- Offentlig rätt
- Aftär rätt, skola
OnLiU
www.liu.se
92
LINKÖPINGS UNIVERSITET



Measurement overview

Active + passive measurements

Popularity dynamics



Just det att kiale
fär markeu sjö
dugarna med s
den paltbrödem
det frunc on lada
Och det frunc m
som blanta riefat
w masonu

ECKEN
ENINU
O3Ubet

MÖBELDESIGN
wards
SQUIS AKAT DÉBÉAT!!
Por jaginte p

odford's daughter
Åvez-vous
männi sko ej
fter vad hon
en vad hon

- Rätt teori och rät
- Affär rät
- Offentlig rät
- Aftal rät, skale

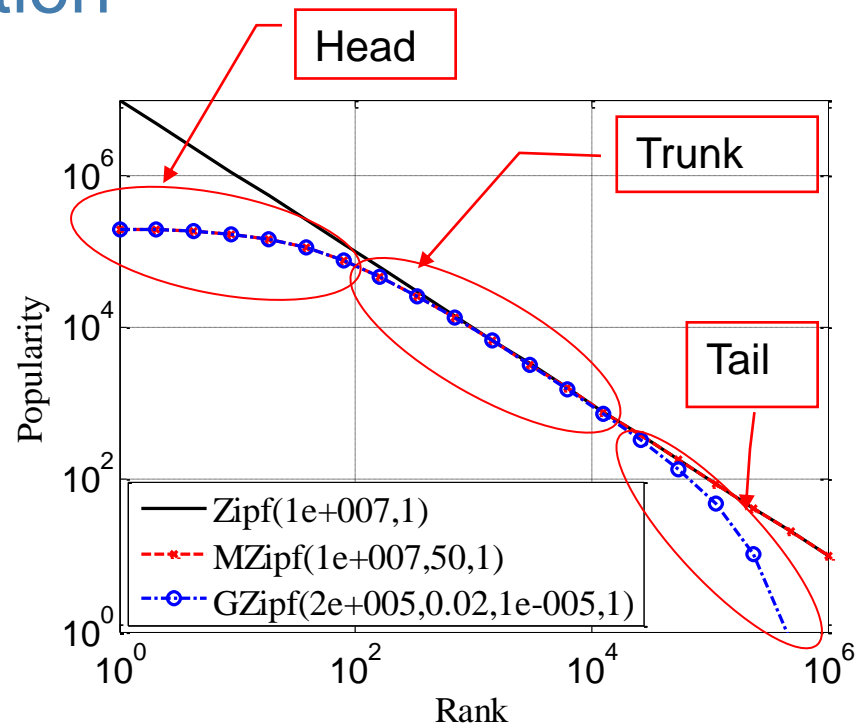
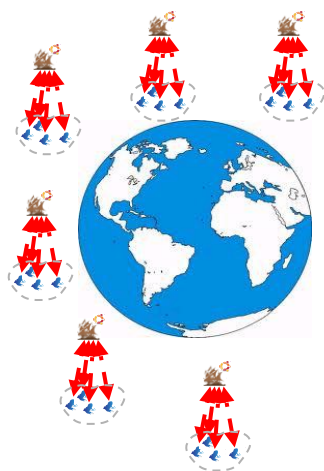
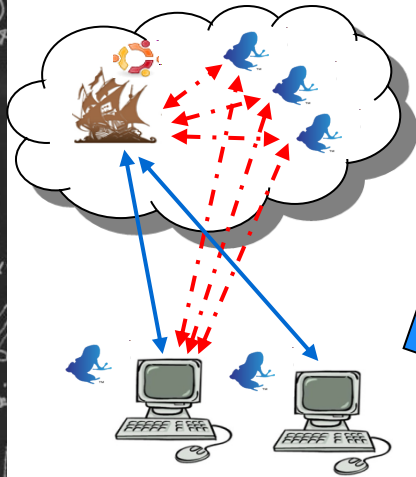
OnLiU
www.liu.se

92: g
Br lidu

LINKÖPINGS UNIVERSITET

Previous work

Popularity distribution



E.g., Dan & Carlsson [IPTPS 2010]

- Popularity distribution statistics
 - Over lifetime
 - Over different time period
 - Different sampling methods

Summary of datasets

Property	University	Global	Mininova
Trackers	2,371	721	1,690
Torrents	56,963	11.2 M	911,687
Downloads	1.73 M	37.0 B	--
HTTP requests	249 M	--	--
Start date	Sep. 15, 2008	Sep. 15, 2008	Sep., 2008
End date	Aug. 17, 2009	Aug. 17, 2009	Aug., 2009
Frequency	All requests	Weekly scrapes	Twice



Summary of datasets

Property	University	Global	Mininova
Trackers	2,371	721	1,690
Torrents	56,963	11.2 M	911,687
Downloads	1.73 M	37.0 B	--
HTTP requests	249 M	--	--
Start date	Sep. 15, 2008	Sep. 15, 2008	Sep., 2008
End date	Aug. 17, 2009	Aug. 17, 2009	Aug., 2009
Frequency	All requests	Weekly scrapes	Twice

- 48 weeks of overlapping longitudinal data



Summary of datasets

Property	University	Global	Mininova
Trackers	2,371	721	1,690
Torrents	56,963	11.2 M	911,687
Downloads	1.73 M	37.0 B	--
HTTP requests	249 M	--	--
Start date	Sep. 15, 2008	Sep. 15, 2008	Sep., 2008
End date	Aug. 17, 2009	Aug. 17, 2009	Aug., 2009
Frequency	All requests	Weekly scrapes	Twice

- Many torrents (and downloads) ...

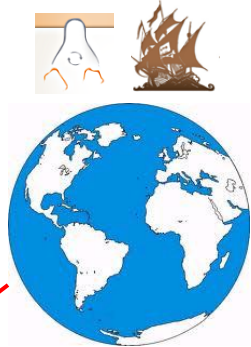


Dataset summary

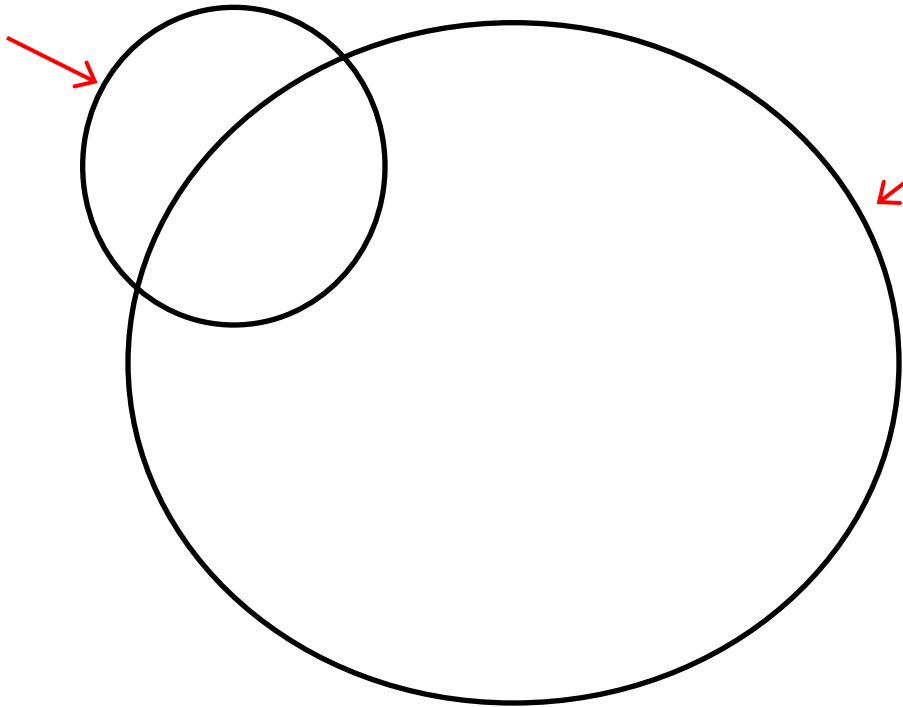
Torrents observed



56,963



11.2 M

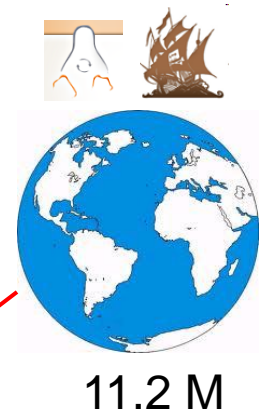
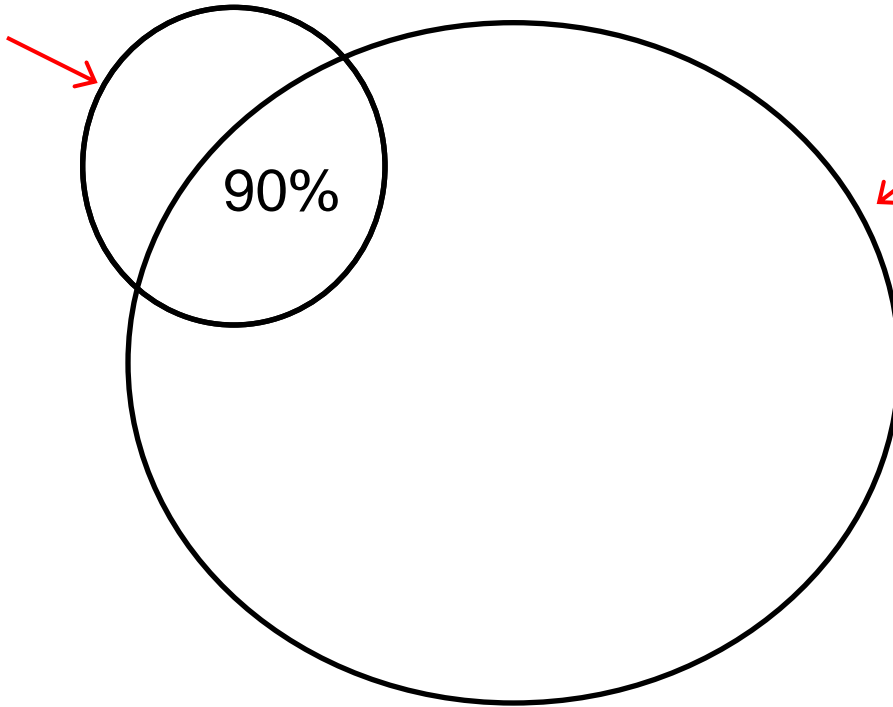


Dataset summary

Torrents observed



56,963



Just det att kiale
får marben sjä
dugarna med s
den paltbrödem
det frunc on lada
Och det frunc m
som blanta riefat
w masonu

ECKEN
ENINU
O3Ubet

MÖBELDESIGN

wards
QUIS AHAT DÉBÉAT!!
Por juscinte p

odford's daughter

DP x KN P x
P x P x Auez -vous
P x

minni sko ej
fter vad hon
en vad hon

- Rätt teori och r
- Åtalsrätt
- Offentlig rätt
- Åtalsrätt, skade

OnLiU
www.liu.se

92: g
B. l
LINKÖPINGS UNIVERSITET

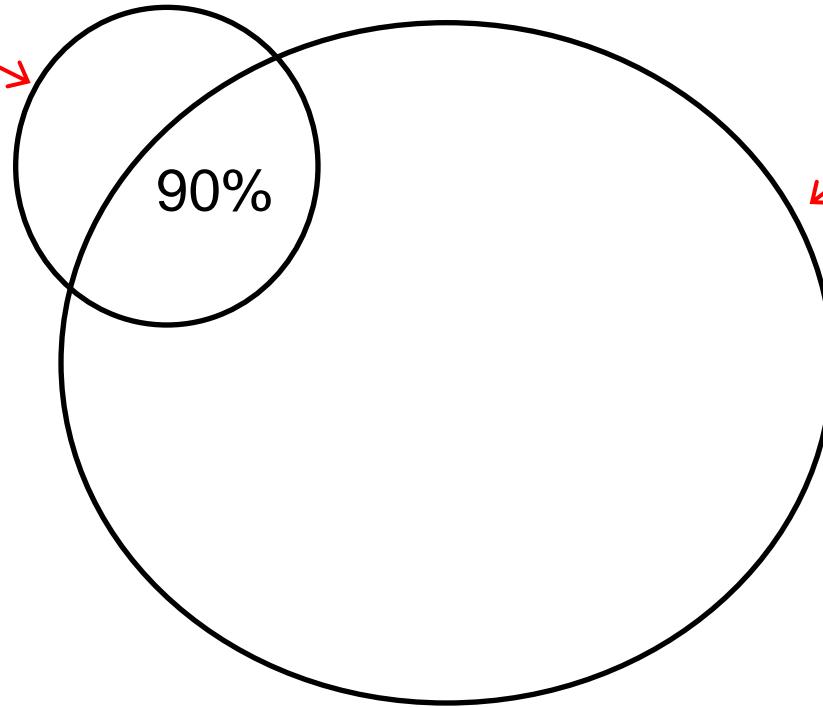


Dataset summary

Torrents observed



56,963



- Most of the files observed locally are also observed in the global dataset

Dataset summary

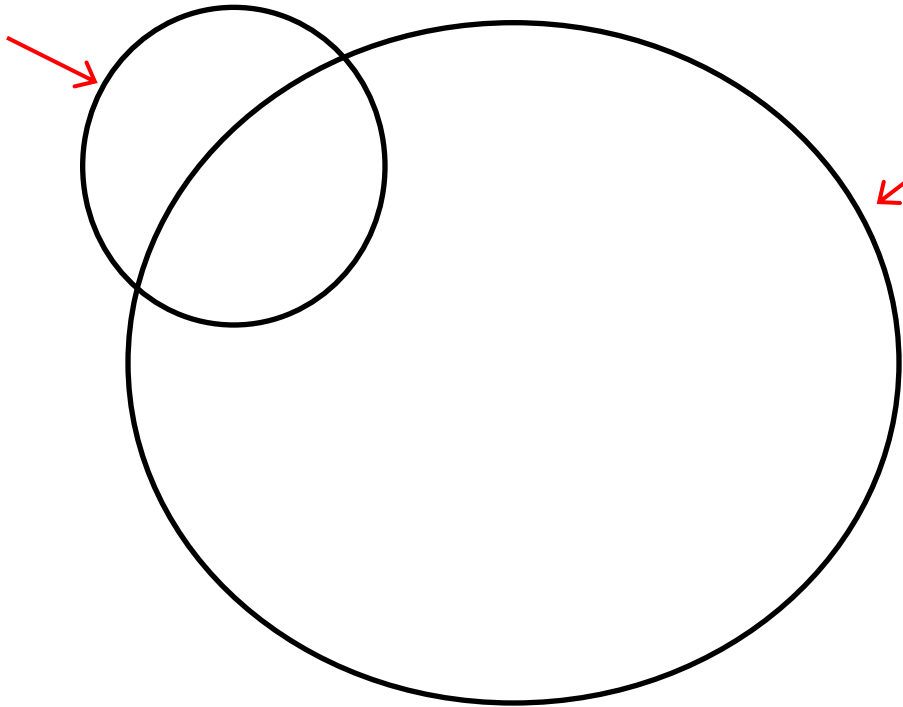
Torrents observed



56,963



11.2 M



Dataset summary

Torrents observed



56,963

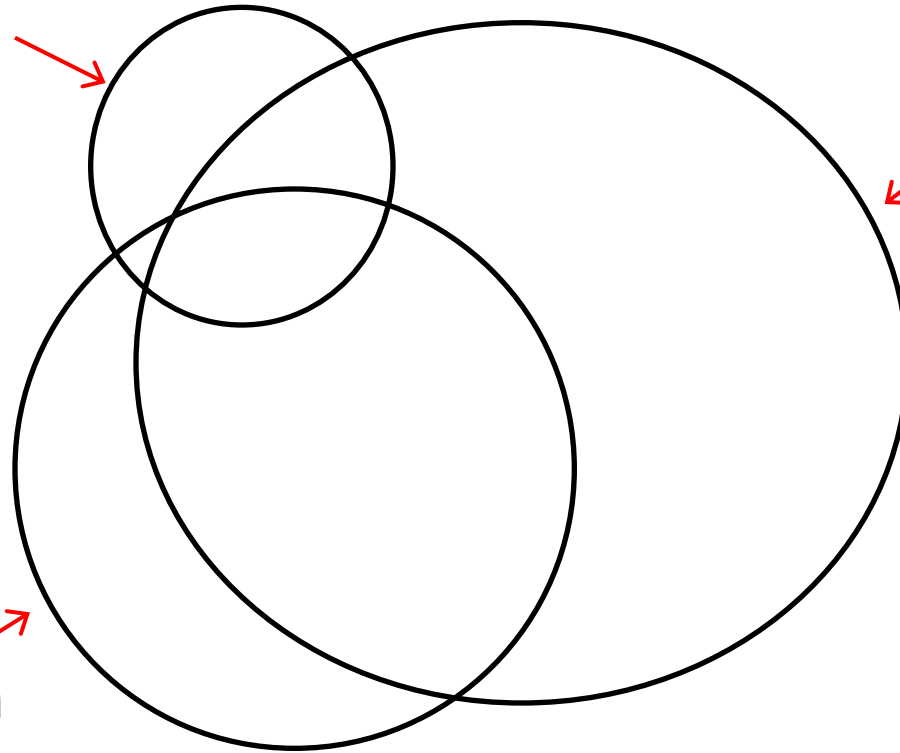
mininova



911,687



11.2 M



Dataset summary

Torrents observed



56,963

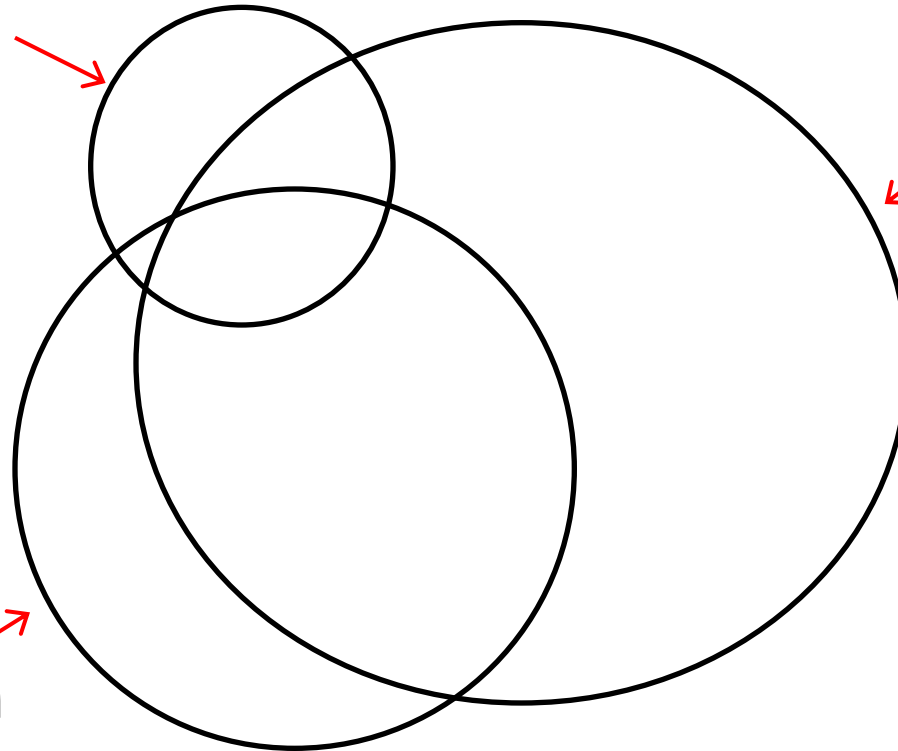
mininova



911,687



11.2 M



- Mininova screen scrapes also provide us with **size** and **category** information for some of these files

Dataset summary

Torrents observed



56,963

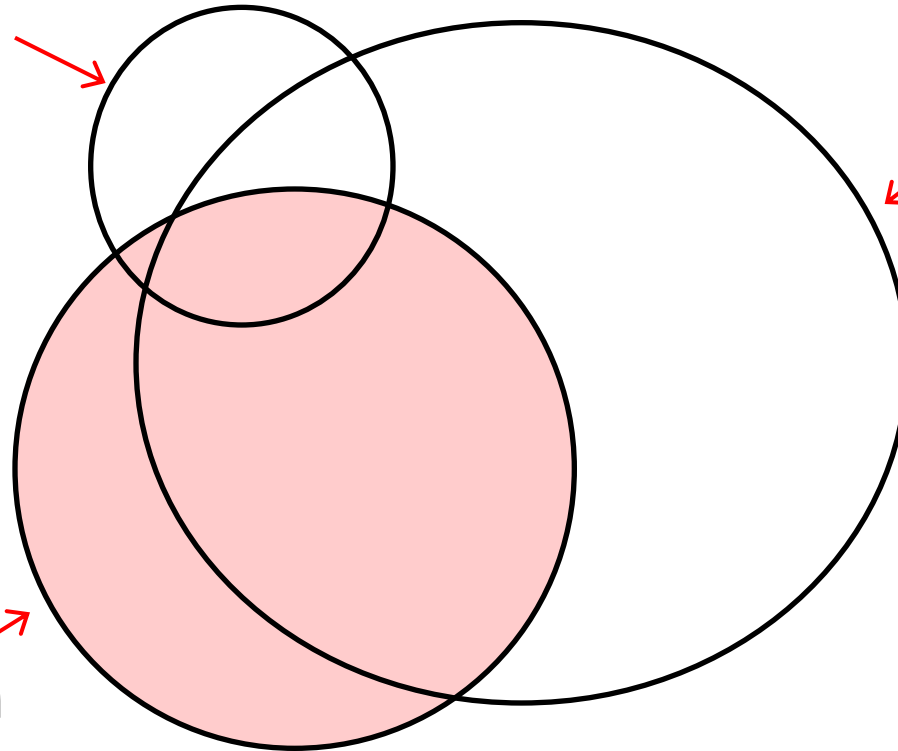
mininova



911,687



11.2 M



- Mininova screen scrapes also provide us with **size** and **category** information for some of these files

Dataset summary

Torrents observed

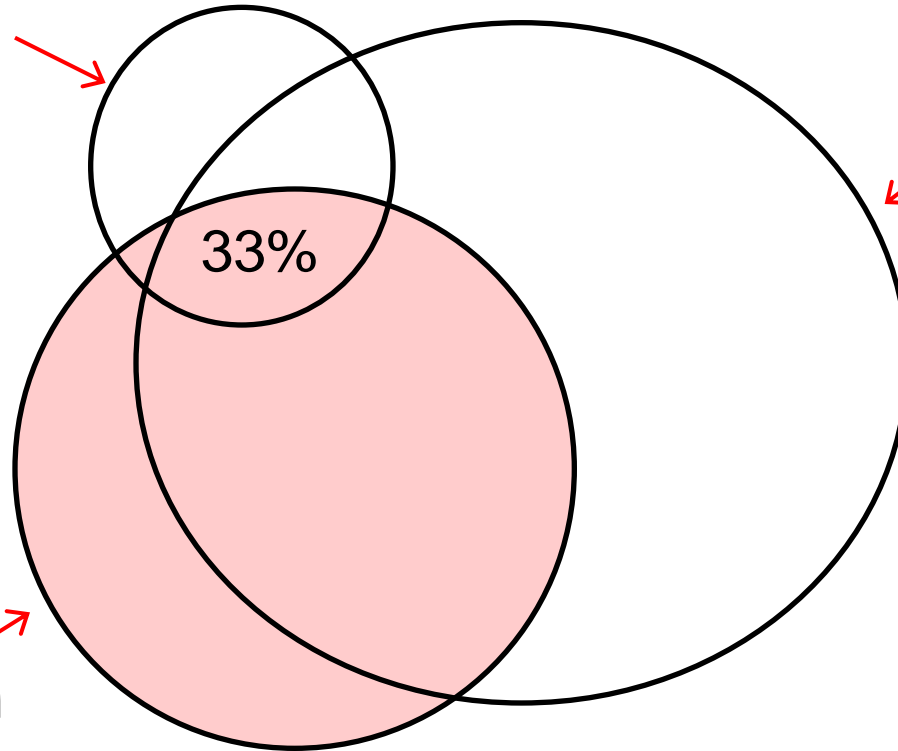


56,963

mininova



911,687

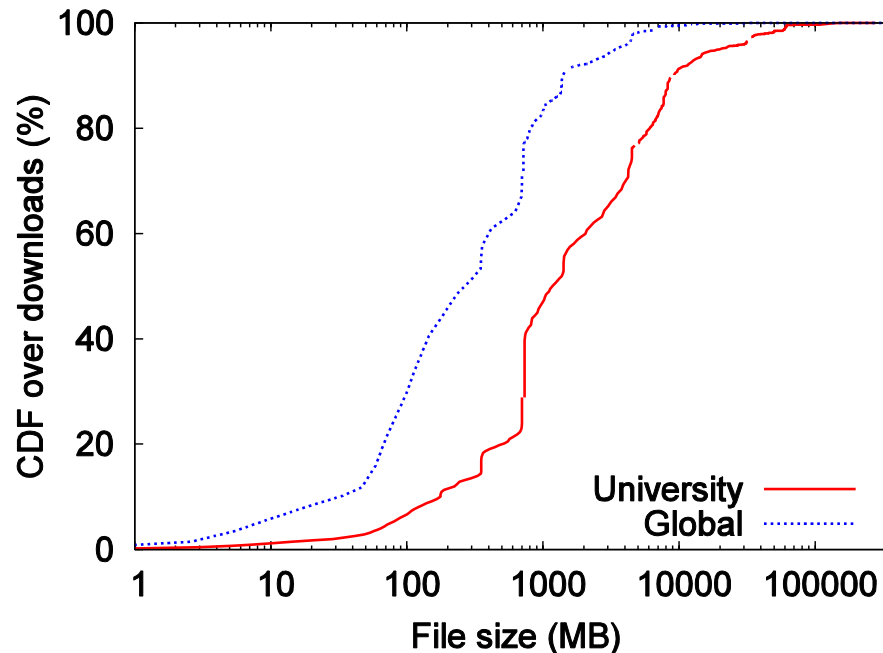


11.2 M

- Mininova screen scrapes also provide us with **size** and **category** information for some of these files

Content download characteristics

File size distribution, per download

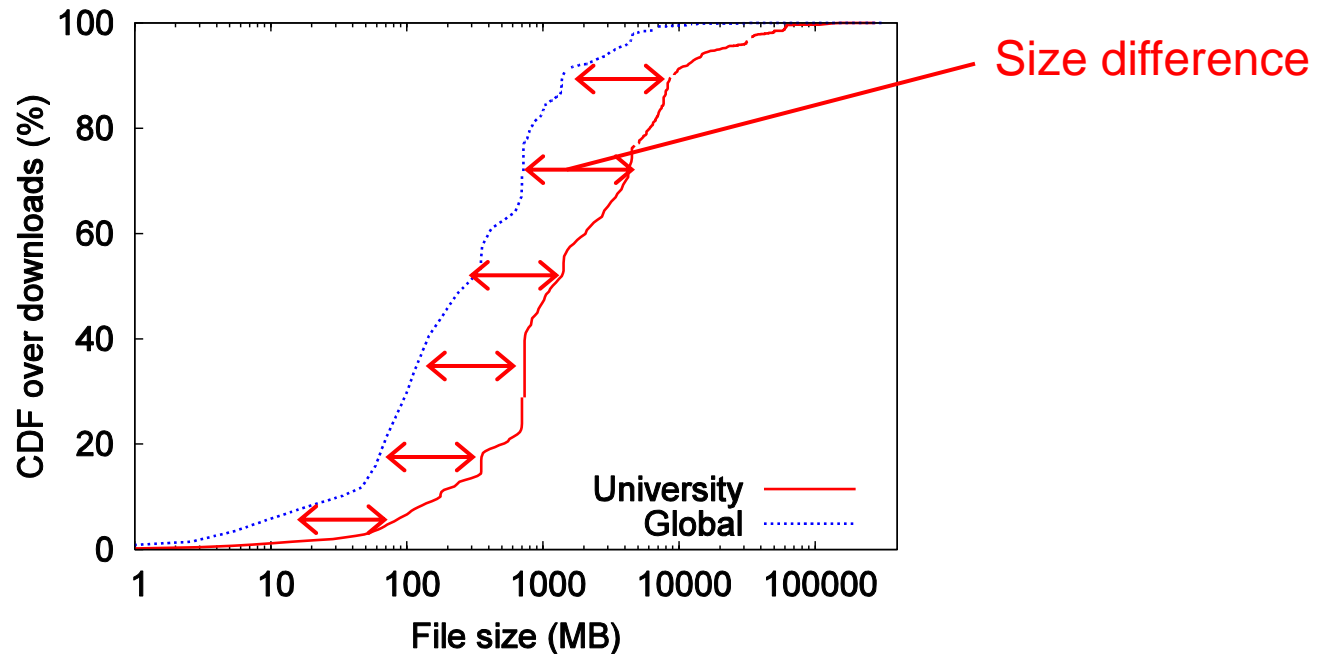


- Campus users download larger files



Content download characteristics

File size distribution, per download

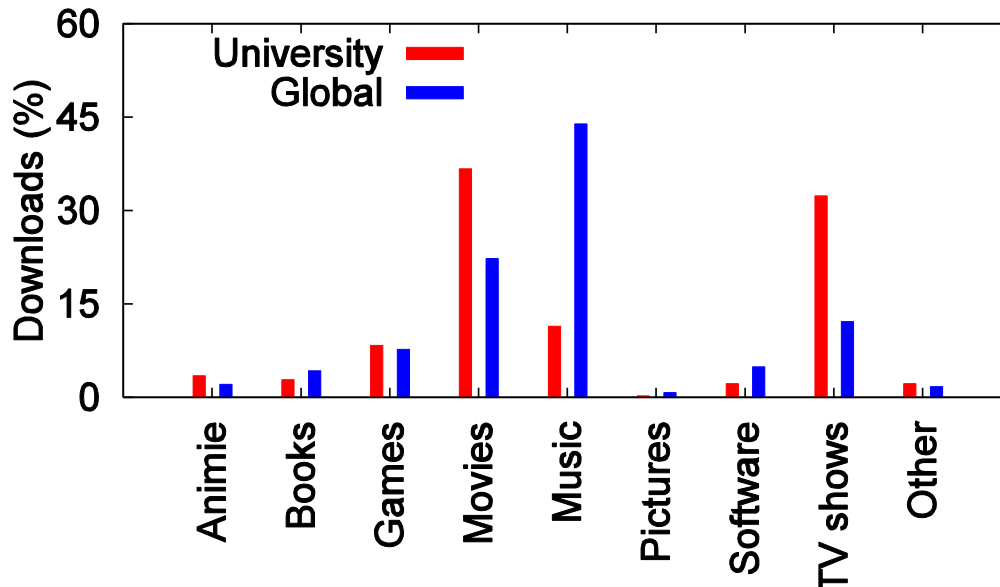


- Campus users download **larger** files



Content download characteristics

Breakdown per category

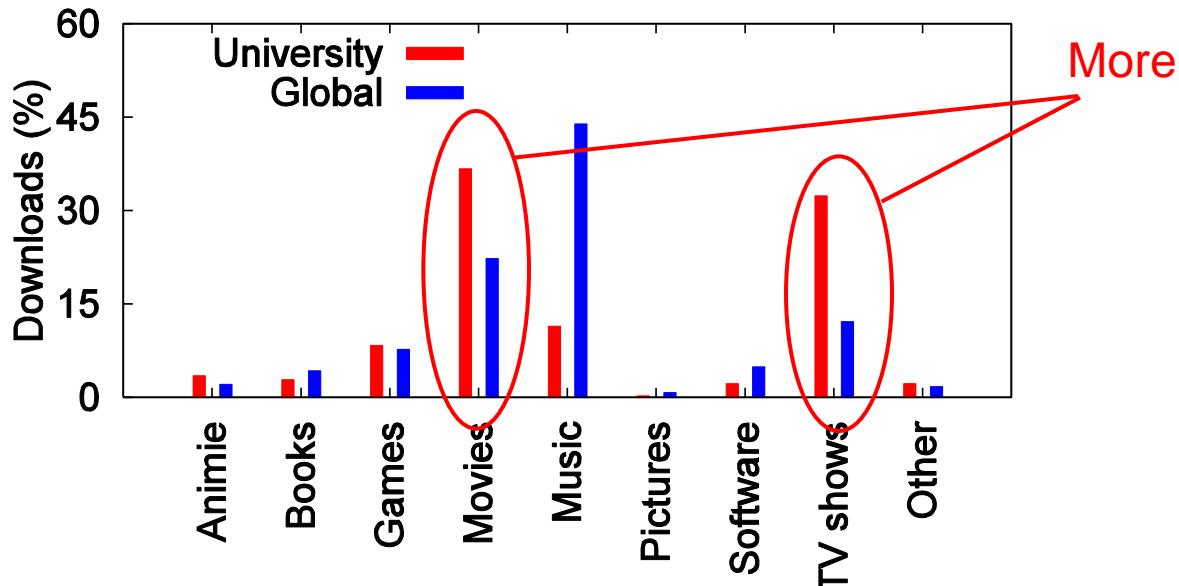


- Campus users download
 - More movies and TV shows
 - Less music



Content download characteristics

Breakdown per category

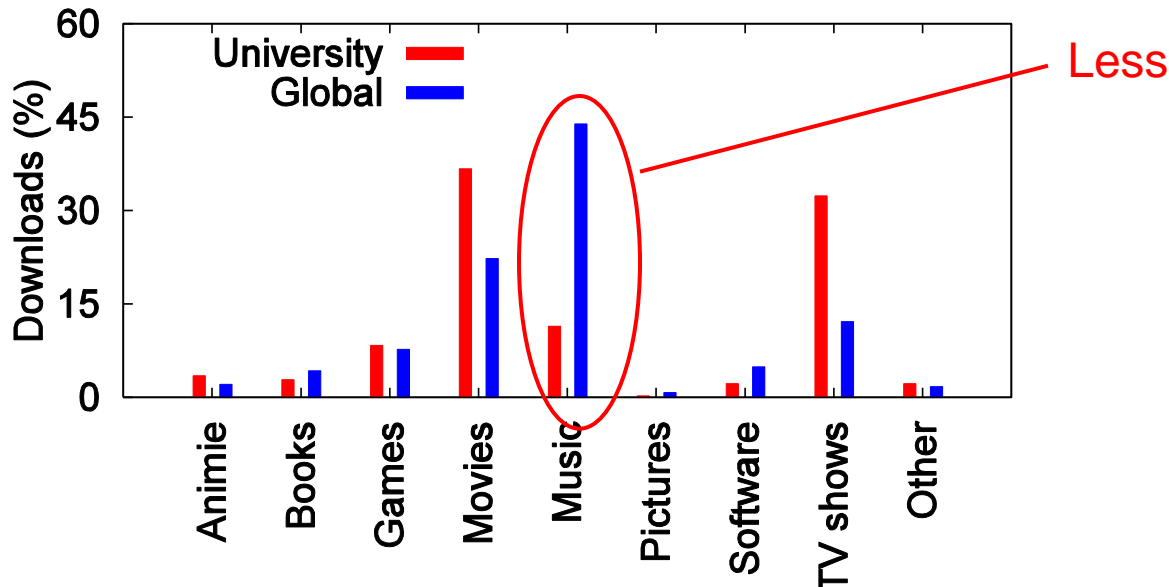


- Campus users download
 - More movies and TV shows
 - Less music



Content download characteristics

Breakdown per category

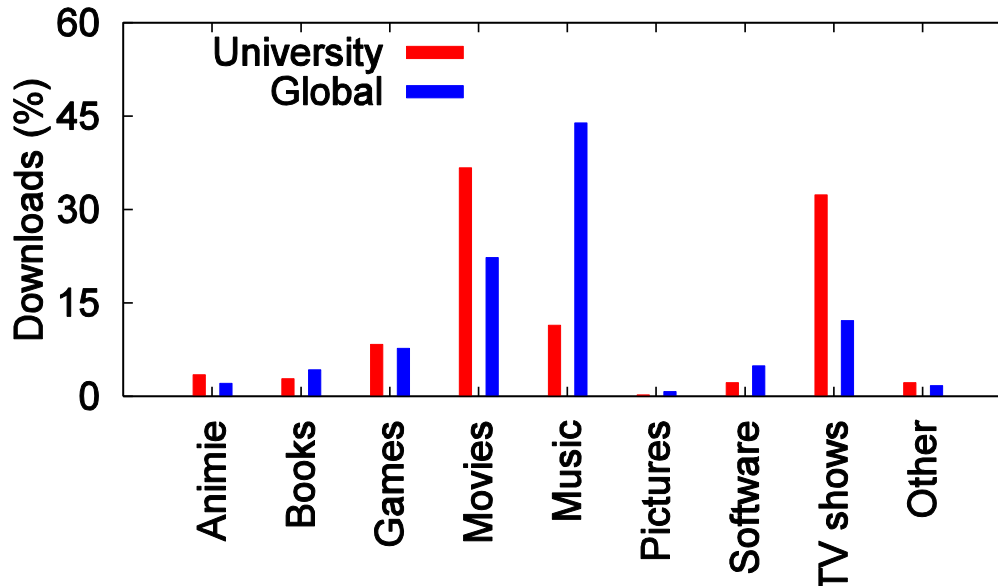


- Campus users download
 - More movies and TV shows
 - **Less music**

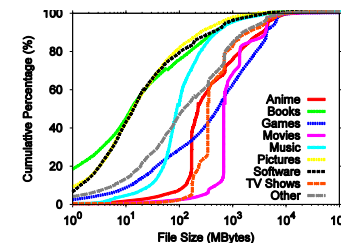


Content download characteristics

Breakdown per category

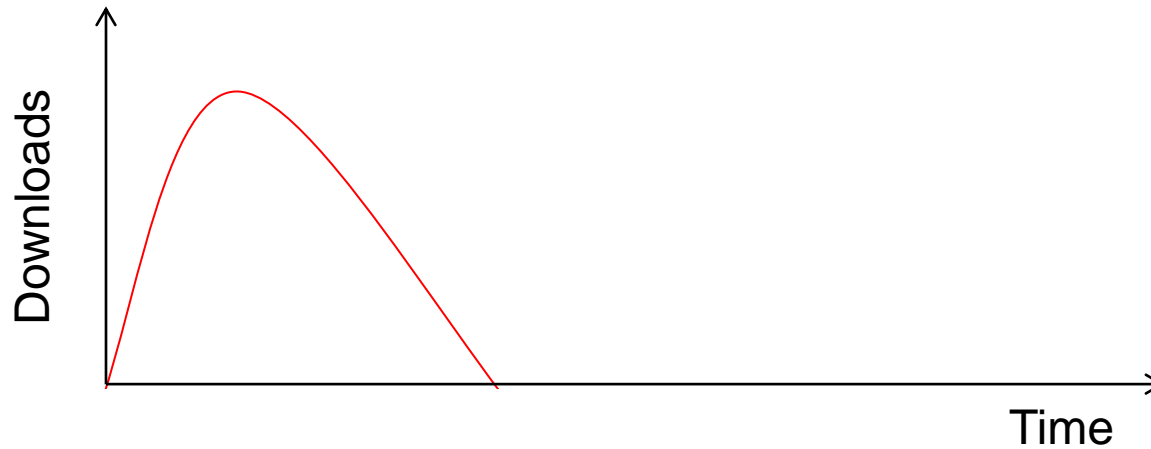


- Campus users download
 - More movies and TV shows
 - Less music
- Again, biased towards larger contents ...



Early adopters

Terminology



Just det att kiale
fär marben sjä
dugarna med sif
den paltbröden
det finns en lada
och det finns m
som blanda riefat
u masonu

EKEN
ENINU
OBJET

MÖBELDESIGN

wards
SQUIS AHAT DÉBÉAT!!
Por jicinte p

odford's daughter

DP x KN Px
P x PN Avez-vous


männi sko ej
fter vad hon
en vad hon

- Rätt teori och rät
- Affär rätt
- Offentlig rätt
- Aralar rätt, skade

OnLiU
www.liu.se

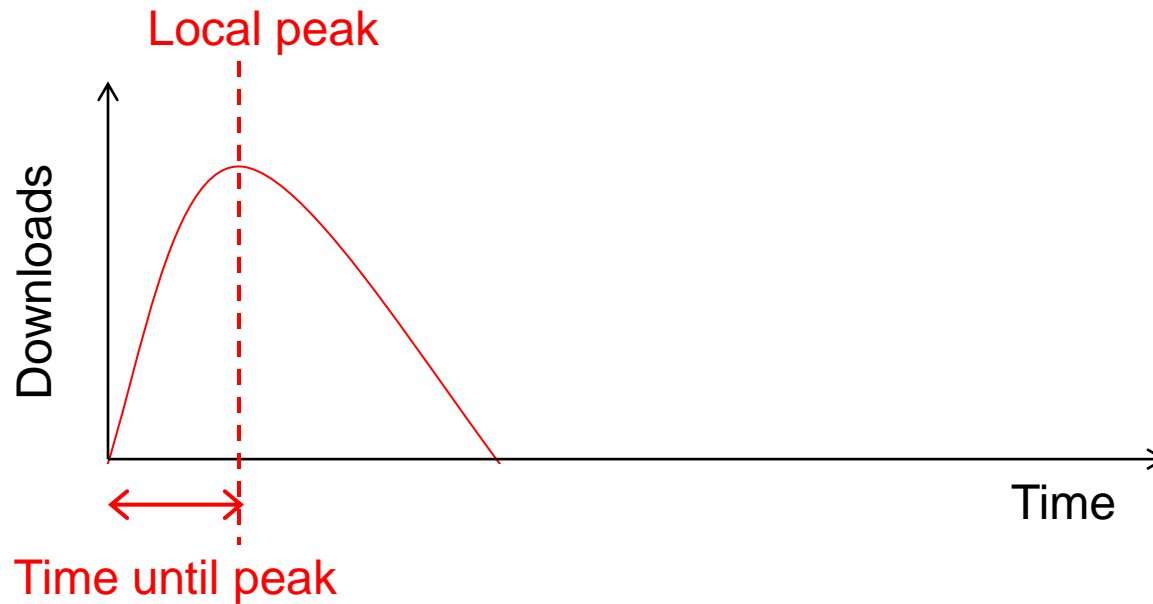
92: 92

LINKÖPINGS UNIVERSITET



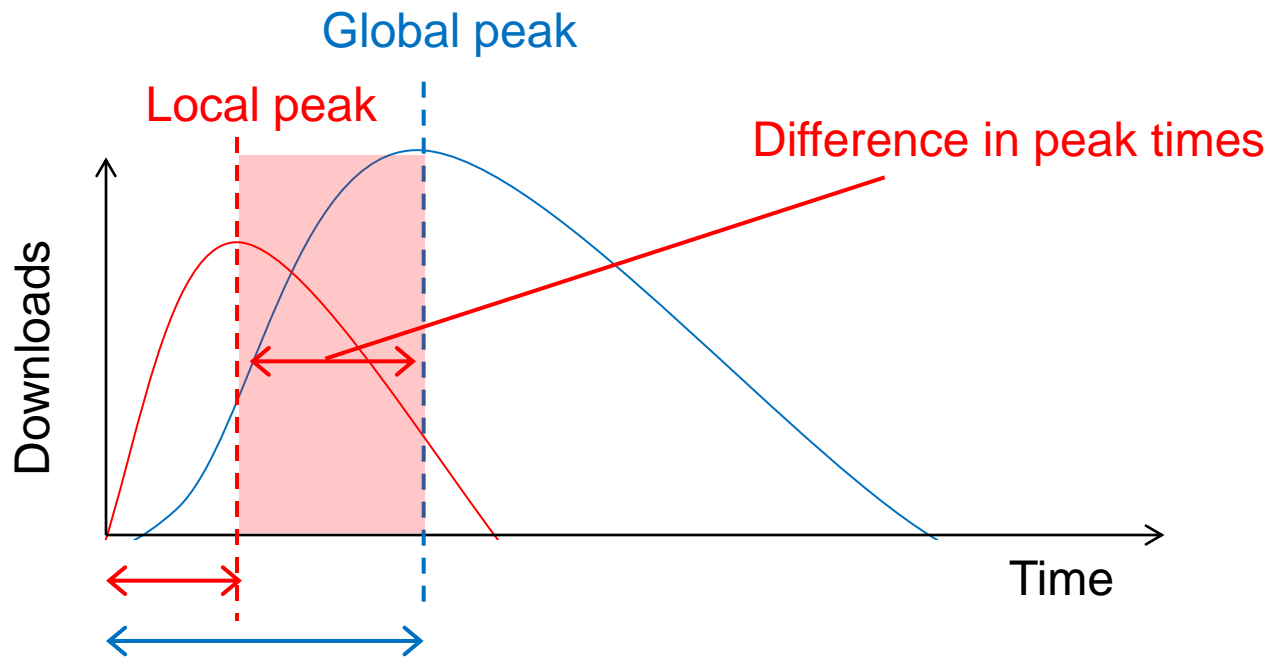
Early adopters

Terminology



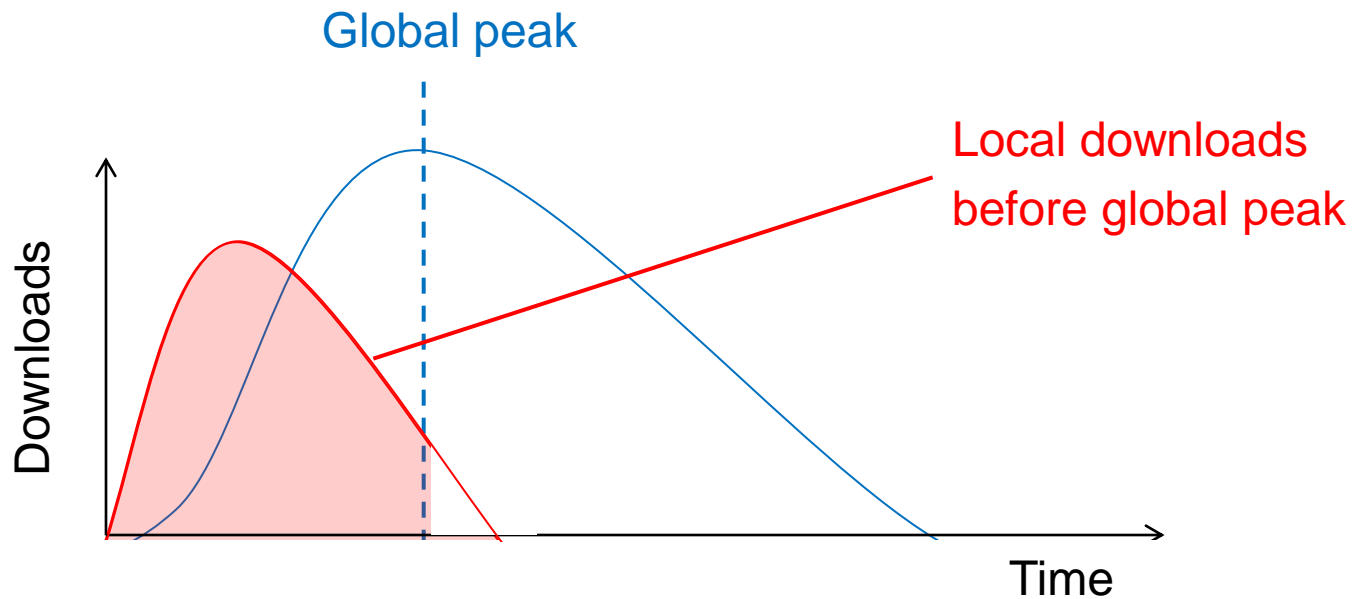
Early adopters

Terminology



Early adopters

Terminology



Just det att kiale
fär marben själ
dugarna med sif
den paltbröden
det finns en lada
och det finns en
som blanda riefat
u masonu

ECKEN
ENINU
OBJET

MÖBELDESIGN

wards
SQUIS AKAT DÉBÉAT!!
Por juginte p

odford's daughter

DP x KN P x
P x P x Avez-vous

minni sko ej
fter vad hon
er vad hon

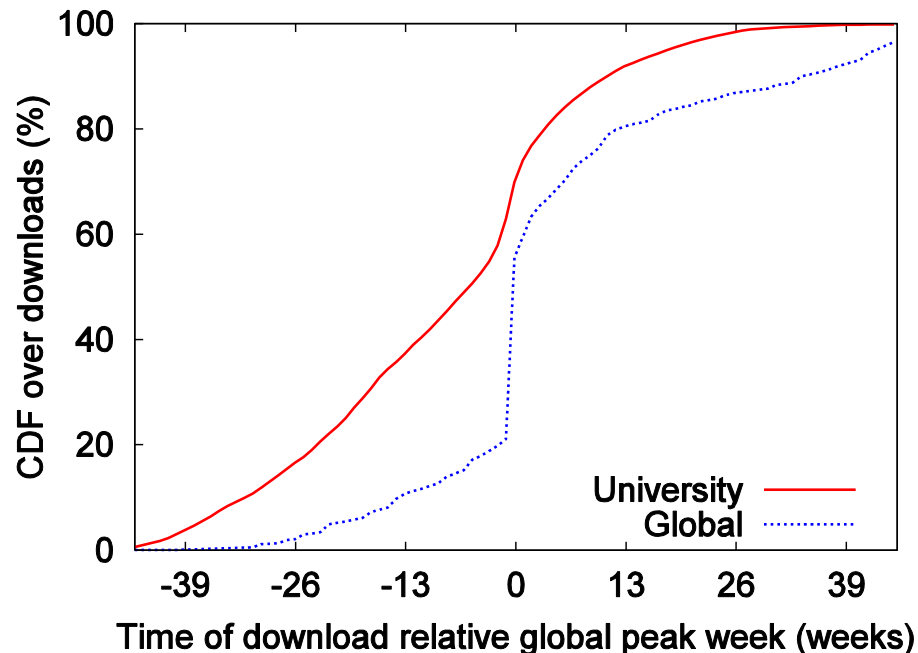
- Rätt teori och rät
- Affärsrätt
- Offentlig rätt
- Arterätt, skade

OnliU
www.liu.se

92: g
B. l
LINKÖPINGS UNIVERSITET

Early adopters

Downloads relative to global peak

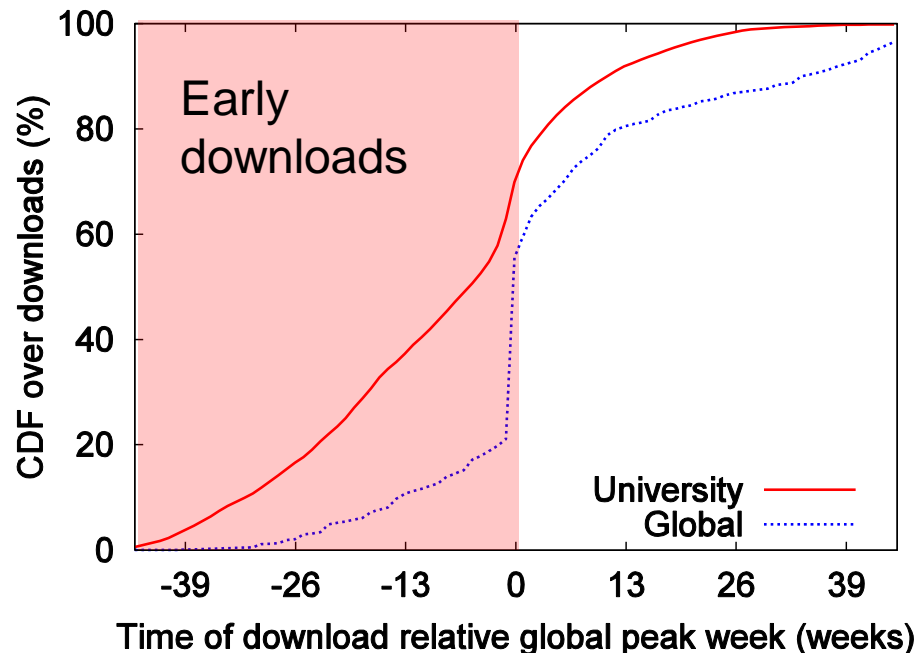


- Campus users are generally early adopters of content
 - 70% of downloads before global peak
 - 40% of downloads at least 10 weeks before global peak



Early adopters

Downloads relative to global peak

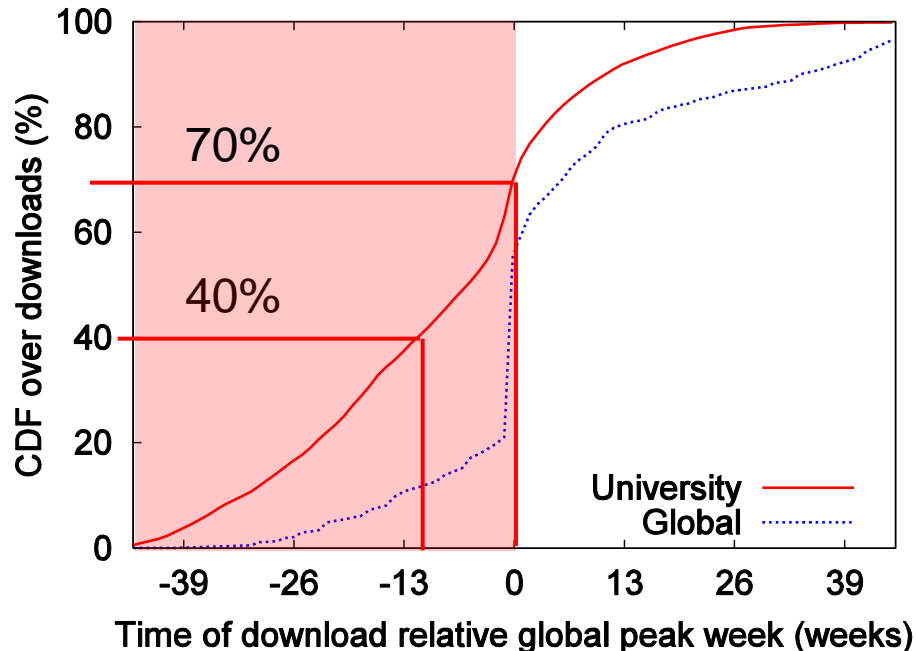


- Campus users are generally early adopters of content
 - 70% of downloads before global peak
 - 40% of downloads at least 10 weeks before global peak



Early adopters

Downloads relative to global peak

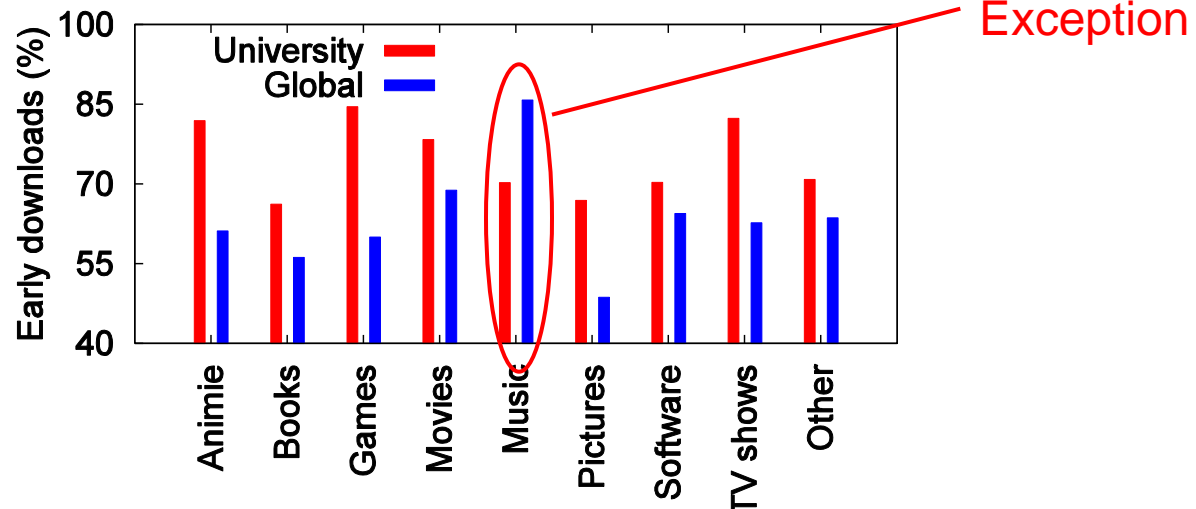


- Campus users are generally early adopters of content
 - 70% of downloads before global peak
 - 40% of downloads at least 10 weeks before global peak



Early adopters

Downloads relative to global peak

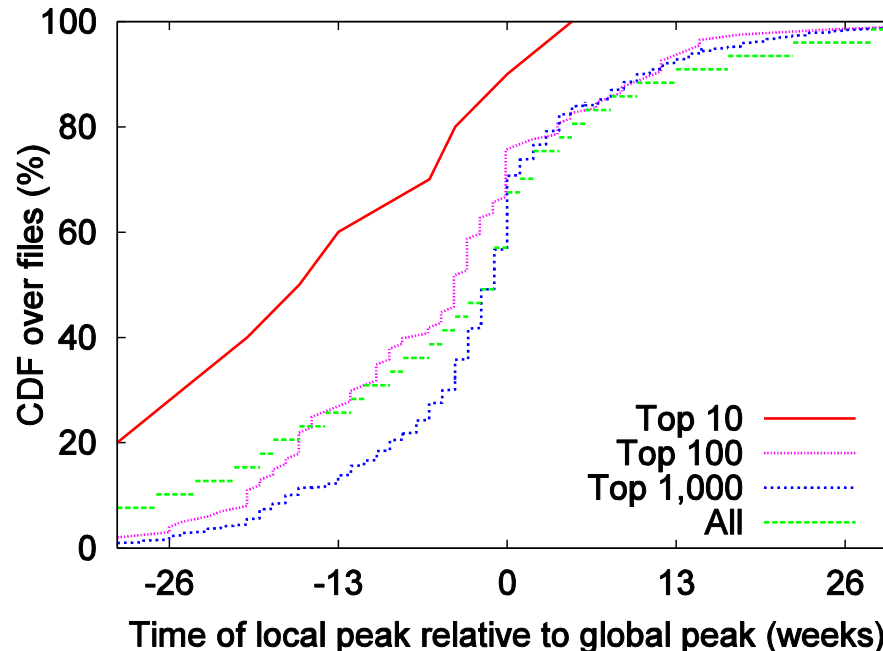


- Campus users are generally early adopters of content
 - Except for music
- Perhaps campus users can be used to predict some future popularity ...
 - And used for seeding such content



Early adopters

Downloads relative to global peak

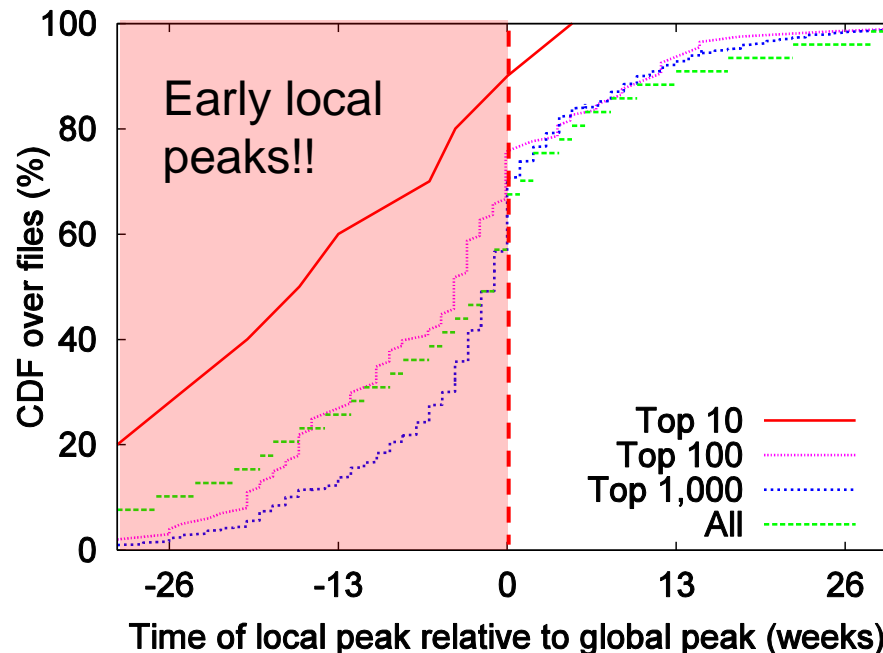


- Better predictor the more popular the content becomes
 - As well as for some niche content ...



Early adopters

Downloads relative to global peak

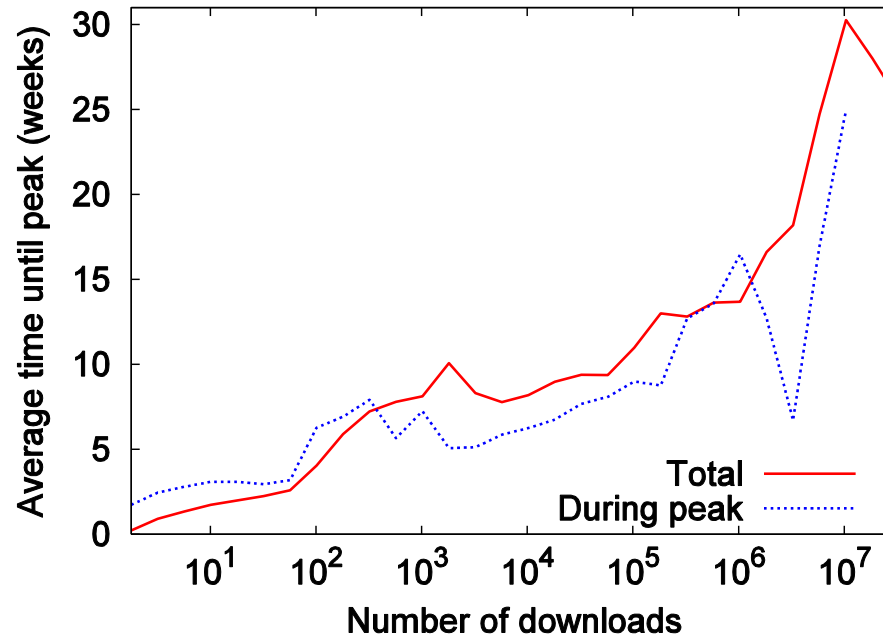


- Better predictor the more popular the content becomes
 - As well as for some niche content ...



Time until peak

Global popularity peaks ...

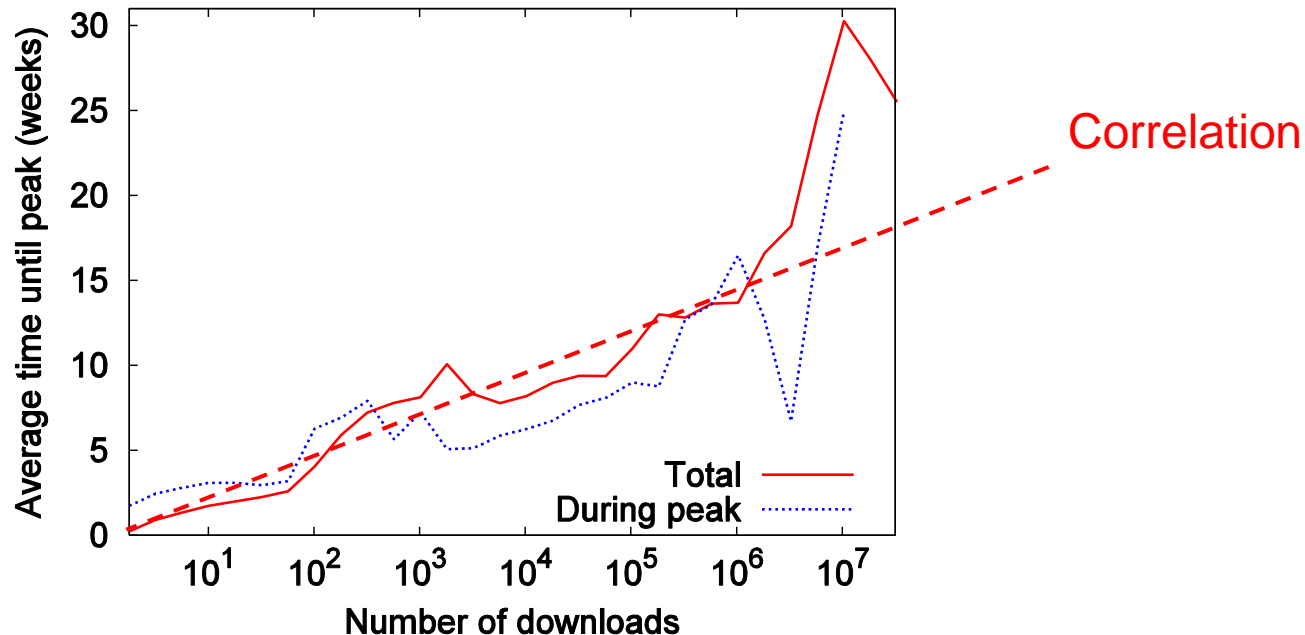


- The global popularity often peak late for popular content
 - Early flash crowds do not dominate the popularity
 - Perhaps a sign that rich-gets-richer a better model ...



Time until peak

Global popularity peaks ...

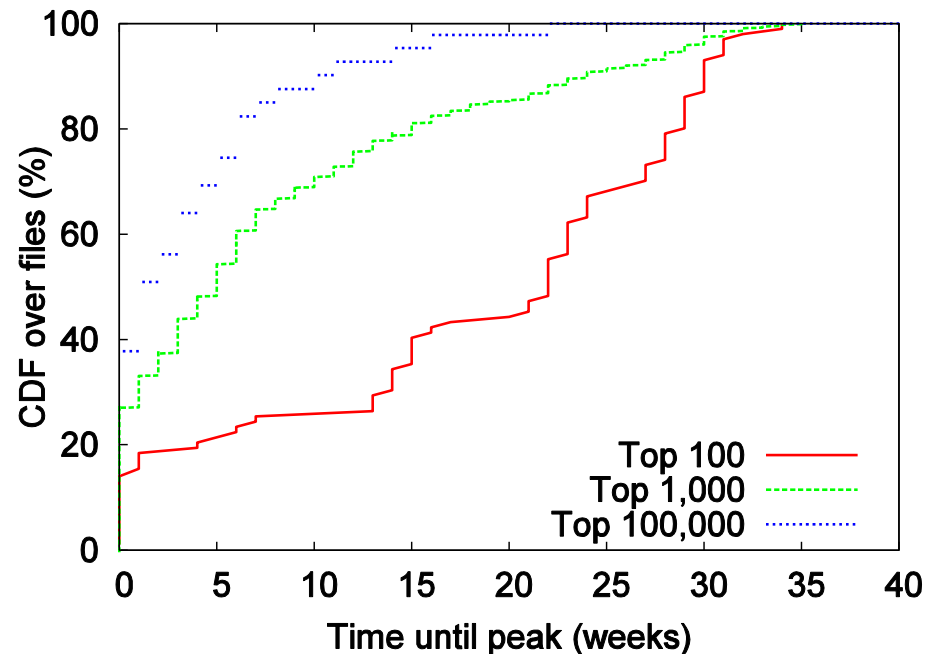


- The global popularity often peak later for popular content
 - Early flash crowds do not dominate the popularity
 - Perhaps a sign that rich-gets-richer a better model ...



Time until peak

Global popularity peaks ...

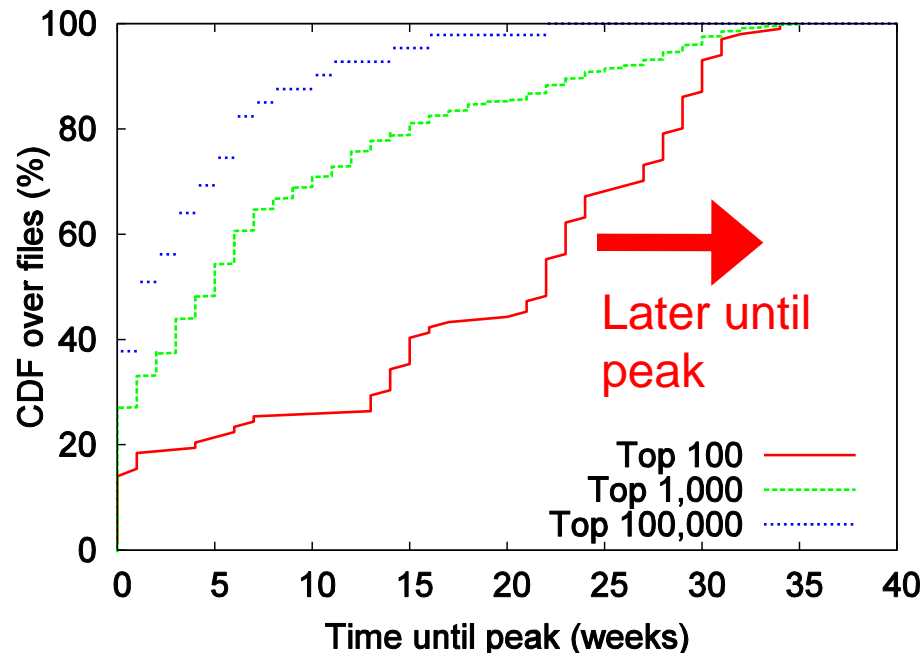


- The more popular the content
 - The later it peaks ...



Time until peak

Global popularity peaks ...

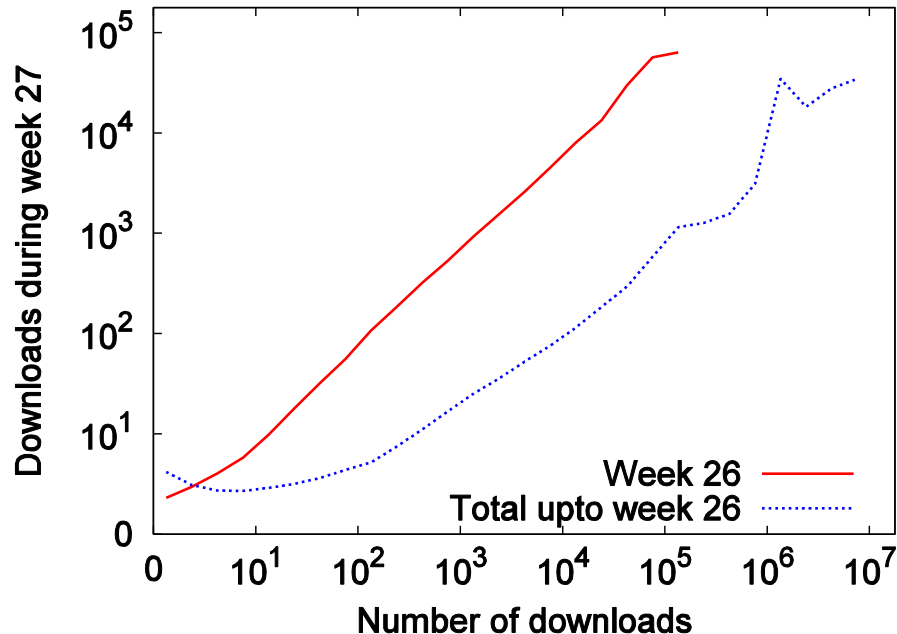


- The more popular the content
 - The later it peaks ...

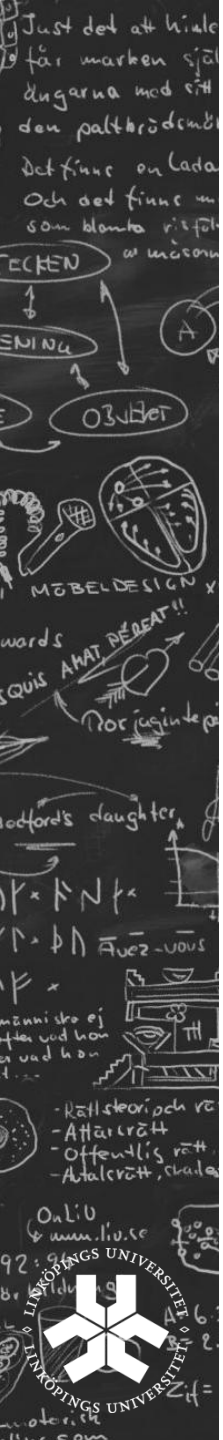


Time until peak

Global popularity peaks ...

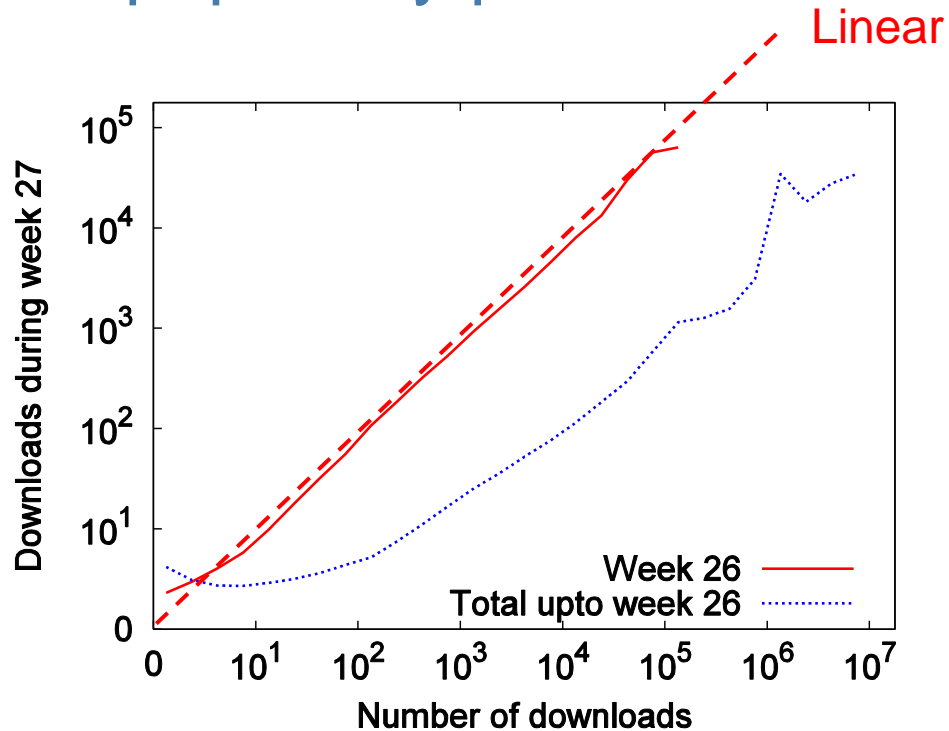


- Rich-gets-richer
 - Close to linear from week-to-week
 - Cumulative total downloads show weaker (sub-linear) rich-gets-richer behavior



Time until peak

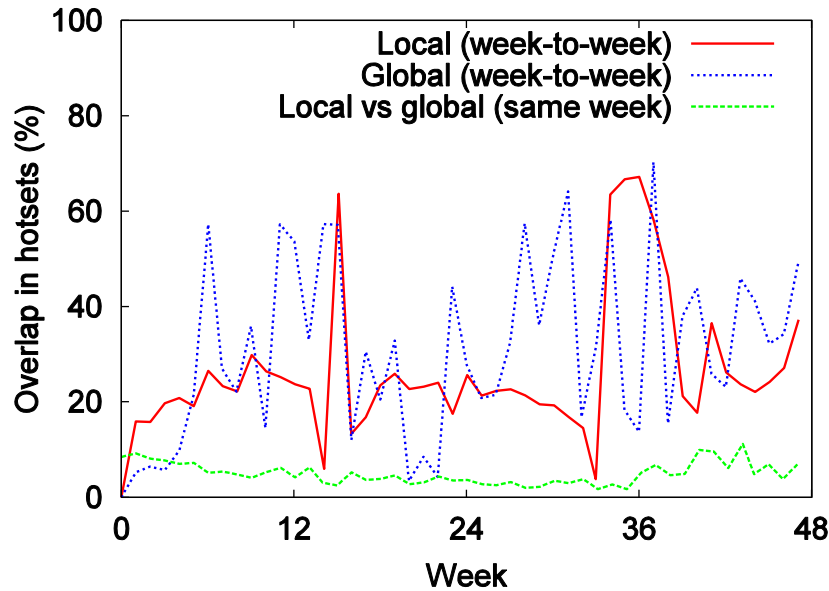
Global popularity peaks ...



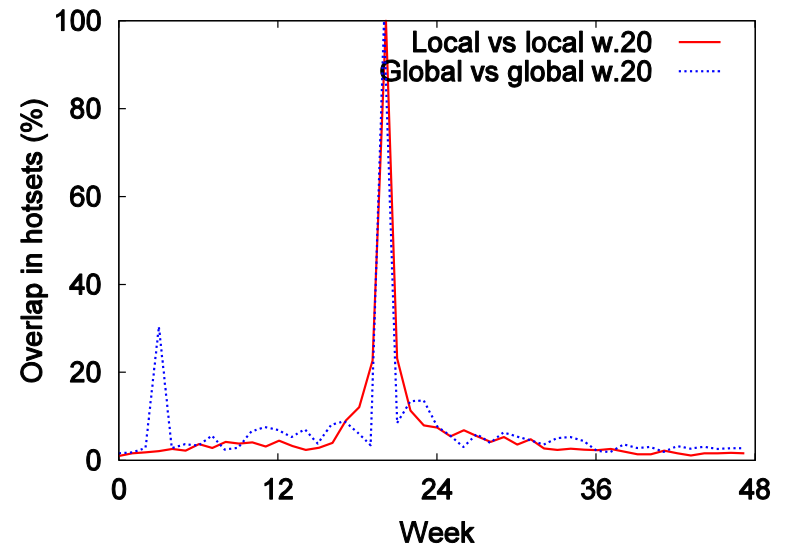
- Rich-gets-richer
 - Close to linear from week-to-week
 - Cumulative total downloads show weaker (sub-linear) rich-gets-richer behavior

Hotset analysis

Popularity churn

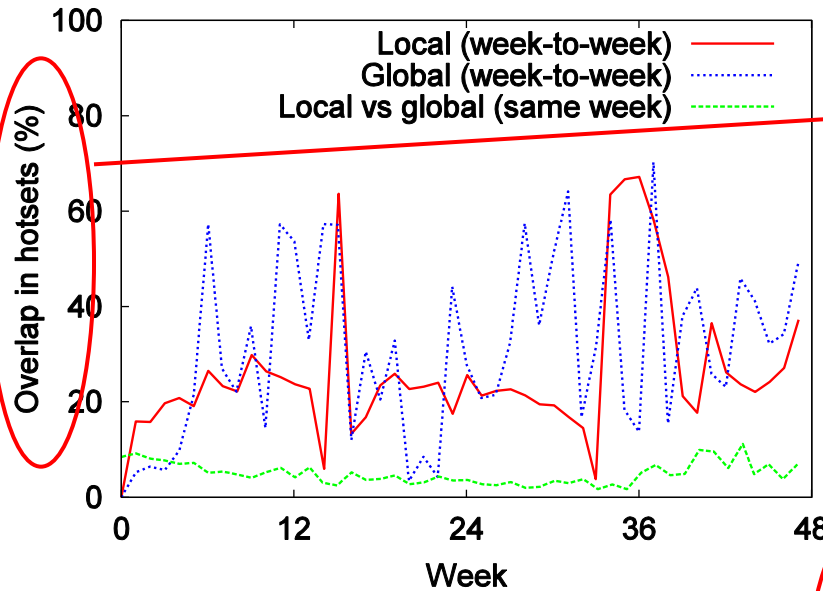


- High popularity churn
- Roughly 50-60% new videos each week
- Some files reoccur
- Some video reoccur in hotset



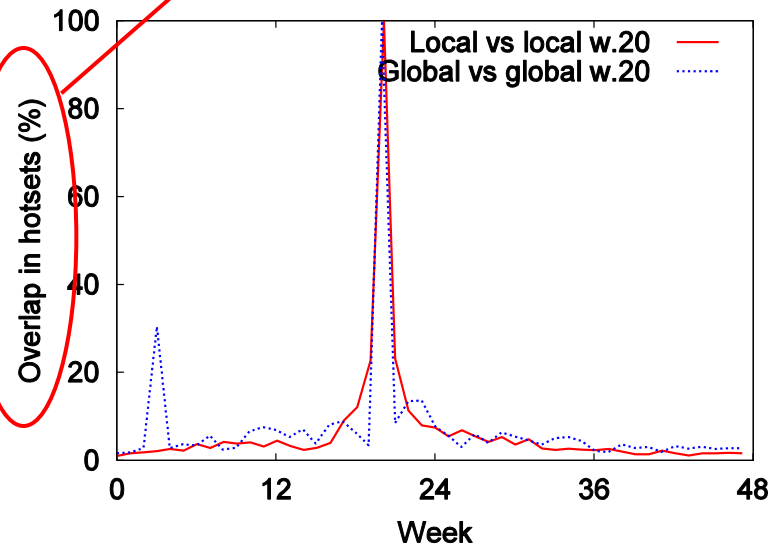
Hotset analysis

Popularity churn



Similarity between weeks

Similarity with week 20



- High popularity churn
- Roughly 50-60% new videos each week
- Some files reoccur
- Some video reoccur in hotset



Conclusions

- Large-scale longitudinal multi-torrent analysis
 - University campus
 - Global trackers
- Campus users download more large files (TV shows and movies) and a smaller fraction of music
- Campus users are “early adopters”
 - Except for music
- High weekly churn in set of popular files
- Most of the popular files peak well after their initial use
 - Signs of rich-gets-richer behavior



Thank you!

- Niklas Carlsson Linköping University
- György Dan KTH Royal Institute
- Anirban Mahanti NICTA
- Martin Arlitt HP Labs and University of Calgary

