

# A Longitudinal Characterization of Local and Global BitTorrent Workload Dynamics

Niklas Carlsson  
György Dan  
Anirban Mahanti  
Martin Arlitt

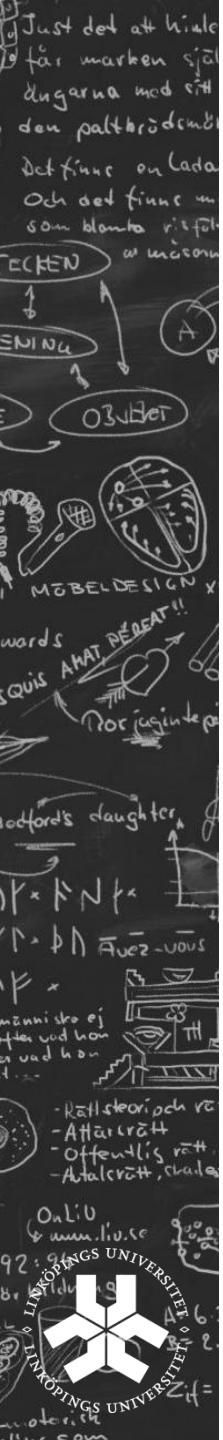
Linköping University  
KTH Royal Institute of Technology  
NICTA  
HP Labs and University of Calgary



March 14, 2012

# Motivation

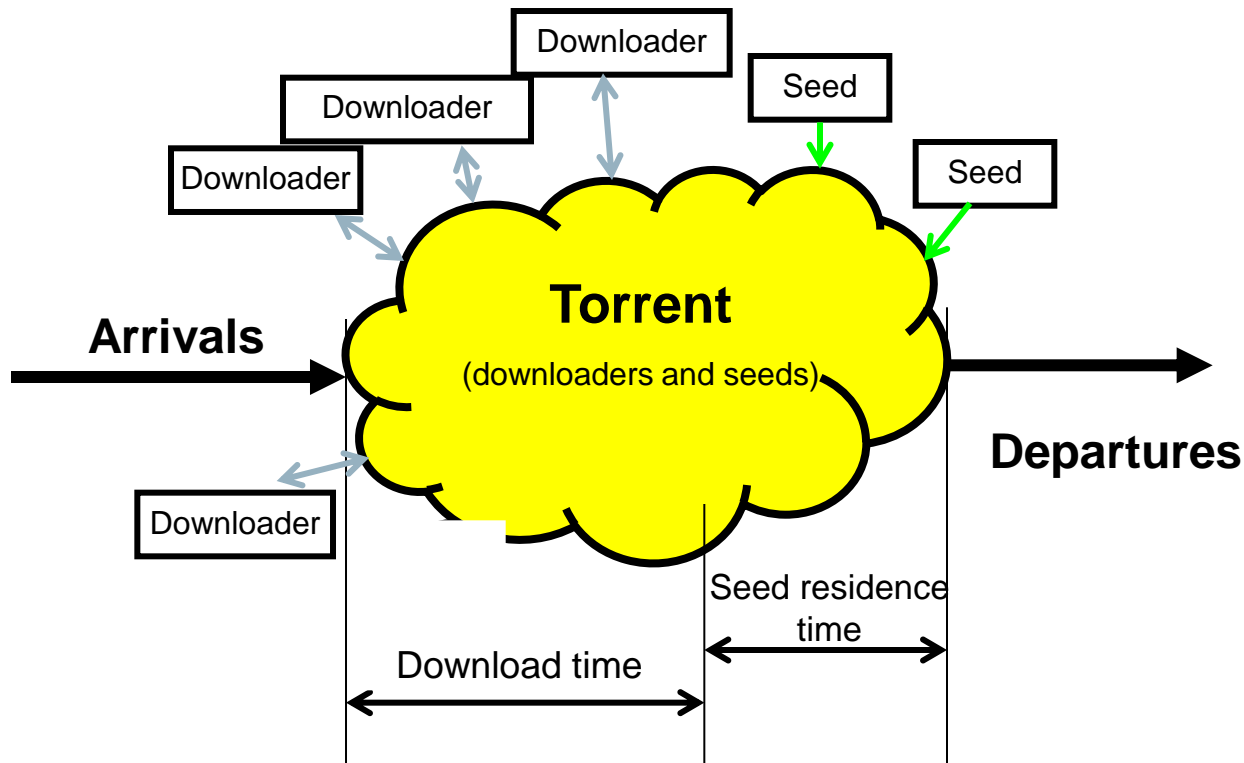
- Use of Internet for content delivery is massive ... and becoming more so
  - How to make scalable and efficient?
  - Server-based and peer-to-peer
- Chunk-based approach proven scalable
  - Files split into smaller chunks
  - Clients can download from both servers and other clients (peers)
- How to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



# Background: BitTorrent

## Single file download

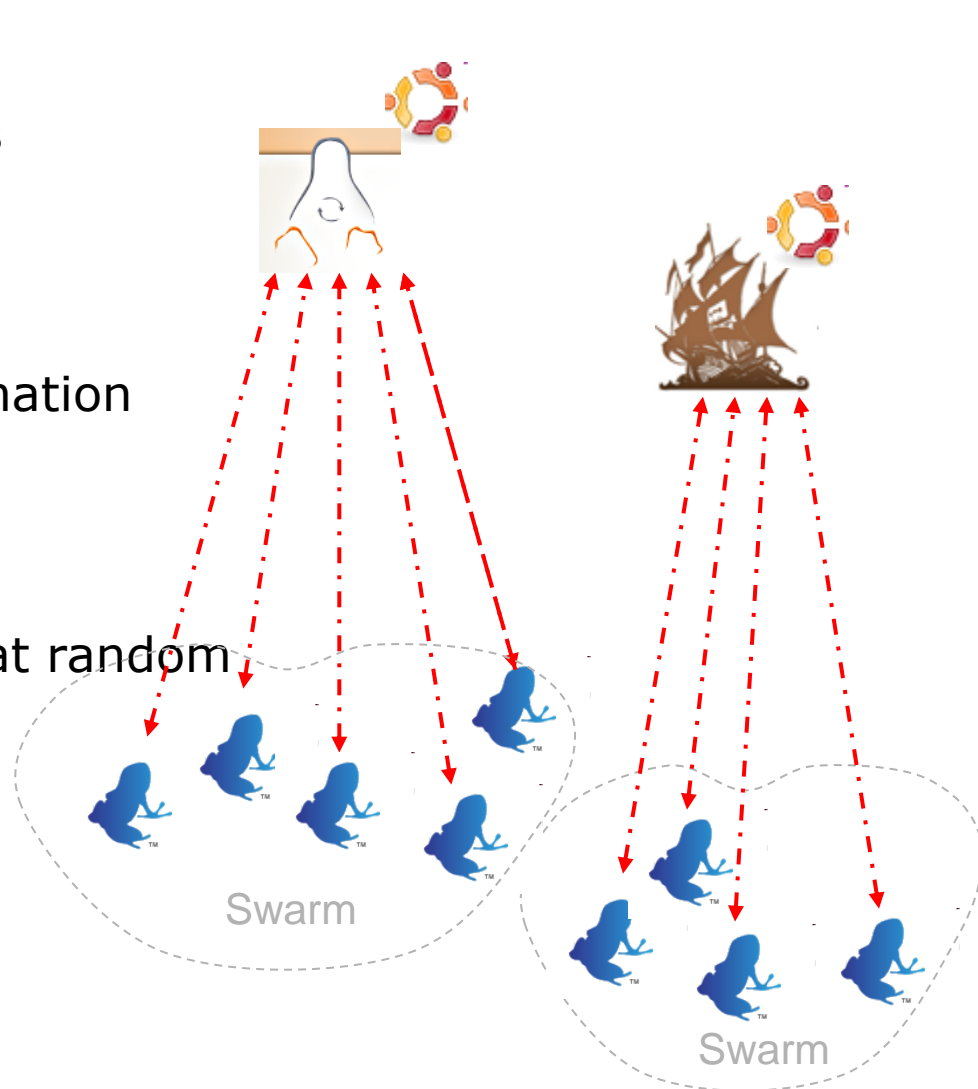
- File split into many smaller chunks
- Downloaded from both seeds and downloaders
- Distribution paths are dynamically determined
  - Based on data availability



# Background: BitTorrent






## Multi-tracked torrents

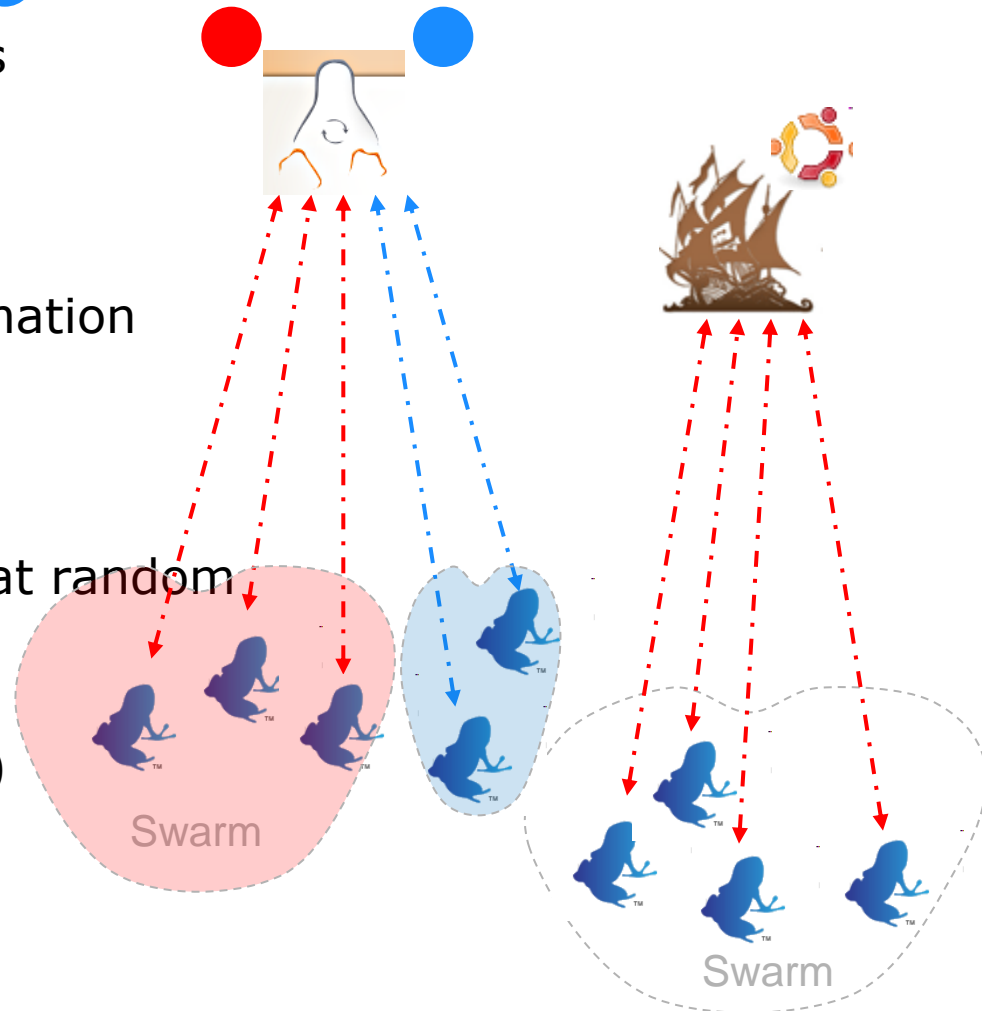
- Torrent file
  - "announce-list" URLs
- Trackers
  - Register torrent file
  - Maintain state information
- Peers
  - Obtain torrent file
  - Choose one tracker at random
  - Announce
  - Report status
  - Peer exchange (PEX)



# Background: BitTorrent

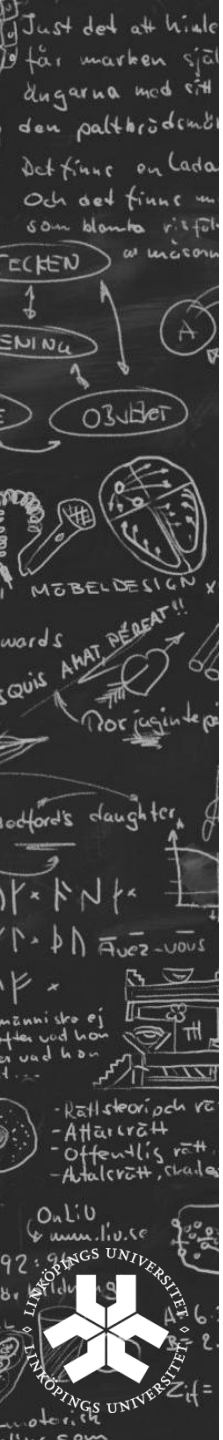
## Multi-tracked torrents

- 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)



# 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

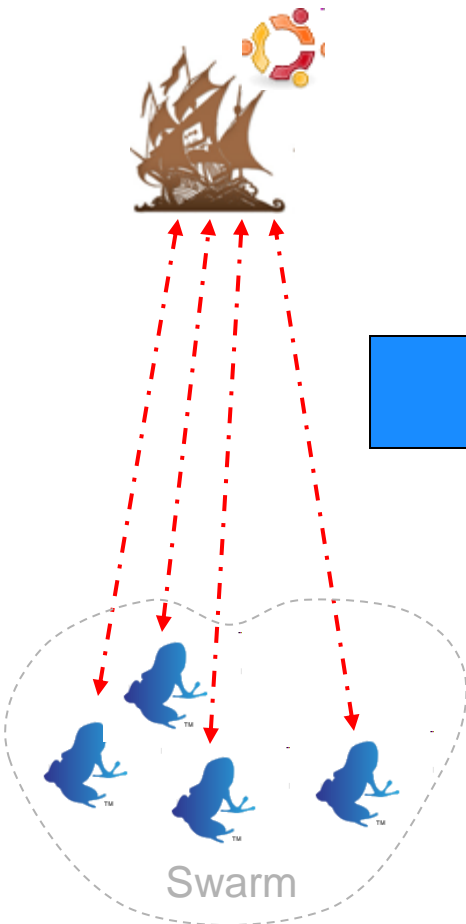




# Measurement overview

## Active + passive measurements

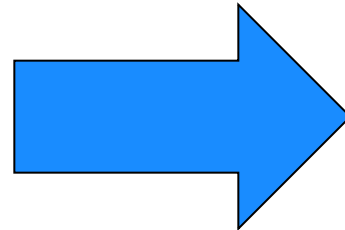
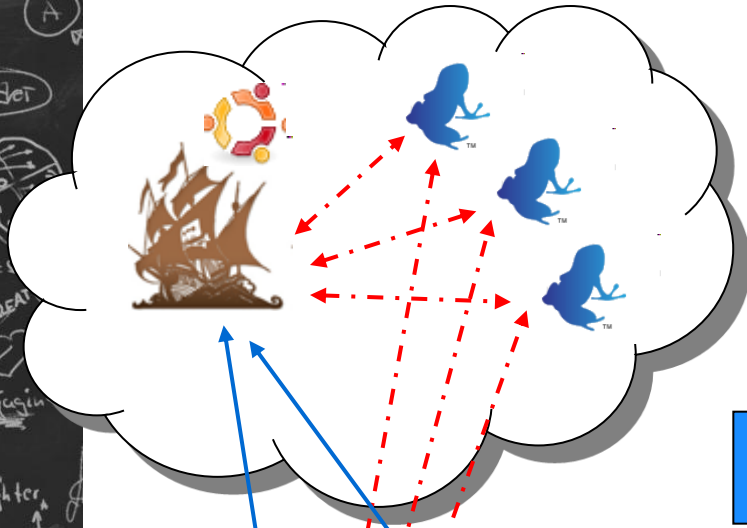
### Popularity dynamics



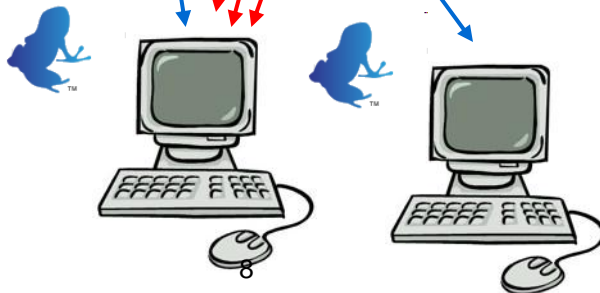
- Longitudinal data
- Two vantage points
  - University campus (ingress/egress)
  - Global trackers

# University: tracker communication

## Passive measurements



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



Just det att hiale  
får marcken sjä  
dugarna med s  
den paltbröden  
det frunc on lada  
Och det frunc m  
som blanta riefat  
w mason

ECKEN  
ENINU  
O3Ubet

MÖBELDE  
wards  
SQUIS AKAT DÉREAT  
Por iugin

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

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

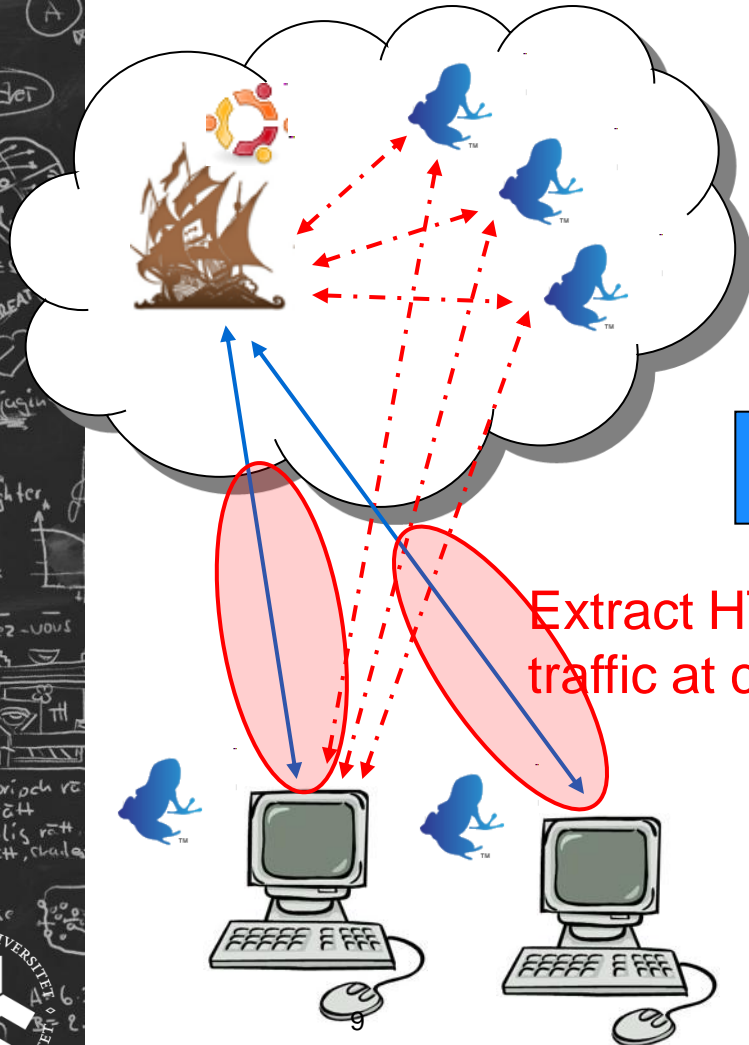
OnLiU  
www.liu.se

LINKÖPINGS UNIVERSITET

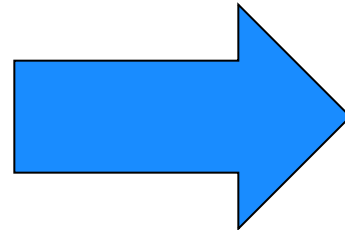


# University: tracker communication

## Passive measurements



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

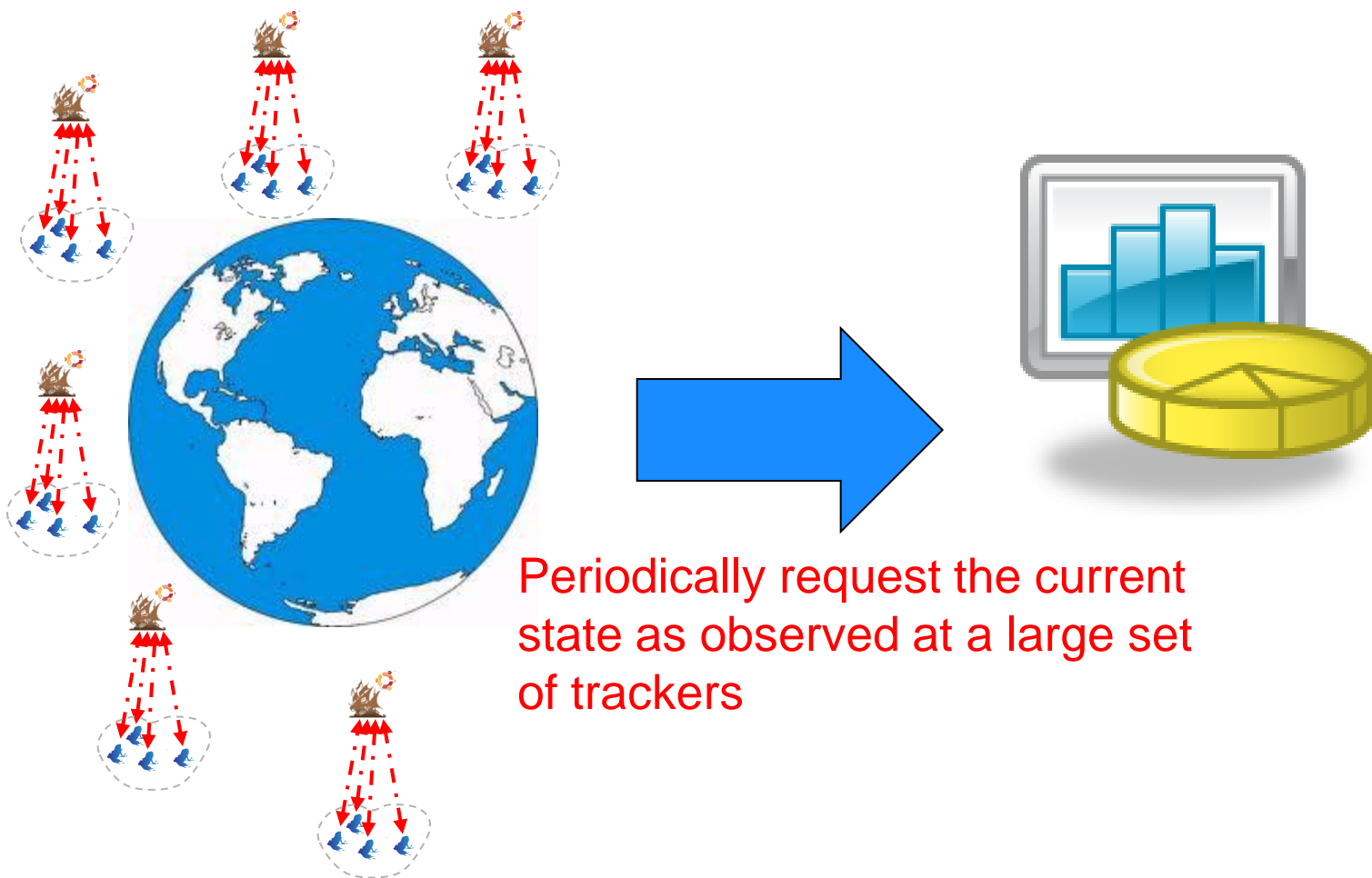


Just det att kiale  
får marke sjö  
dugarna med s  
den paltbröden  
det frunc on lada  
Och det frunc m  
som blanta riefat  
w masonu  
EKEN  
ENINU  
O3Ubet  
MÖBELDE  
wards  
SQUIS AHAT DEBEAT  
Por iugin  
odford's daughter  
Åvez-vous  
minni sko ej  
fter vad hon  
en vad hon  
- Rell teor och va  
- AHAICRÄTT  
- Offentlig rätt  
- AHAICRÄTT, skade  
OnLiU  
www.liu.se  
92: g  
LINKÖPINGS UNIVERSITET  
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 marben sjö  
dugarna med sif  
den paltbröden  
det finns en lada  
och det finns m  
som blanda riefat  
u masonu

ECKEN  
ENINU  
OBJET

MÖBELDESIGN  
wards  
SQUIS AKAT DÉBÉAT!!  
Por jaginte p  
ford's daughter  
Åvez-vous  
- Rell teori och va  
- Affär rätt  
- Offentlig rätt  
- Attal rätt, skola

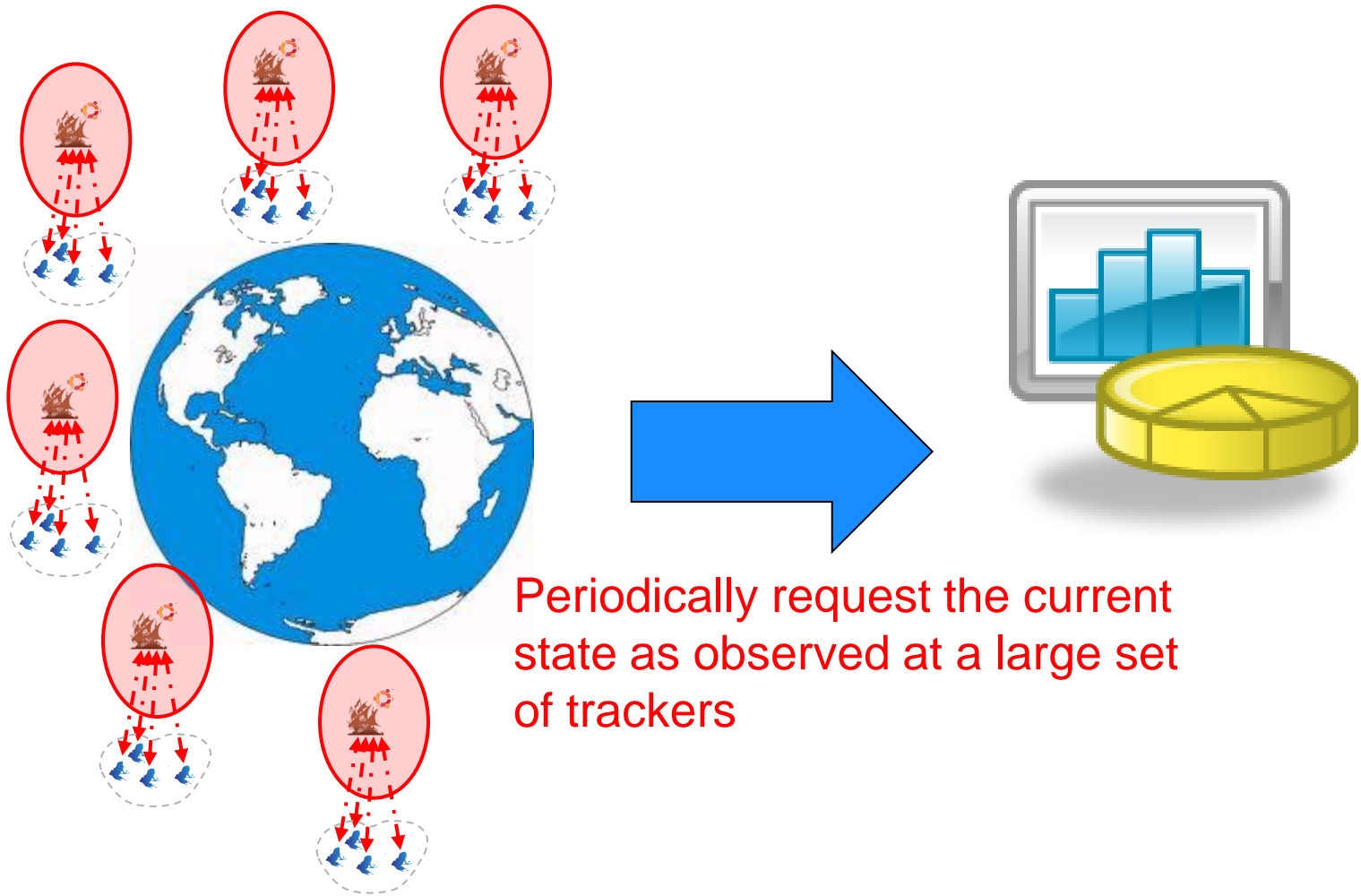
OnLiU  
www.liu.se

LINKÖPINGS UNIVERSITET



# Global: Tracker scrapes

## Active measurements



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

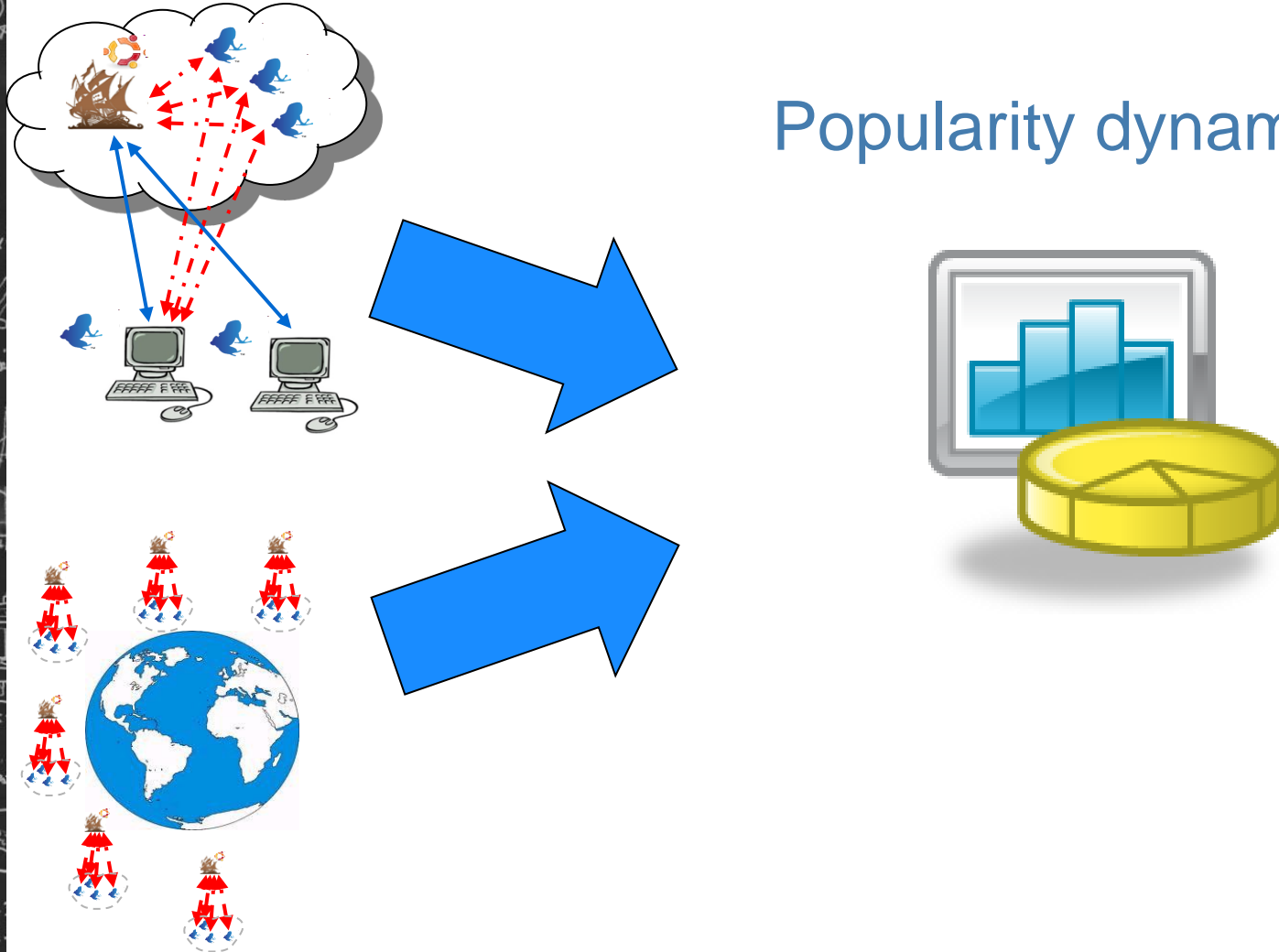
Just det att kiale  
får marken sjö  
dugarna med sig  
den paltbröden  
det finns en lada  
och det finns en  
som blanda riefat  
u masonu  
ECKEN  
ENINU  
O3Ubet  
MÖBELDESIGN  
swords  
SQUIS AHAT DÉBÉAT!!  
Por jaginte p  
mofford's daughter  
Åvez-vous  
- Rell teori och va  
- Aftalcrätt  
-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



# Measurement overview

## Active + passive measurements

Popularity dynamics

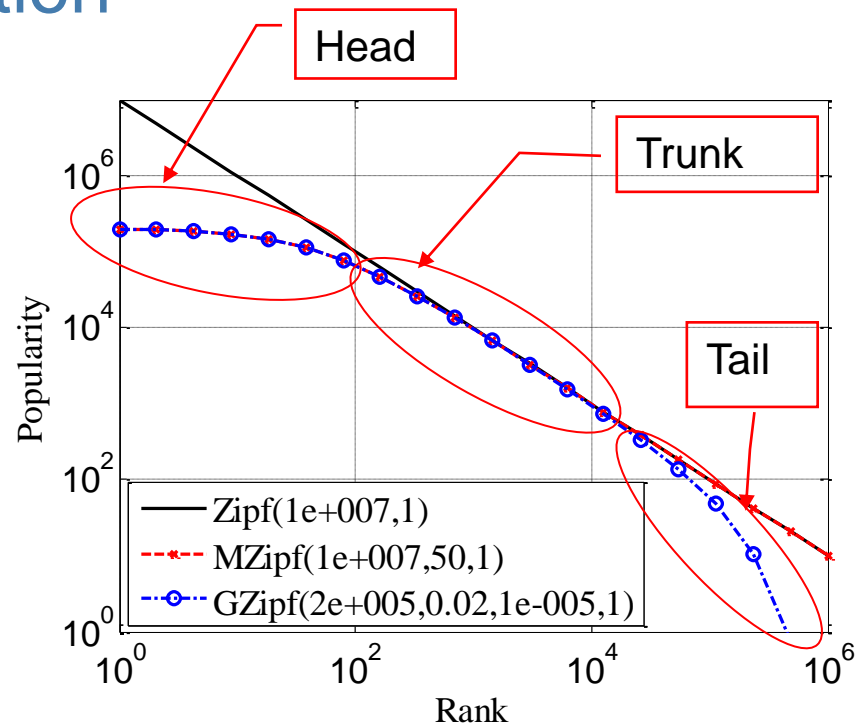
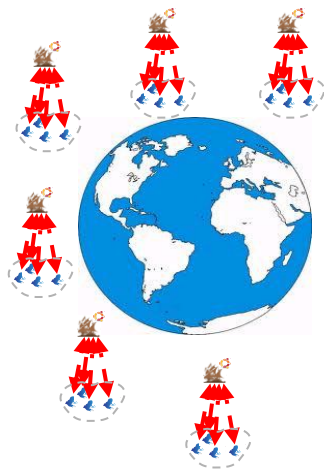
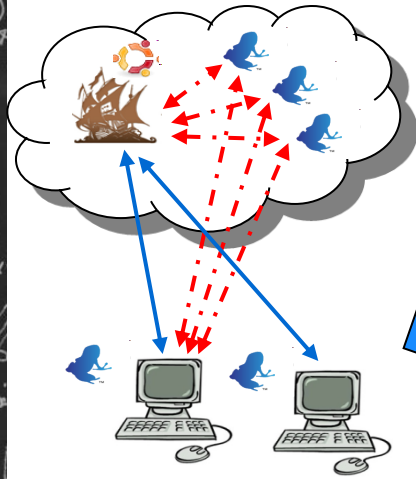


Just det att kiale  
får marke sjö  
dugarna med s  
den paltbröden  
det finns en lada  
och det finns m  
som blanda r  
EKEN  
ENINU  
O3Ubet  
MÖBELDESIGN  
squis AMAT DÉBATE!!  
Por jag inte p  
mofford's daughter  
Åvez-vous  
- Rätt teori och r  
- Affär rätt  
- Offentlig rätt  
- Aktierätt, skade  
OnLiU  
www.liu.se  
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) ...

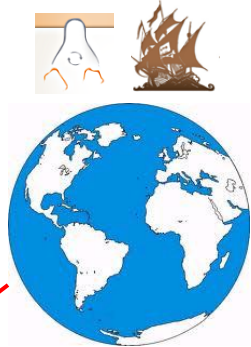


# Dataset summary

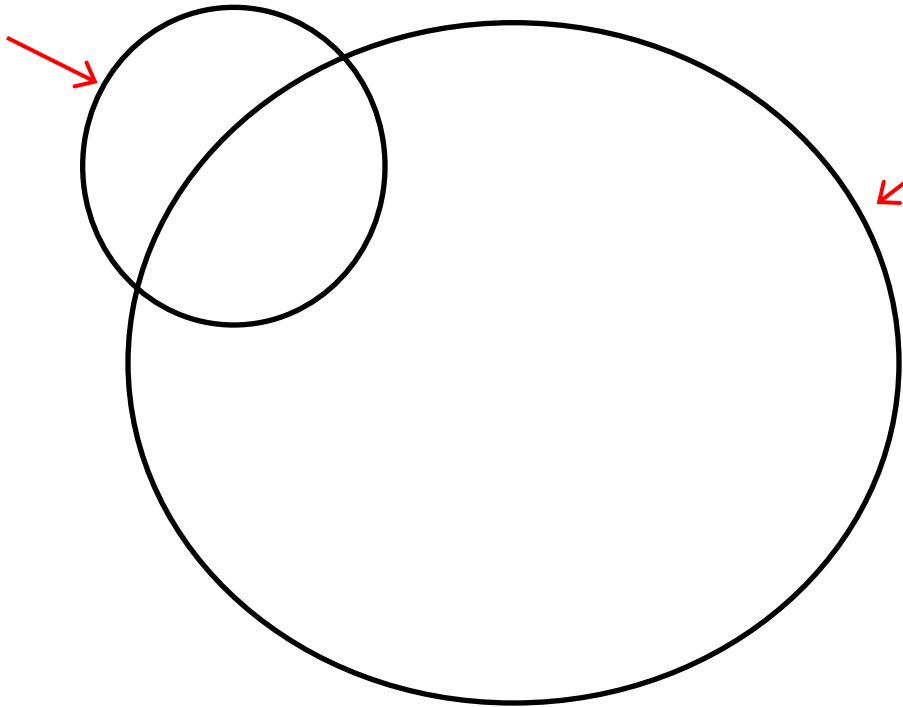
## Torrents observed



56,963



11.2 M

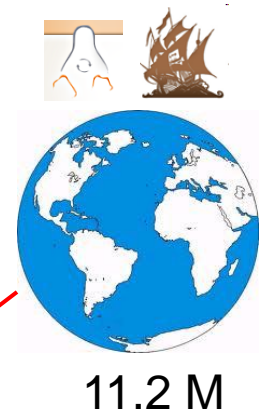
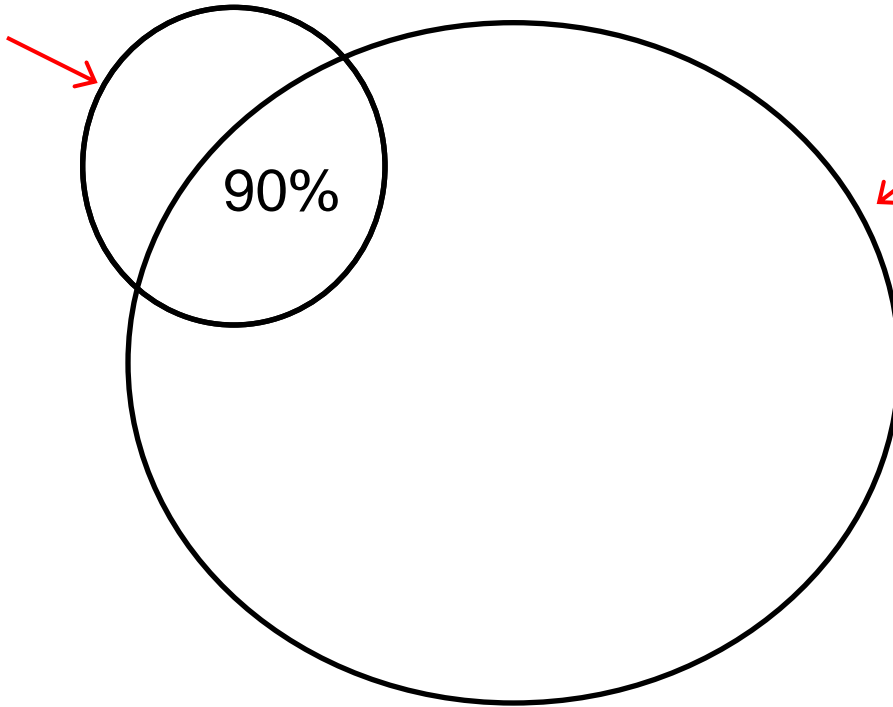


# Dataset summary

## Torrents observed



56,963



Just det att kiale  
får marcken sjä  
dugarna med s  
den paltbrödem  
det finns en lada  
och det finns m  
som blanda rief  
w masonu

ECKEN  
ENINU  
O3Ubet

MÖBELDESIGN

WARDS  
QUIS AHAT DEEAT!!  
Por jaginte p

odford's daughter

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

männi sko ej  
fter vad hon  
en vad hon

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

OnLiU  
www.liu.se

LINKÖPINGS UNIVERSITET

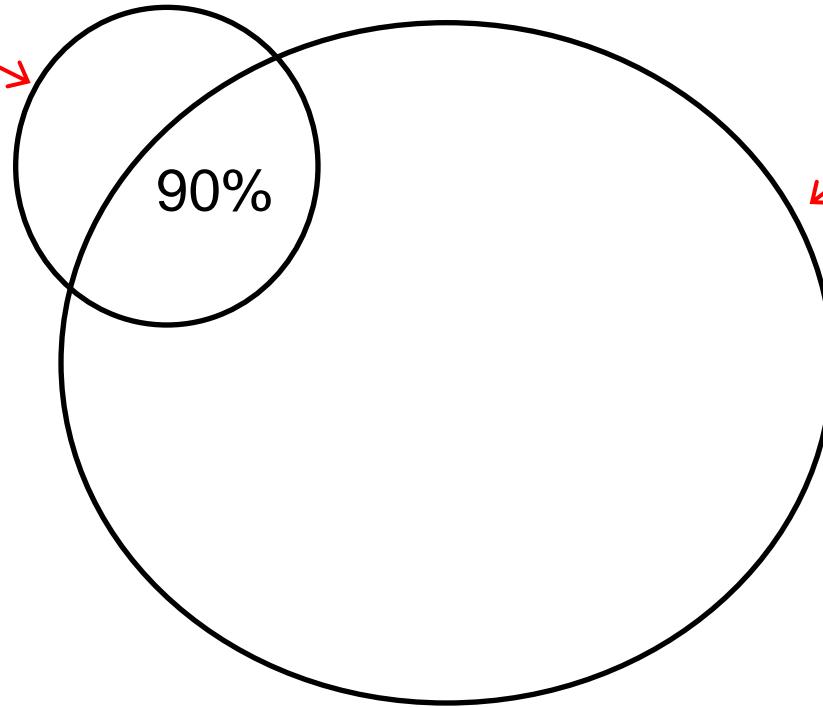


# Dataset summary

## Torrents observed



56,963



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

# Dataset summary

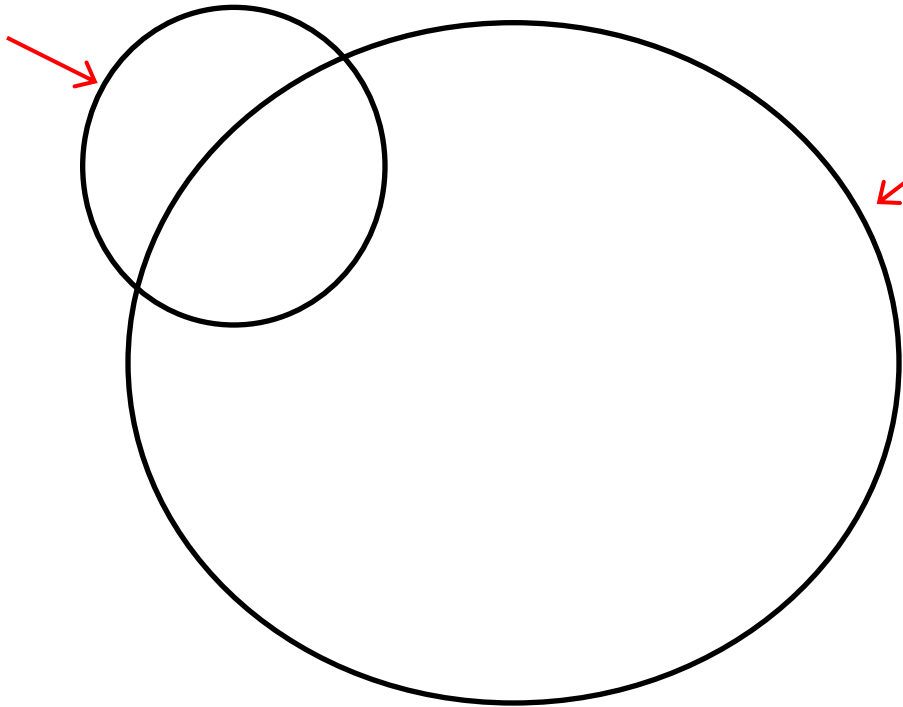
## Torrents observed



56,963



11.2 M



Just det att kiale  
får marben själ  
dugarna med sig  
den paltbrödem  
det finns en lada  
och det finns en  
som blanda riefat  
w masonu

ECKEN  
ENINU  
O3Ubet

MÖBELDESIGN

wards  
SQUIS AKAT DEEAT!!  
Por jaginte p

odford's daughter

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

männi sko ej  
fter vad hon  
en vad hon

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

OnLiU  
www.liu.se

92: g  
B. ldu  
LINKÖPINGS UNIVERSITET





# Dataset summary

## Torrents observed



56,963

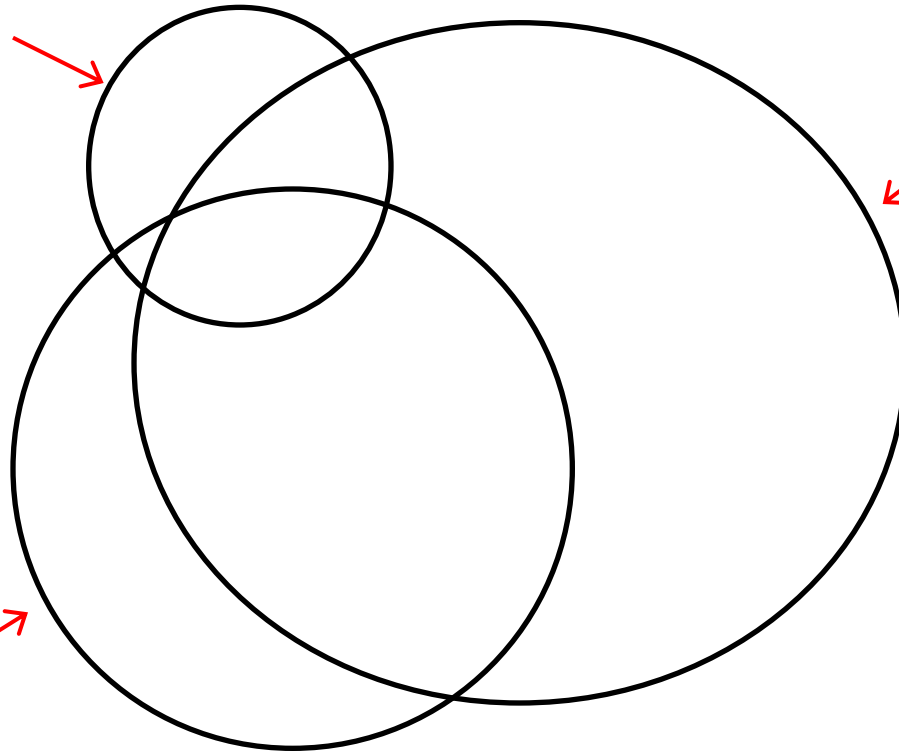
mininova



911,687



11.2 M



# Dataset summary

## Torrents observed



56,963

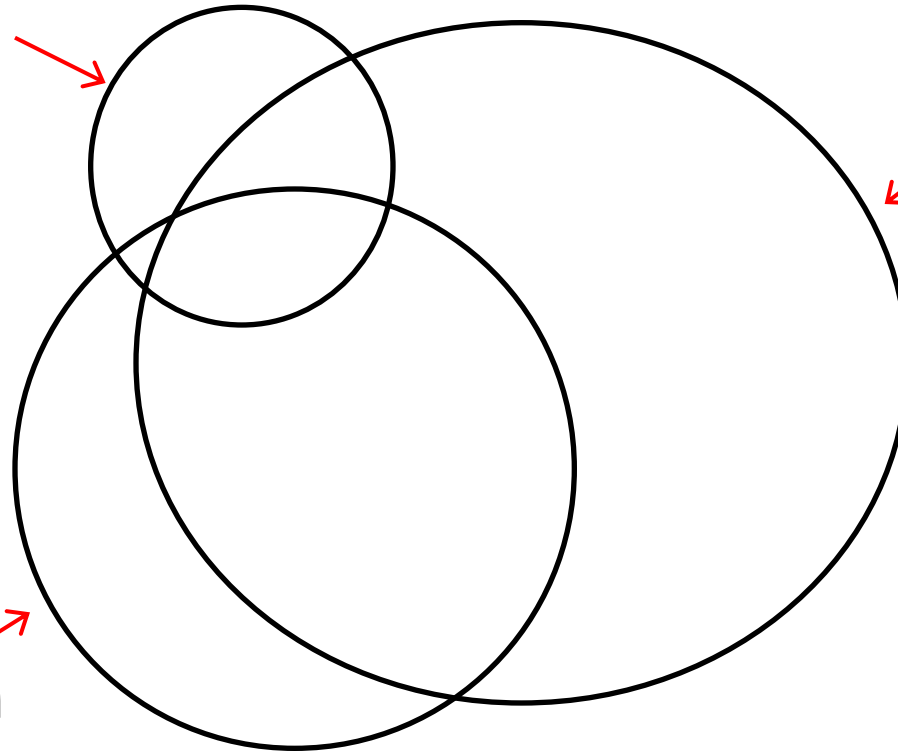
mininova



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

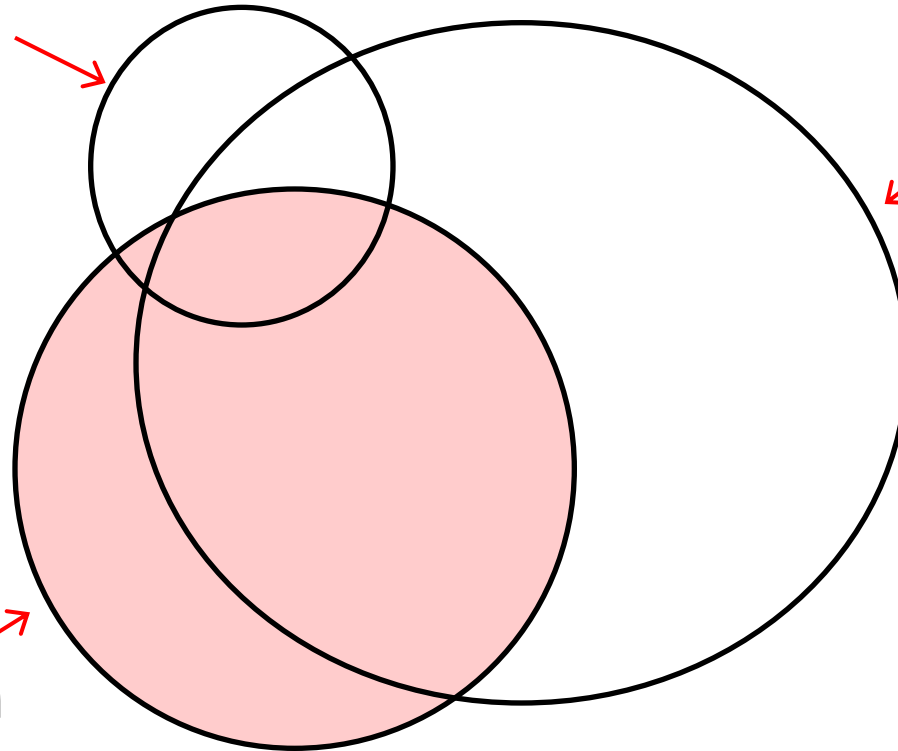
mininova



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

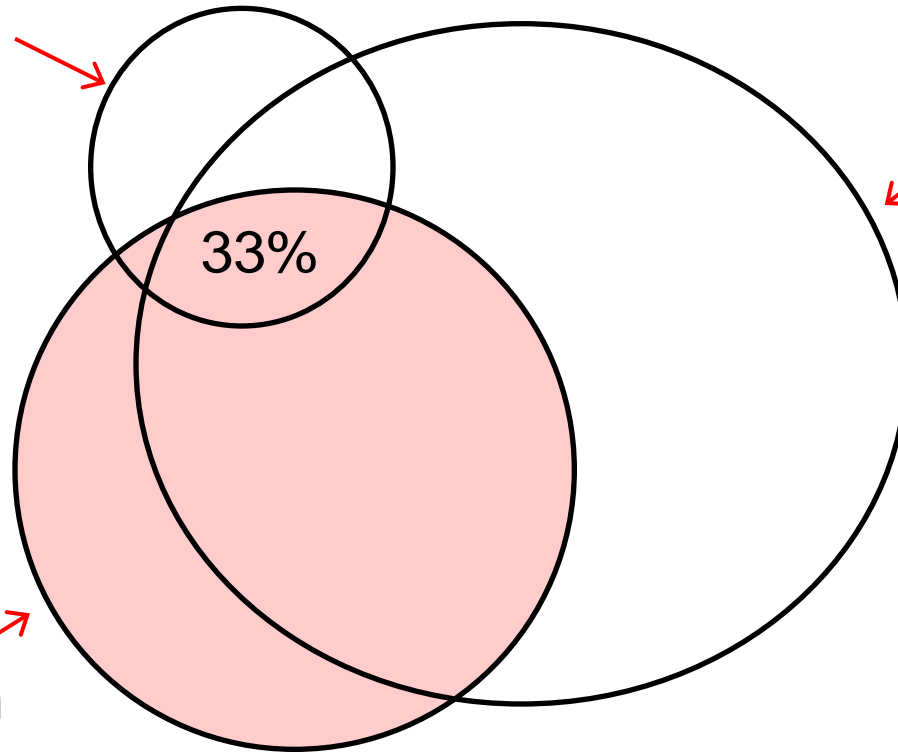


56,963

mininova



911,687

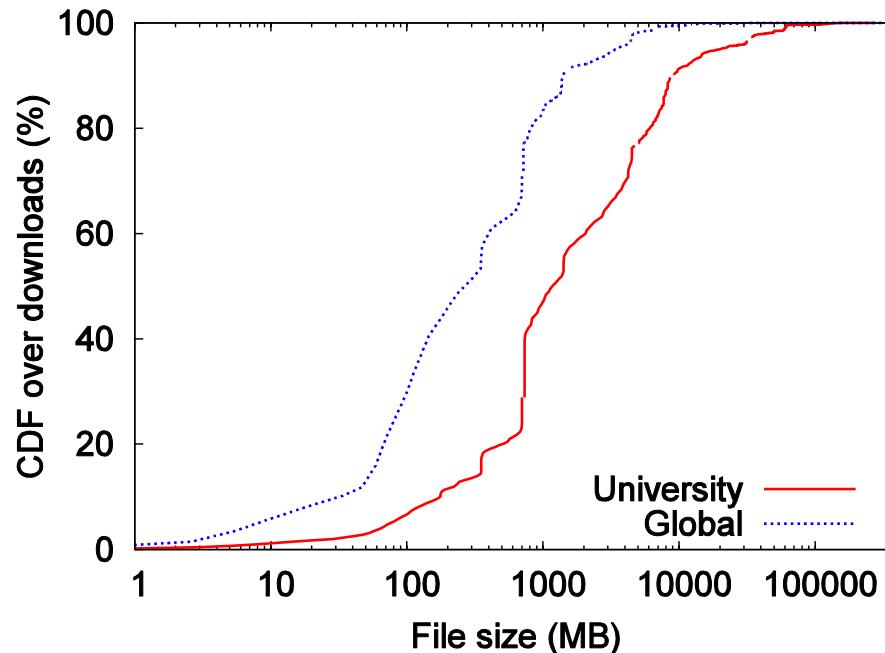


11.2 M

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

# Content download characteristics

## File size distribution, per download

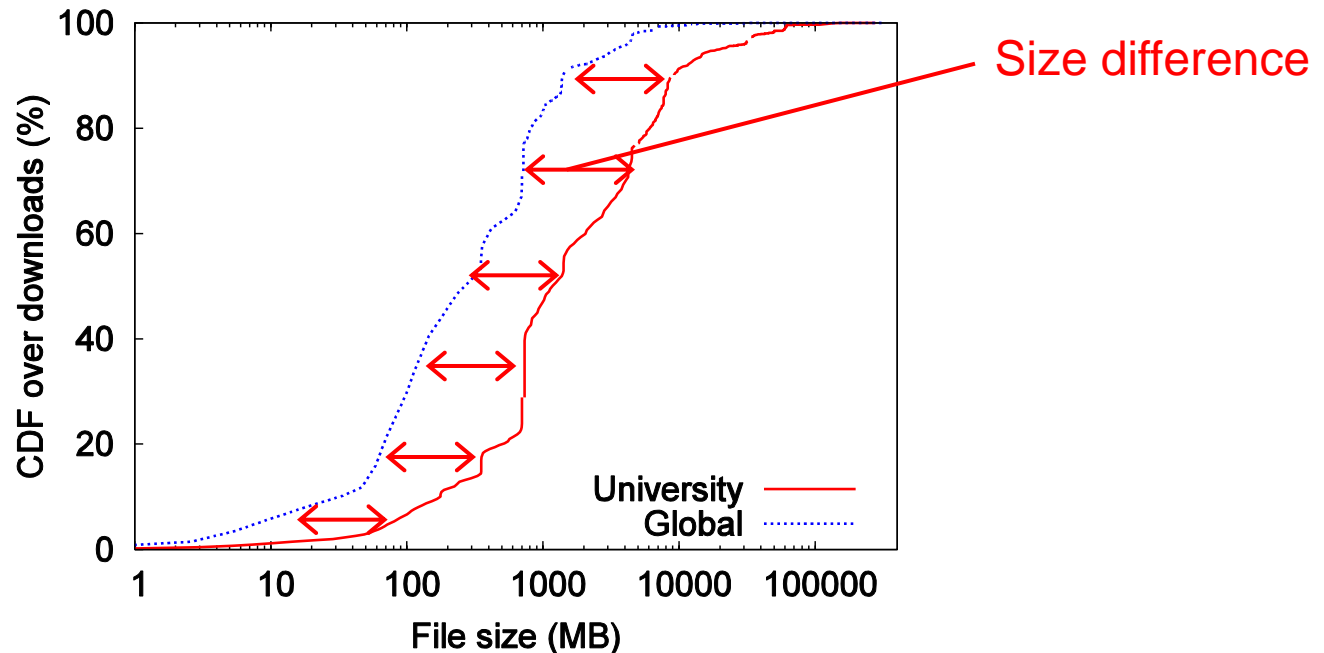


- Campus users download larger files



# Content download characteristics

## File size distribution, per download



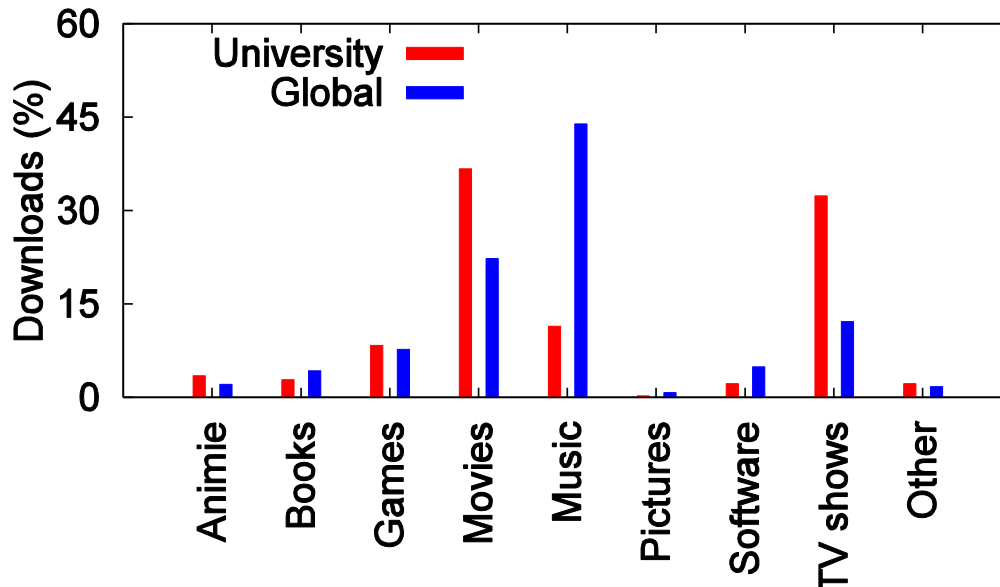
- Campus users download **larger** files





# Content download characteristics

## Breakdown per category

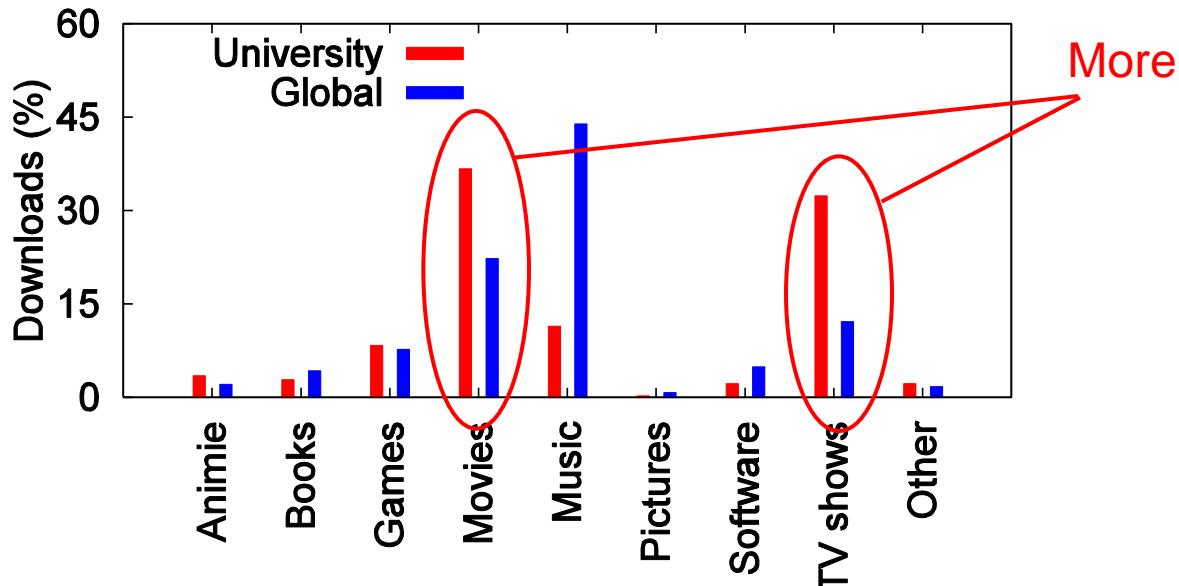


- Campus users download
  - More movies and TV shows
  - Less music



# Content download characteristics

## Breakdown per category

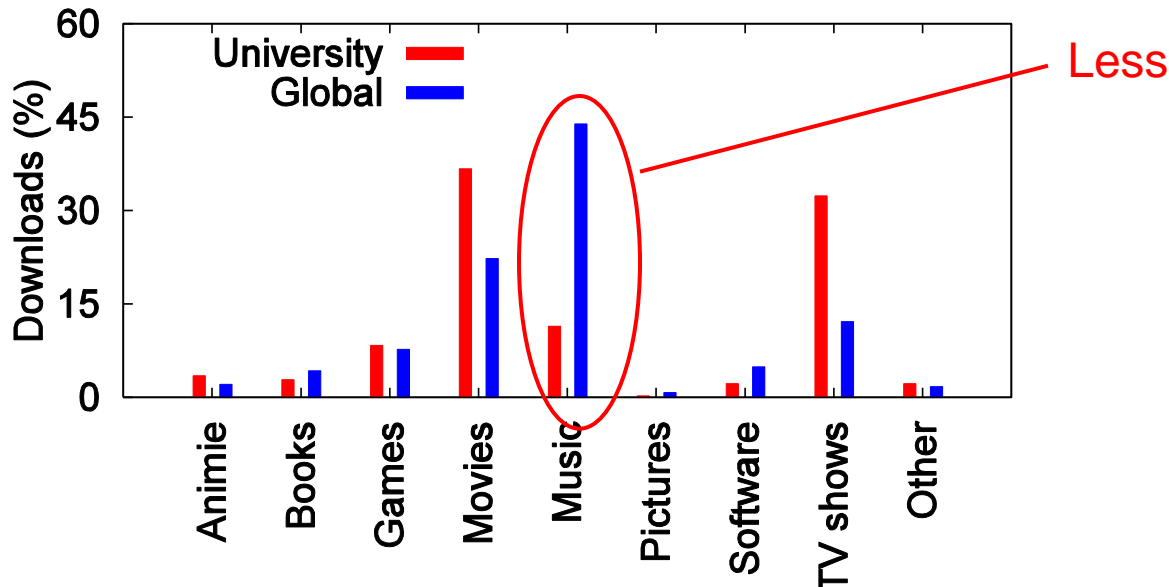


- Campus users download
  - More movies and TV shows
  - Less music



# Content download characteristics

## Breakdown per category

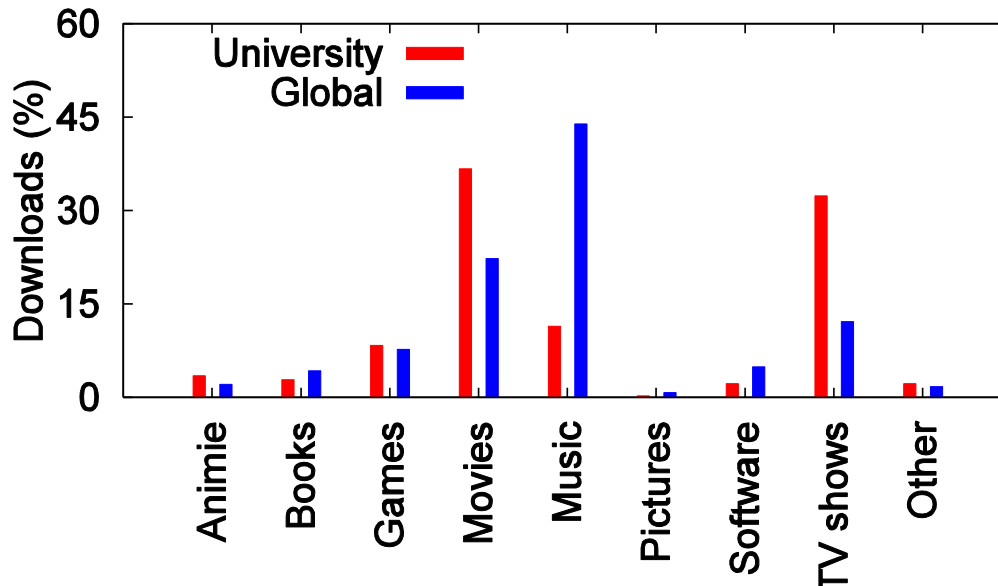


- Campus users download
  - More movies and TV shows
  - **Less music**

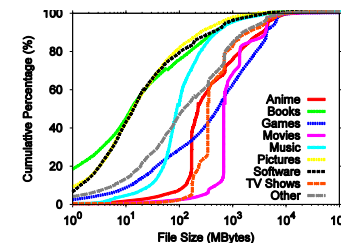


# Content download characteristics

## Breakdown per category

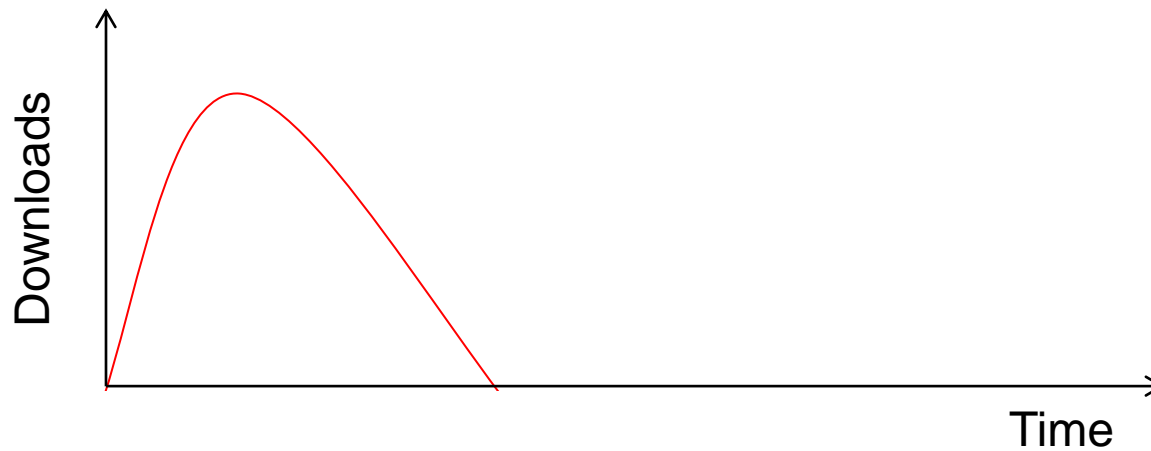


- Campus users download
  - More movies and TV shows
  - Less music
- Again, biased towards larger contents ...



# Early adopters

## Terminology



Just det att kiale  
fär marben sjä  
dugarna med s  
den paltbrödem  
det finne en lada  
Och det finne m  
som blanta riefat  
u masonu

EKEN  
ENINU  
OBJET

MÖBELDESIGN

wards  
SQUIS AHAT DÉBÉAT!!  
Por juginte p

odford's daughter

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

minni sko ej  
fter vad hon  
en vad hon

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

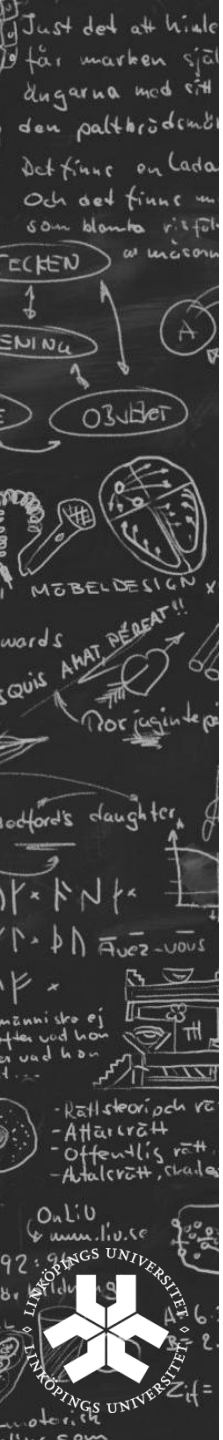
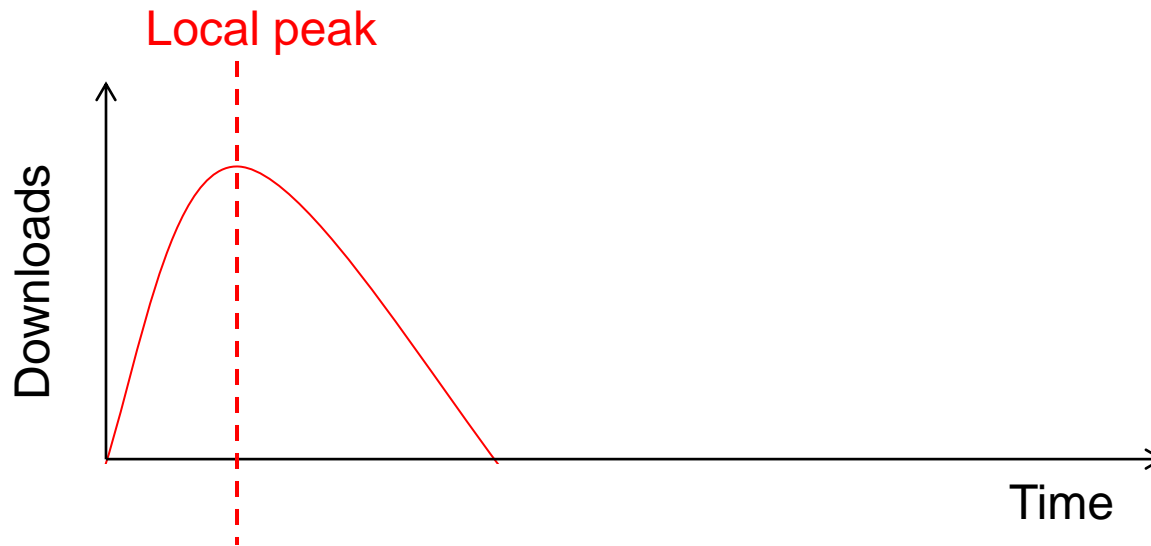
OnLiU  
www.liu.se

92: 9  
B. ldu

LINKÖPINGS UNIVERSITET

# Early adopters

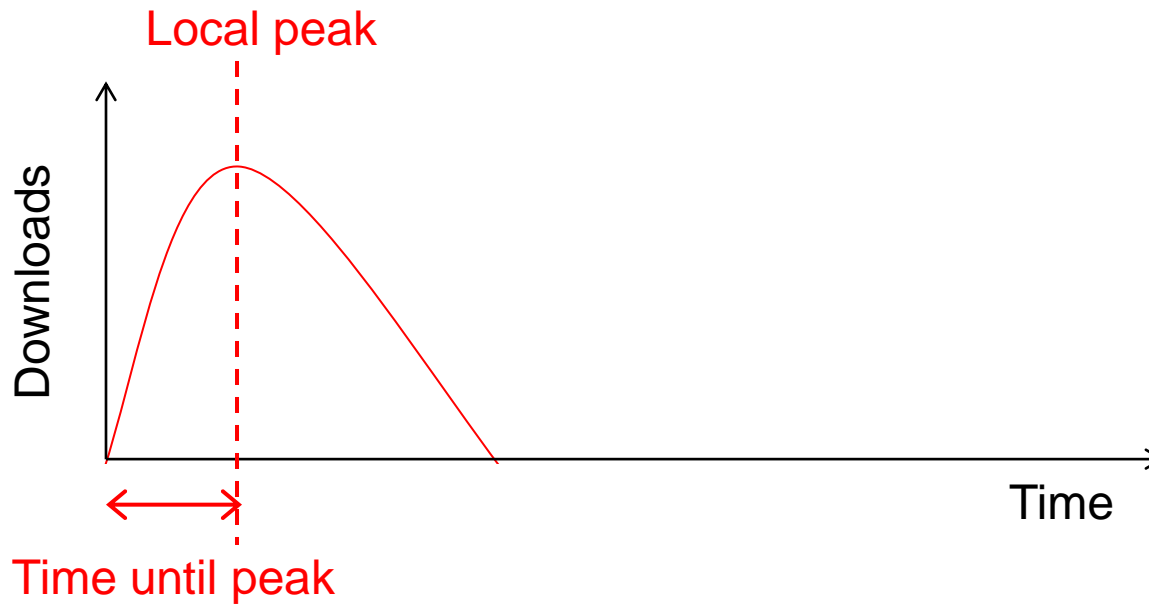
## Terminology





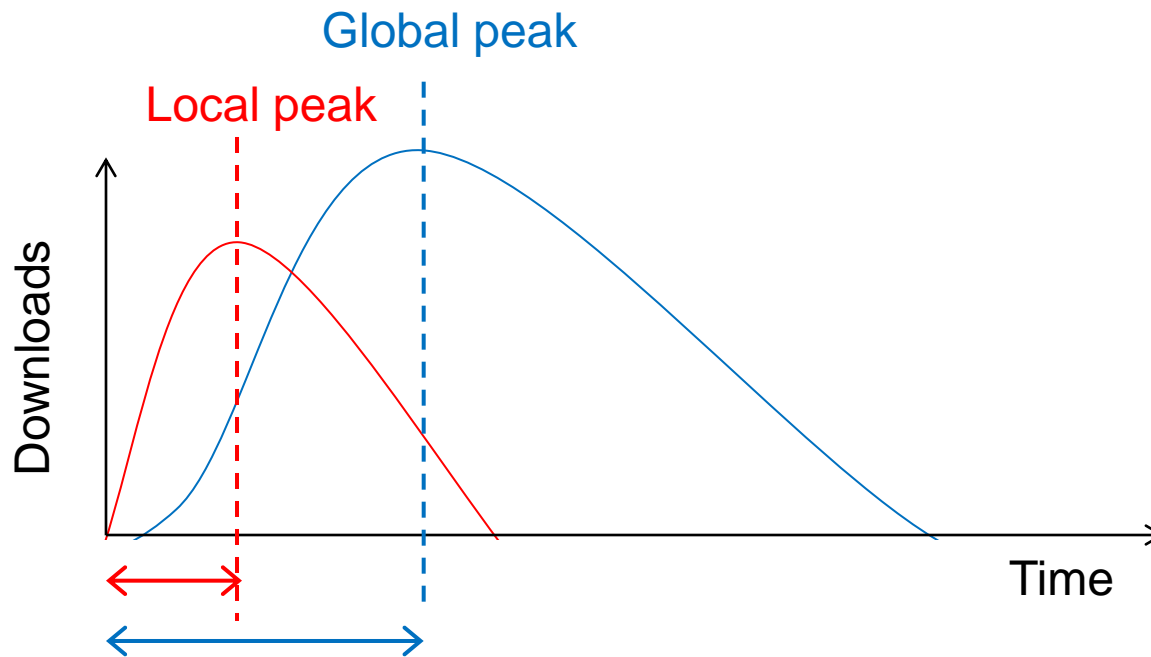
# Early adopters

## Terminology



# Early adopters

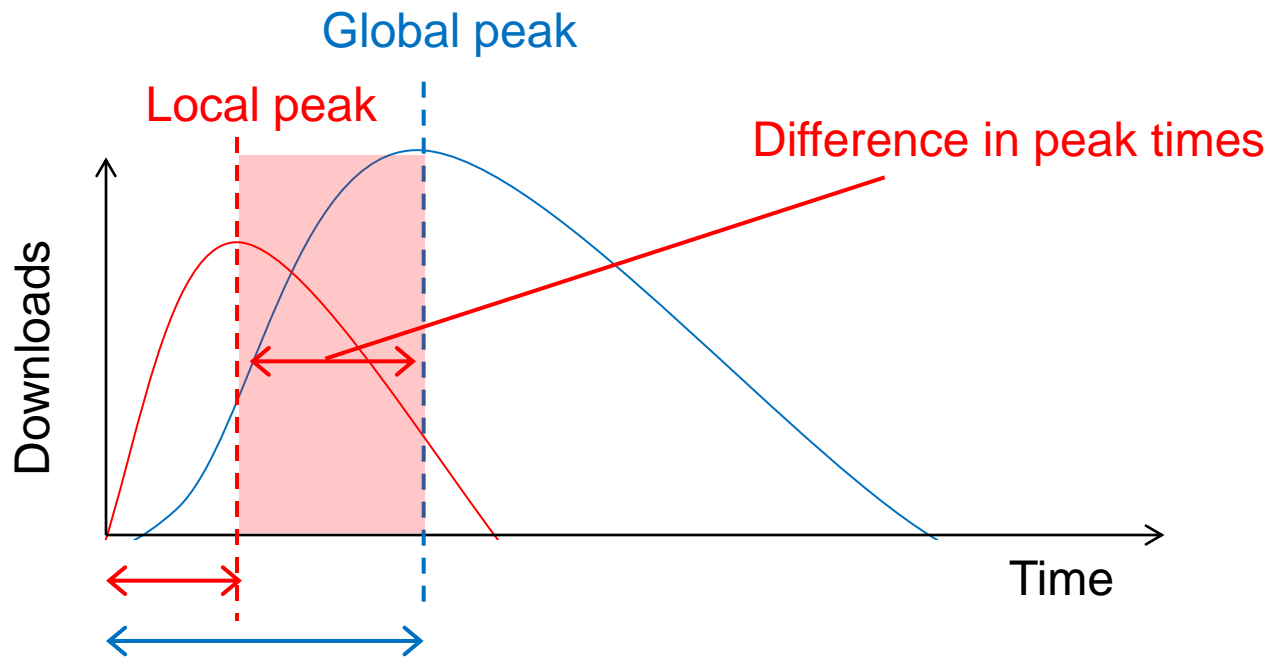
## Terminology



Just det att kiale  
får marben sjä  
dugarna med s  
den paltbröden  
det finns en lada  
och det finns m  
som blanda r  
EKEN  
ENINU  
OBJET  
MÖBELDESIGN  
wards  
SQUIS AKAT DÉBÉAT!!  
Por jüginte p  
ford's daughter  
KIN K  
Avez-vous  
männi sko ej  
fter vad hon  
er vad hon  
- Rätt teori och r  
- Affär rätt  
- Offentlig rätt  
- Aftal rätt, skade  
Onli  
www.liu.se  
92  
LINKÖPINGS UNIVERSITET  
LINKÖPINGS UNIVERSITET

# Early adopters

## Terminology



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

ECKEN  
ENINU  
OBJET

MÖBELDESIGN

wards  
SQUIS AKAT DÉCAT!!  
Por jüante p

odford's daughter

BY x KN x  
V x N 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: g  
B. ldu

LINKÖPINGS UNIVERSITET

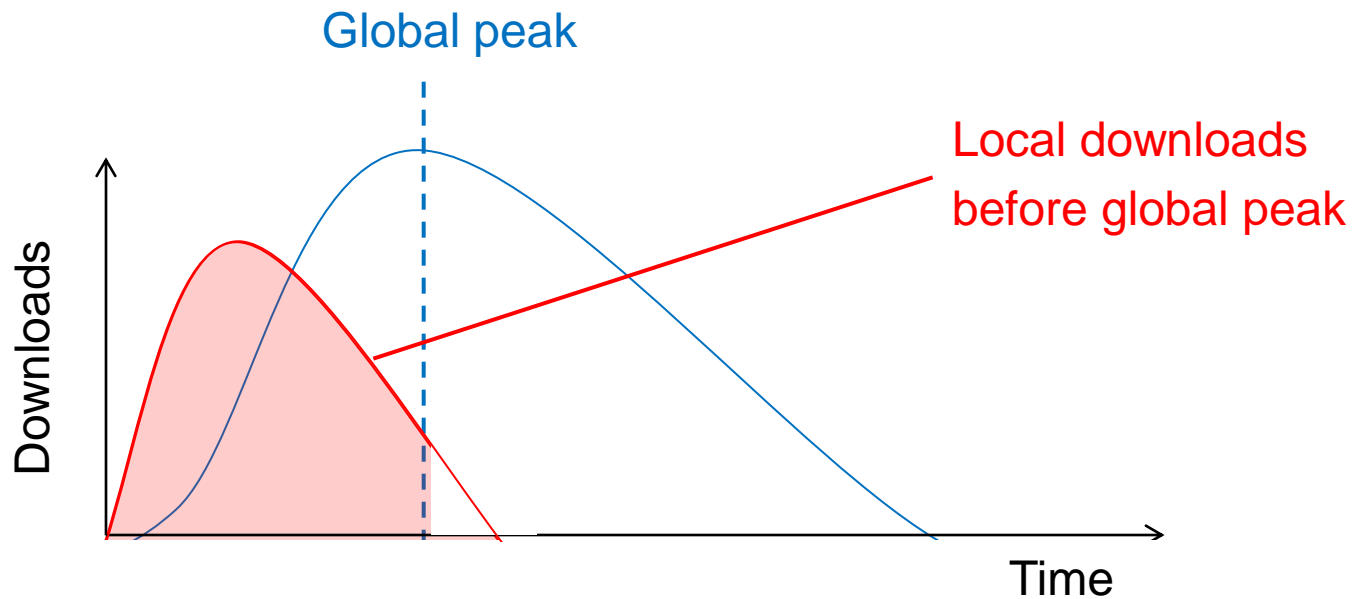
LINKÖPINGS UNIVERSITET

ote r  
lliv sam



# Early adopters

## Terminology



Just det att kiale  
fär marben själ  
dugarna med sif  
den paltbrödem  
det frunc on lada  
Och det frunc m  
som blanta 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 P x  
P x P x P x

minni sko ej  
fter vad hon  
er vad hon

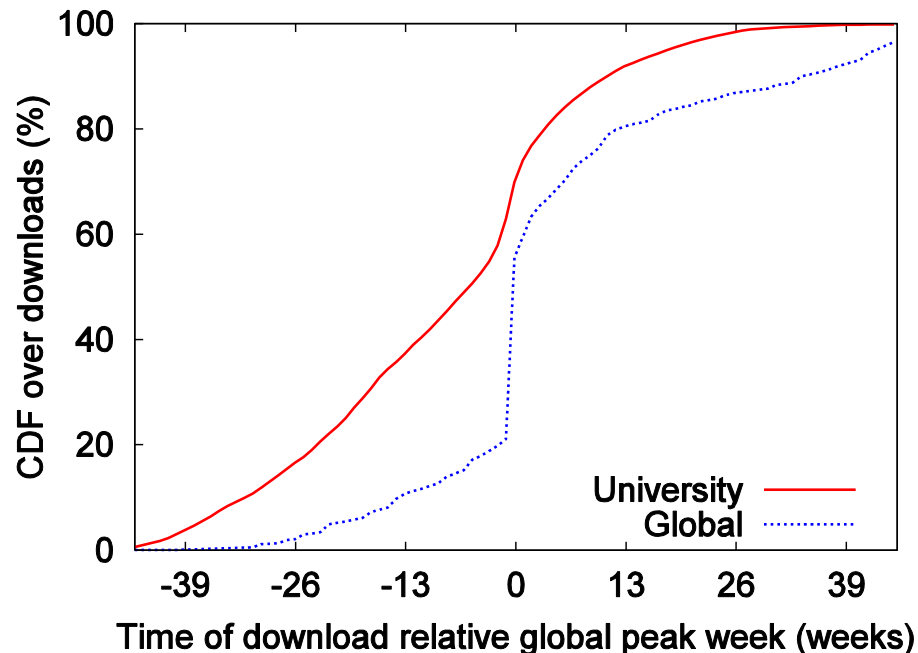
- Rätt teori och rät
- Affärsrätt
- Offentlig rätt
- Arterätt, skade

OnLiU  
www.liu.se

92: g  
B. l  
LINKÖPINGS UNIVERSITET

# Early adopters

## Downloads relative to global peak

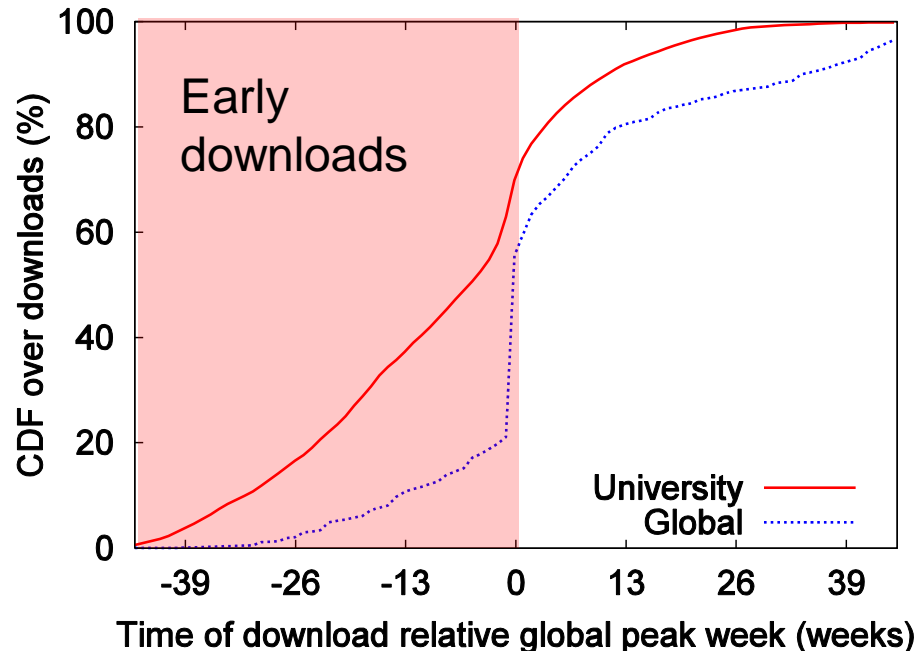


- Campus users are generally early adopters of content
  - 70% of downloads before global peak
  - 40% of downloads at least 10 weeks before global peak



# Early adopters

## Downloads relative to global peak



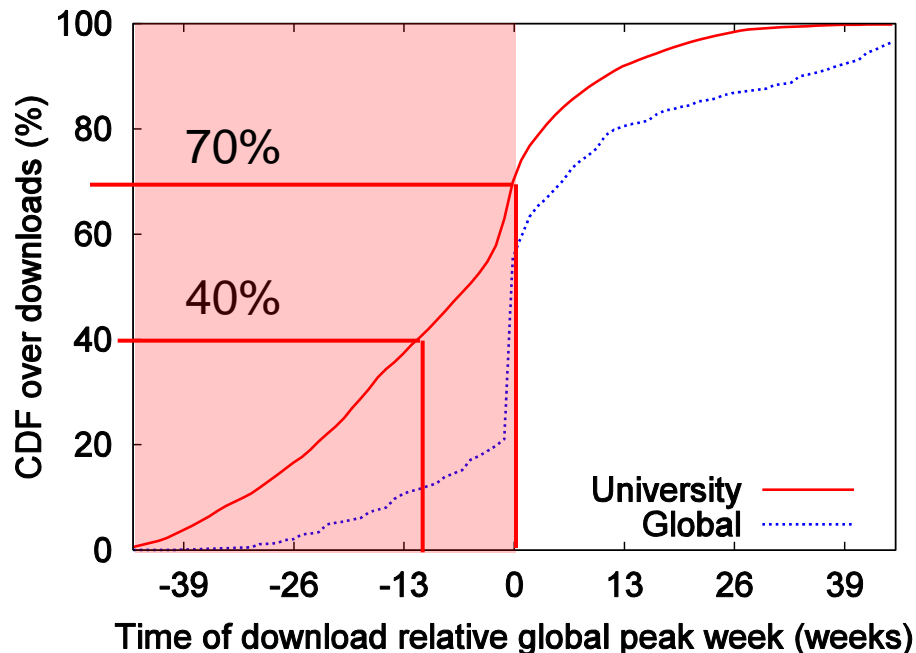
- Campus users are generally early adopters of content
  - 70% of downloads before global peak
  - 40% of downloads at least 10 weeks before global peak





# Early adopters

## Downloads relative to global peak

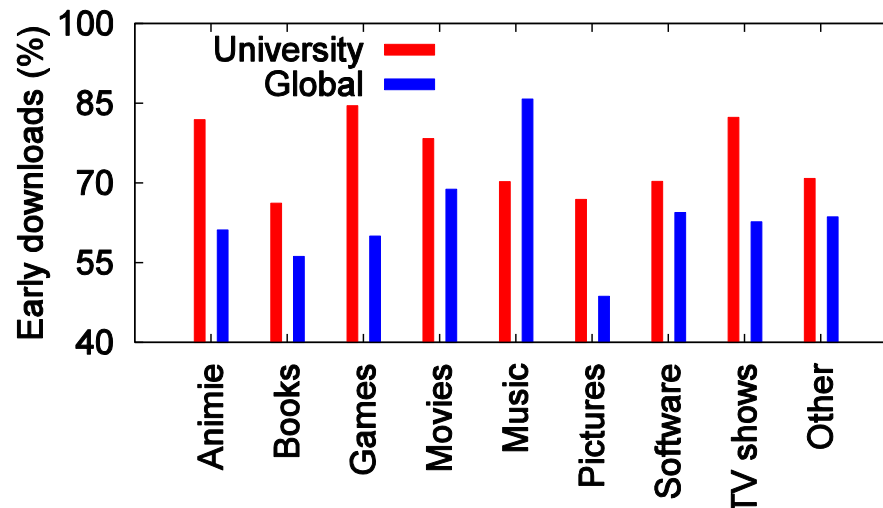


- Campus users are generally early adopters of content
  - 70% of downloads before global peak
  - 40% of downloads at least 10 weeks before global peak



# Early adopters

## Downloads relative to global peak

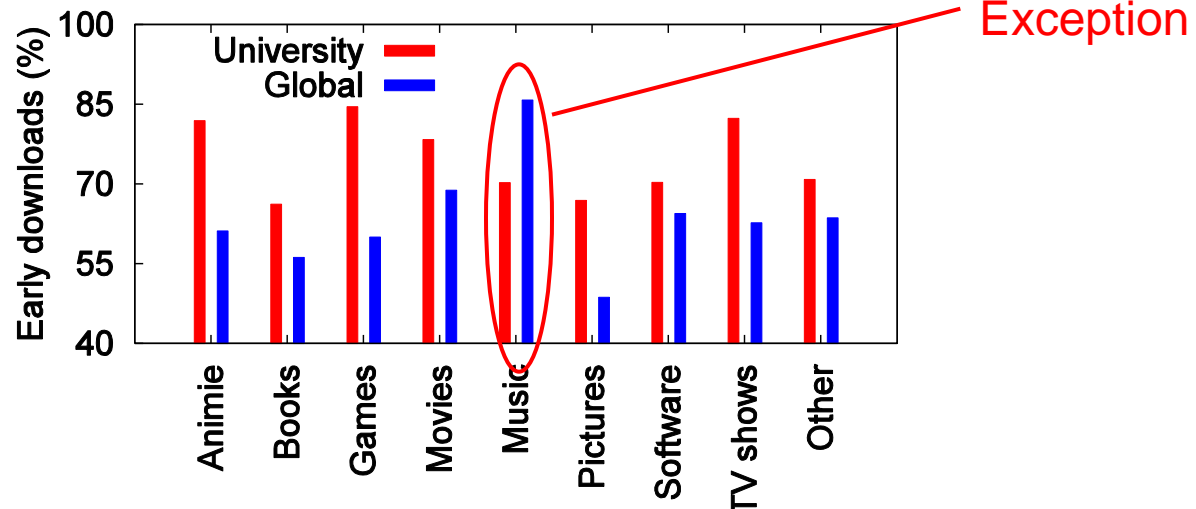


- Campus users are generally early adopters of content
  - Except for music
- Perhaps campus users can be used to predict some future popularity ...
  - And used for seeding such content



# Early adopters

## Downloads relative to global peak

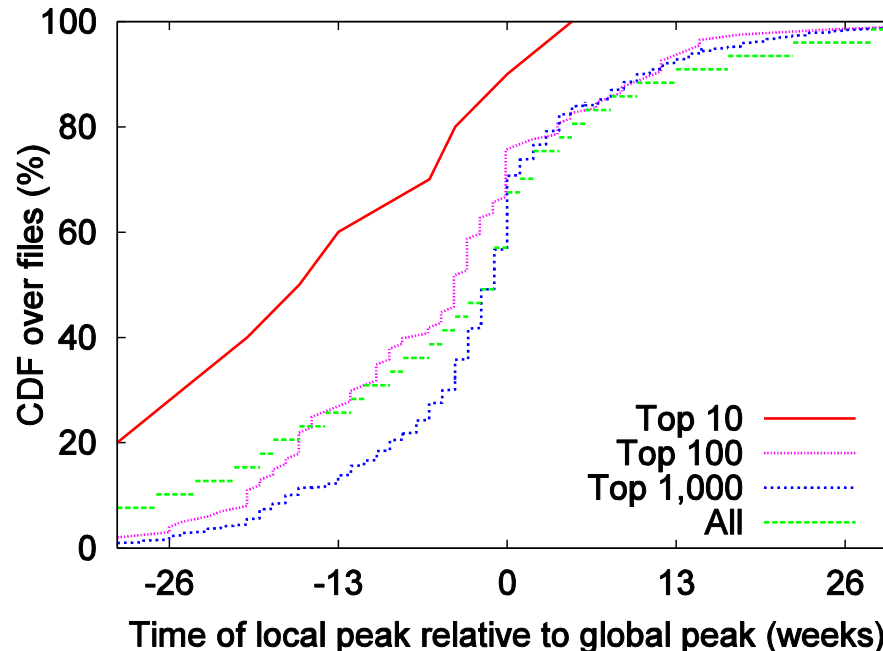


- Campus users are generally early adopters of content
  - Except for music
- Perhaps campus users can be used to predict some future popularity ...
  - And used for seeding such content



# Early adopters

## Downloads relative to global peak

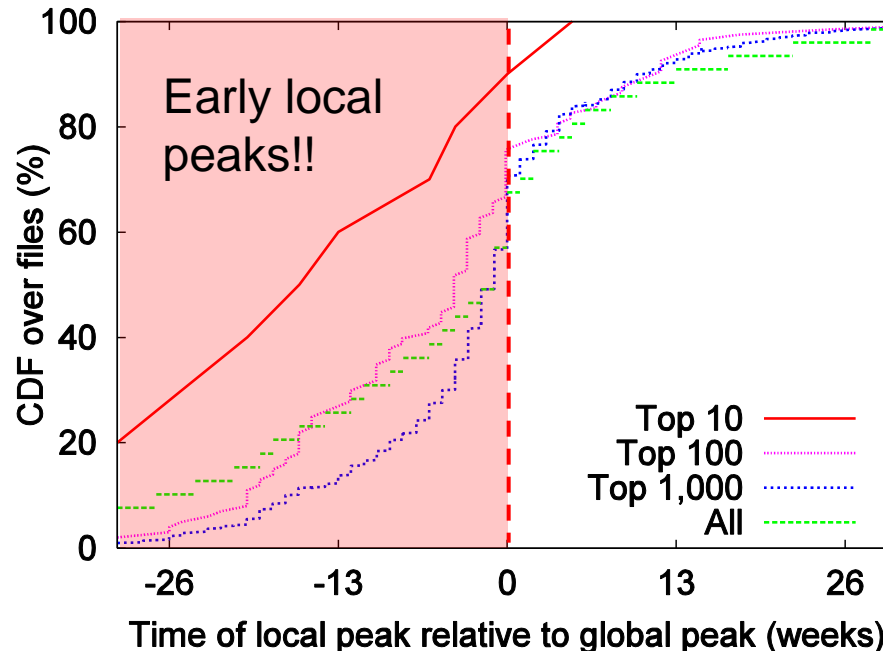


- Better predictor the more popular the content becomes
  - As well as for some niche content ...



# Early adopters

## Downloads relative to global peak

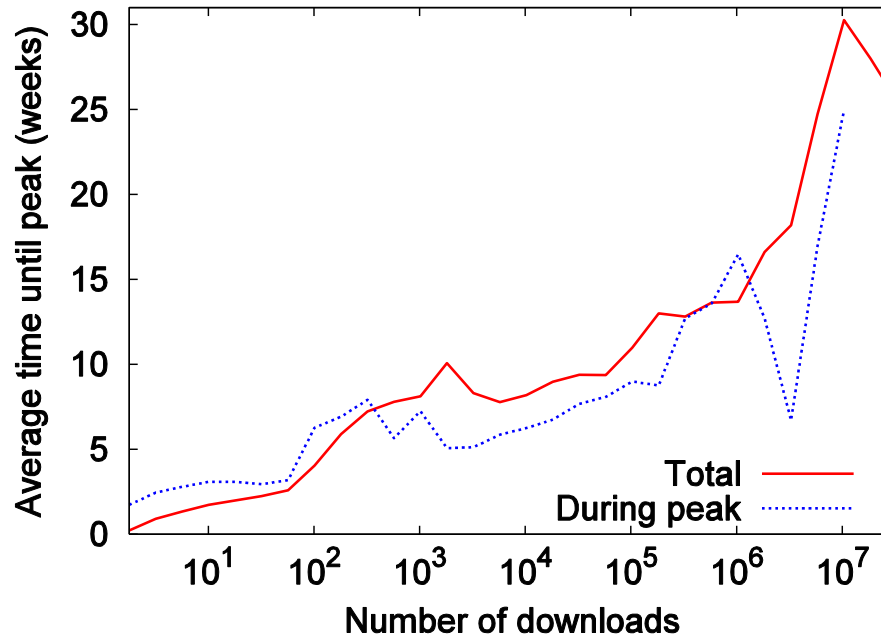


- Better predictor the more popular the content becomes
  - As well as for some niche content ...



# Time until peak

## Global popularity peaks ...



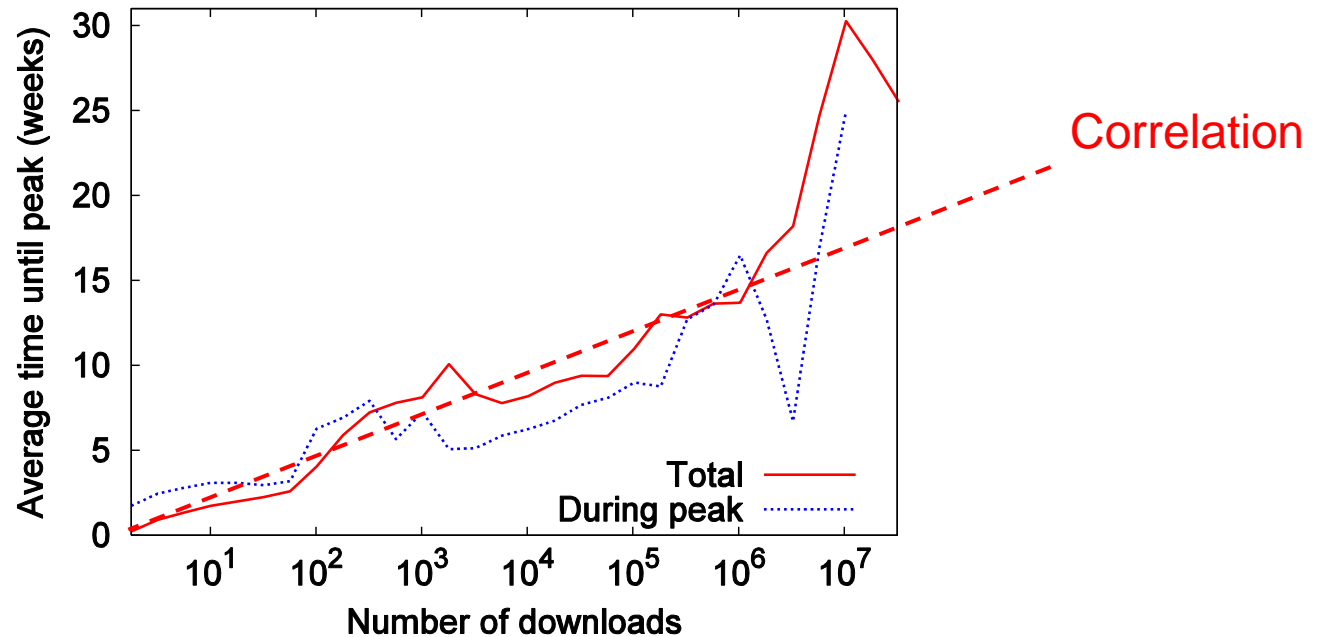
- The global popularity often peak late for popular content
  - Early flash crowds do not dominate the popularity
  - Perhaps a sign that rich-gets-richer a better model ...





# Time until peak

## Global popularity peaks ...

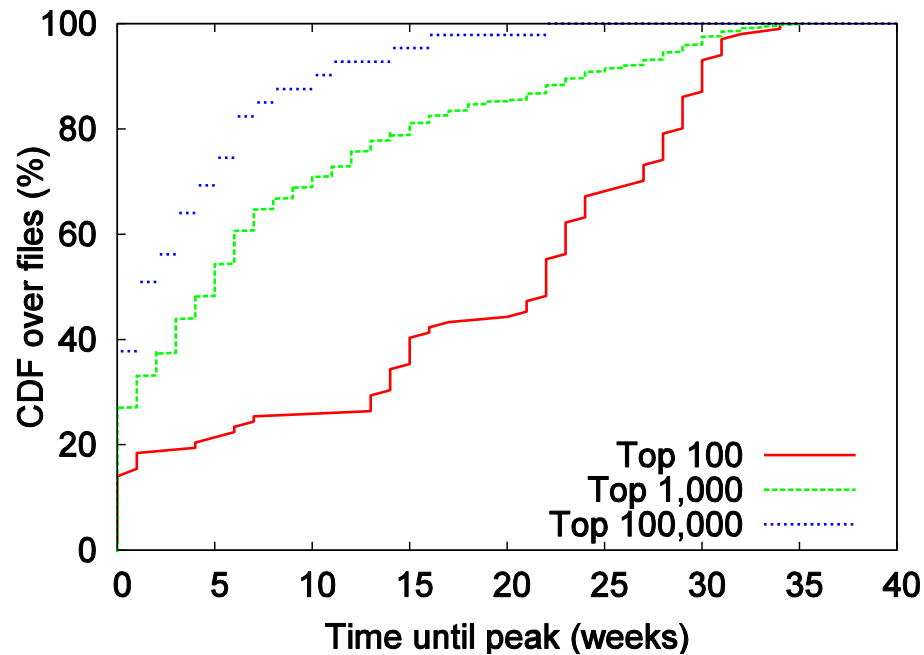


- The global popularity often peak later for popular content
  - Early flash crowds do not dominate the popularity
  - Perhaps a sign that rich-gets-richer a better model ...



# Time until peak

## Global popularity peaks ...

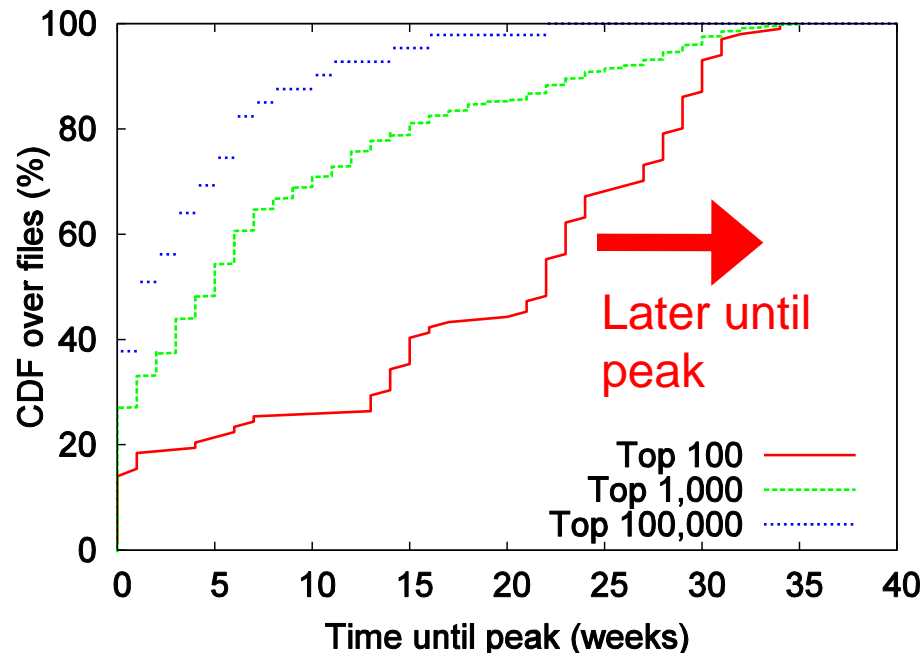


- The more popular the content
  - The later it peaks ...



# Time until peak

## Global popularity peaks ...

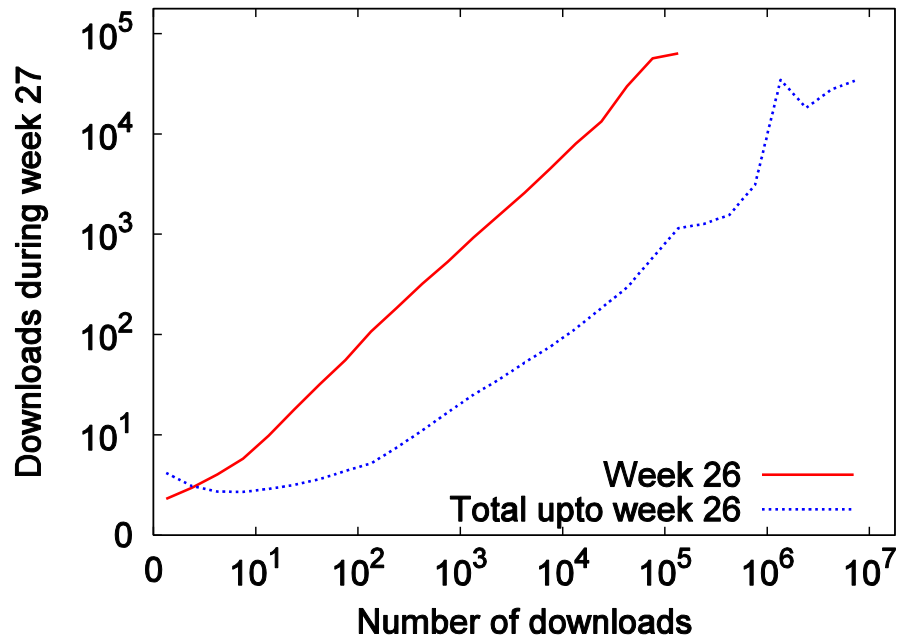


- The more popular the content
  - The later it peaks ...



# Time until peak

## Global popularity peaks ...

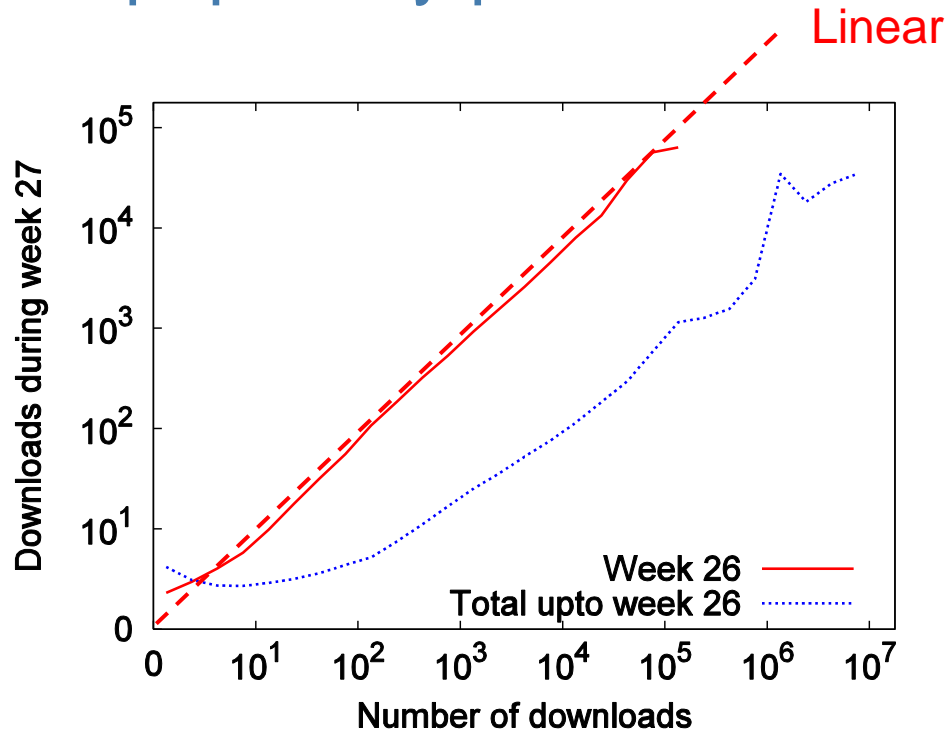


- Rich-gets-richer
  - Close to linear from week-to-week
  - Cumulative total downloads show weaker (sub-linear) rich-gets-richer behavior



# Time until peak

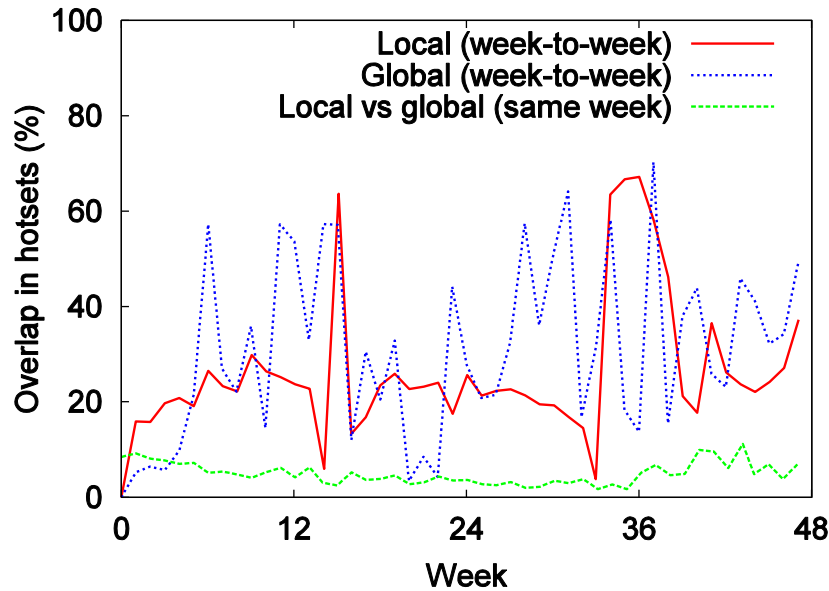
## Global popularity peaks ...



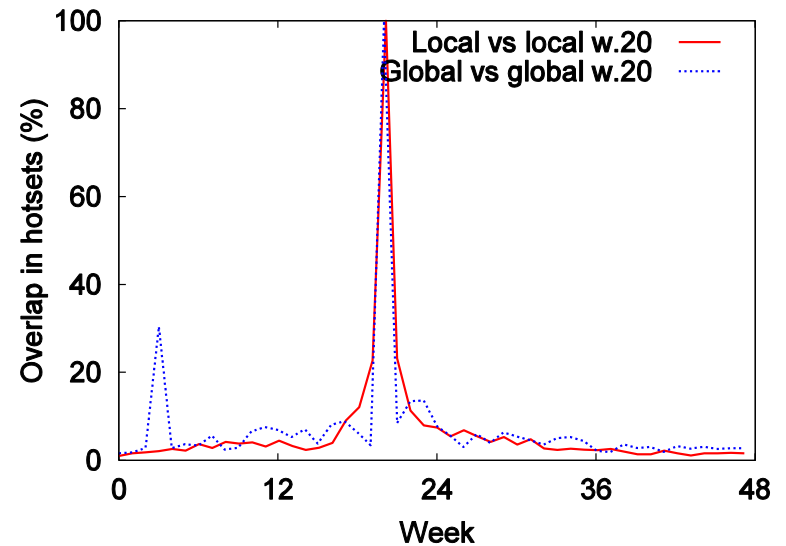
- Rich-gets-richer
  - Close to linear from week-to-week
  - Cumulative total downloads show weaker (sub-linear) rich-gets-richer behavior

# Hotset analysis

## Popularity churn



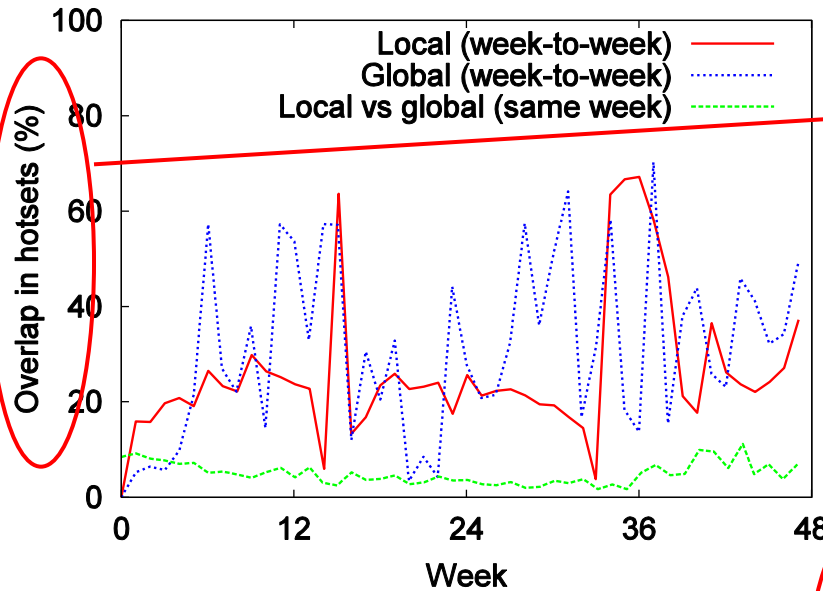
- High popularity churn
- Roughly 50-60% new videos each week
- Some files reoccur
- Some video reoccur in hotset





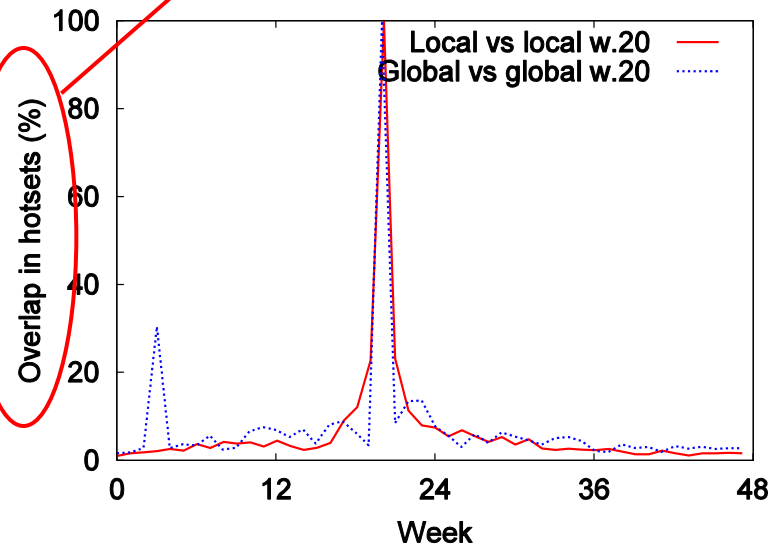
# Hotset analysis

## 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

Just det att hiale  
fär marben sjä  
dugarna med s  
den paltbröden  
det frue on lada  
Och det frue m  
som blanta riefat  
u masonu

ECKEN

ENINU

OZUBET

MÖBELDESIGN

wards

QUIS AKAT DÉBATE!!

Por jicinte

odford's daughter

AYEZ-VOUS

minni sko ej  
fter vad hon  
er vad hon

- Kell teor och ra  
- AHA crätt  
- Offentlig rätt  
- AHA crätt, skade

OnLiU  
www.liu.se

LINKÖPINGS UNIVERSITET

# 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

