

# The Hidden Locality in Swarms

**John S. Otto**

Fabián E. Bustamante

*Northwestern U.*



**AquaLab**

# Locality in P2P systems

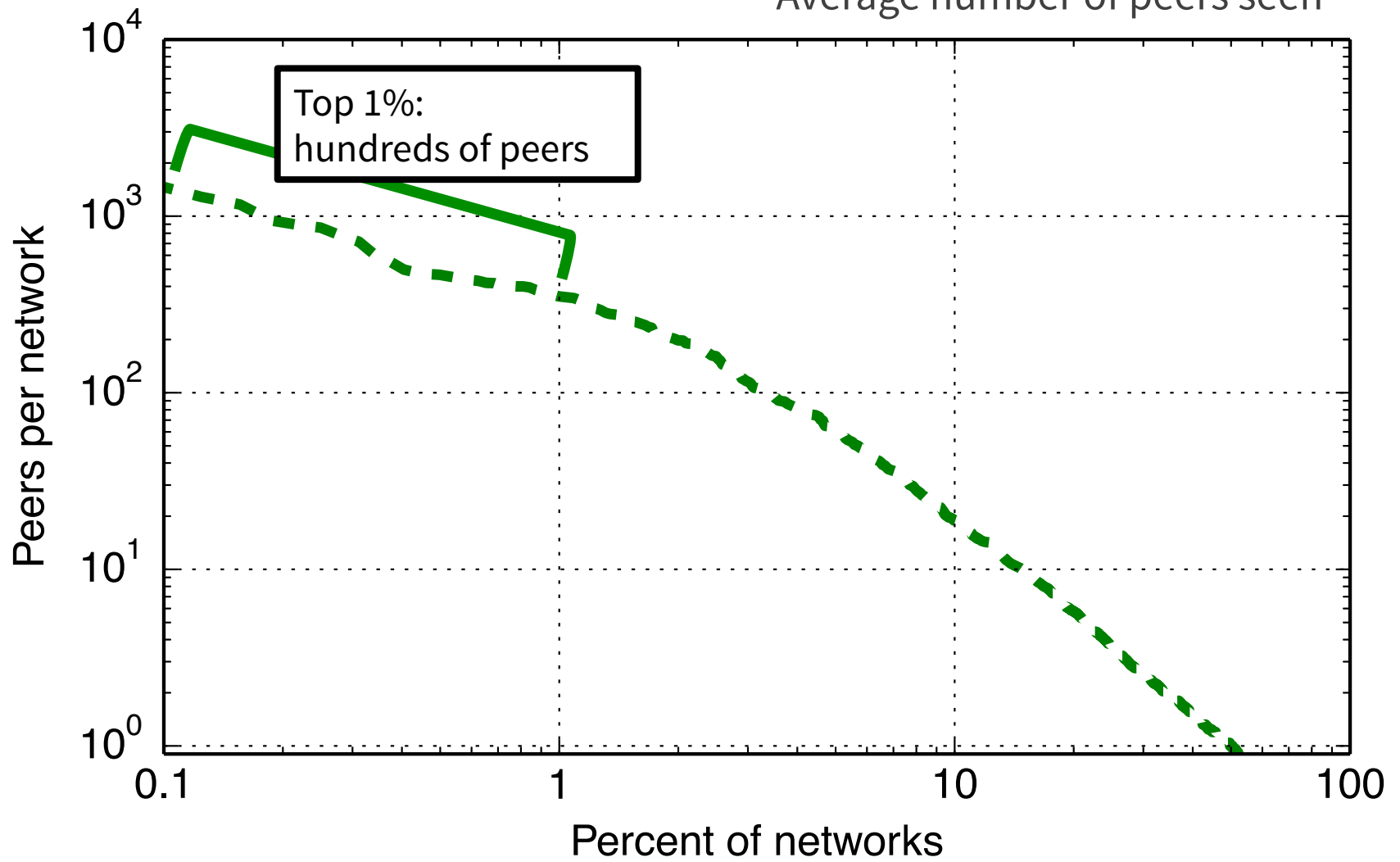
- Avoid cross-ISP traffic
- Better performance
- Lower ISP costs, reduced peering link congestion
  
- Much work toward finding and connecting to local peers
  - Ono, P4P, ALTO
  
- Why are we still talking about this?
  - Does much locality exist?
  - How can a peer find it?

# Widespread locality

- Focus on popular torrents → where challenge of finding locality is hardest
- Use Dasu platform of widely-distributed nodes
- Repeated requests to all trackers → approximate full knowledge of the swarm

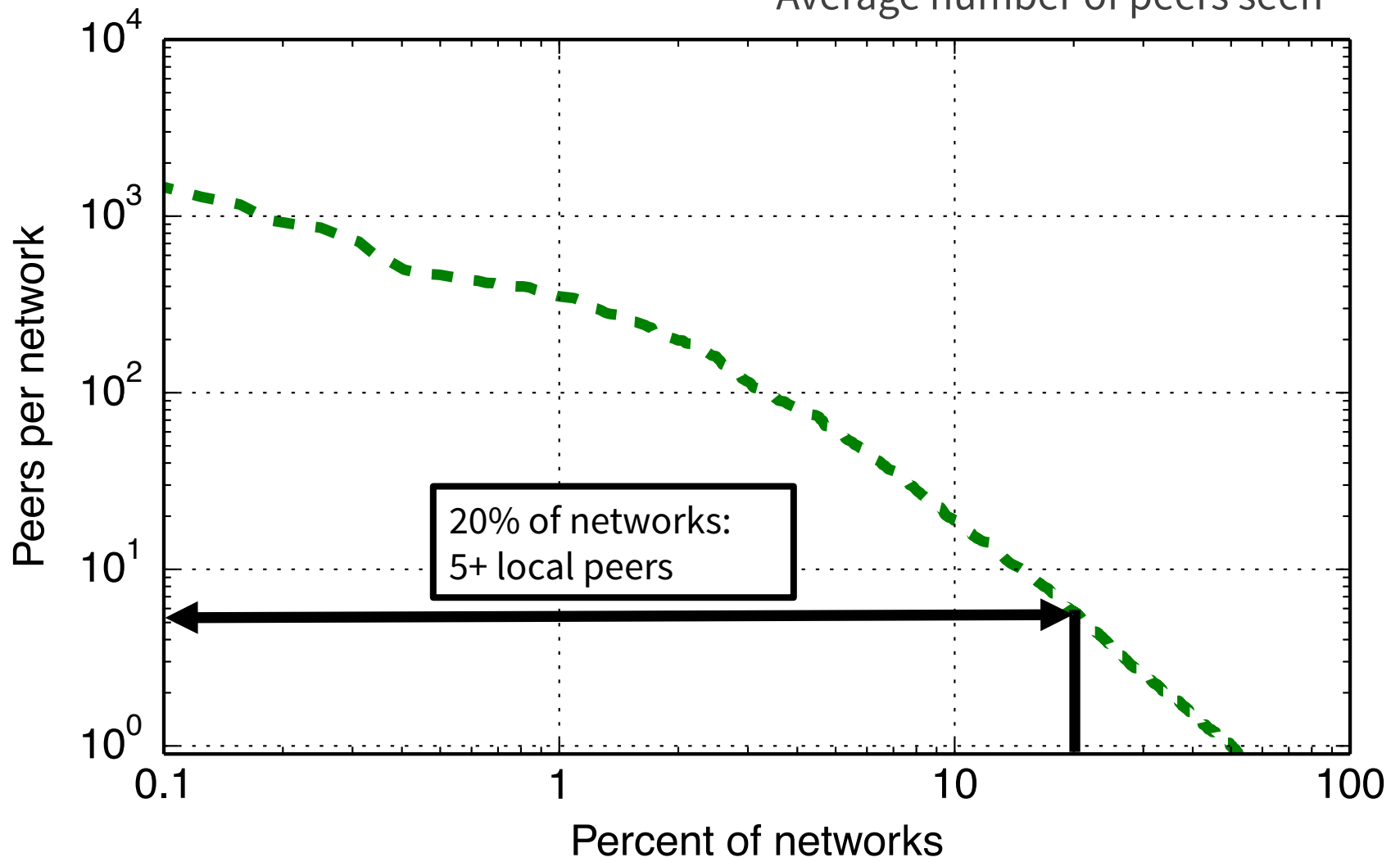
# Widespread locality

Average number of peers seen



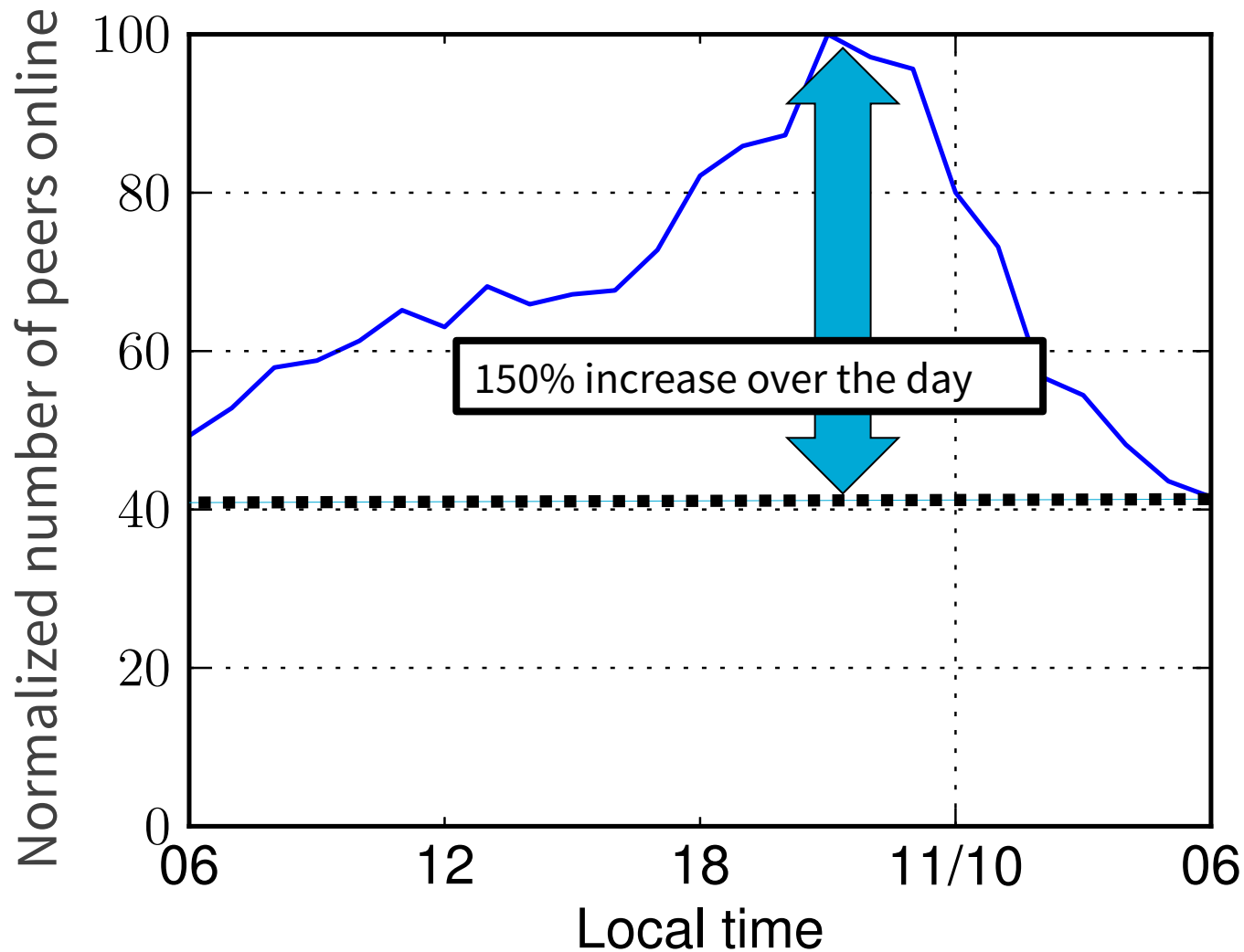
# Widespread locality

Average number of peers seen

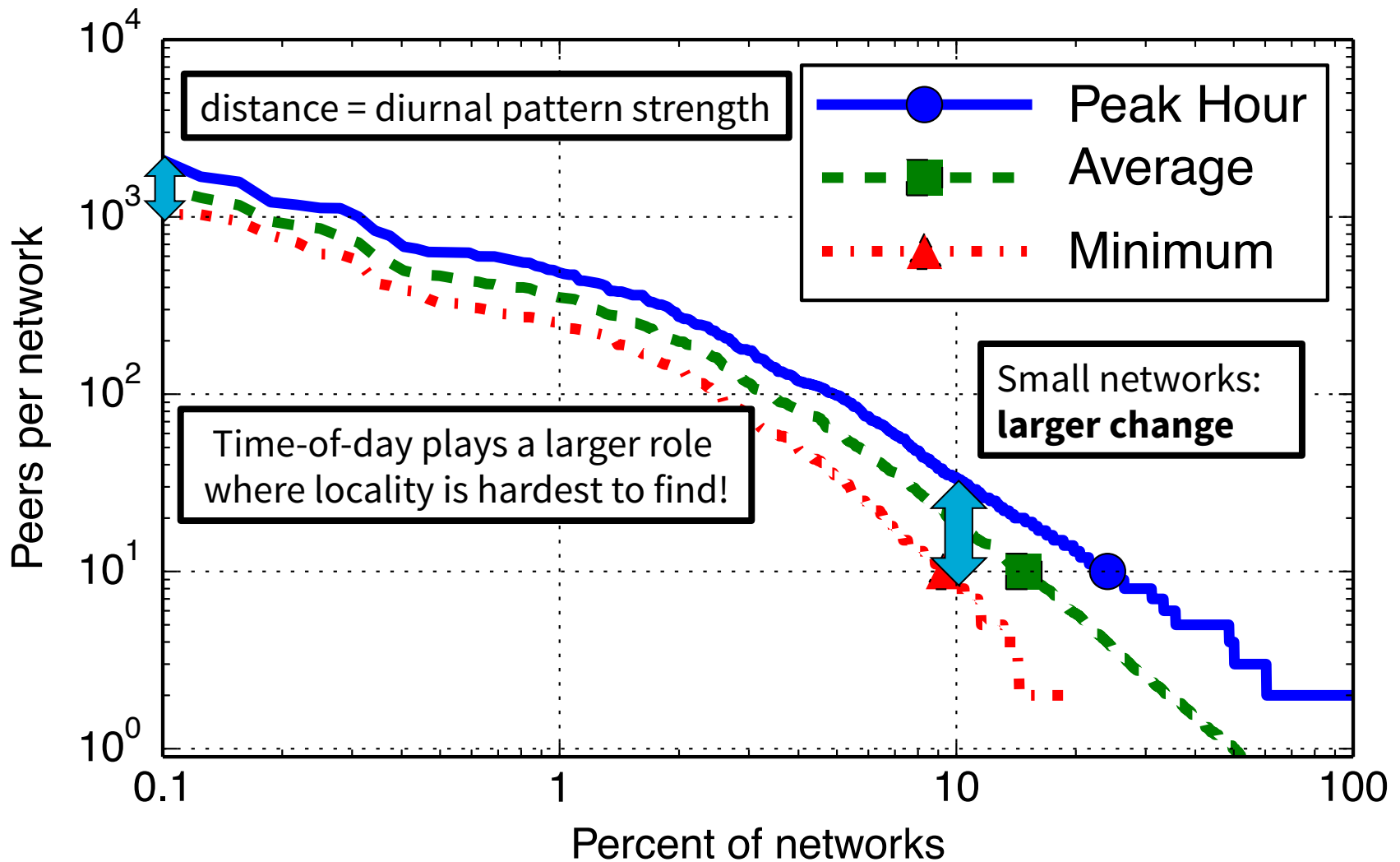


# ...but it matters when you look

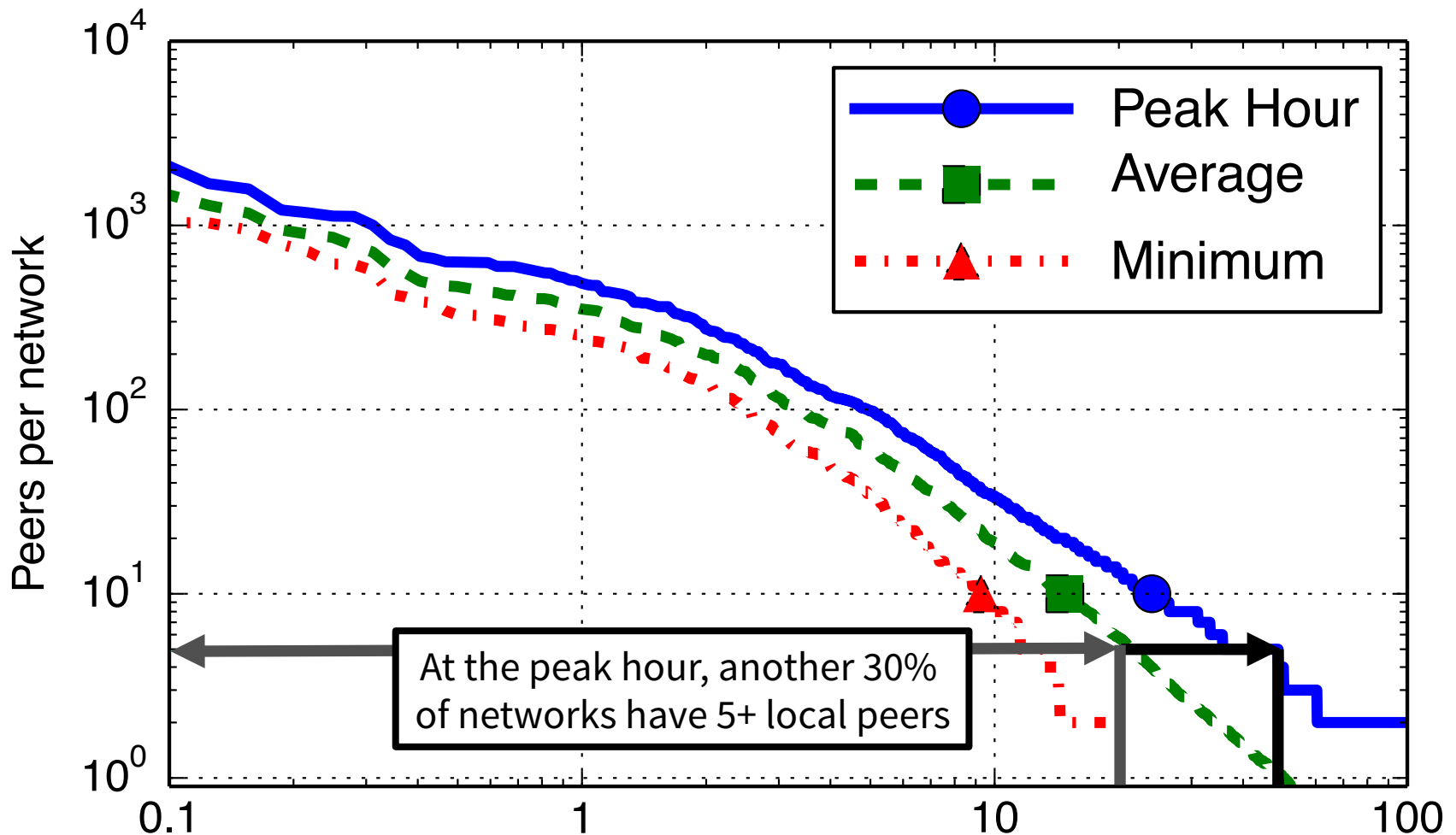
For instance, in a single network (Virgin UK)



# Widespread locality, but it matters when you look



# Widespread locality, but it matters when you look



*Locality exists, but can peers find it?*



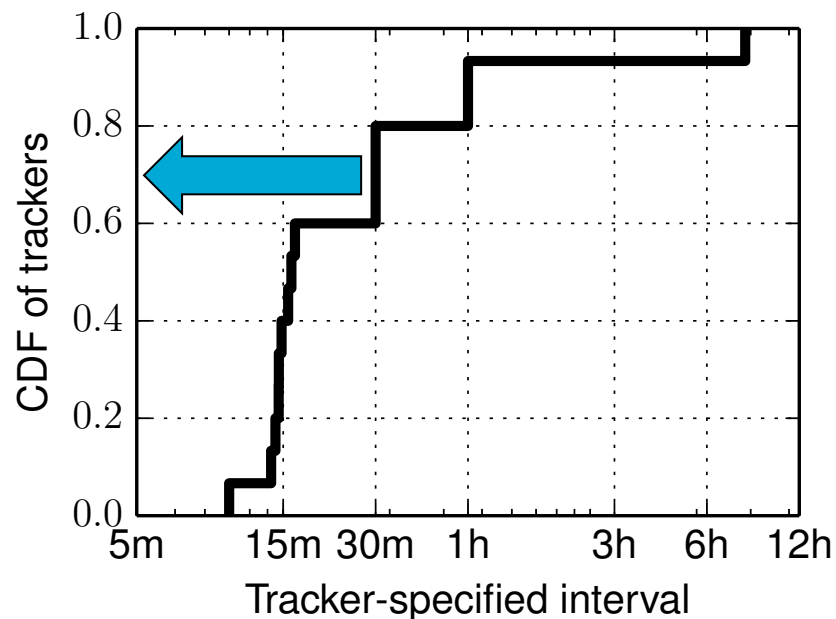
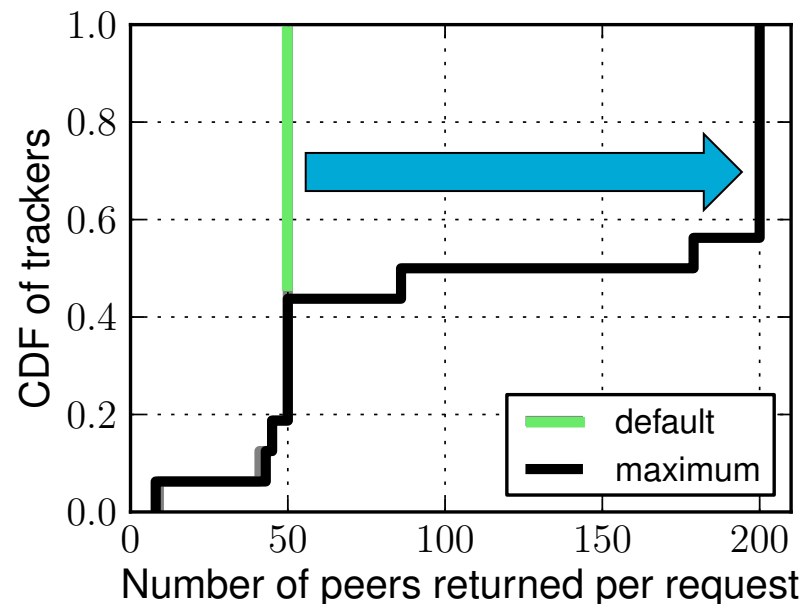
# Peer discovery uses random sampling

- Trackers
- Distributed Hash Table (DHT)
- Peer EXchange (PEX)
- *All use random sampling*, making it difficult to find locality
  
- Why focus on trackers?
  - DHT, PEX disabled in darknets, private torrents
  - We turn a key disadvantage of trackers to our benefit!
  
- If you can't beat 'em...
  - Can we work around the limitations of random sampling?

# More is better

- Ask for larger sample
  - 2x increase in peers
- Query more frequently
  - Trackers did not rate-limit
  - Order-of-magnitude increase

*Client-side ways to speed peer discovery*



# Torrent site diversity works against the user!

The image shows a screenshot of a torrent search results page for the file "Breaking Bad S05E12 HDTV x264-EVOLVE[ettv]". The page lists various torrent sites and their corresponding trackers. Annotations include:

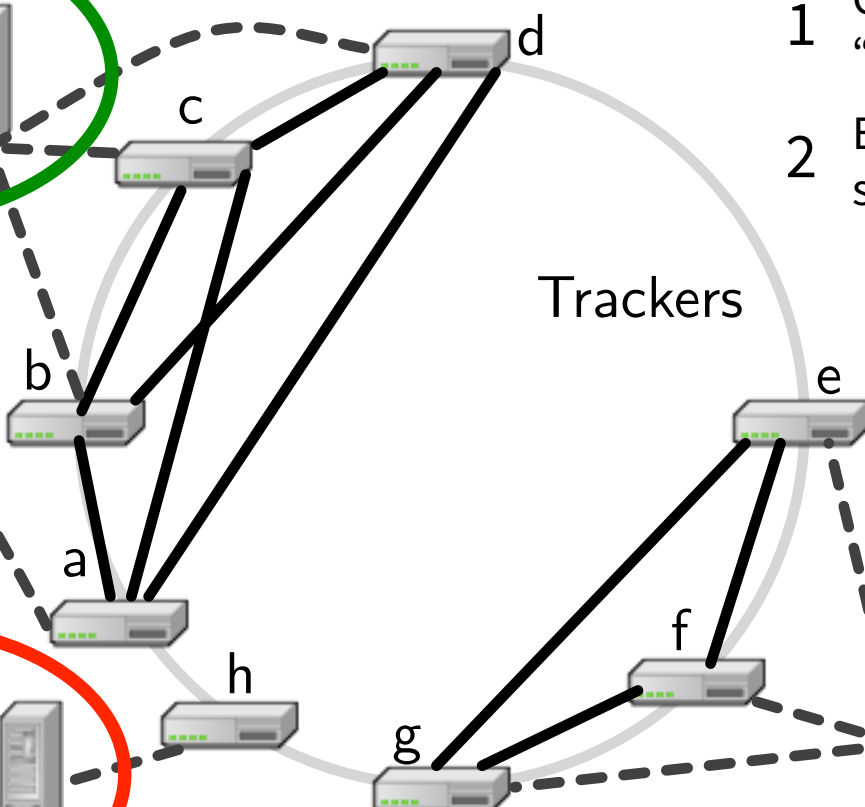
- A blue box highlights the "Download Direct" link and the "Search" button.
- A green box highlights the "extratorrent.com" site in the list of results.
- A blue box highlights the "Torrent Trackers" section, which lists several trackers with their respective URLs and peer counts.
- A green box highlights the "torrentreactor.net" site in the list of results.
- A white callout box with a black border contains the text: "No overlap extratorrent.com users cannot find torrentreactor.net peers".
- A white callout box with a black border contains the text: "Torrent site choice limits to a subset of trackers and the peer population...".
- A white callout box with a black border contains the text: "Solution: use all the trackers!".

Tracker URL	Time	Peers	Seeds
udp://tracker.istole.it:80/announce	1 hour	36,586	3,044
udp://tracker.publicbt.com:80/announce	1 hour	24,779	2,574
udp://tracker.openbittorrent.com:80/announce	1 hour	22,751	2,210
http://9.rarbg.com:2710/announce			
http://121.14.98.151:9090/announce			
http://www.eddie4.nl:6969/announce			
http://explodie.org:6969/announce			
http://p2pdl.com:2710/announce	53 min	1,076	290
http://tracker1.wasabii.com.tw:6969/announce	1 hour	196	62
http://tracker.yibis.com:6969/announce	3 hours	90	28

# Some torrent sites are better than others

Best choice

Search Engine A



Search Engine C



Worst choice

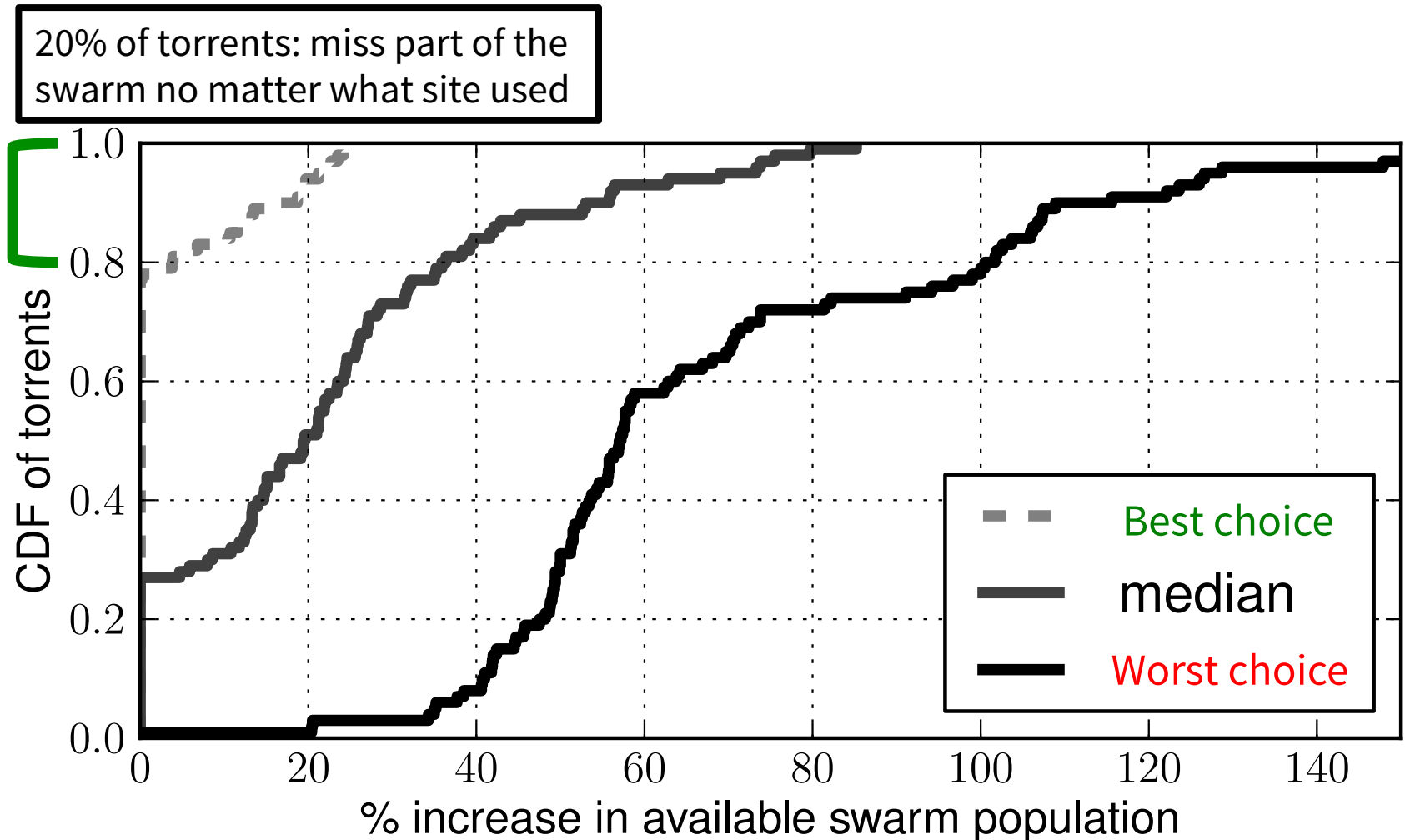
## Graph representation

- 1 Connected trackers are “co-listed” in a .torrent file
- 2 Each component is an isolated subset of the swarm population

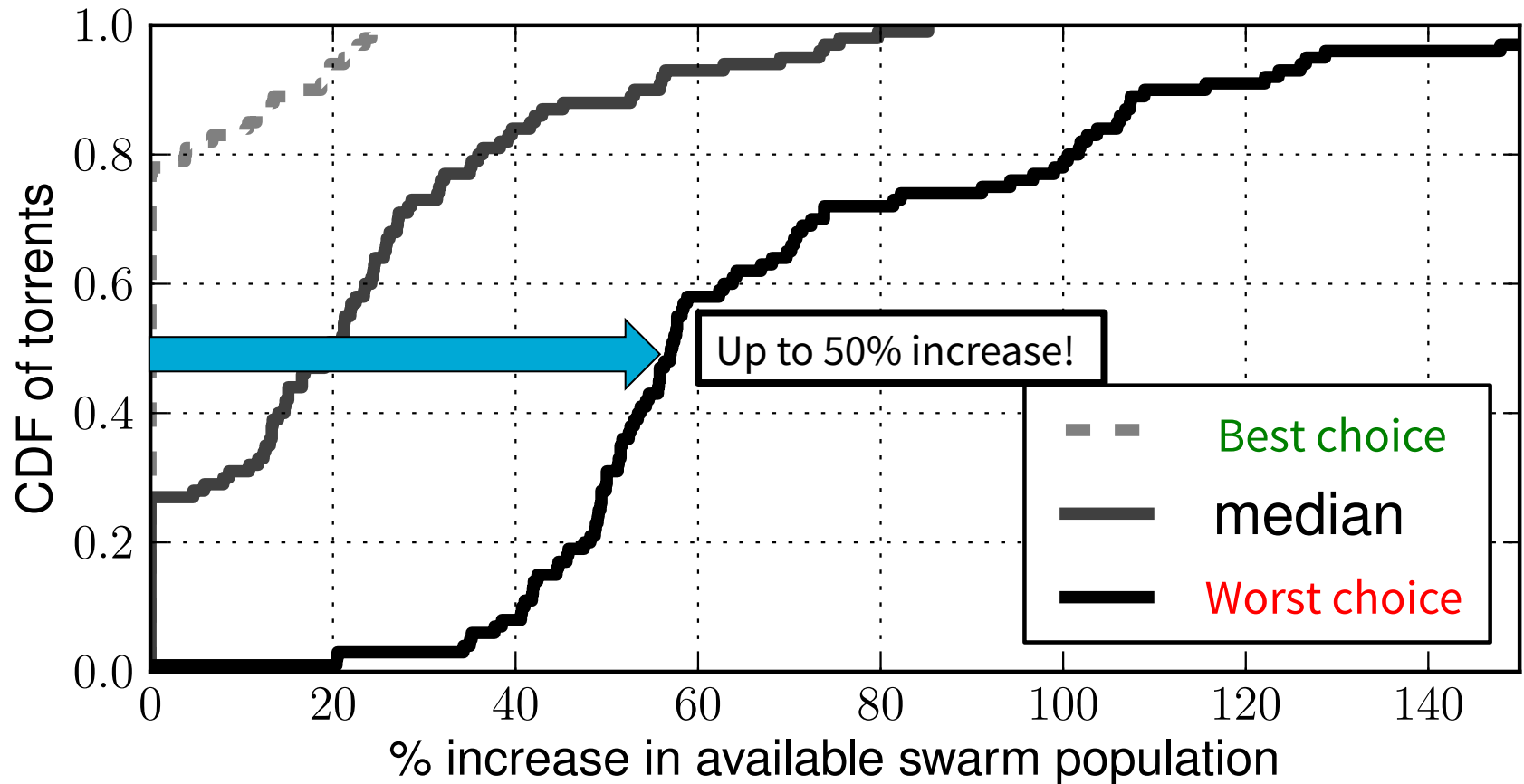
Benefits of using all trackers:

- 1) More sources of peers
- 2) Access to more of the swarm, more local peers

# More trackers = access to more of the swarm



# More trackers = access to more of the swarm



# Summary

- Widespread locality exists
  - 50% of networks have 5+ local peers
  - Time-of-day is a key factor: 200% locality increase during peak hour
- Client-side modifications improve discovery rate by >100x
- Using all trackers is essential to discover available locality
- These techniques boost the efficacy of end-host-based biased neighbor selection

# Why we are still talking about locality

- Sustainable growth requires efficient use of resources
  - local copies of content offload the network
- When systems ossify or are beyond our control, we...
  - cope as best we can
  - build on top – end-host solutions