

Shareable  
How ~~Private~~ are Home  
Directories?

Carlos Maltzahn  
UC Santa Cruz

27 February 2008

# Problem

- People are overwhelmed by their data collection.
- Difficult to manage even with keyword search
  - ▶ Scarce metadata (e.g. few relationships between files)
  - ▶ Small sets of relevant files (high demand for accuracy)
- So what?
  - ▶ Mounting data loss
  - ▶ Digital Age becomes a Dark Age

# Shareability Hypothesis

Idea: Maximize sharing of data management effort

Web has many successful examples:

 Delicious  Digg  reddit  Facebook  StumbleUpon

File systems have enough *shareable files* to make collaborative data management feasible.

# Approach

- *shareable file*:
  - ▶ *should be* managed across file systems and/or users.
  - ▶ not necessarily managed that way, currently.
- How to measure:
  - ▶ compare home directories of a user group
    - ▶ over-estimate due to common system files
    - ▶ under-estimate due to sharing outside the group

# Approach

- *shareable file*:

- ▶ *should be* managed across file systems and/or users.
- ▶ not necessarily managed that way, currently.

- How to measure:

- ▶ ~~compare home directories of a user group~~
  - ▶ ~~over-estimate due to common system files~~
  - ▶ ~~under-estimate due to sharing outside the group~~

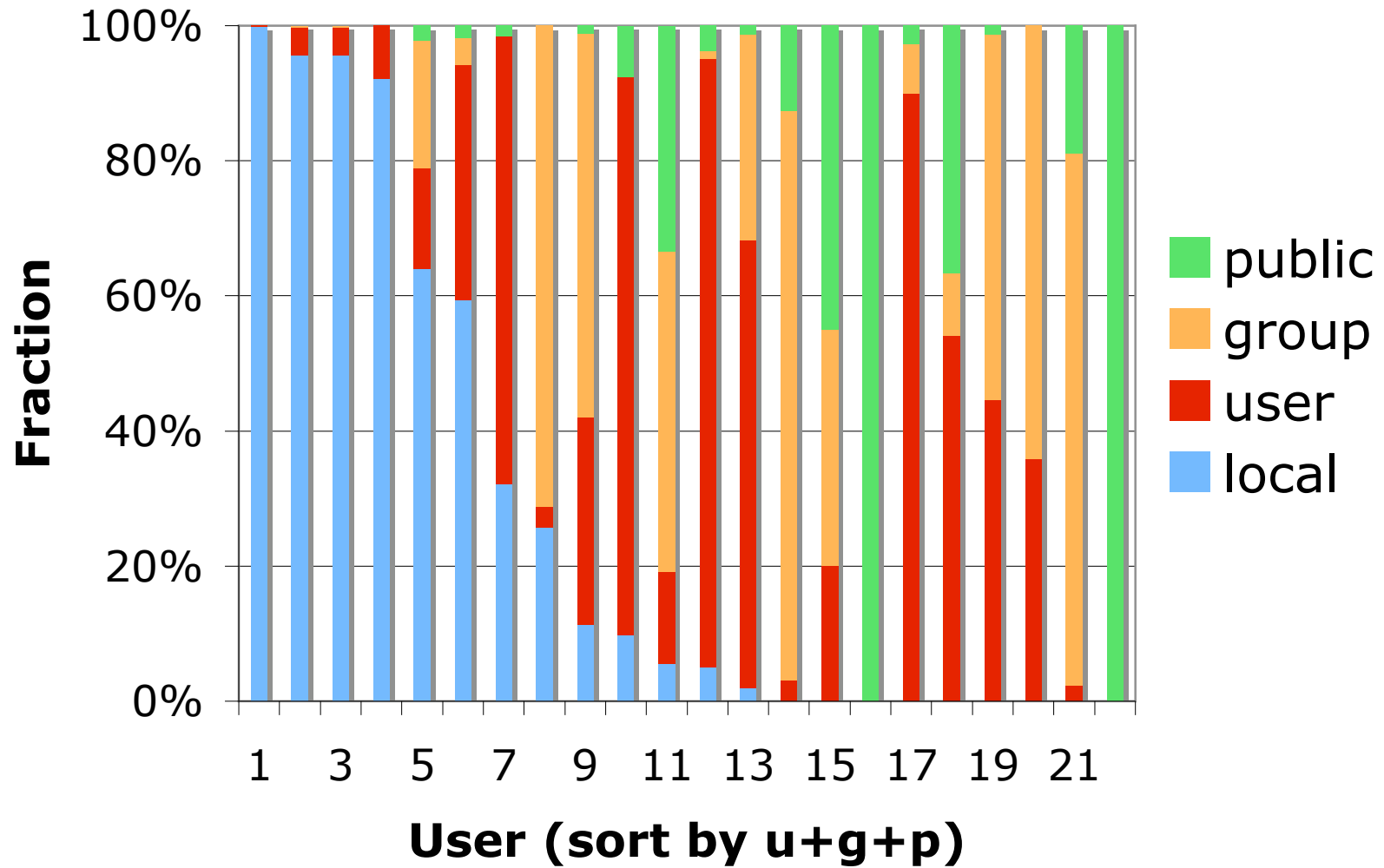
# Approach

- *shareable file*:
  - ▶ *should be managed across file systems and/or users.*
  - ▶ not necessarily managed that way, currently.
- How to measure:
  - ▶ *subjective categorization by user survey*
    - ▶ focus on files user cares about
    - ▶ measurement independent of sample size

# Categories

- **local:** file never leaves this computer (not shareable)
  - ▶ user wants to manage file
  - ▶ not suitable for sharing among computers or users
- **user:** file is private (shareable)
  - ▶ suitable for sharing among computers of same user
- **group:** file is restricted to group (shareable)
- **public:** file is public (shareable)
  - ▶ downloaded or published files

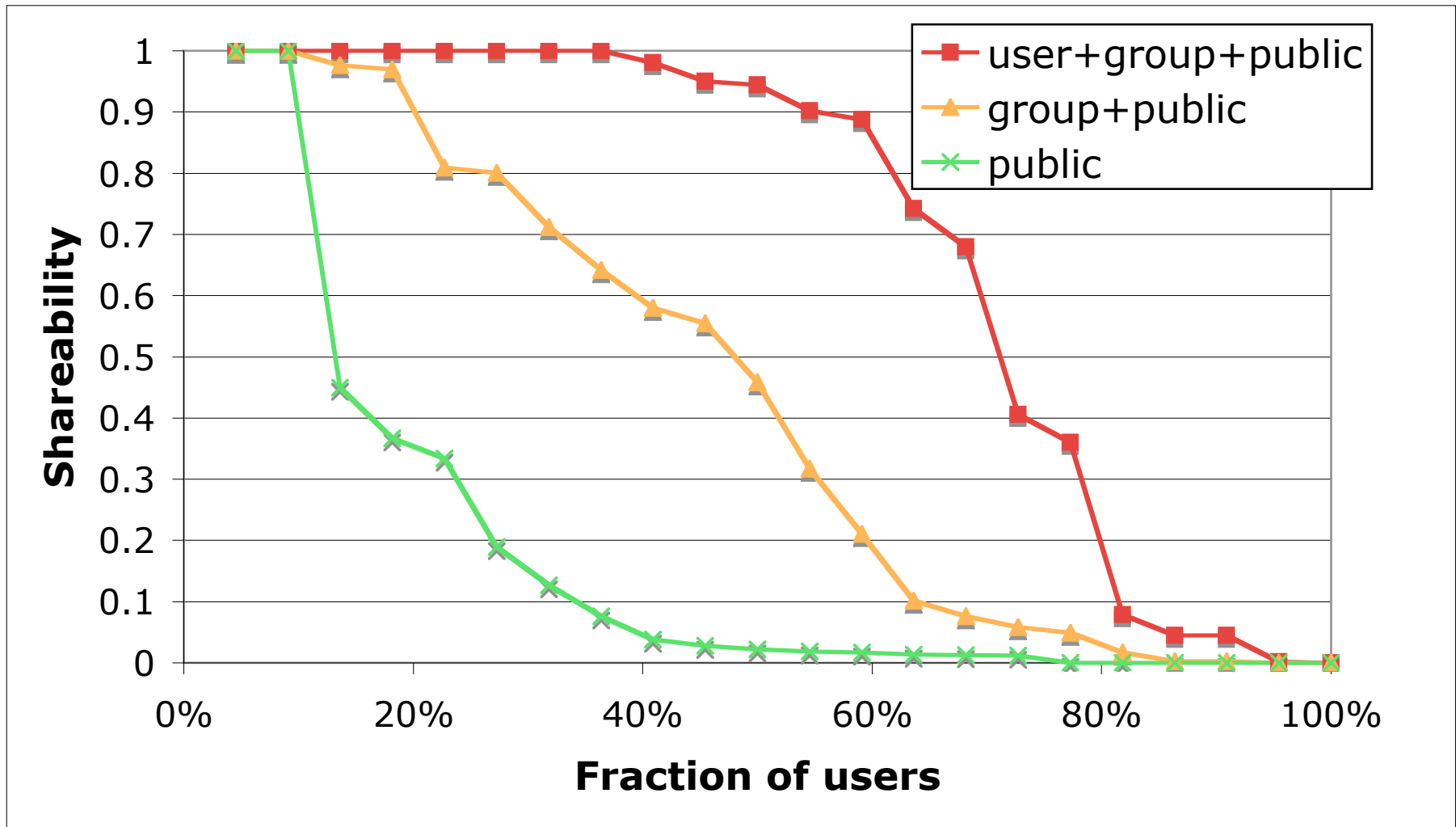
# Results:



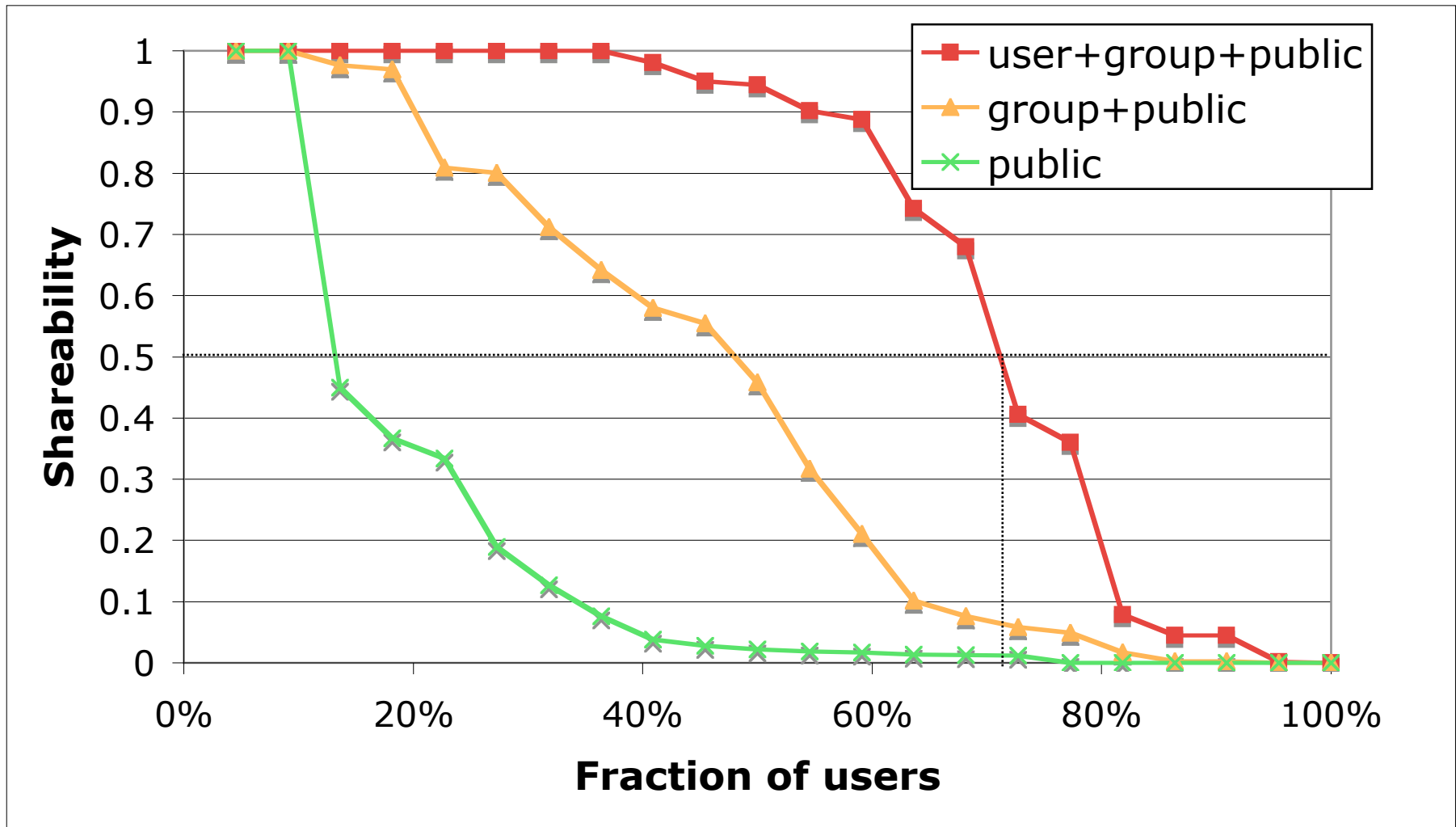
**73% of surveyed & relevant files are shareable!**



# Results:

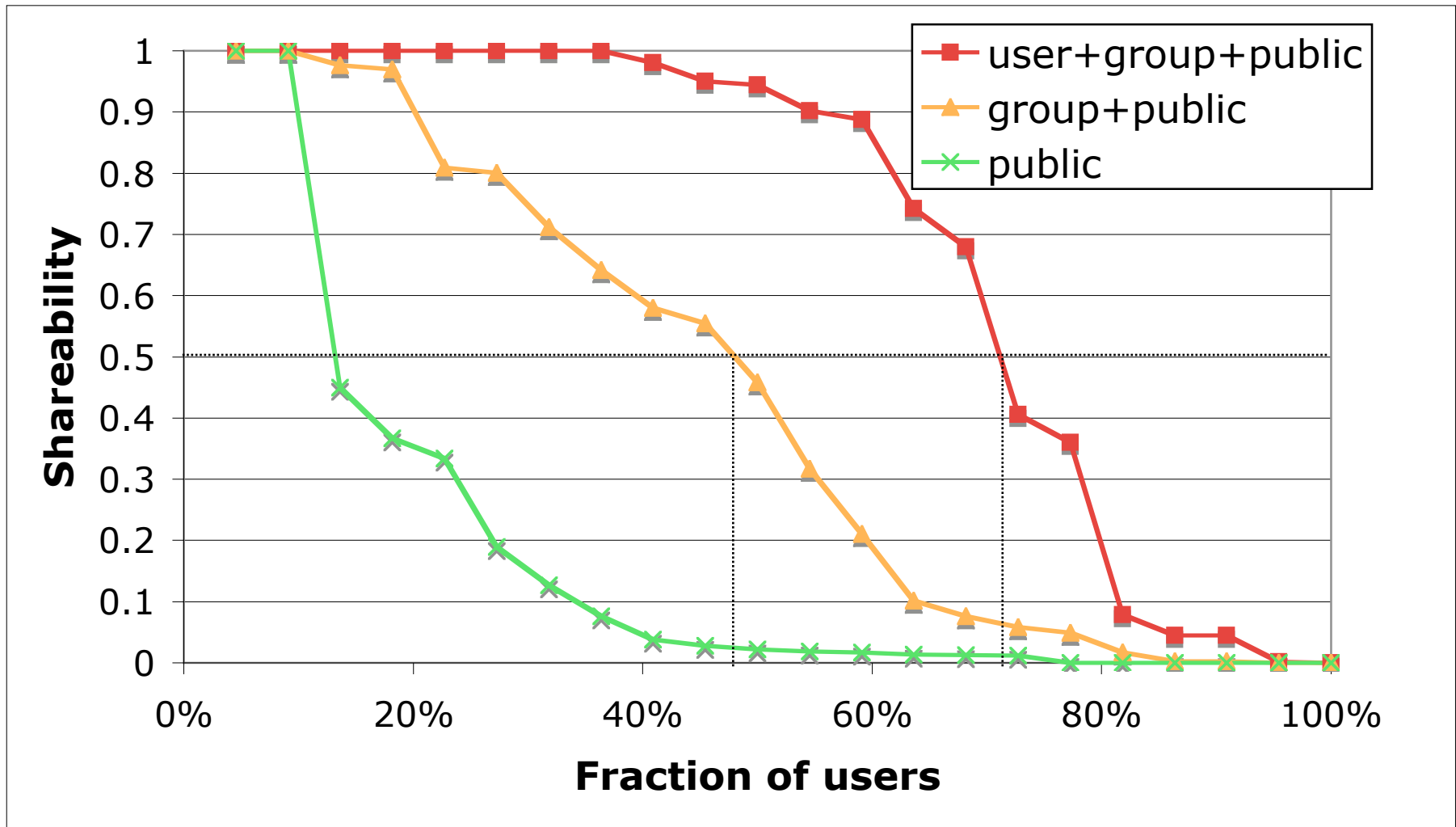


# Results:



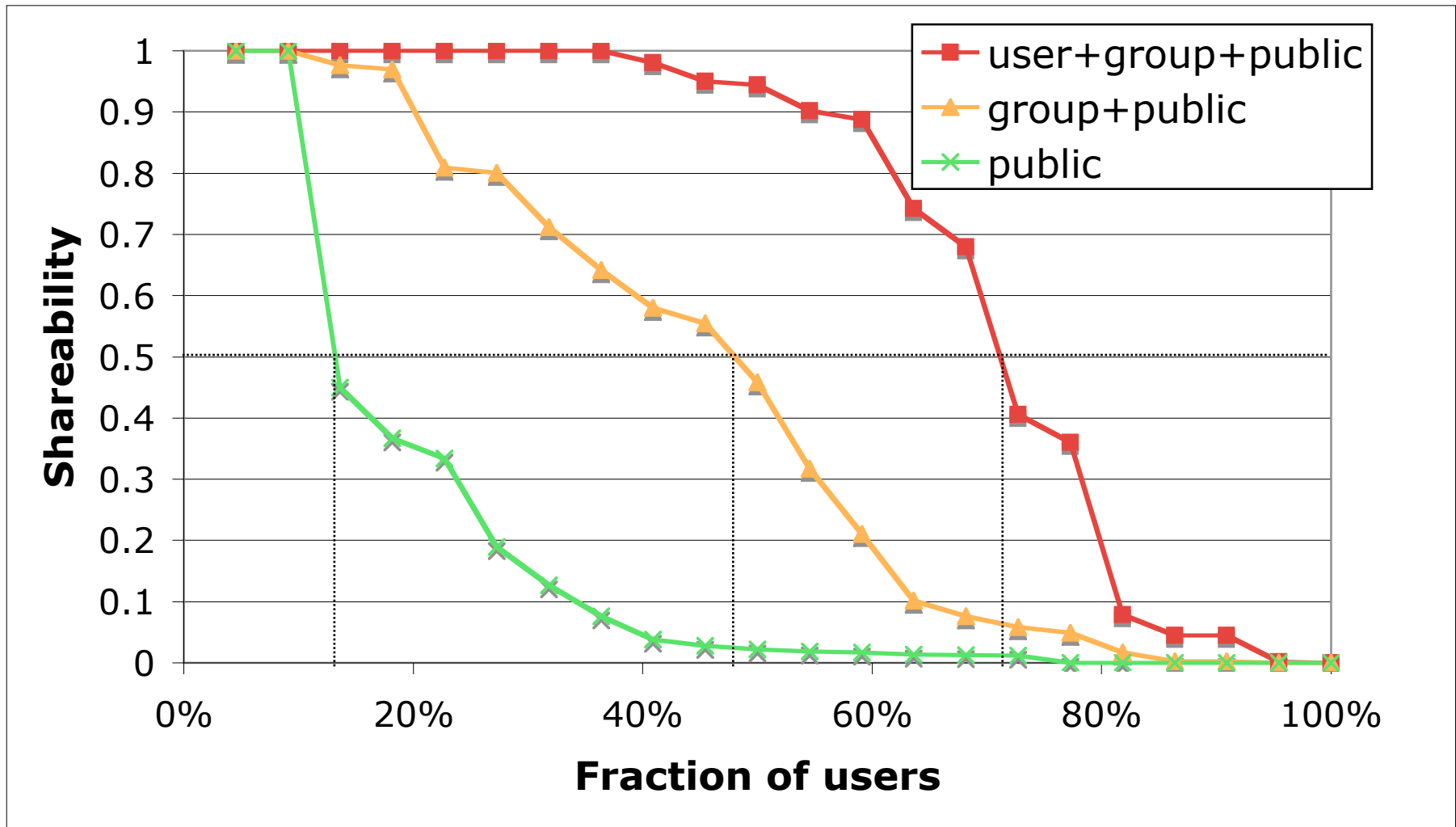
■ 75% of users with  $\geq 50\%$  shareability

# Results:



- 75% of users with  $\geq 50\%$  shareability
- 50% of users with  $\geq 50\%$  shareability across users

# Results:

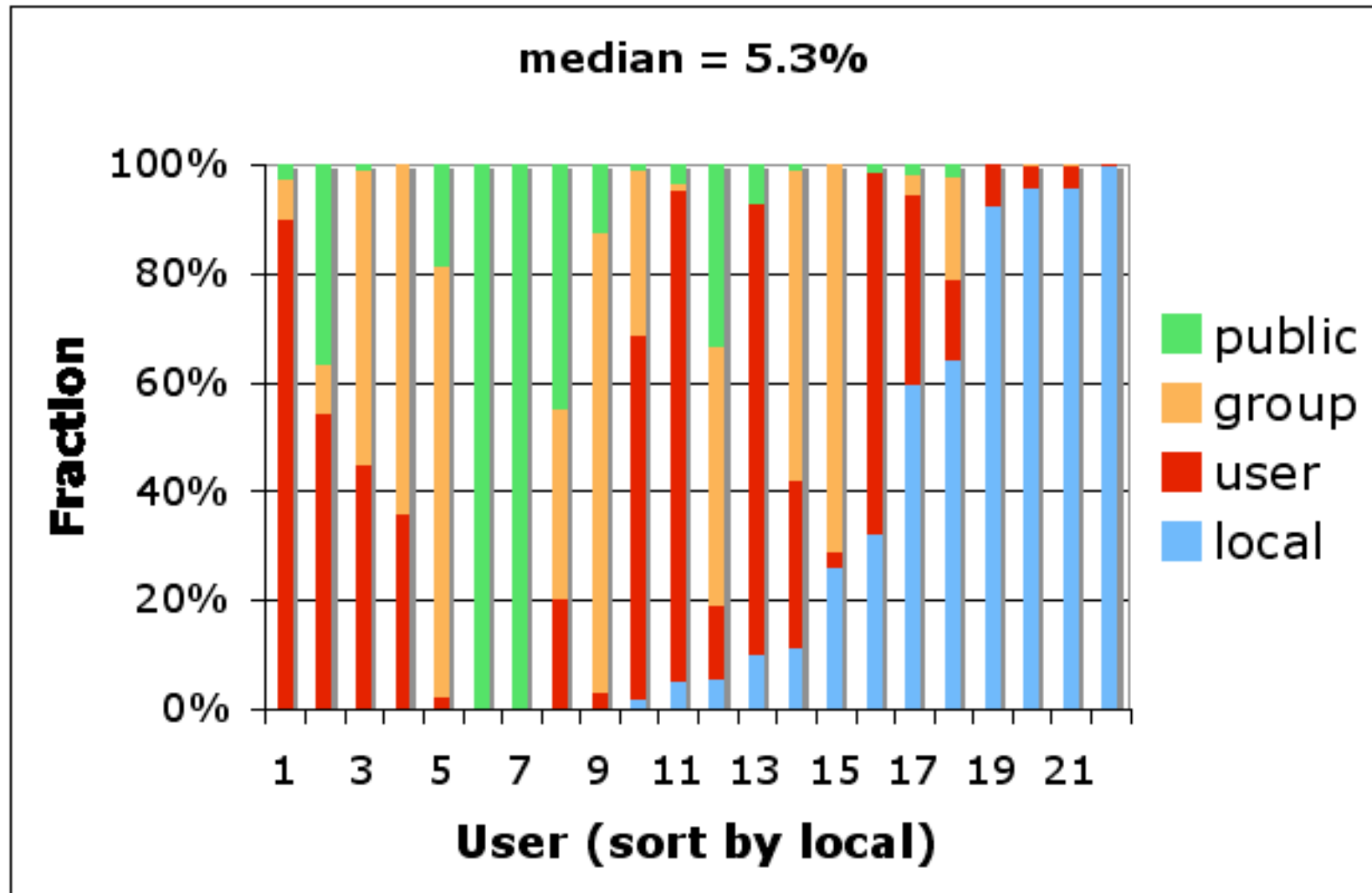


- 75% of users with  $\geq 50\%$  shareability
- 50% of users with  $\geq 50\%$  shareability across users
- 10% of users with  $\geq 50\%$  of files are public

# For more information...

- Ongoing survey:
  - ▶ [www.cs.ucsc.edu/~carlosm/Survey/](http://www.cs.ucsc.edu/~carlosm/Survey/)
- See me at my poster!
- Thank you!

# Results:



# Results:

