Clustering Download Events to Identify Classrooms

Dwight McKay
HUBzero
Purdue University

Michael Zenter
HUBzero
Purdue University

Gerhard Klimeck
Network for Computational Nanotechnology
Purdue University

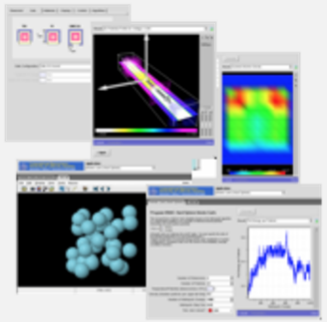
hubzero

Thanks!

The author gratefully acknowledges the Network for Computational Nanotechnology funded by the US National Science Foundation under Grant Nos. EEC-0228390, EEC-1227110, EEC-0228390, EEC-0634750, OCI-0438246, OCI-0832623 and OCI-0721680 for their support of this work. Thanks also to Nathan Denny for editorial assistance and the entire nanoHUB team, especially Lynn Zentner, Krishna Madhaven, and Tanya Faltens for their encouragement and thoughts on presenting this work.

nanoHUB

Simulation & Modeling Tools



Seminars
Tutorials

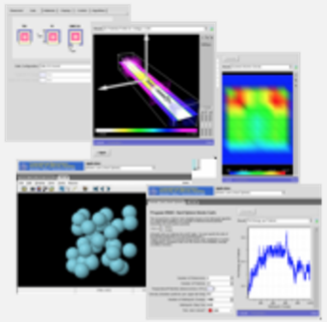
Tech Reports
Teaching Materials

Data from Experiments

hubzero

nanoHUB

Simulation & Modeling Tools



Seminars
Tutorials

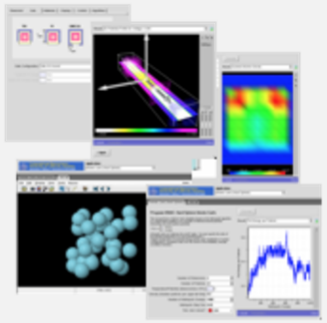
Tech Reports
Teaching Materials

Data from Experiments

hubzero

nanoHUB

Simulation & Modeling Tools



Seminars
Tutorials

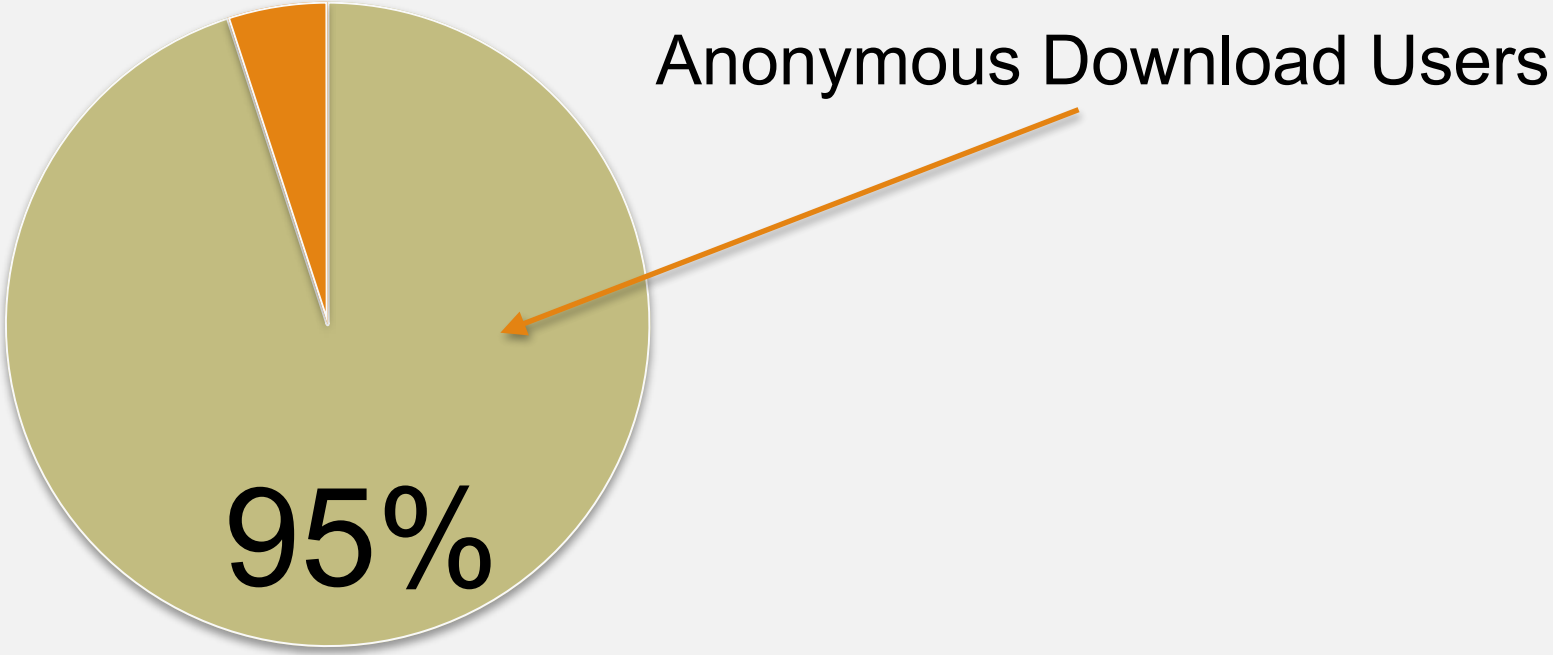
Tech Reports
Teaching Materials

Data from Experiments

hubzero

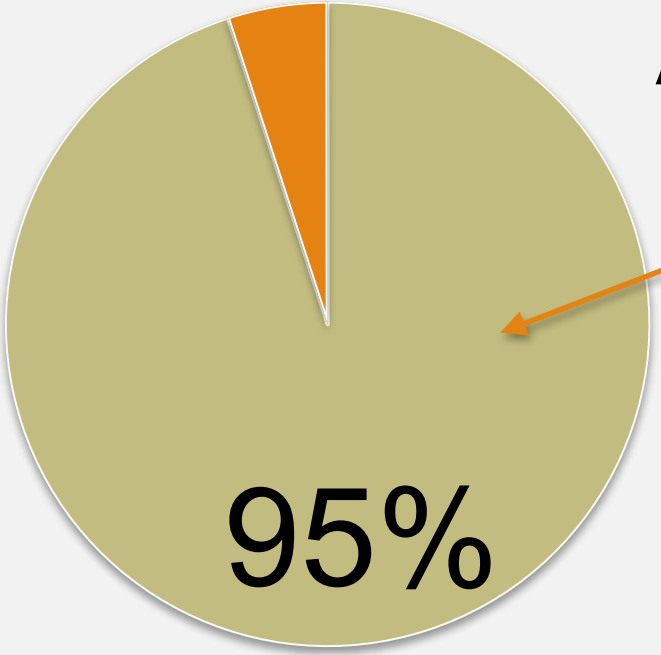
hubzero

nanoHUB



hubzero

nanoHUB

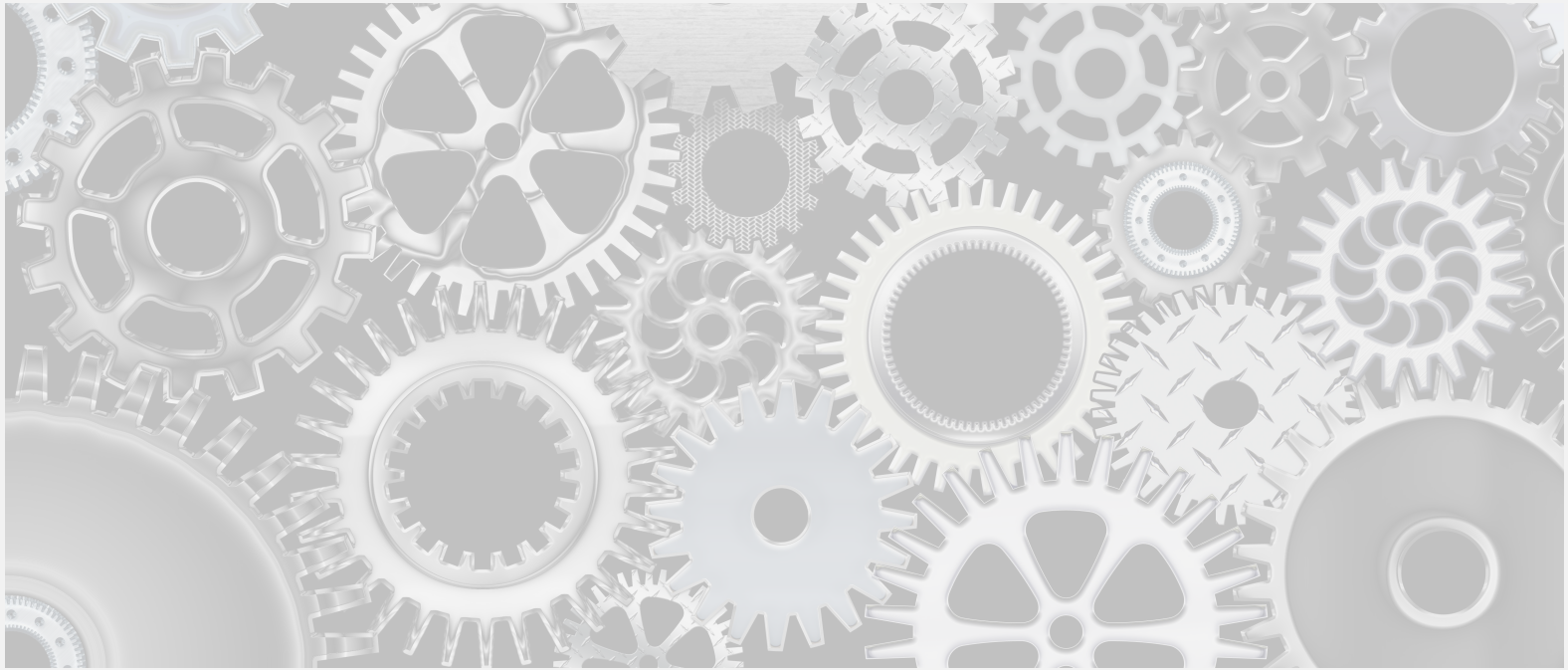


Anonymous Download Users

- Content item or Resource
- Timestamp
- IP address

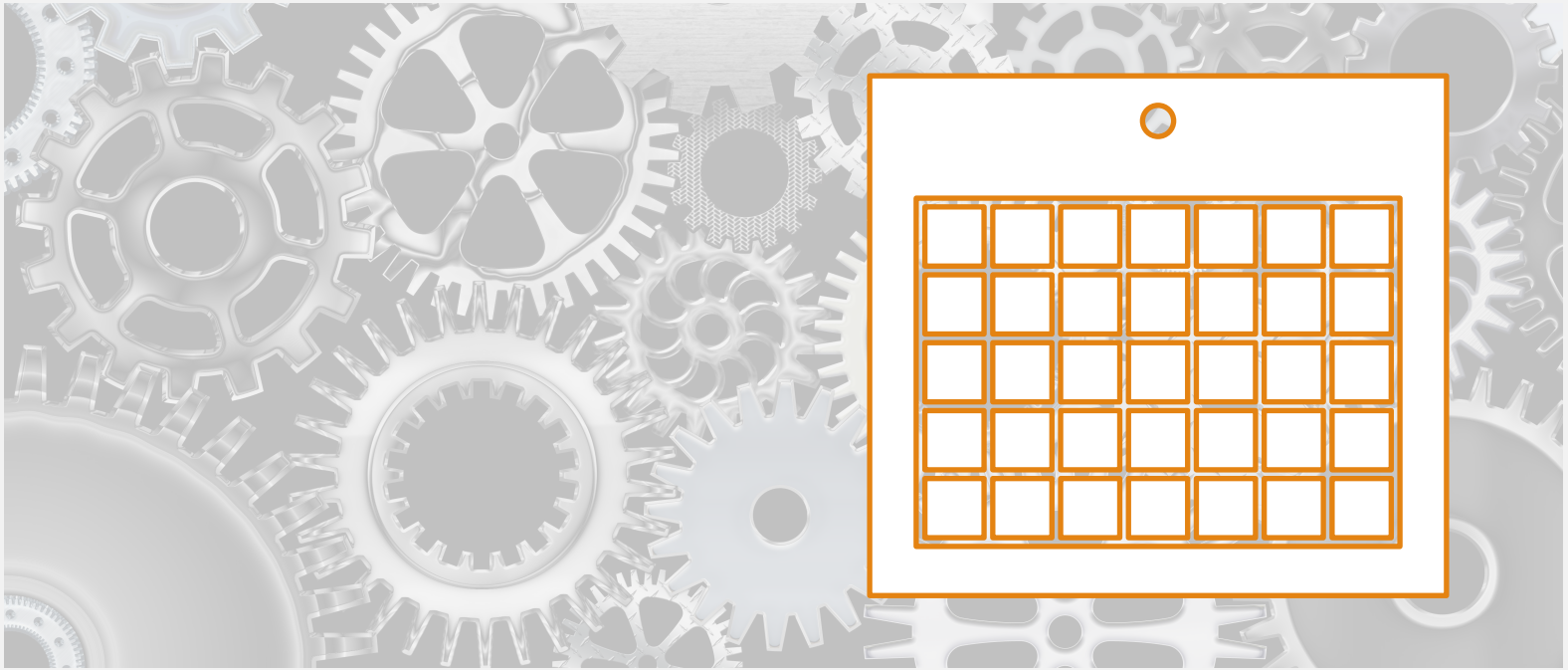
hubzero

Dates



hubzero

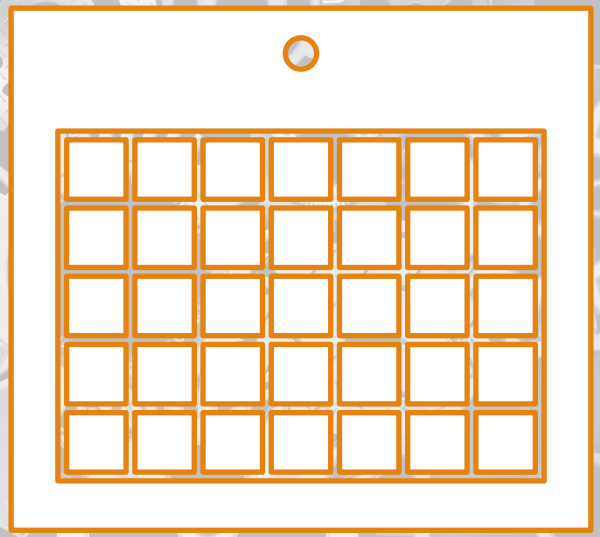
Dates



hubzero

Dates

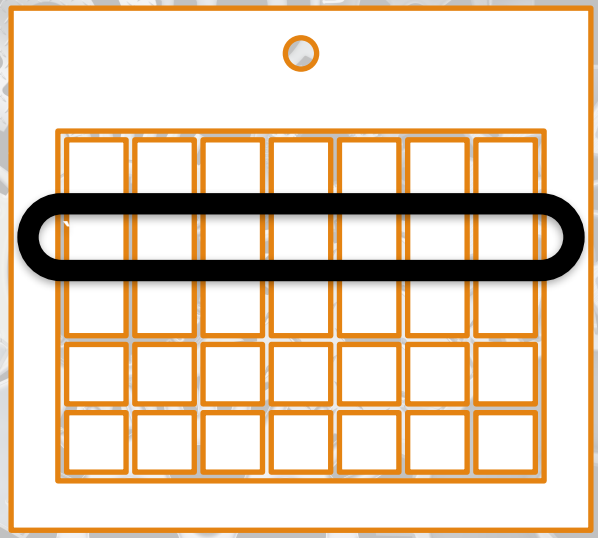
- Natural Groupings
  - Weeks
  - Months
  - Years



hubzero

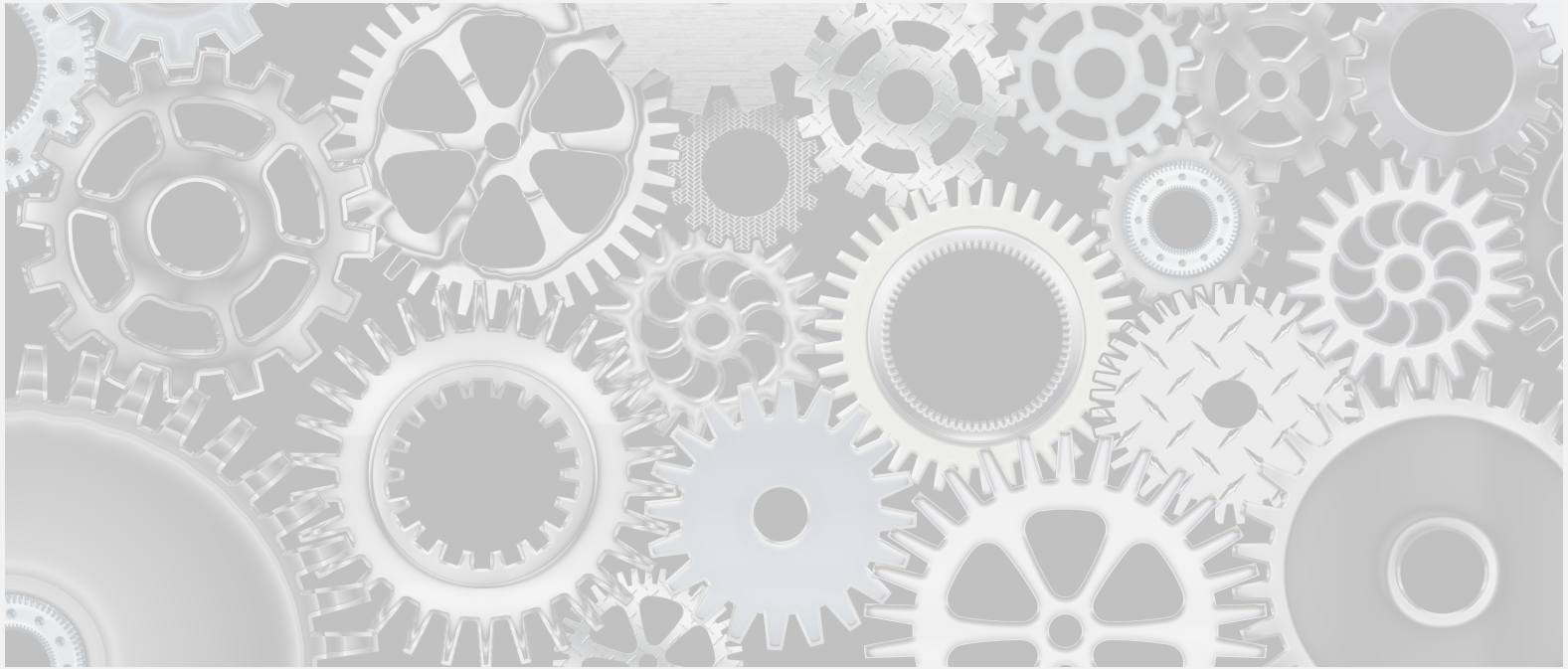
Dates

- Natural Groupings
  - Weeks
  - Months
  - Years



hubzero

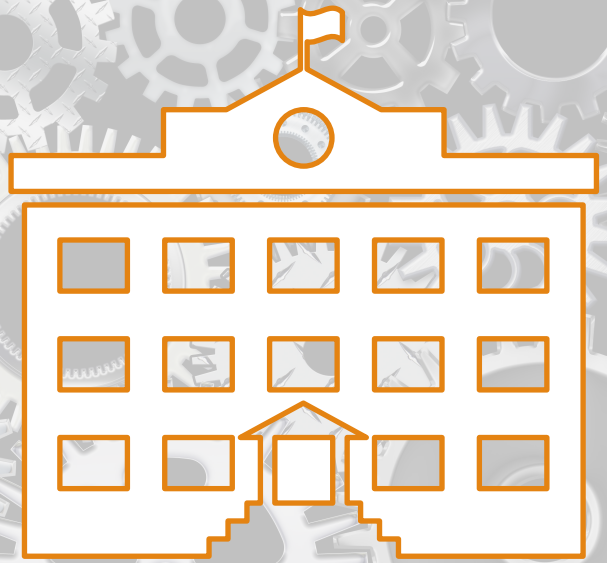
Location



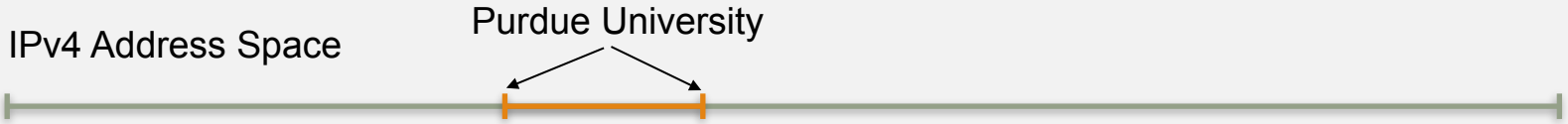
hubzero

Location

- Classrooms bring students together
- Nearness vs. Location



IP Addresses and “nearness”

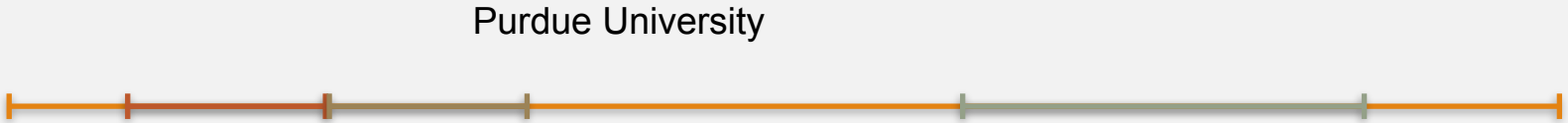


IP Addresses and “nearness”

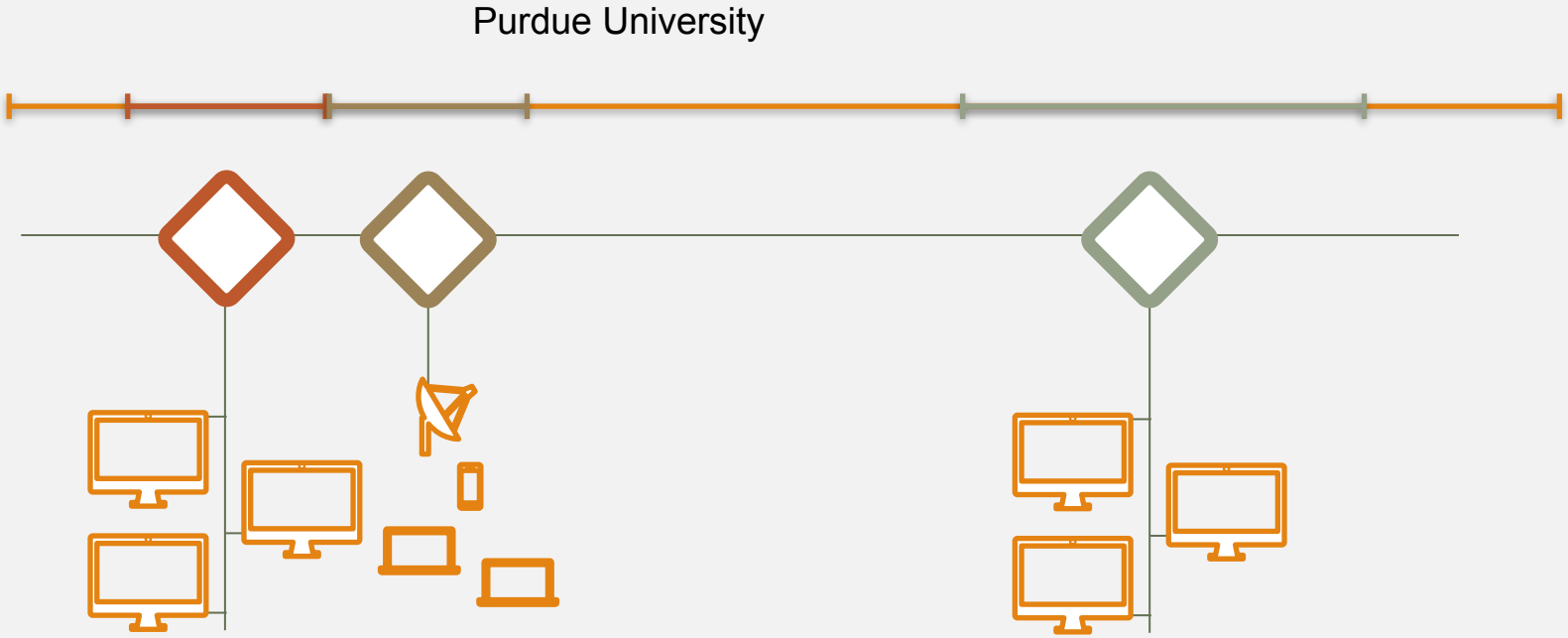
Purdue University

hubzero

IP Addresses and “nearness”

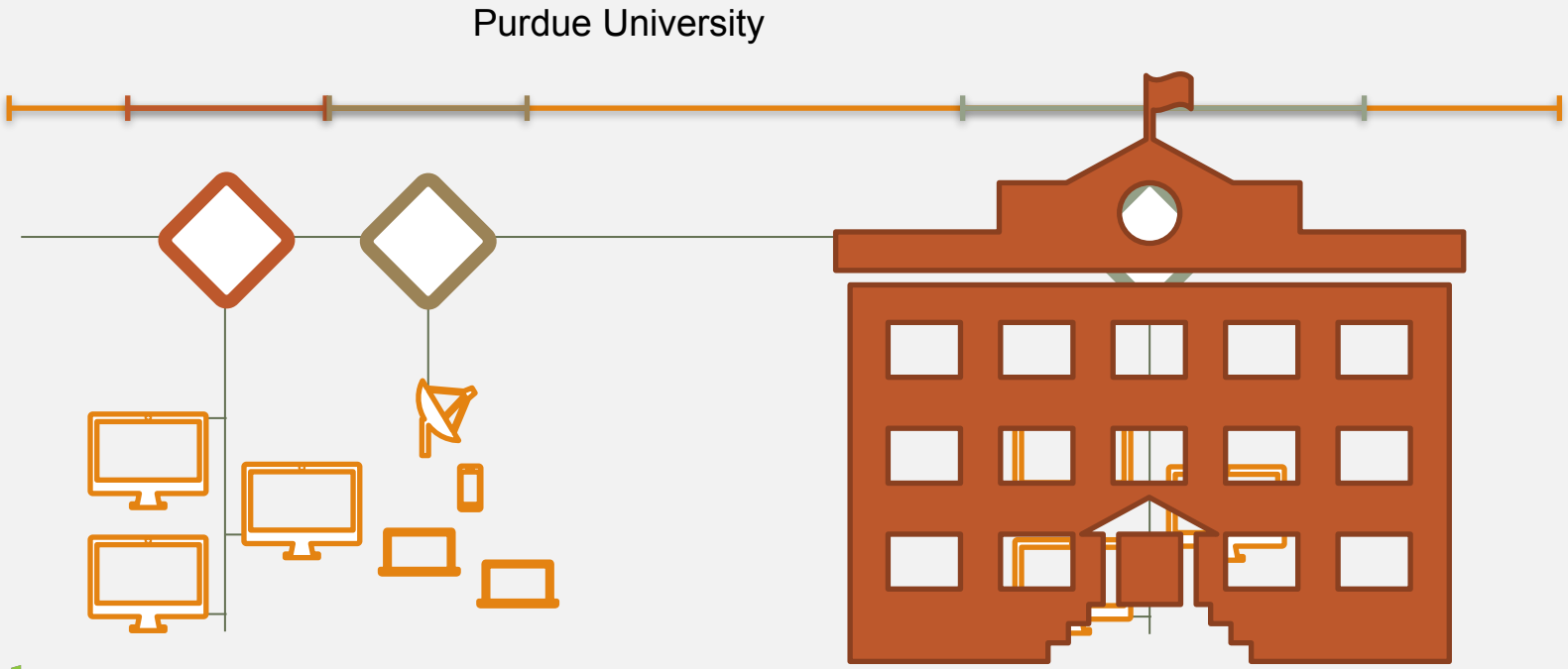


IP Addresses and “nearness”



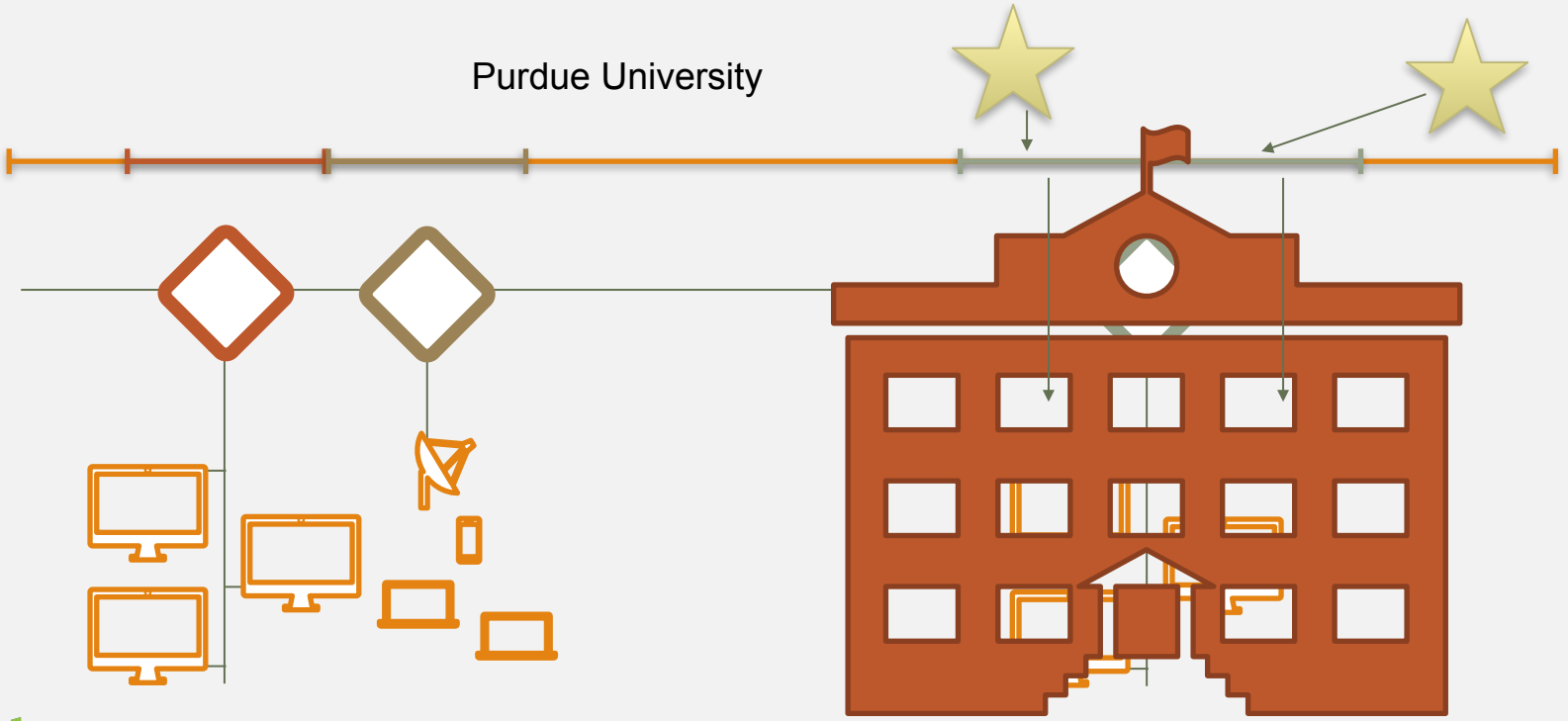
hubzero

IP Addresses and “nearness”



hubzero

IP Addresses and “nearness”



hubzero

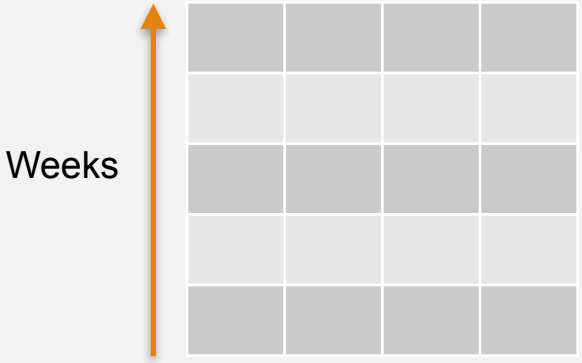
Build a “heatmap” in time & space

hubzero

Build a “heatmap” in time & space

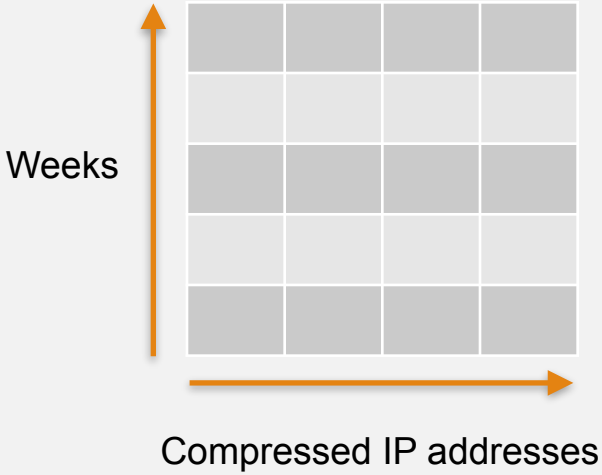
hubzero

Build a “heatmap” in time & space



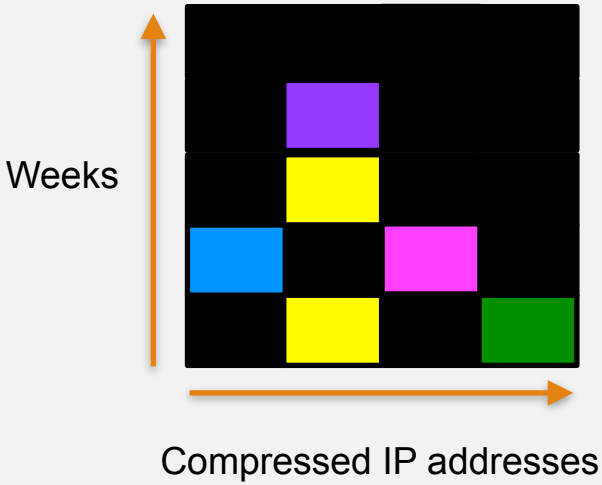
hubzero

Build a “heatmap” in time & space



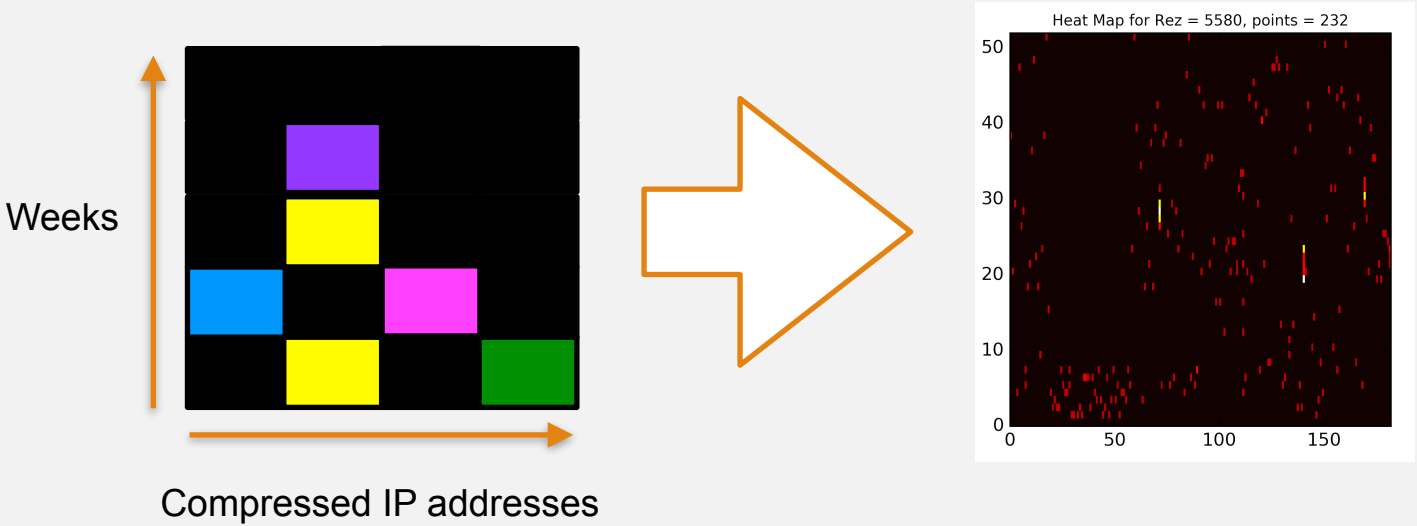
hubzero

Build a “heatmap” in time & space



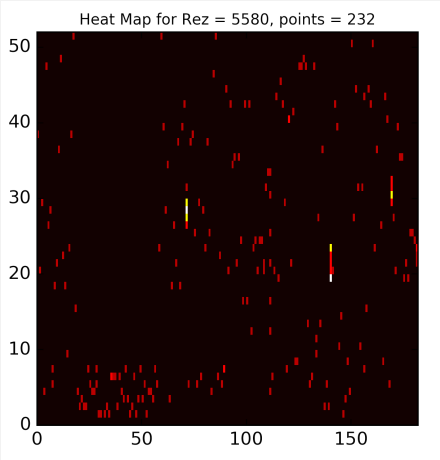
hubzero

Build a “heatmap” in time & space



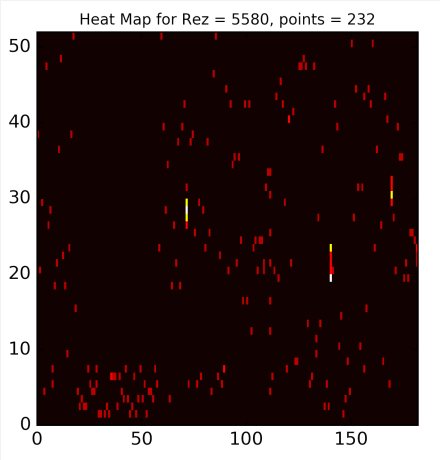
hubzero

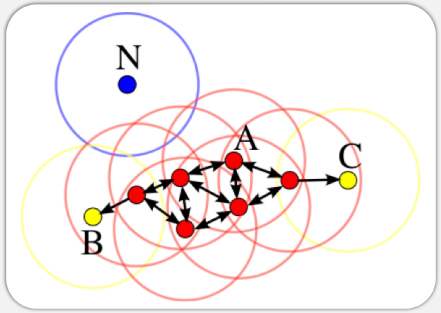
Identify clusters with DBSCAN



hubzero

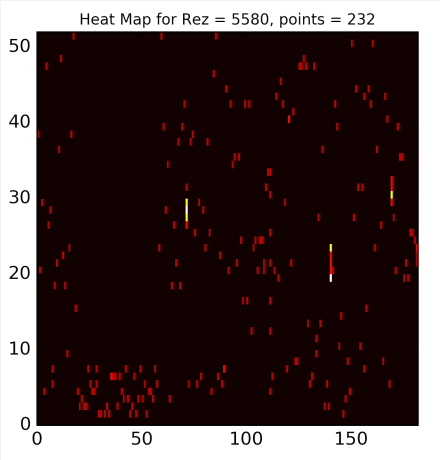
Identify clusters with DBSCAN

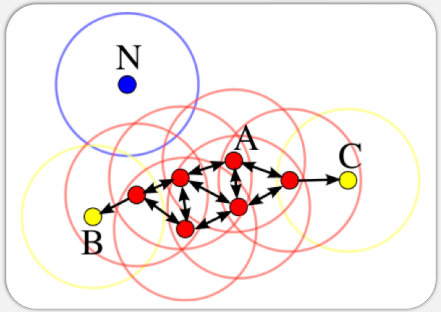


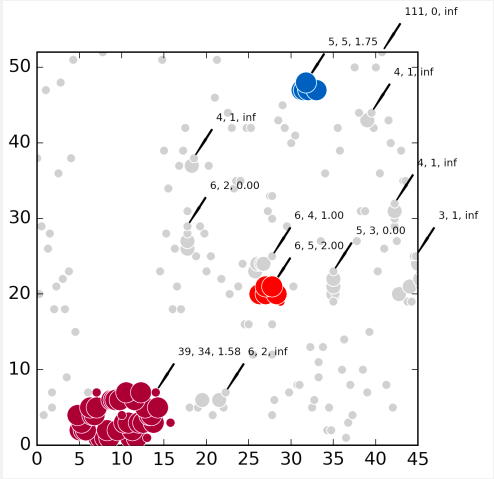


hubzero

Identify clusters with DBSCAN







hubzero

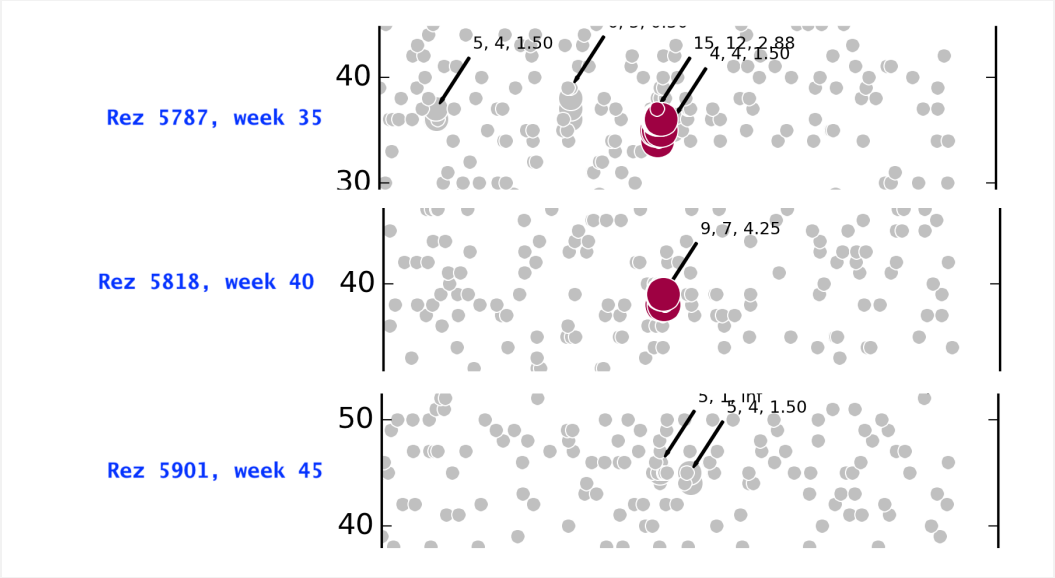
Identify clusters with DBSCAN

Identify clusters with DBSCAN

- 112,290,752 event records
- 2353 classroom clusters detected
- 1460 (23.2%) content items had at least one cluster
- 6296 total content items

Identify clusters with DBSCAN

- 112,290,752 event records
- 2353 classroom clusters detected
- 1460 (23.2%) content items had at least one cluster
- 6296 total content items



hubzero

Nothing’s perfect

hubzero

Nothing’s perfect

- IP Address “nearness” Issues
  - Mobile devices
  - Non-Physical classrooms
  - NAT

Q & A

hubzero