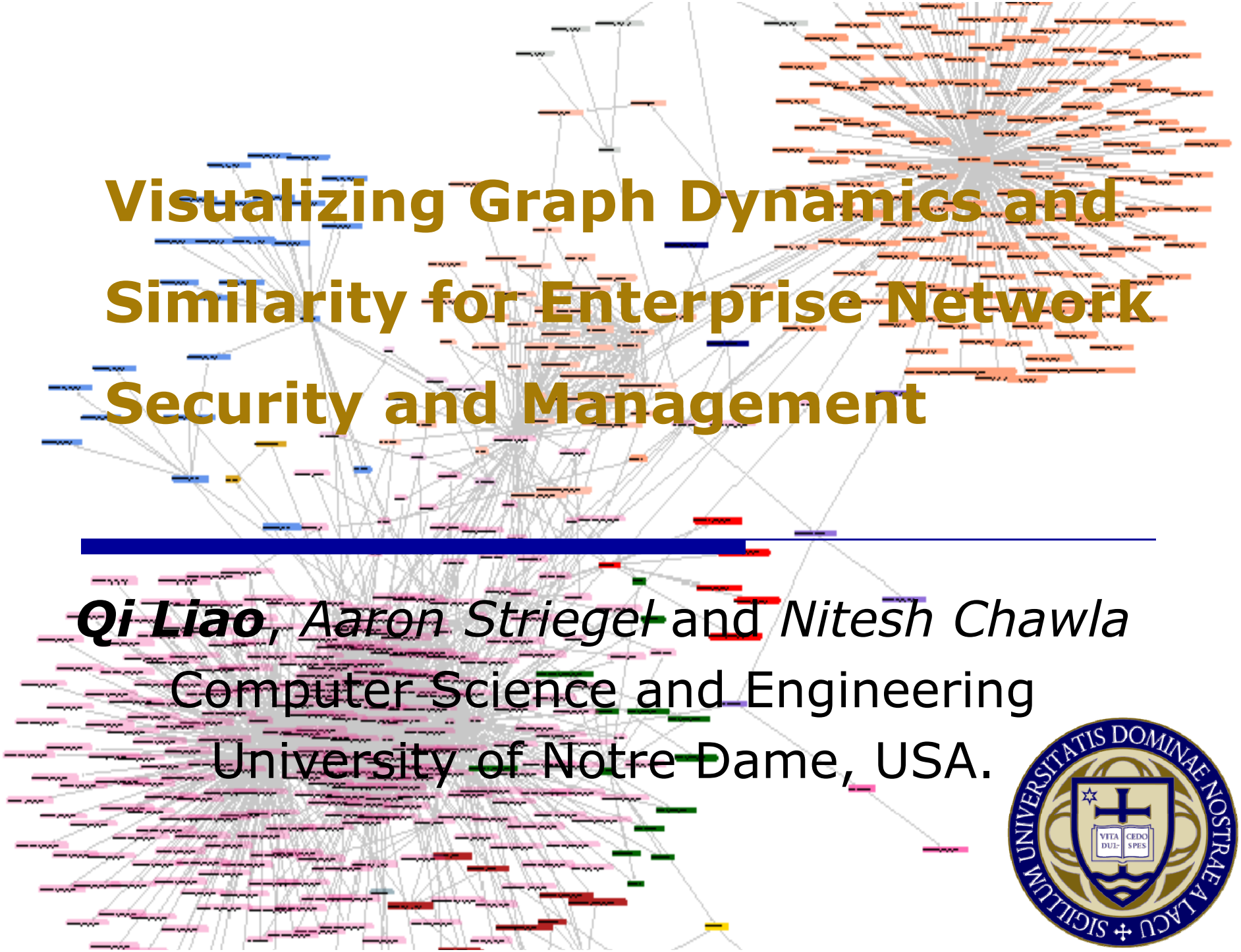
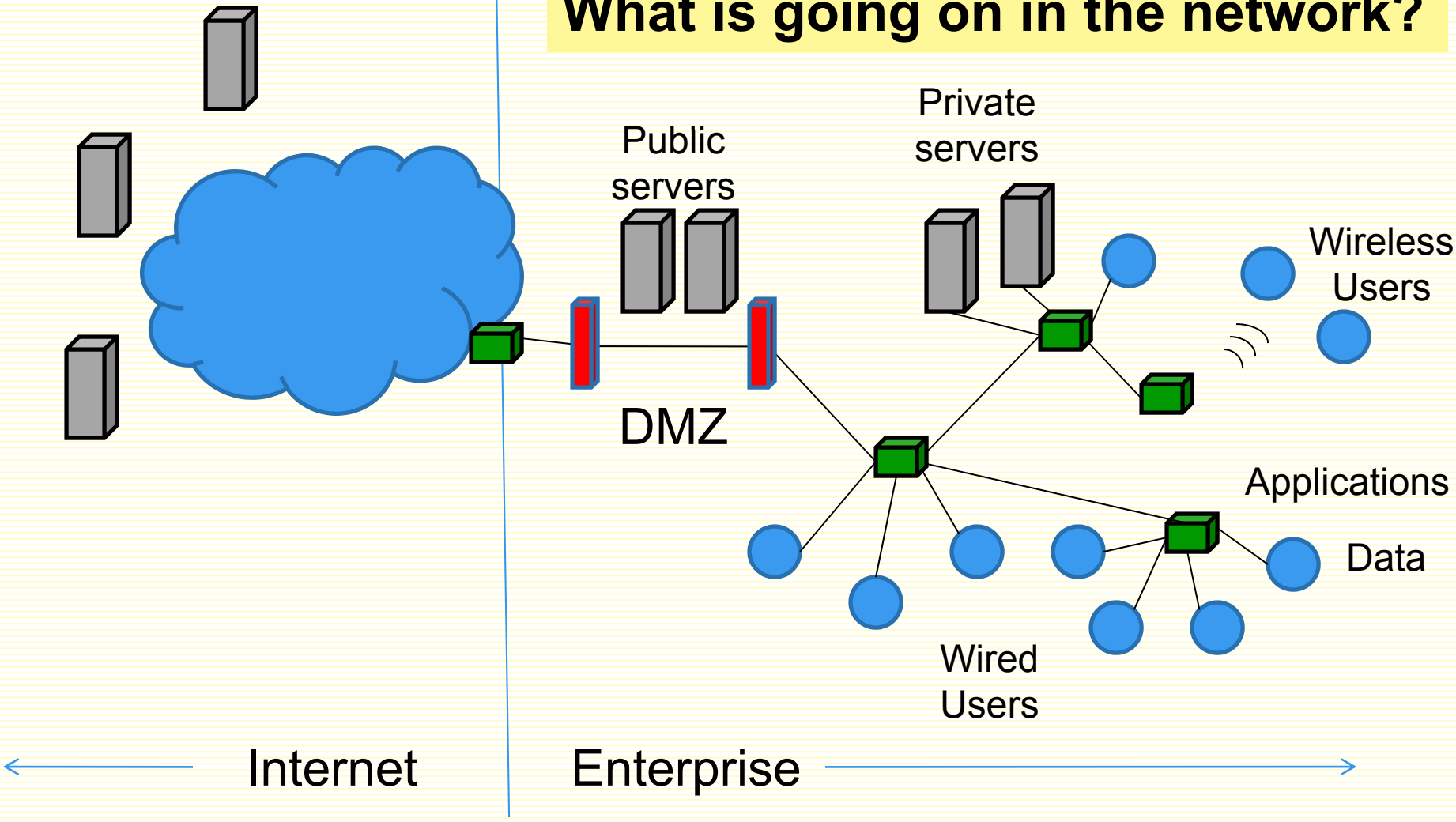


Visualizing Graph Dynamics and Similarity for Enterprise Network Security and Management

Qi Liao, *Aaron Striegel* and *Nitesh Chawla*
Computer Science and Engineering
University of Notre Dame, USA.

Enterprise Network Management

What is going on in the network?



Traditional Logging

- Network connectivity *logging* usually in form of *IP addresses* and *port numbers*.
- Traditional Cisco ***NetFlow*** definition:

Ingress Interface	IP Protocol	IP TOS	Src IP	Dst IP	Src port	Dst port

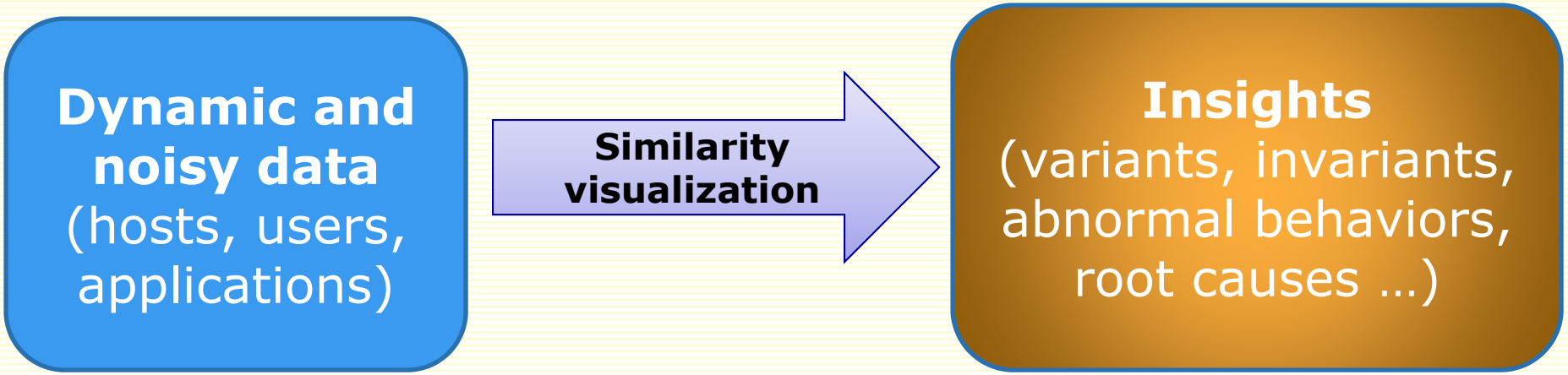
- ***Where***, but not ***Who*** and ***What***
- Visual analysis of hosts, users and applications is more important and harder than traditional IP/ports visualization [Hertzog06, Lalanne07, Liao08]

Introduction

- Security management of enterprise networks is hard
  - Users and applications
  - Complex interrelationships
  - So dynamic, constantly changing
  - No clean signal. Traditional data mining for anomaly detection falls short
- Understanding the **dynamics** / similarities is non-trivial
  - Important step for anomaly detection

Similarity visualization

- First step to understand abnormality
- Key questions:
  - What are the changes?
  - What changes are (ab)normal?
  - How different (or similar) from day-to-day activities?
  - How to effectively visualize them?



Hierarchical Similarity Visualization

Top-down manner
(overview + context)

Graphs

- HUA connectivity graphs
- Bipartite graphs
- Similarity graphs

Inter-graphs

- Among network graphs across multiple timelines.

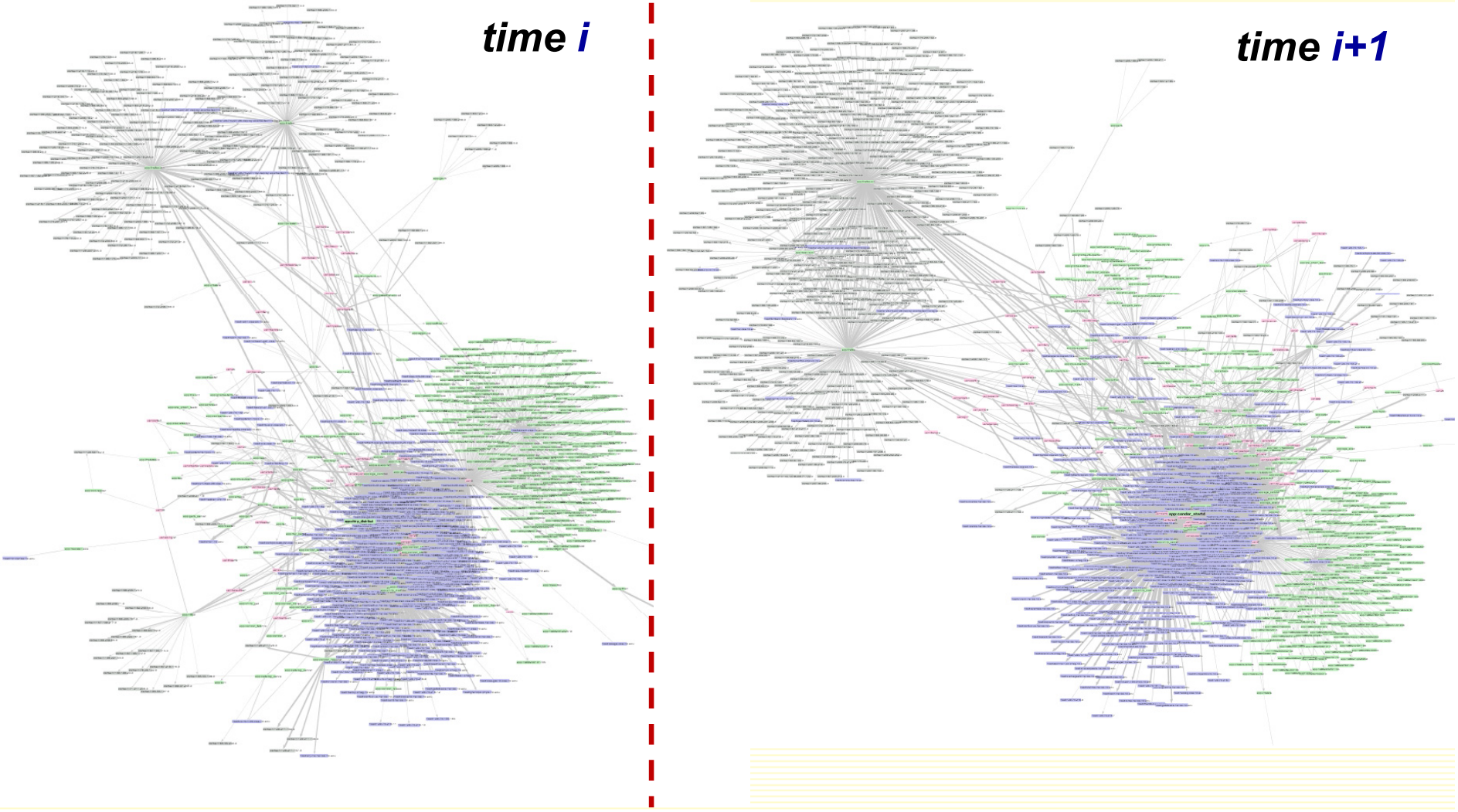
Intra-graph

- Identify similar structure within each individual

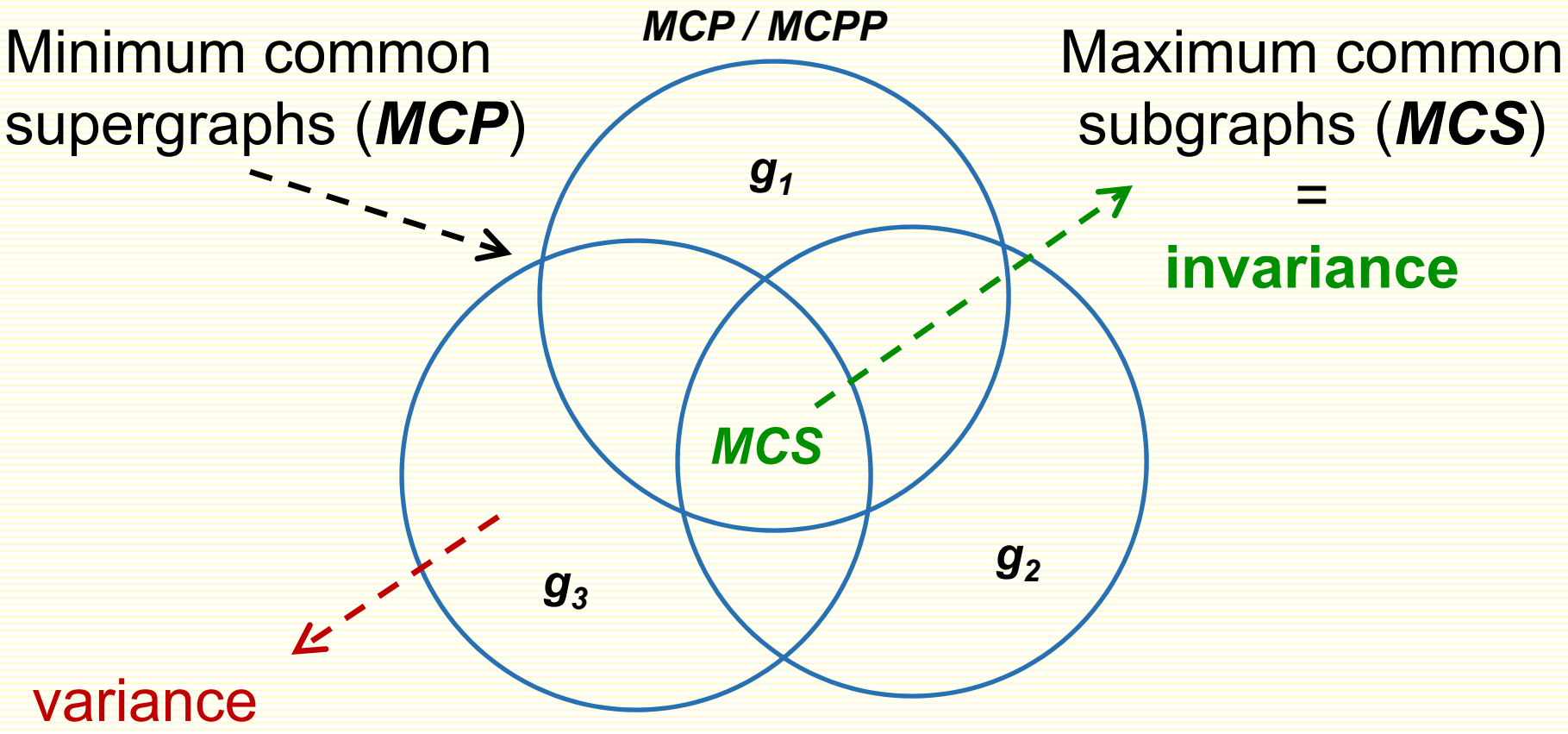
nodes

- Dynamic of neighborhoods changes at each individual node level.

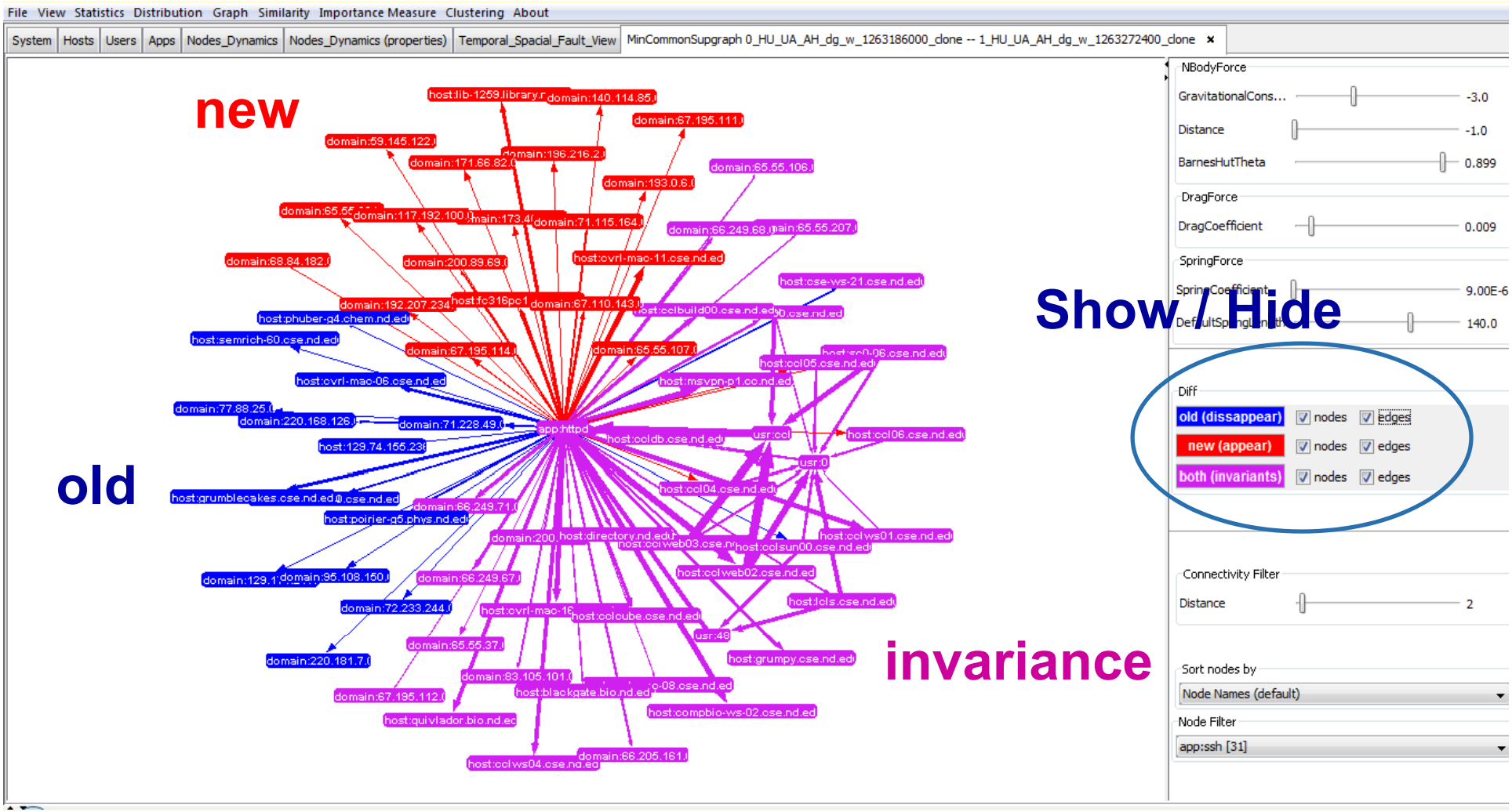
Similarity/Difference Visualization



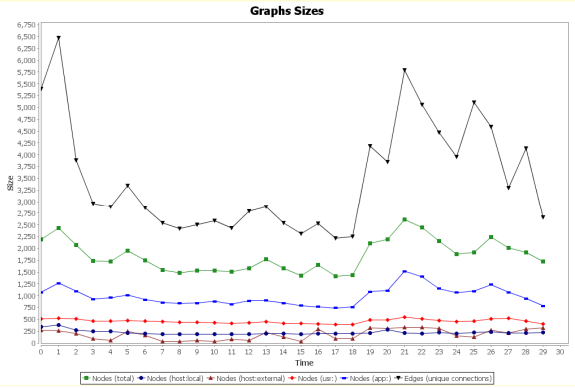
Variance vs. invariance



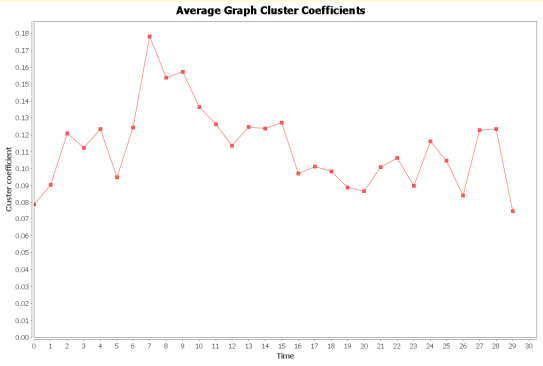
Differential view



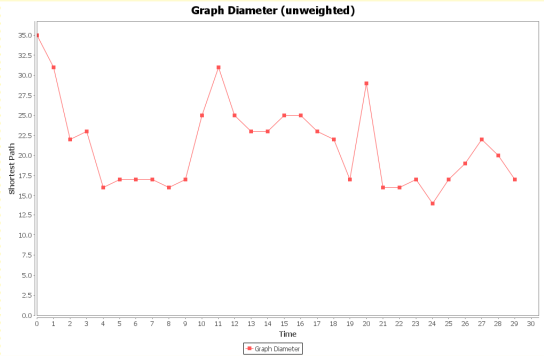
Visualizing graph property changes



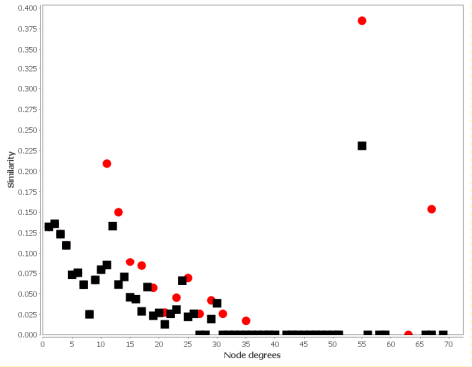
Graphs sizes



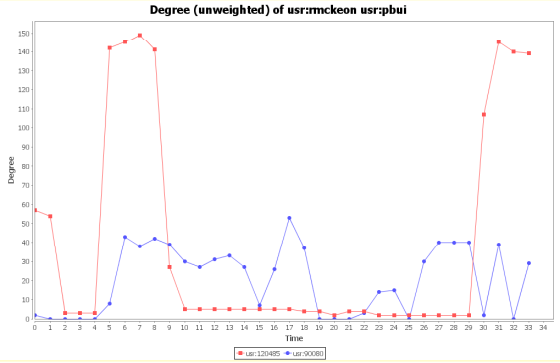
Cluster coefficients



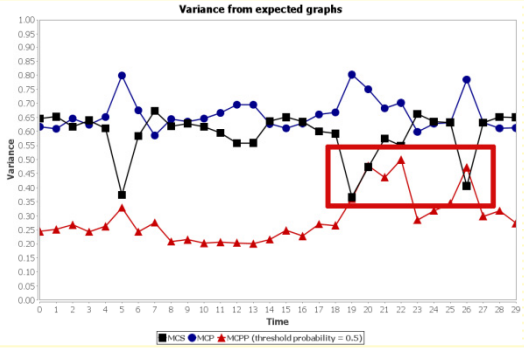
Graphs diameters



Degree distributions



Graphs distances



Graphs variance scores

Graph distance

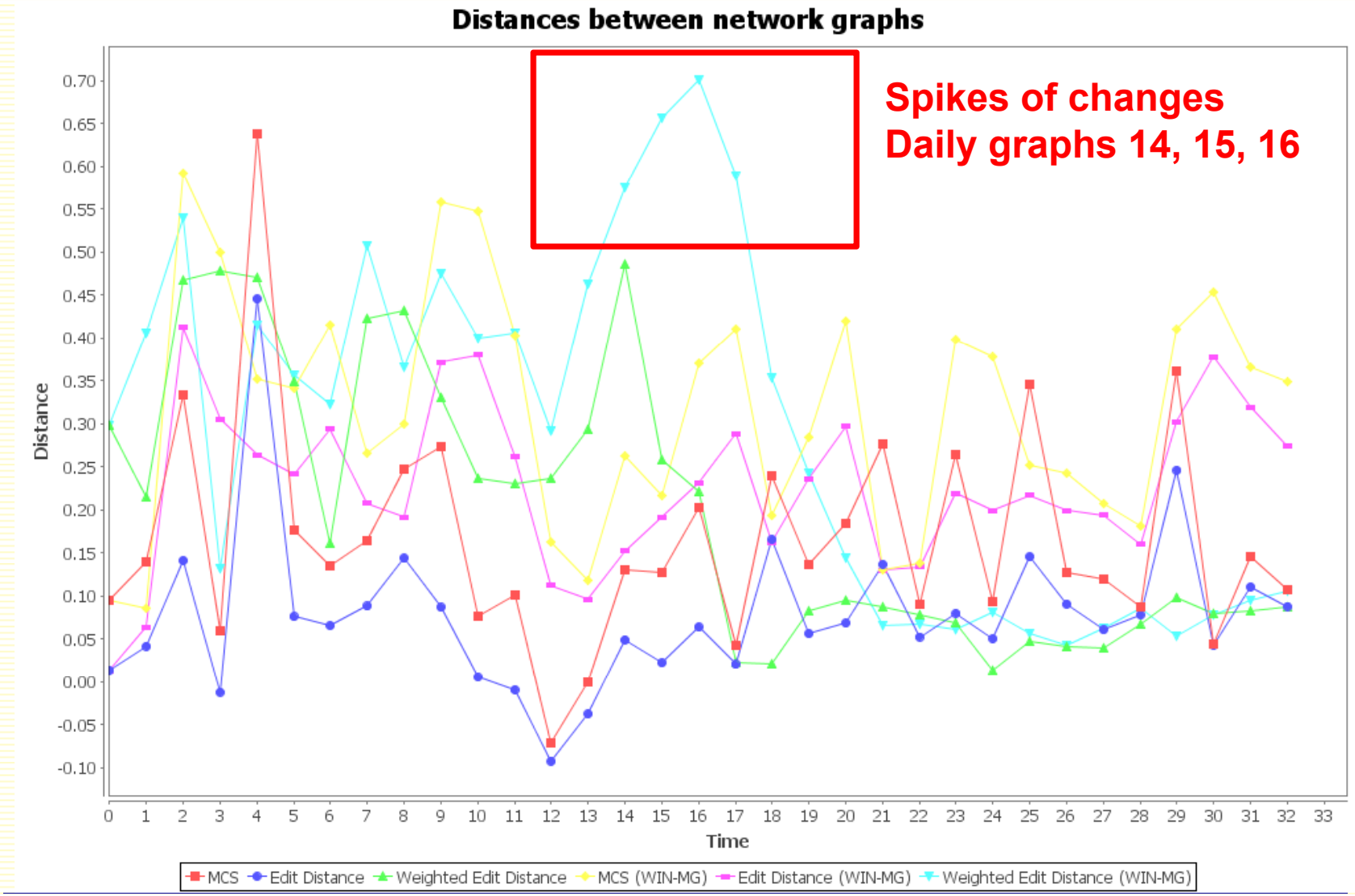
- Quantification through normalized distance functions:
- Schenker:2003, Bunke 1998, 2007.
- MCS based:

$$d(g_1, g_2) = 1 - \frac{|mcs(g_1, g_2)|}{\max(|g_1|, |g_2|)}$$

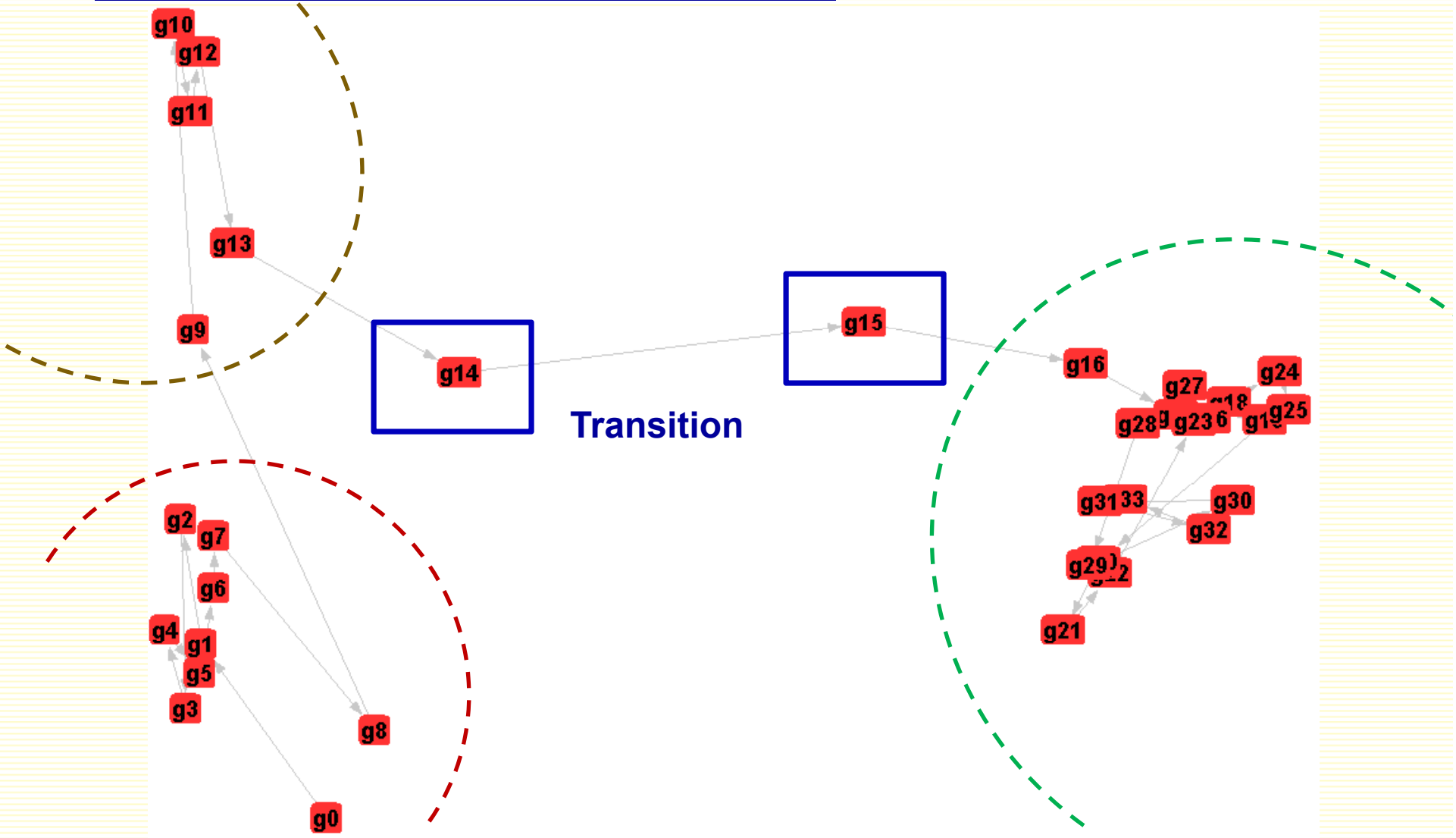
- Graph edit distance (GED) based:

$$d(g_1, g_2) = \frac{|g_1| + |g_2| - 2|mcs(g_1, g_2)|}{|g_1| + |g_2|}$$

Graph distance



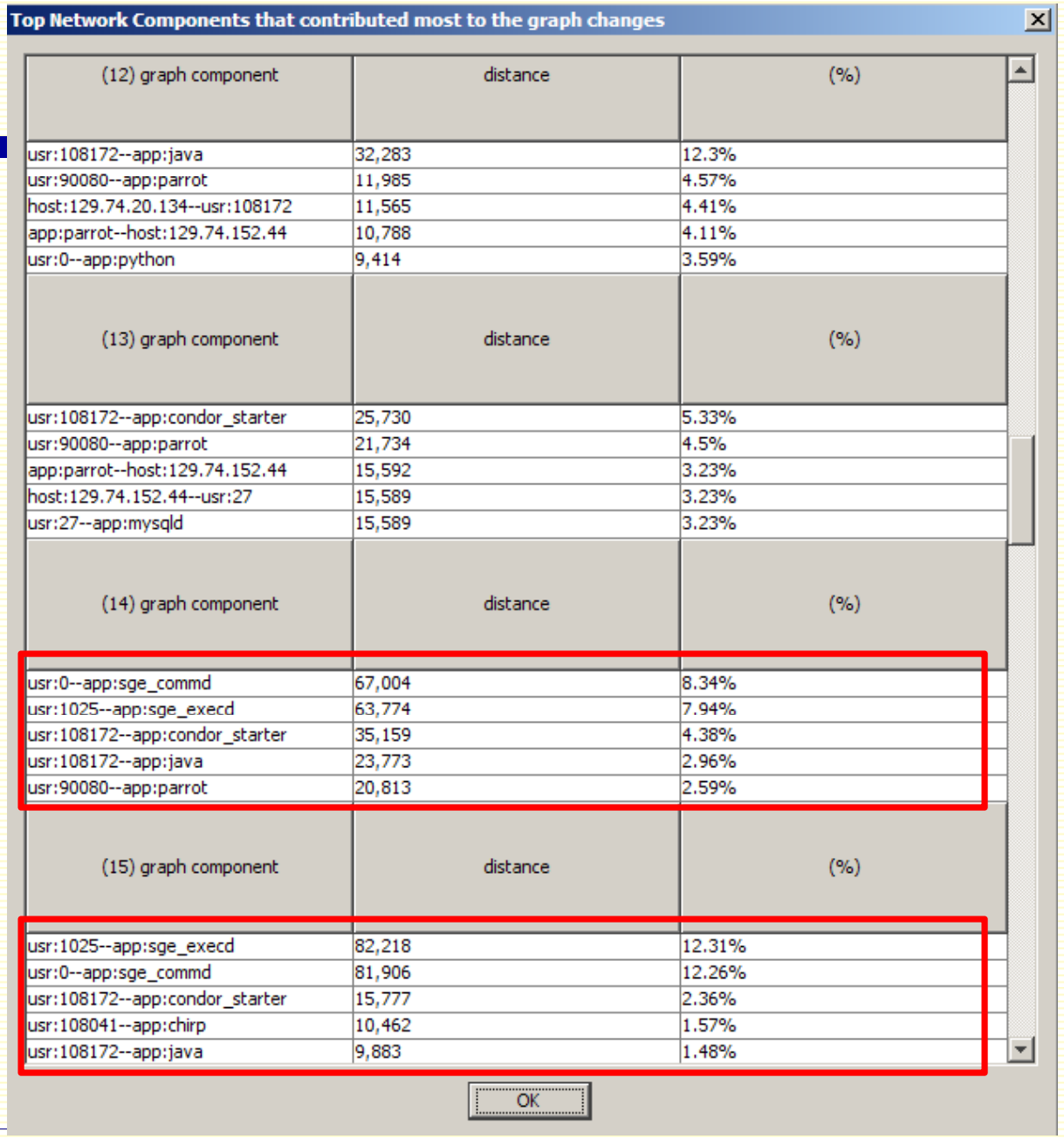
MDS view (cluster evolution)



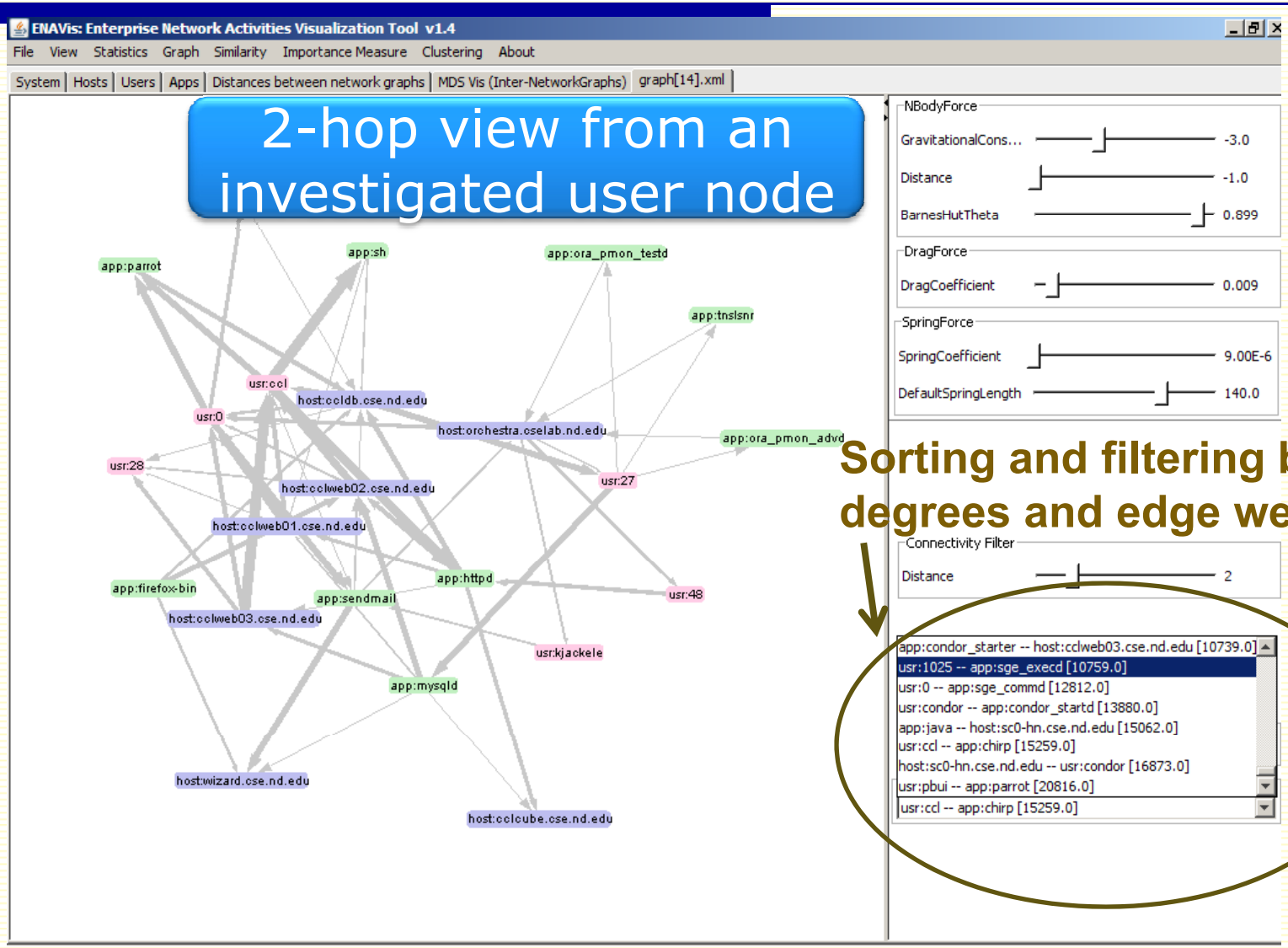
Top network components responsible

Day 14

Day 15



Dynamic interactive and exploration



Queries for degree trend

one-hop view from an investigated user node

app:sge_qmaster

host:iss-node023.cse.nd.edu

usr:1025

host

User

app

domain

Scores

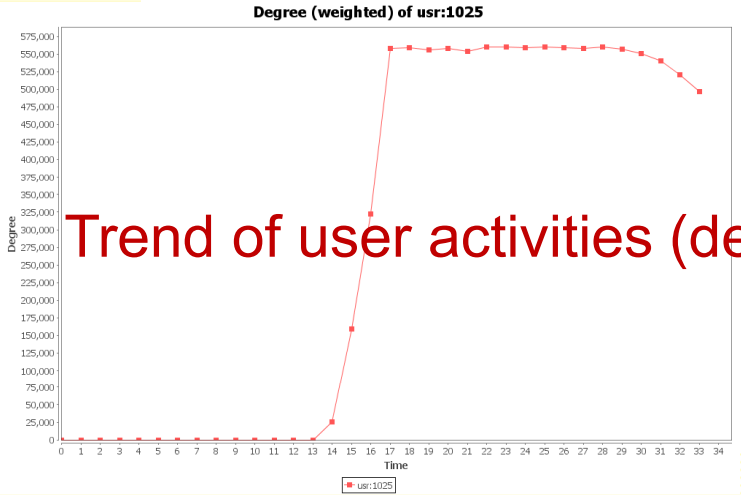
trend

degree

host:head.cse.nd.edu

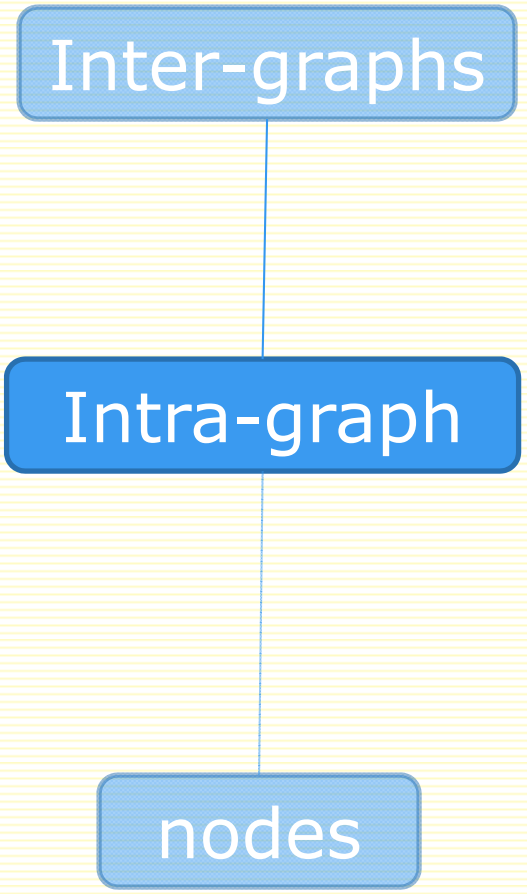
app:sge_execd

host:iss-node024.cse.nd.edu



Trend of user activities (degrees)

Hierarchical Similarity Visualization



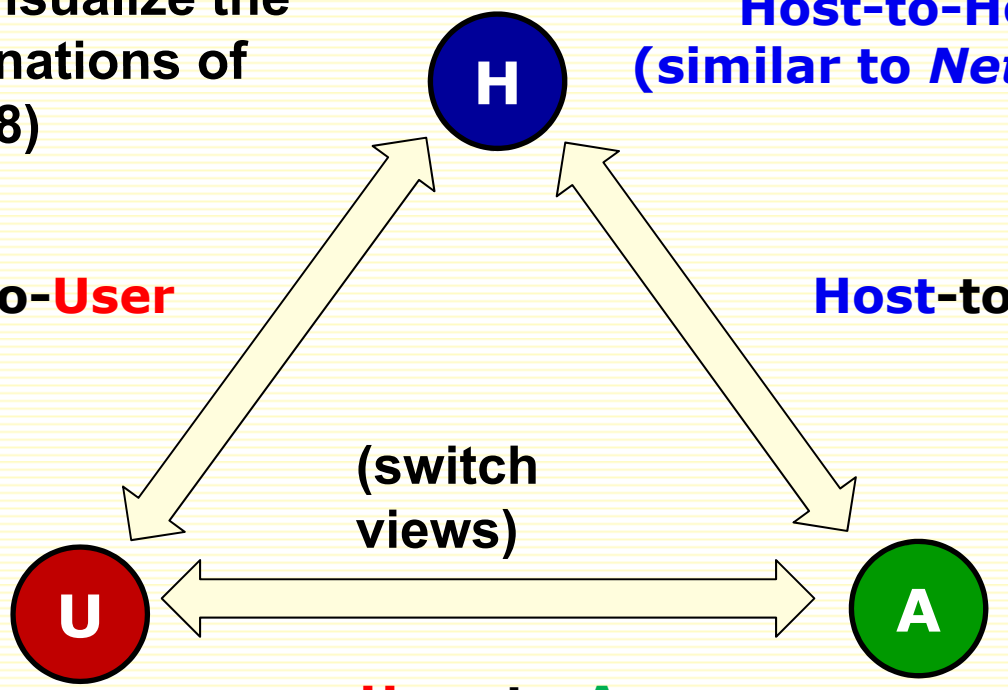
Intra-graph clustering visualization

- HUA connectivity graphs
- (Multi-)bipartite graphs
- Similarity graphs

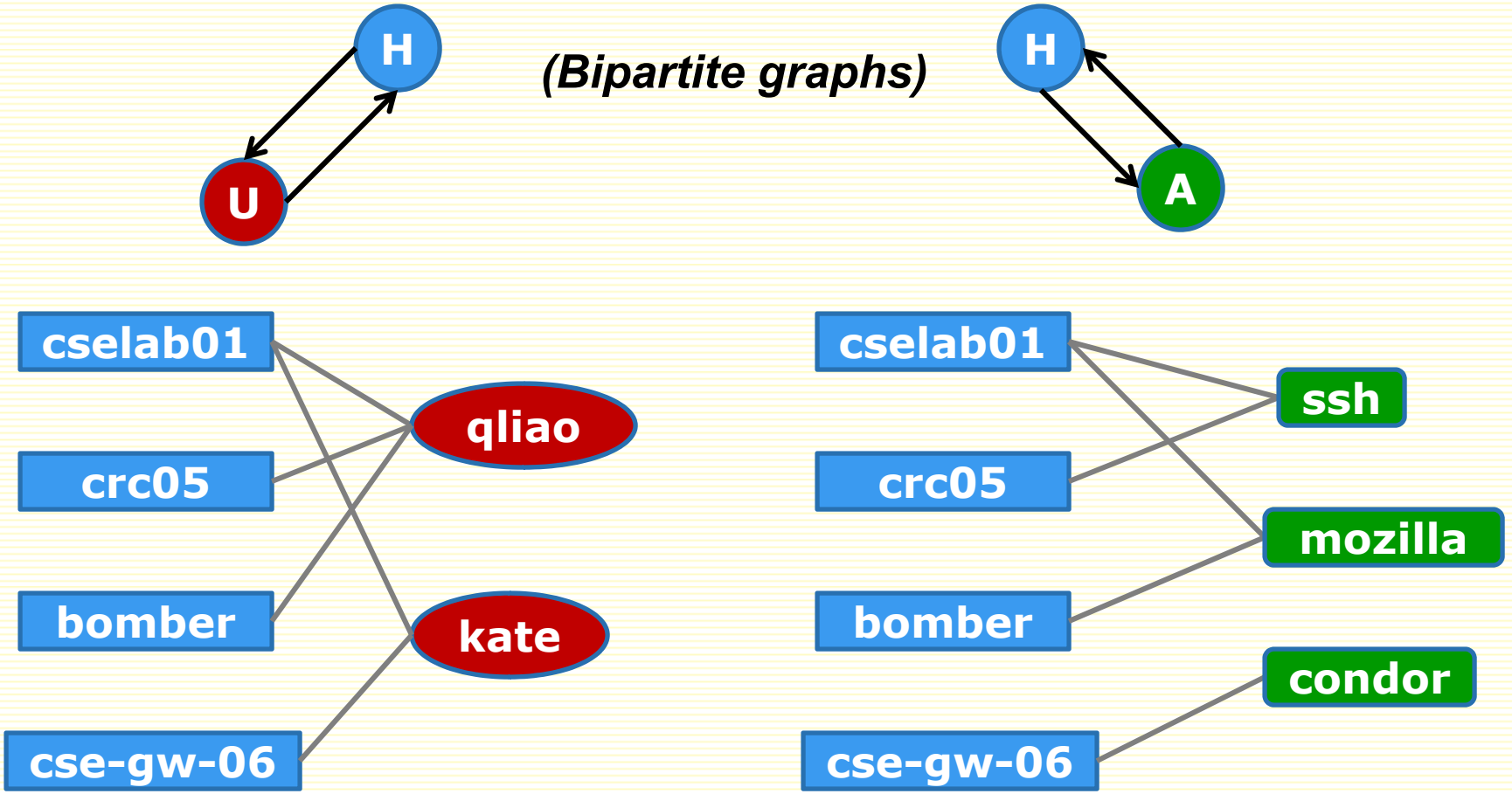
HUA Graph Model

- ***Heterogeneous*** graph
- 4D space
  - Hosts, Users, Applications (***HUA***), Time

The ability to visualize the various combinations of HUA (Liao:2008)

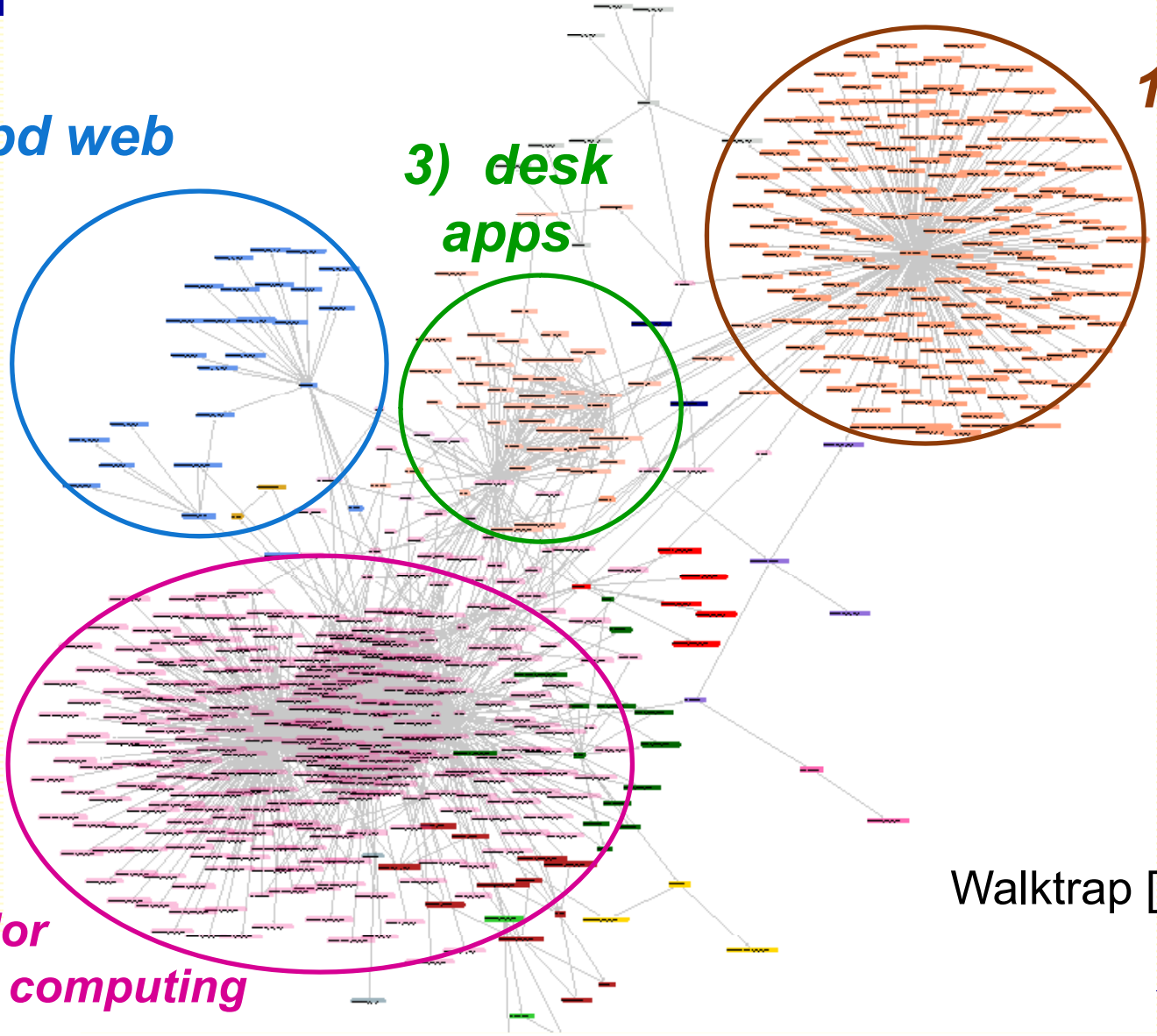


Examples (HU, HA)



Contribution: the ability to visualize the various combinations of HUA

Intra-graph clusters visualization



Walktrap [Pons:2006]

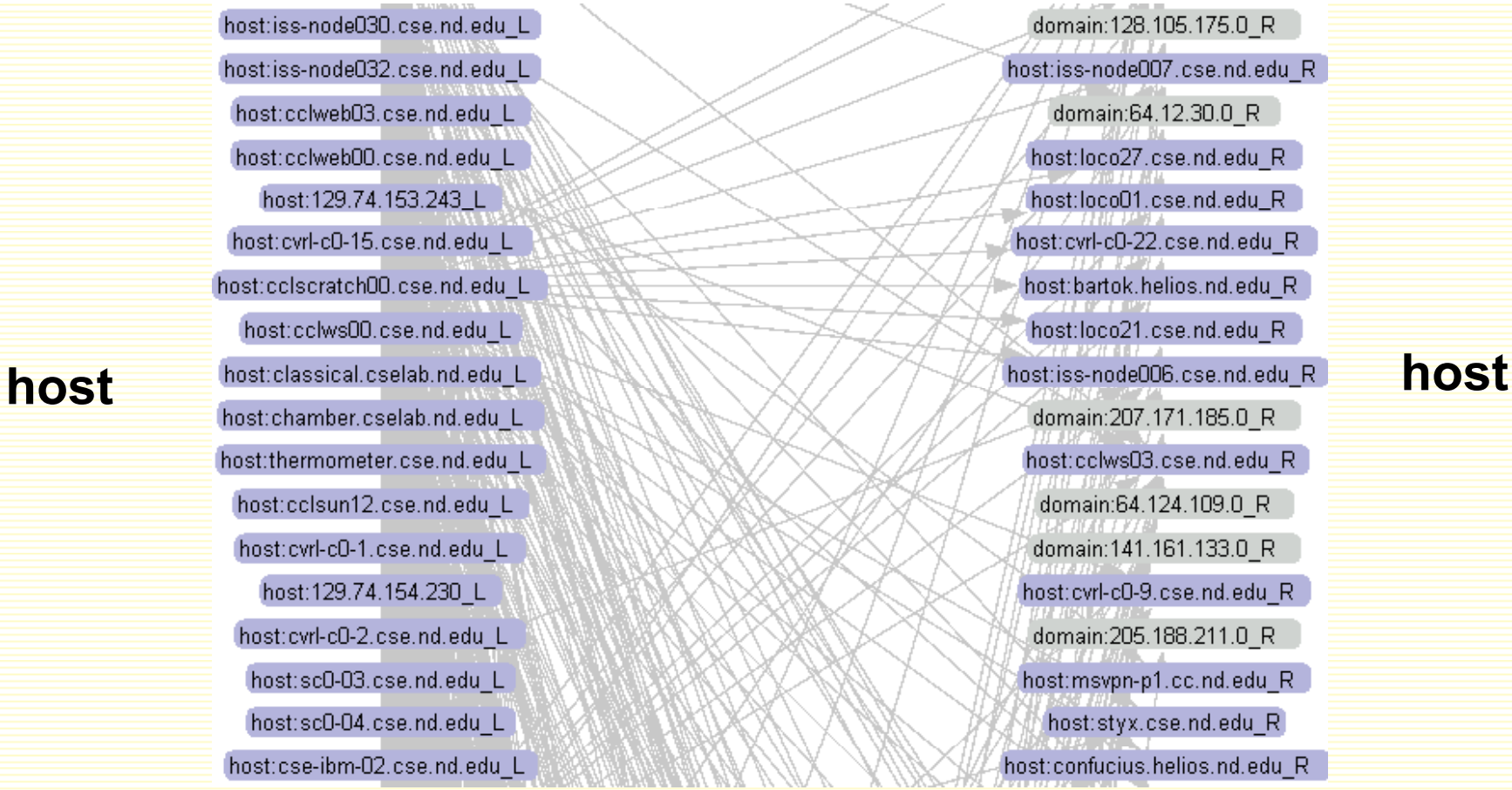
Cluster similarity visualization

- Visualizing the intra-graph clusters can provide:
  - Understanding of network usage ***pattern***
  - Closely related ***community*** formed by similar *hosts*, *users*, *apps*.
  - ***Insight*** and potential ***anomaly*** analysis
- Quantify graphs changes through ***cluster distance***
  - similar to Rand Index [Rand 1971]

$$dist(C_1, C_2) = 1 - \frac{SS + DD}{SS + SD + DD + DS}$$

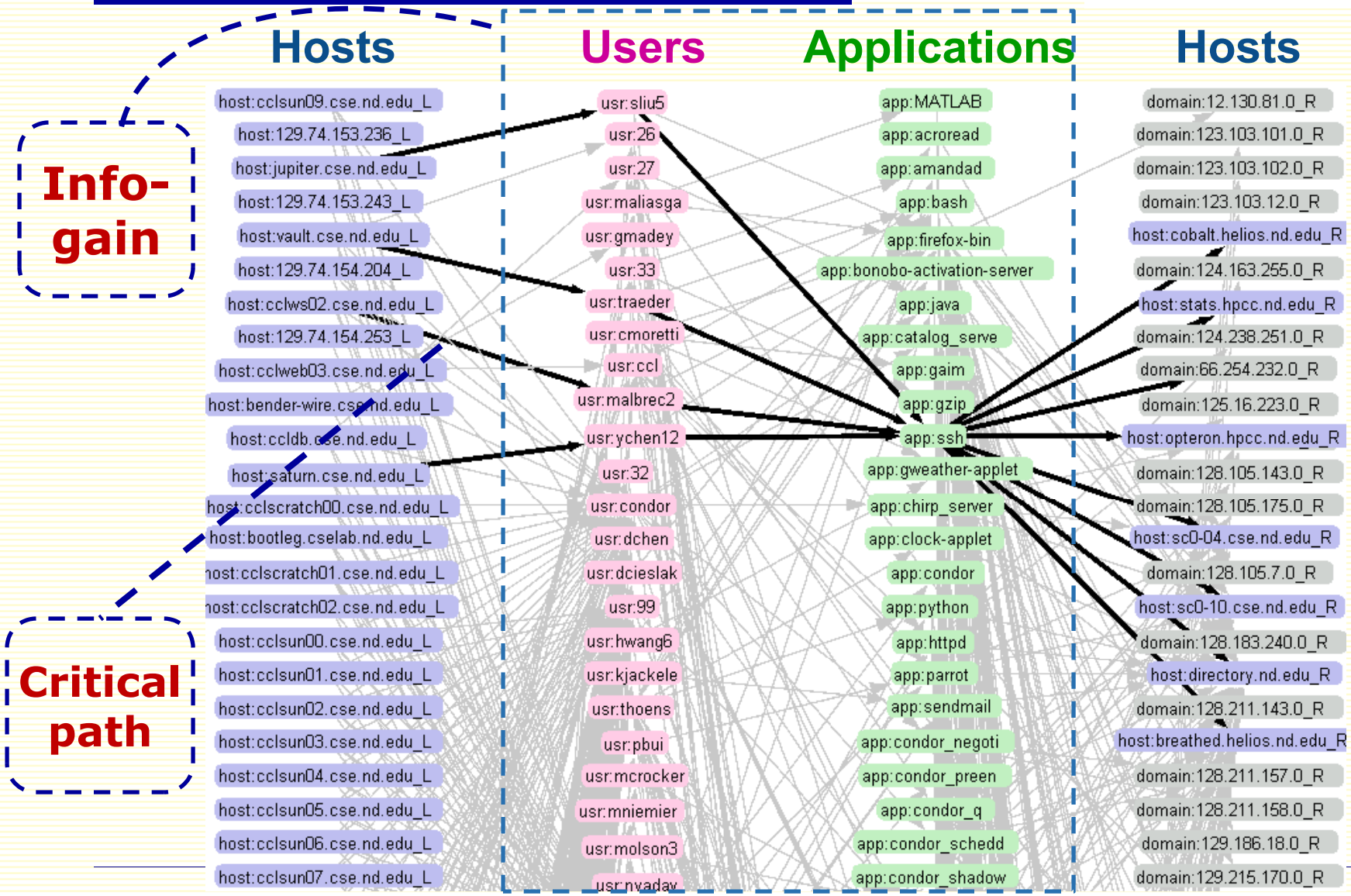
Bipartite graphs

- The general ***HUA connectivity graphs*** can be separated into ***(multi-)bipartite graphs***.



Multi-bipartite graphs

Quadripartite graph



Biclique communities

- **Membership changes**
- *JDoe*

ONE

↓

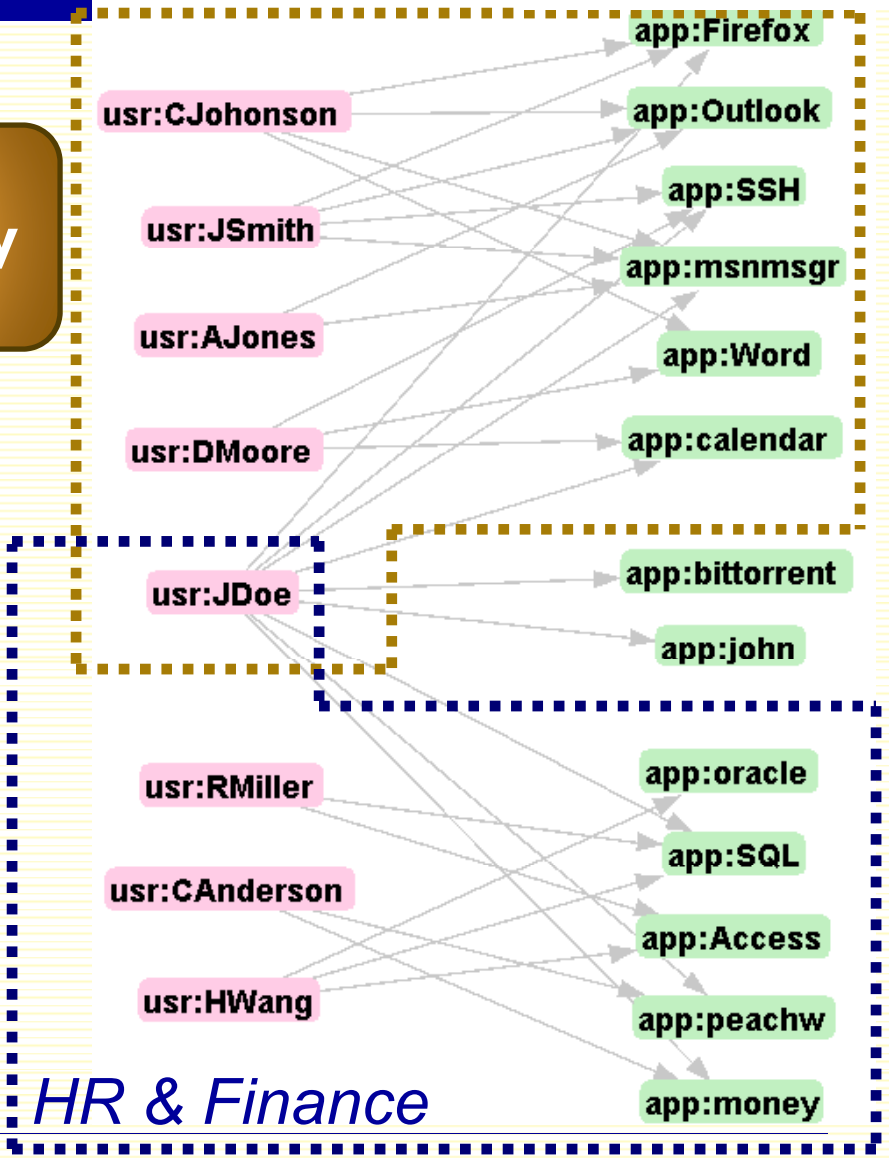
ONE & TWO

overlapping communities

Biclique community ONE

Biclique community TWO

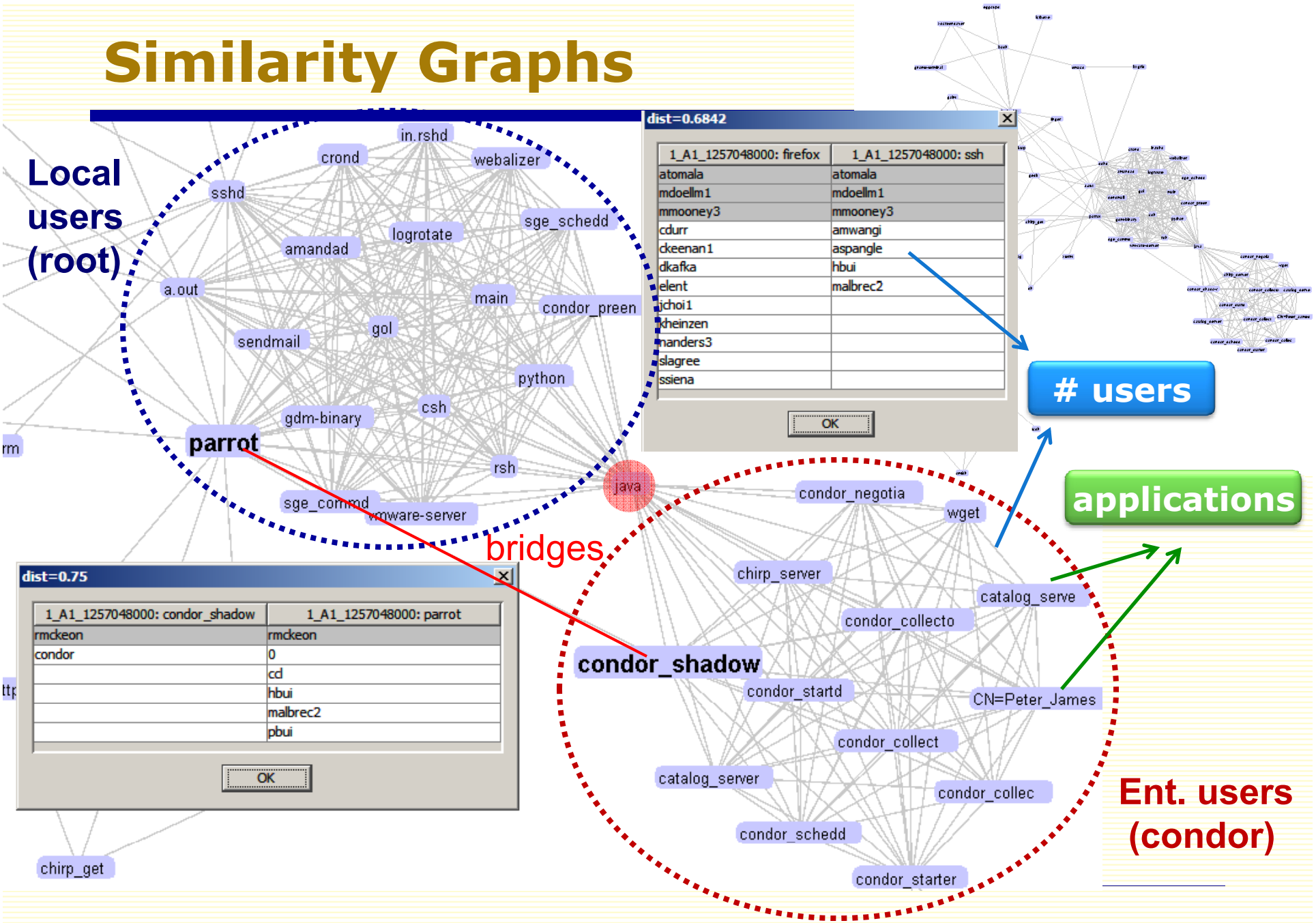
- Biclique communities [Lehmann:2008]



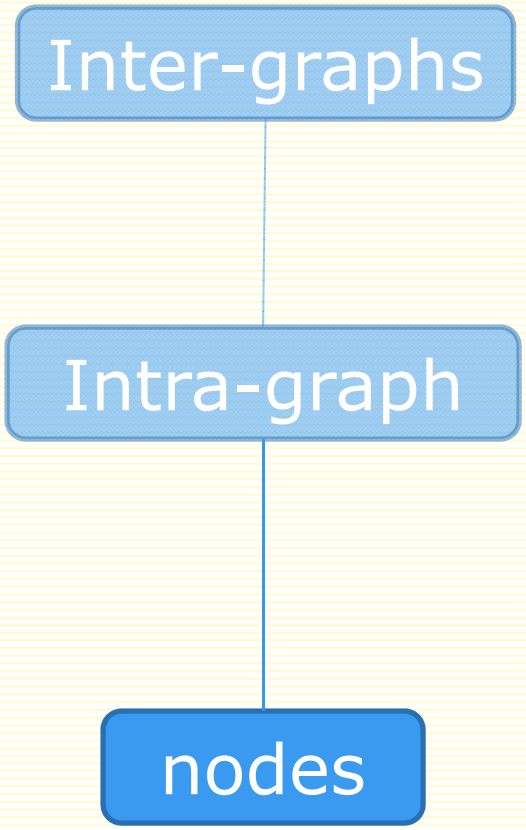
Similarity graphs

- Previous
  - HUA heterogeneous graphs
  - (multi-)bipartite graphs
- ***Similarity graphs***
  - Heterogeneity → Homogeneity
  - Push ***similarity*** into the ***edge weights***
    - (Example)
    - nodes = users
    - edge weights = number of applications they share.

Similarity Graphs

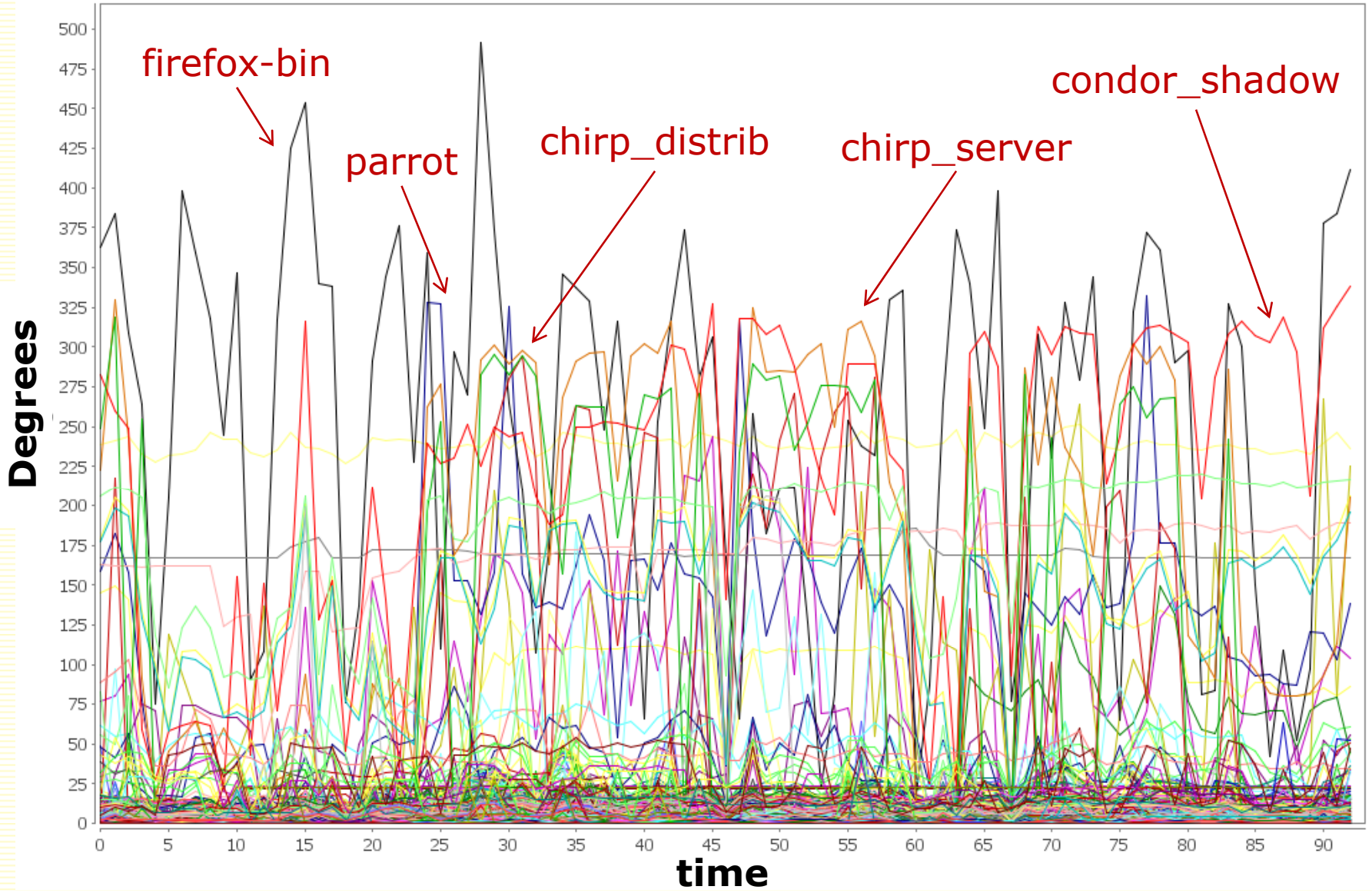


Hierarchical Similarity Visualization



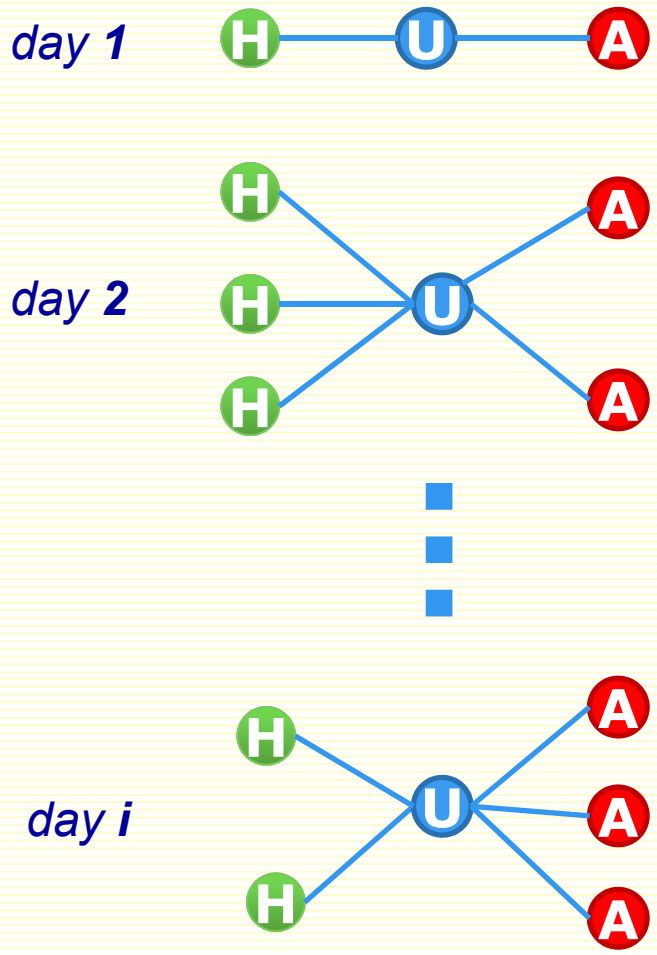
- Dynamics / similarity of each individual nodes
- Visualization of neighborhood changes over time
  - Quick visual analysis on node's anomalous behaviors

Dynamics of Node Degrees

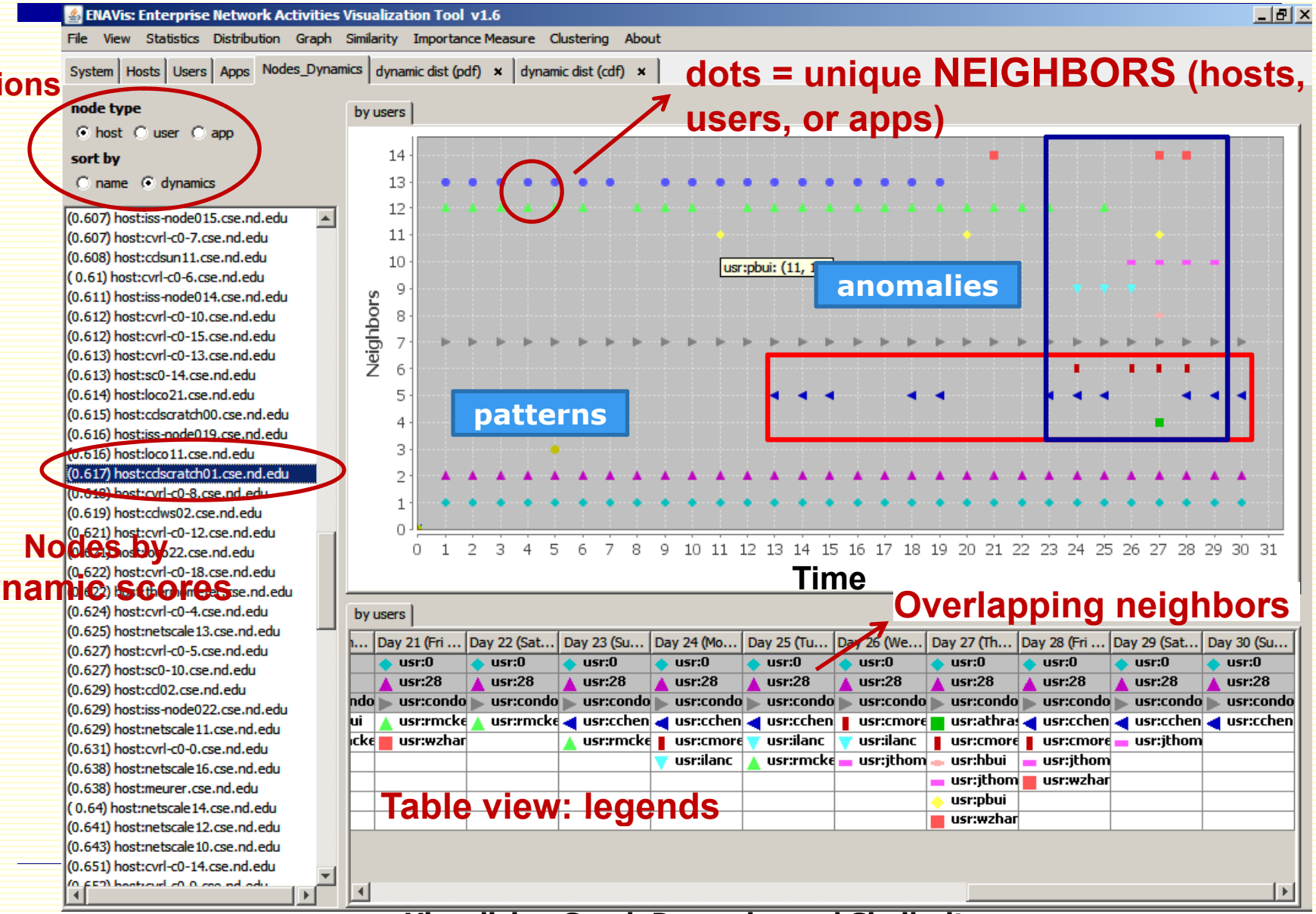


Node Dynamics/Similarity Visualization

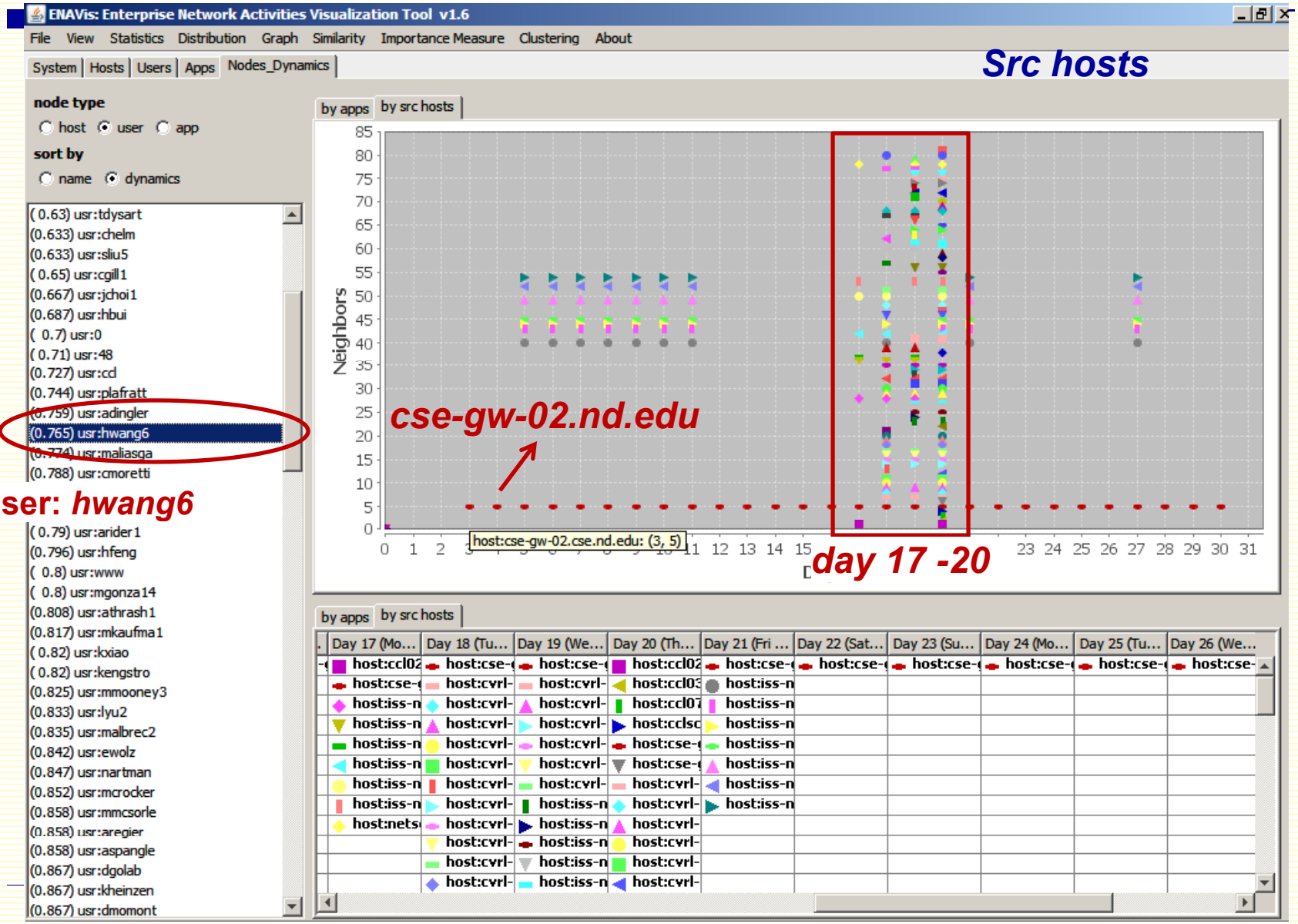
- Neighborhood changes
- Hosts:
  - users
- Users:
  - hosts
  - applications
- Applications:
  - users
  - src/dst hsots
- Quick and easy visualization
  - 2D scatter plots



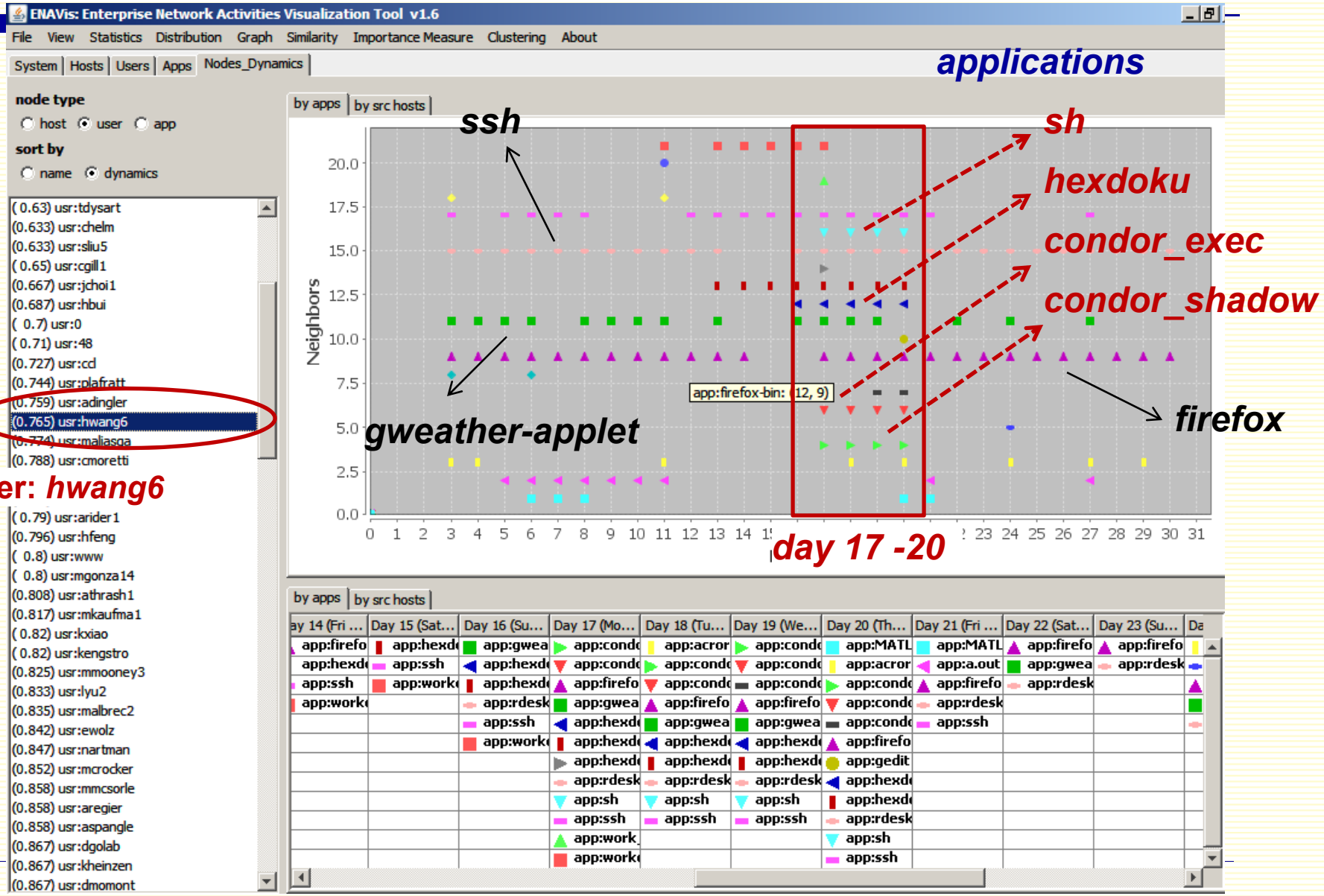
Node Similarity Visualization



Node Similarity Visualization



Node Similarity Visualization



Conclusion

- Visualizing *dynamic relationships* among ***hosts***, ***users***, and ***applications*** vs. traditional IP/port-based *Netflow* monitoring.
- Importance and challenges of ***similarity*** / ***differences*** of network graphs
  - Security / forensics / policy audit
  - Network management, troubleshoot
  - Anomaly analysis
- ***Similarity visualization*** : a promising approach.
- Novel transformation of graphs
  - HUA connectivity graphs, MDS graphs, (multi-)bipartite graphs, similarity graphs
- Hierarchical similarity visualization framework
  - Inter-graphs, intra-graphs, nodes
- More info available at http://netscale.cse.nd.edu/Lockdown