VANDERBILT

Creating Interpretable Collaborative Patterns to Detect Insider Threats

You Chen

Department of Biomedical Informatics,
Vanderbilt University

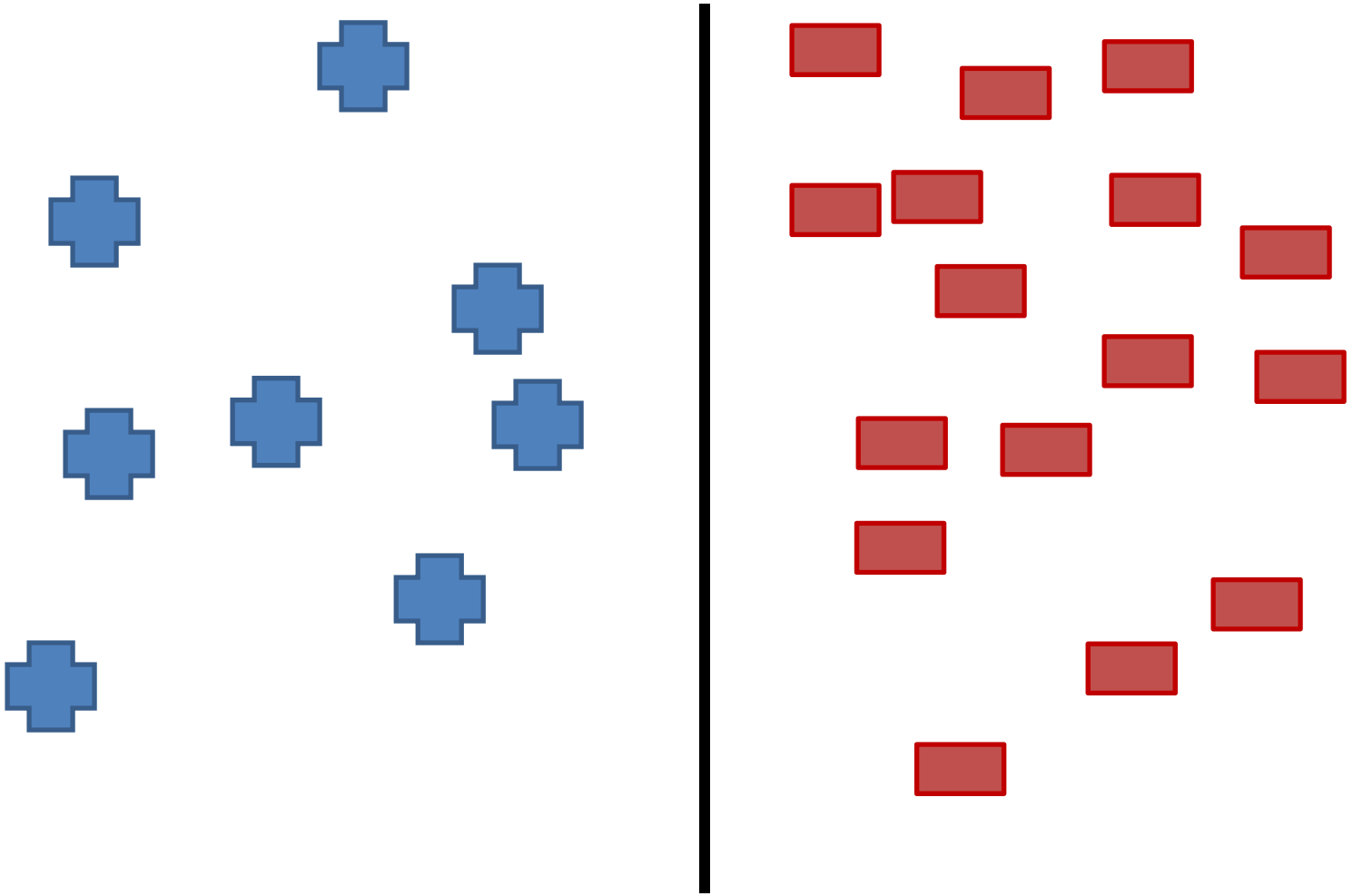
You.chen@vanderbilt.edu

http://hiplab.org/~ychen

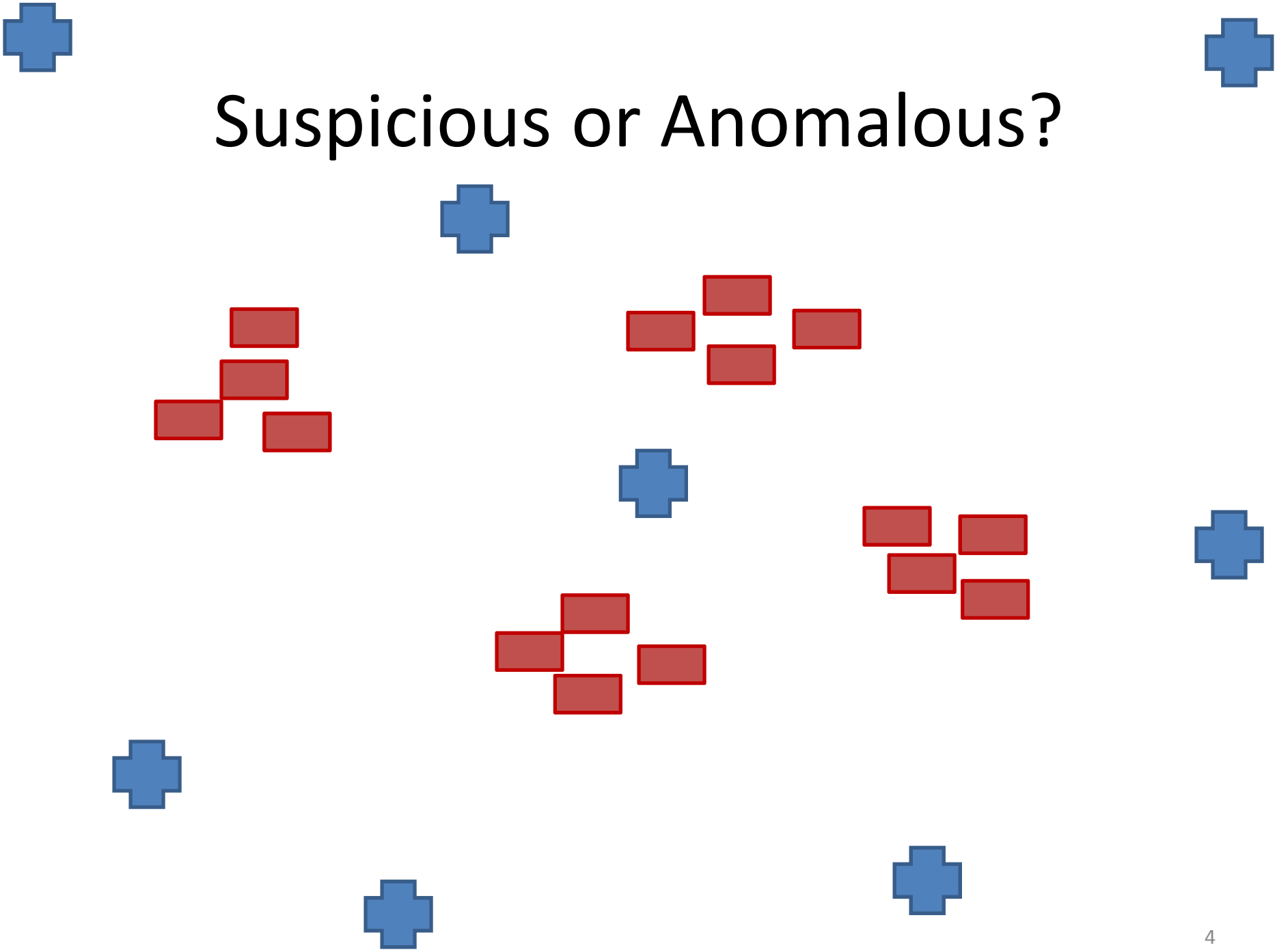
What Makes Sense?

- Dr. Smith's access of Peggy Johnson's medical record was strange
- Dr. Smith's access was 10 standard deviations away from normal behavior in his hospital
- Dr. Smith's access was strange because he is a neonatologist and he accessed the record of a 100 year-old woman who, for the past year, has only been treated by gerontologists

Suspicious or Anomalous?



Suspicious or Anomalous?



How Did We Get Here?

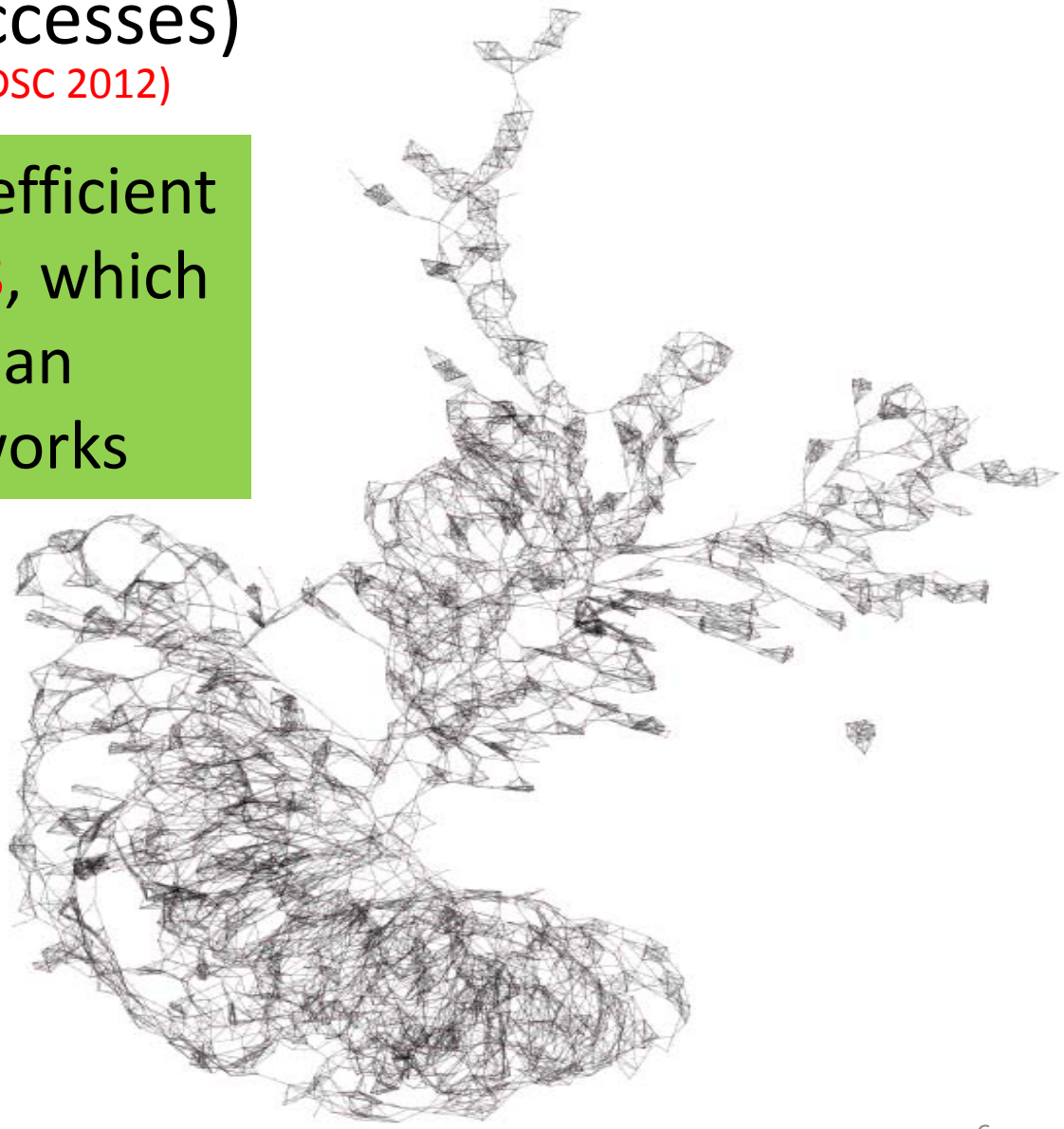
- Collaborative systems are about social phenomena
- People *should* form communities
- We should be able to measure deviation from community structure

6-Nearest Neighbor Network-Vanderbilt Medical Center (1 day of accesses)

(Chen, Nyemba, & Malin – IEEE TDSC 2012)

The average cluster coefficient for this network is 0.48, which is significantly larger than 0.001 for random networks

Users exhibit collaborative behavior in the Vanderbilt StarPanel System



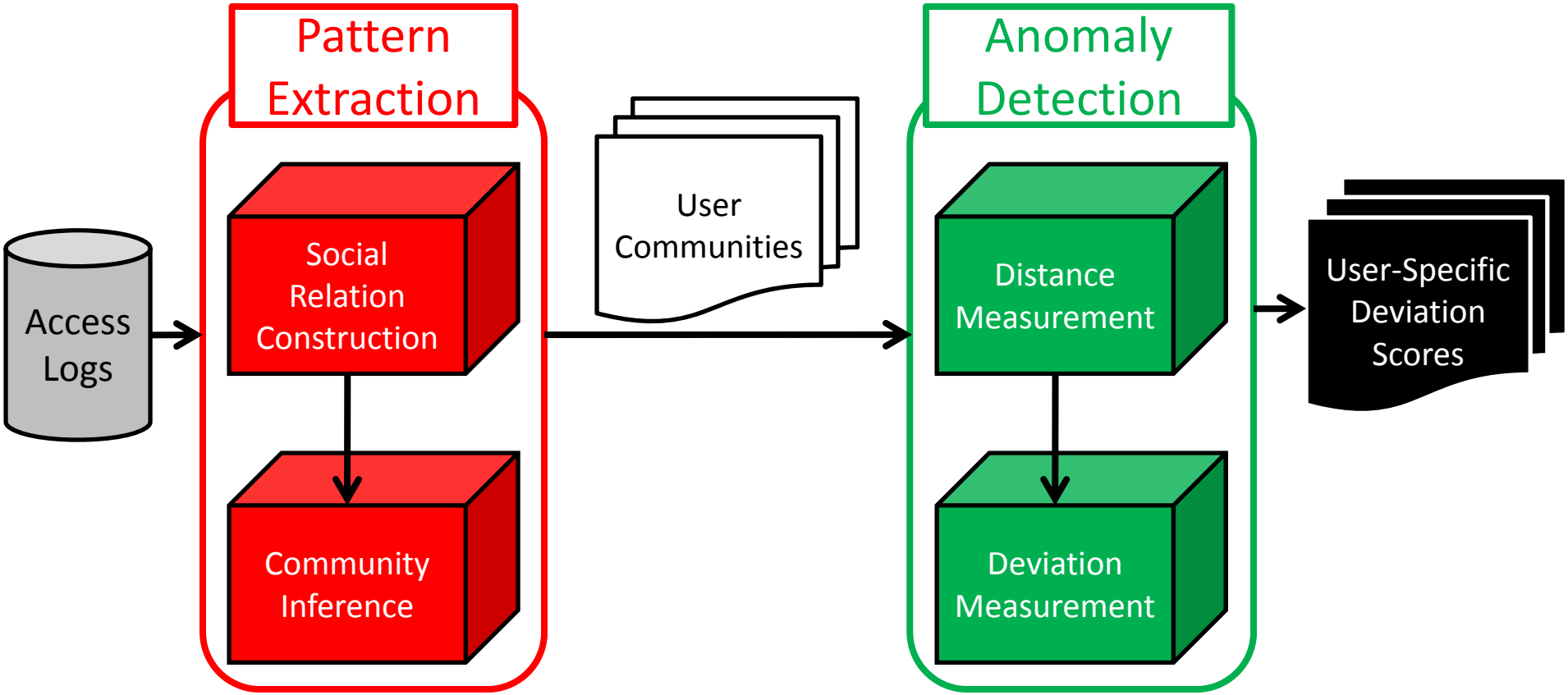
Automatically Learning of Collaborative Patterns

- User Level
  - Interaction relations of users
- Department Level
  - Interactions relations of departments

User Level

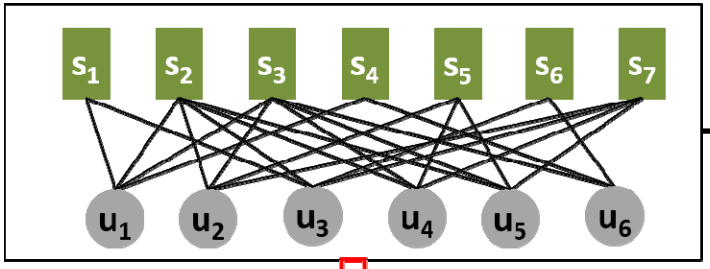
Detecting Anomalous Insiders through Community of Users

Community-Based Anomaly Detection (CADS)



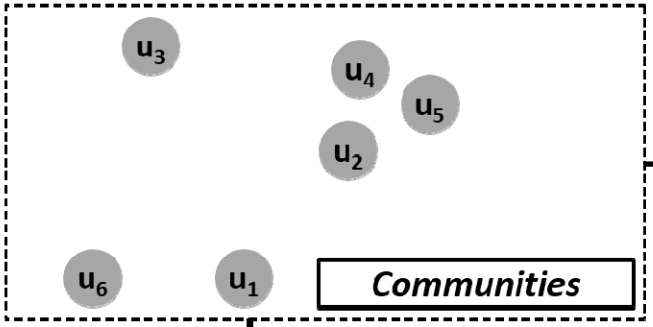
(Chen & Malin – ACM CODASPY 2011)

User Level



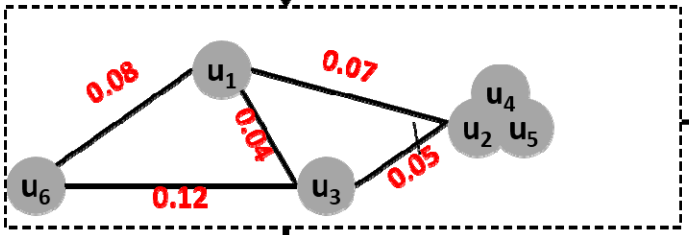
Bipartite Graph->Access Network of Users

Community Patterns



Communities via Singular Value Decomposition

Anomaly Detection

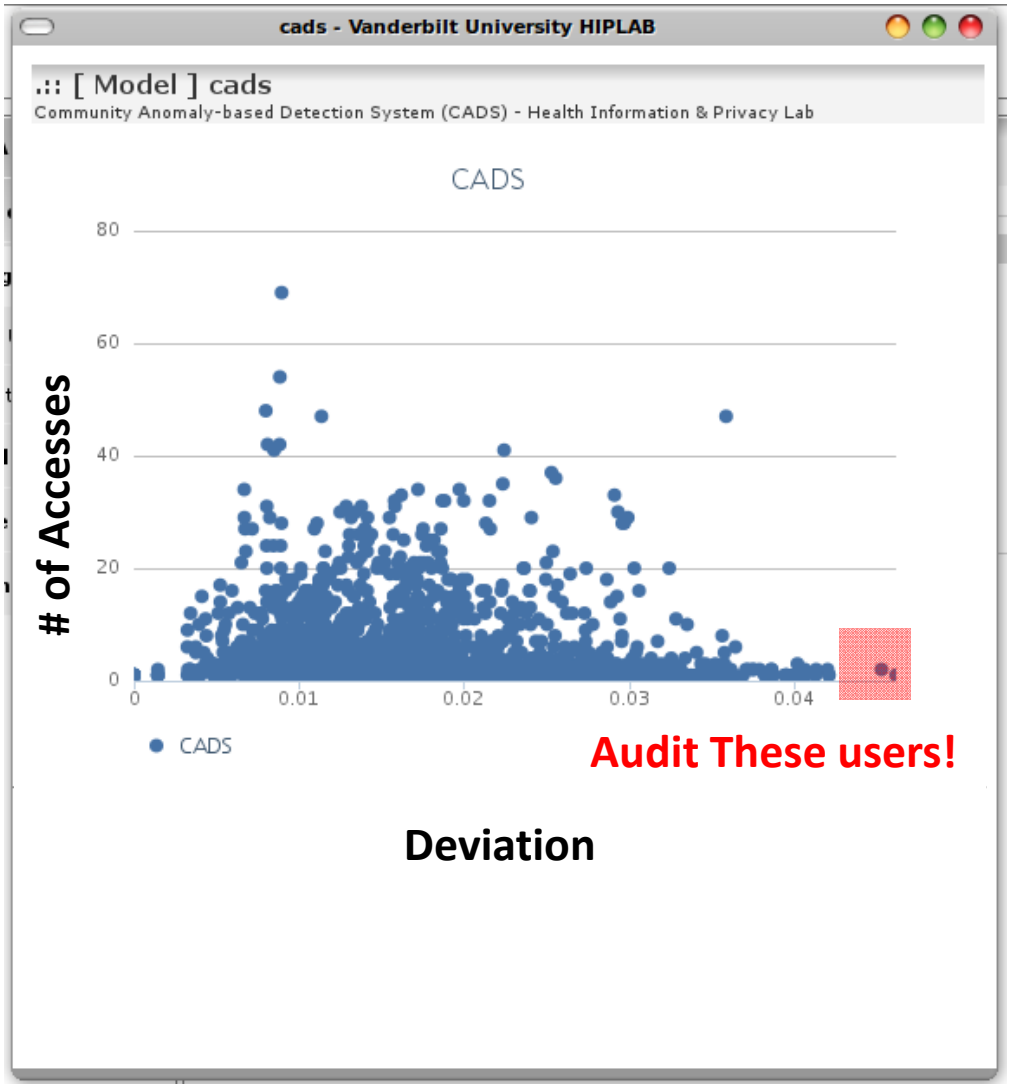


Nearest Neighbor Network

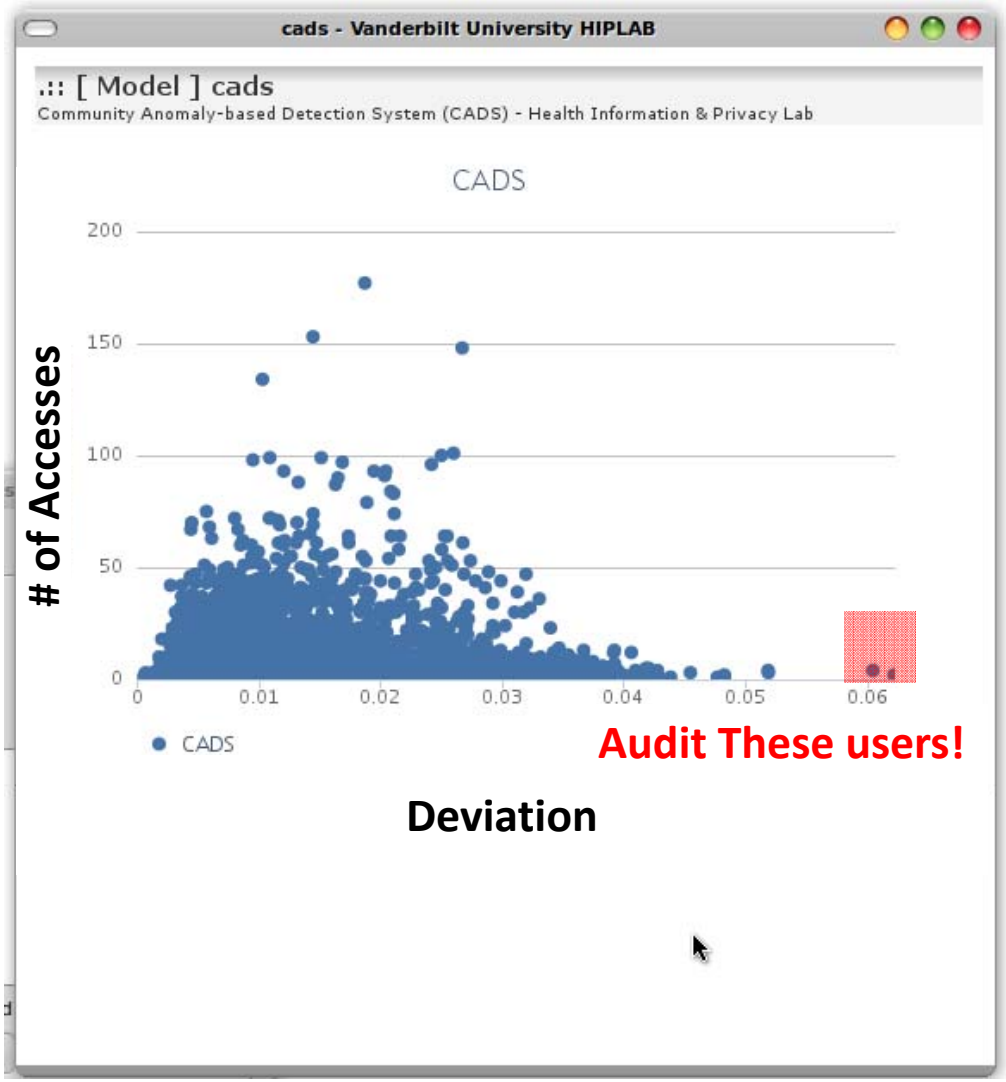
User	2-NN		Deviation
u_1	u_2	u_3	0.0405
u_2	u_4	u_5	0
u_3	u_1	u_2	0.0366
u_4	u_2	u_5	0
u_5	u_2	u_4	0
u_6	u_1	u_3	0.0427

Deviation Scores

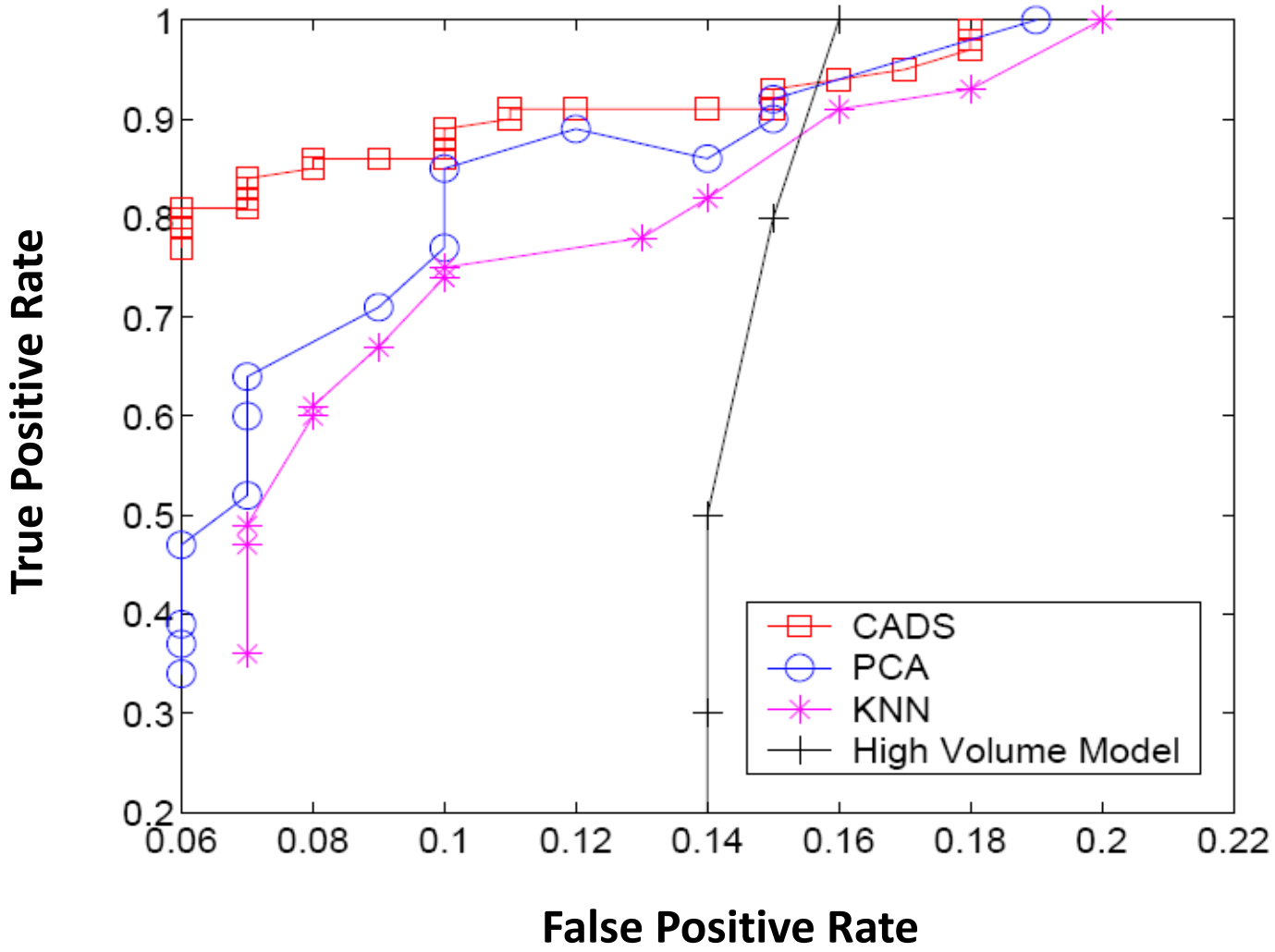
CADS on Vanderbilt Dataset



CADS on Northwestern Dataset

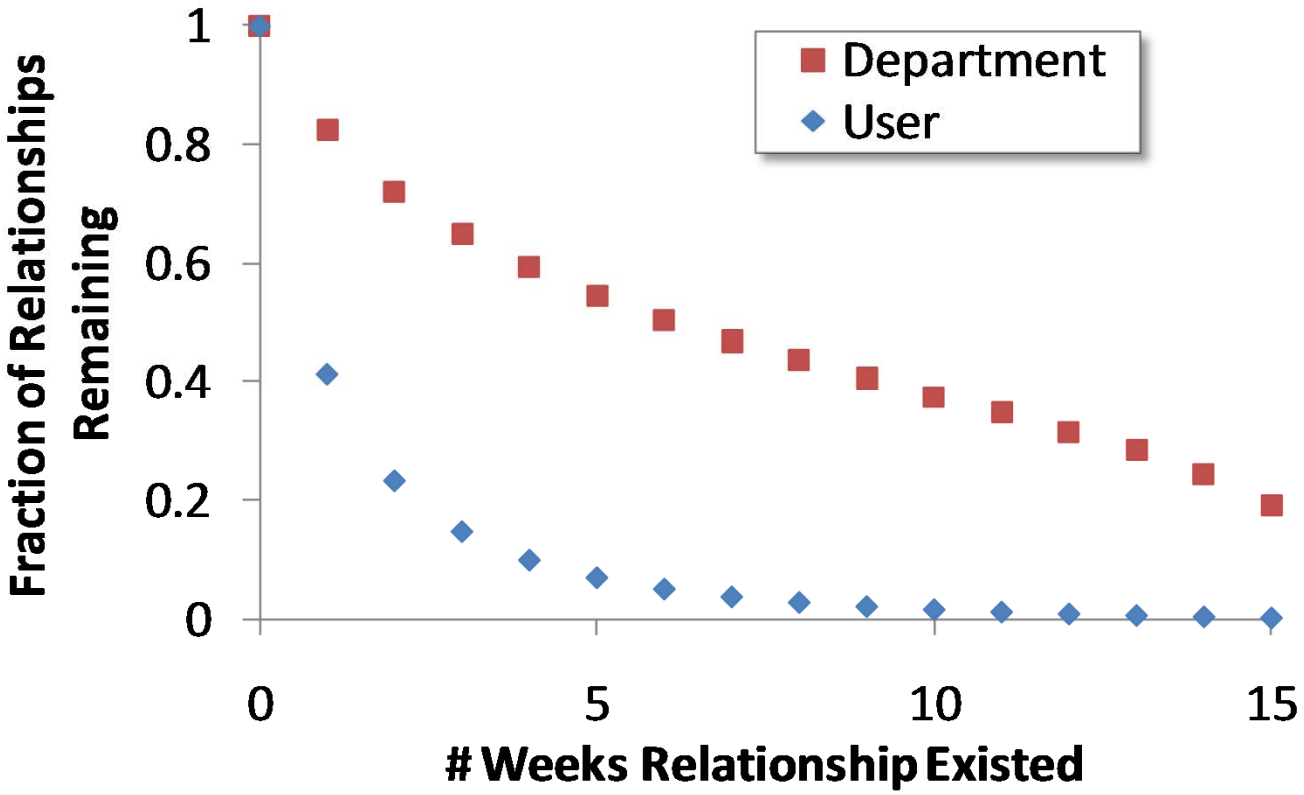


CADS Outperforms Competitors



But Relationships Decay…

(Malin, Nyemba, & Paulett – JBI 2011)



EMR <user, user> relationships

Department Level

Auditing Medical Record Accesses through Healthcare Interactions

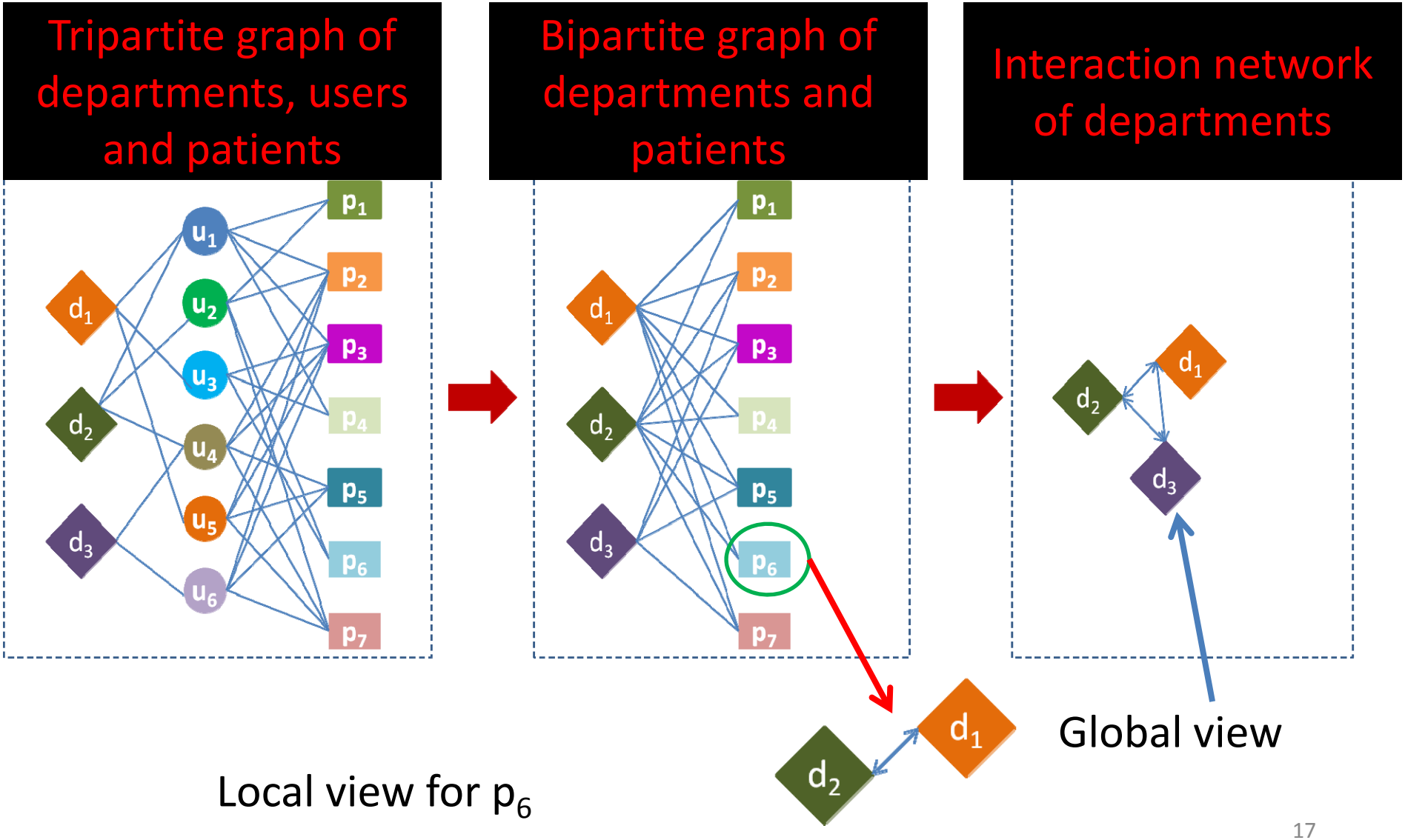
Hospital Departmental Relations Can Be Inferred

(Chen, Nyemba, & Malin - AMIA 2012)

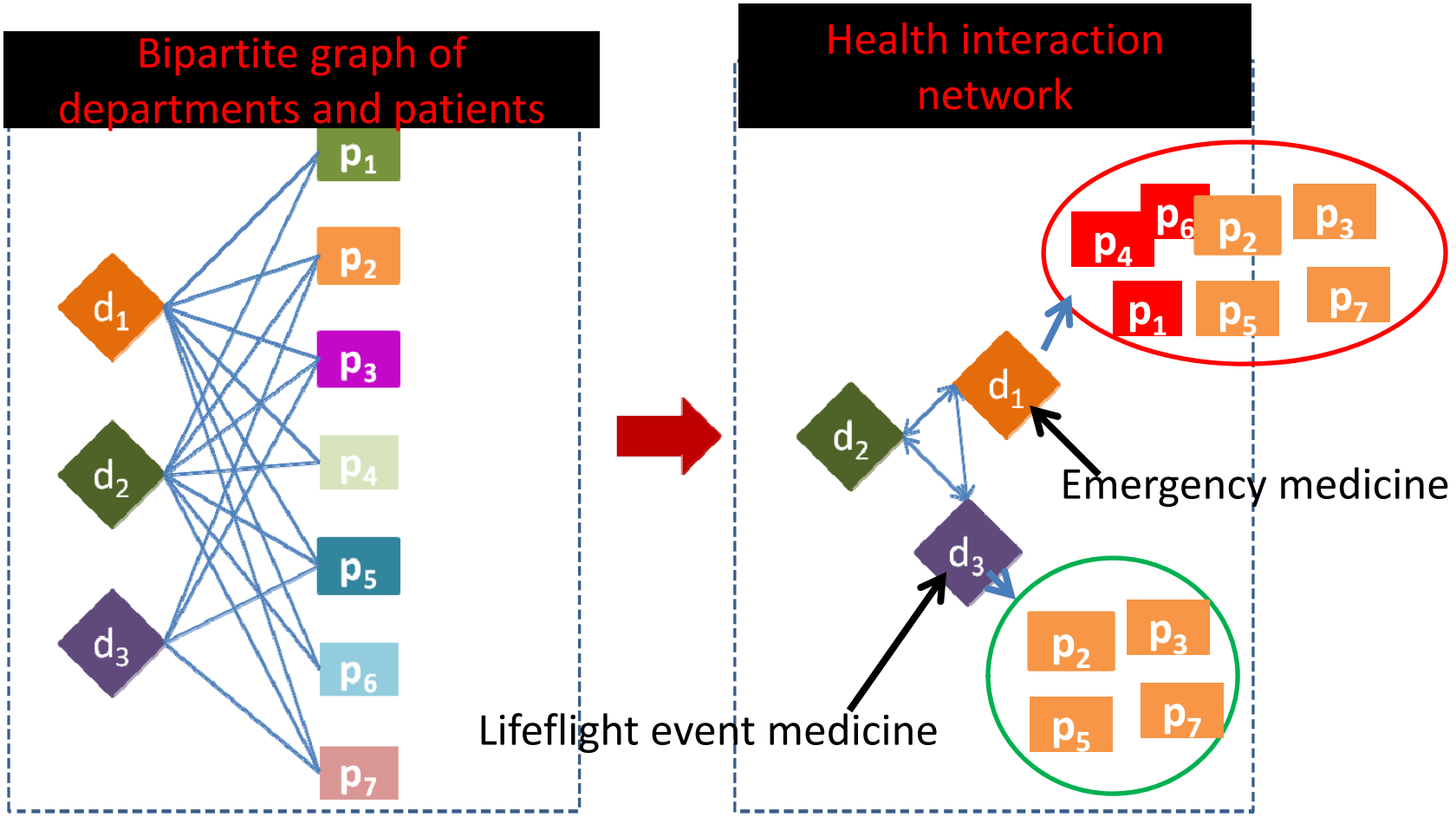
- Probability department d_i accesses a patient's record, given department d_j accessed the record.

Department (d_i)	Department (d_j)	Min Certainty	Max Certainty
Intradepartmental Relations			
4East OB/GYN	4East OB/GYN	0.74319	0.7669
Adult Emergency Medicine	Adult Emergency Medicine	0.74024	0.78453
Cancer Infusion Center	Cancer Infusion Center	0.73171	0.844
8N Inpatient Medicine	8N Inpatient Medicine	0.7197	0.80909
Newborn Nursery	Newborn Nursery	0.70406	0.72727
Interdepartmental Relations			
DOT Radiology	Orthopaedics	0.99621	1
Nursing Education and Development	Medical Information Services	0.95833	1
Main OR - Trauma/Renal	Medical Information Services	0.94444	1
Life Flight Event Medicine	Emergency Medicine	0.90805	1
Emergency Medicine Admin	Adult Emergency Medicine	0.91489	0.94186

Organization Level-Department



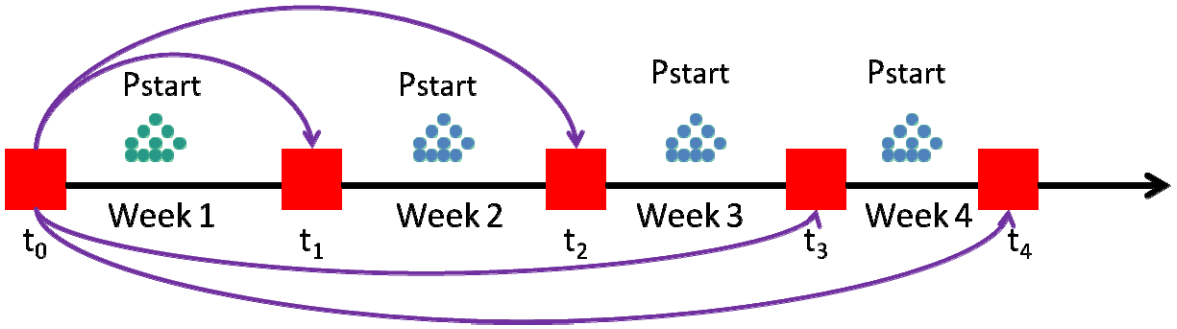
Certainty to Model Relationship Among Departments



Cert(Emergency medicine (d_1)->Lifeflight event medicine (d_3)) = 4/7

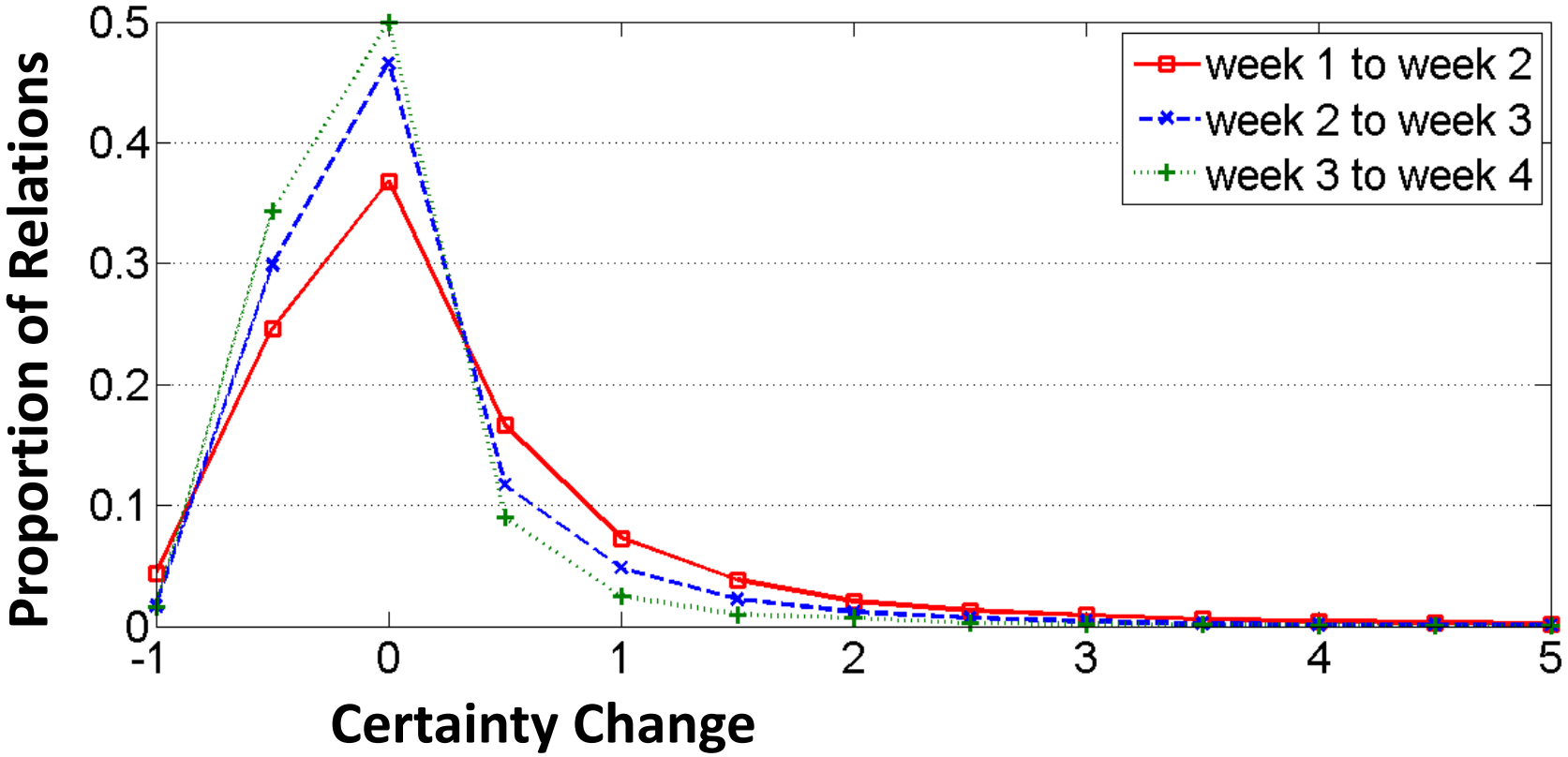
Lifeflight event medicine (d_3)-> Cert(Emergency medicine (d_1)) = 4/4

Evolution of Local Network Relations Can be Used Detect “Strange” Behavior



Each point in the P_{start} corresponds to a local network

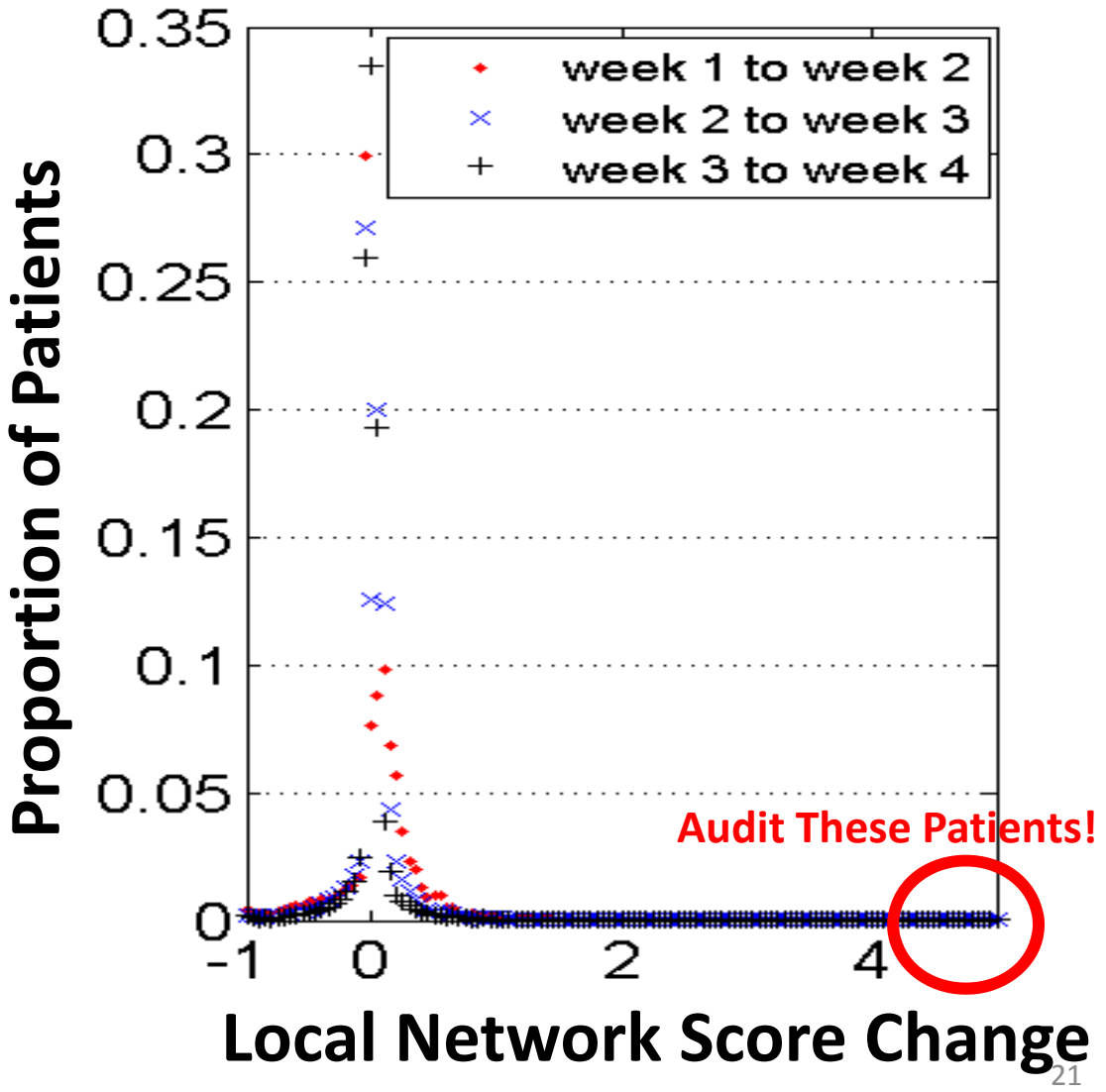
The changes become smaller over time (centralization: green > blue > red)



Degree of relations between departments changes little over time

>82.5% of the change resides in [-0.25, 0.25]

Most Patients Network Suggest They Are "Normal"

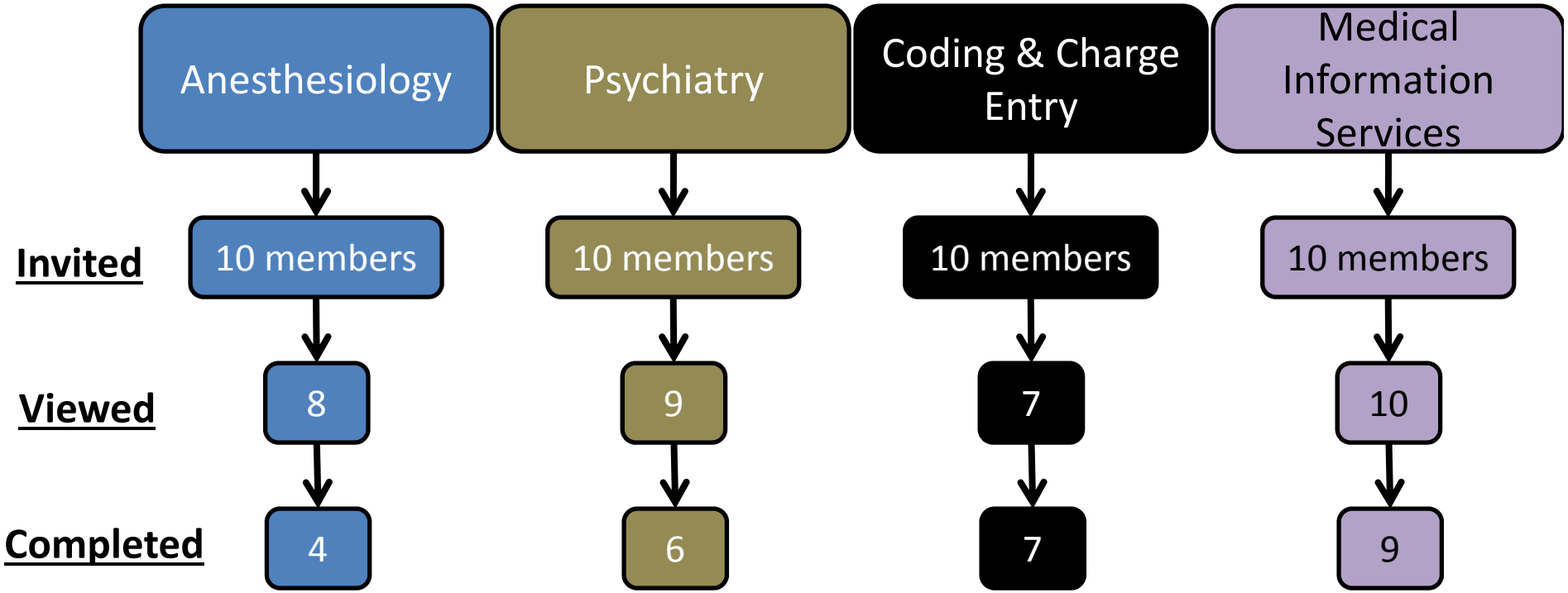


But Do You Believe the Data?

Survey Population

(Chen, Lorenzi, Nyemba, Schildcrout & Malin – IJMI 2014)

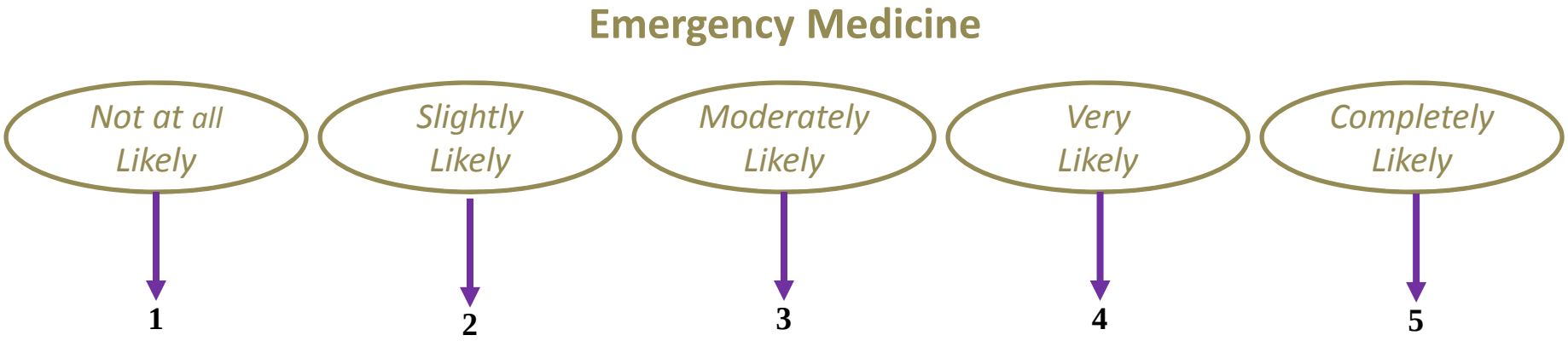
- Vanderbilt University Medical Center areas



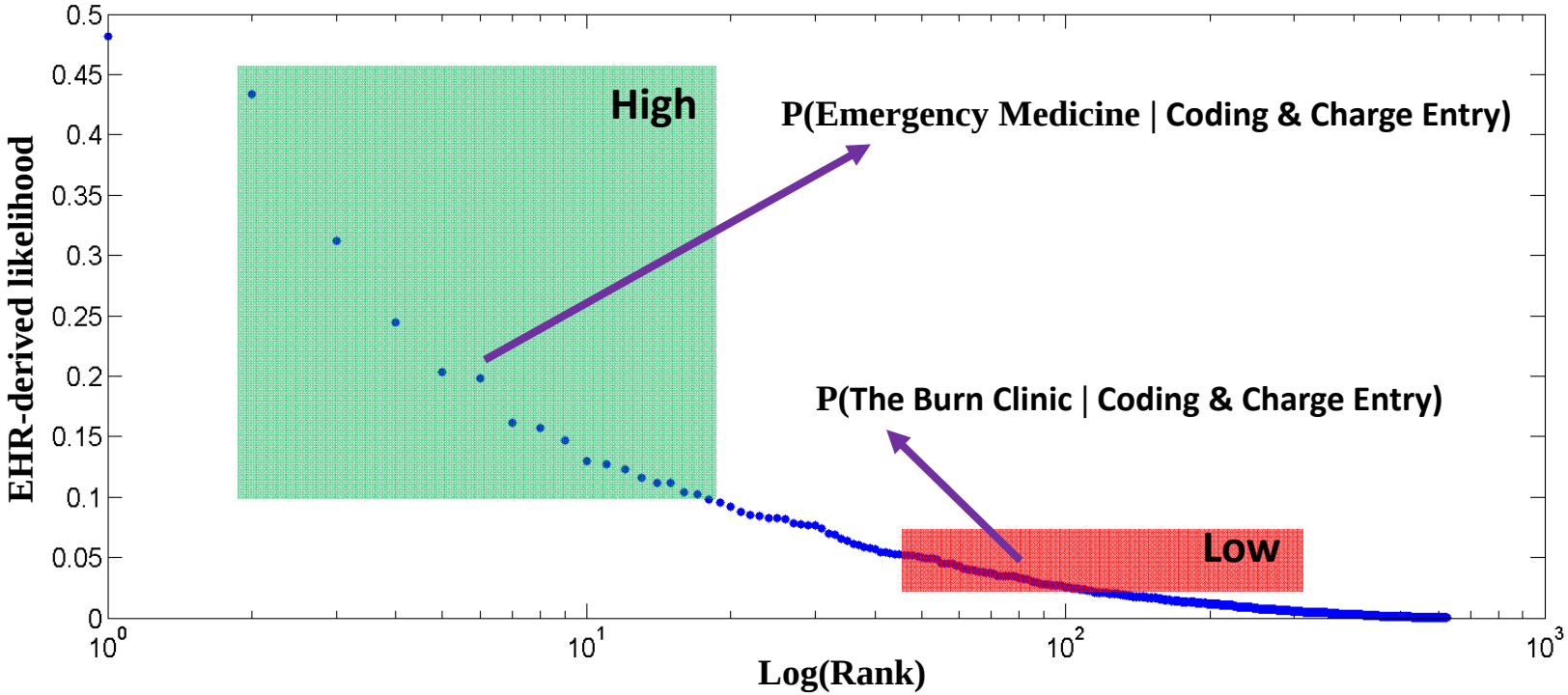
34 respondents did the survey and 26 of them are valuable

Survey Questions

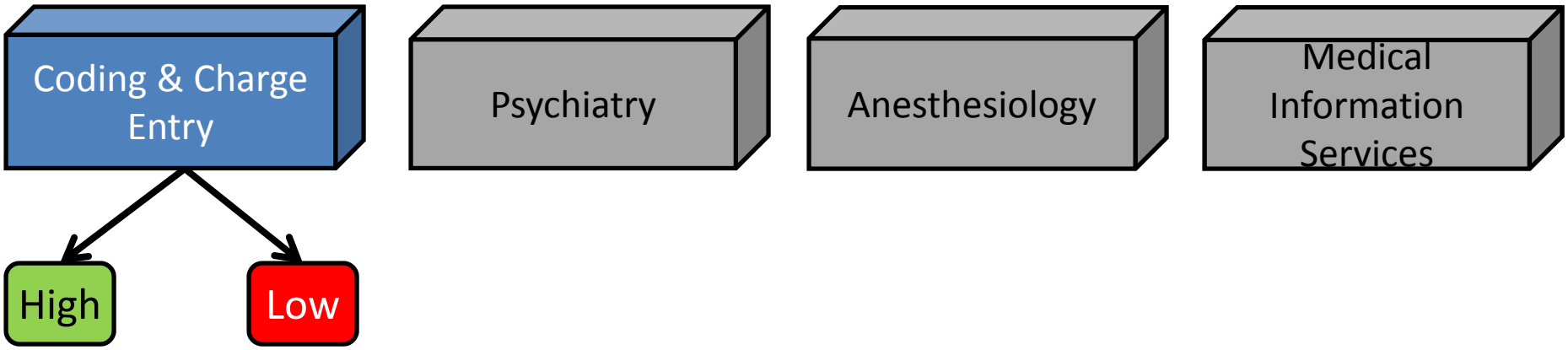
- Departmental interactions
- Conditional probabilities of accessing a record (conditioned on the HCO area)
- "Given someone from Coding & Charge Entry accessed the record, *what's the chance someone from the following Area accessed the record*?



Coding & Charge Entry Interactions
(one week, ~620 points)

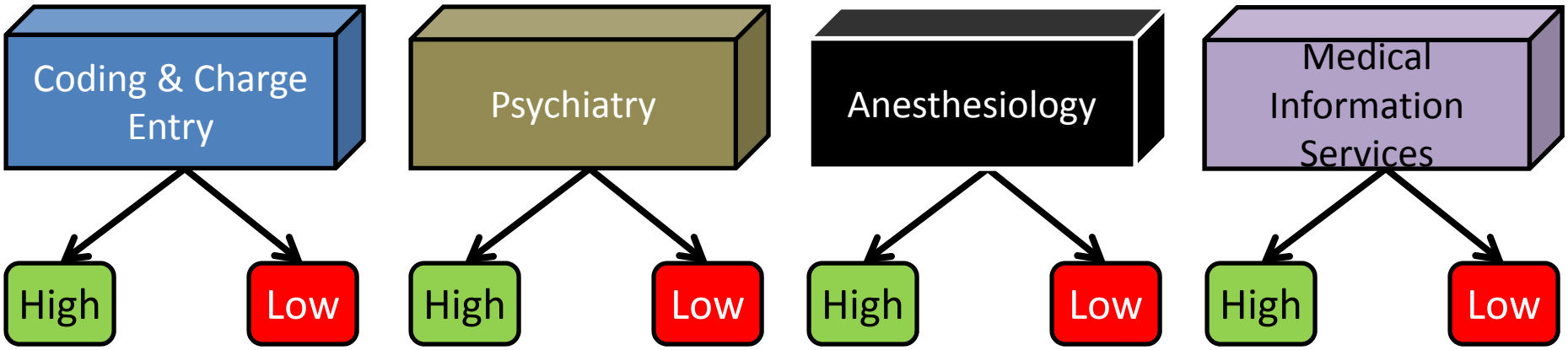


Survey Questions



20 rules

Survey Questions



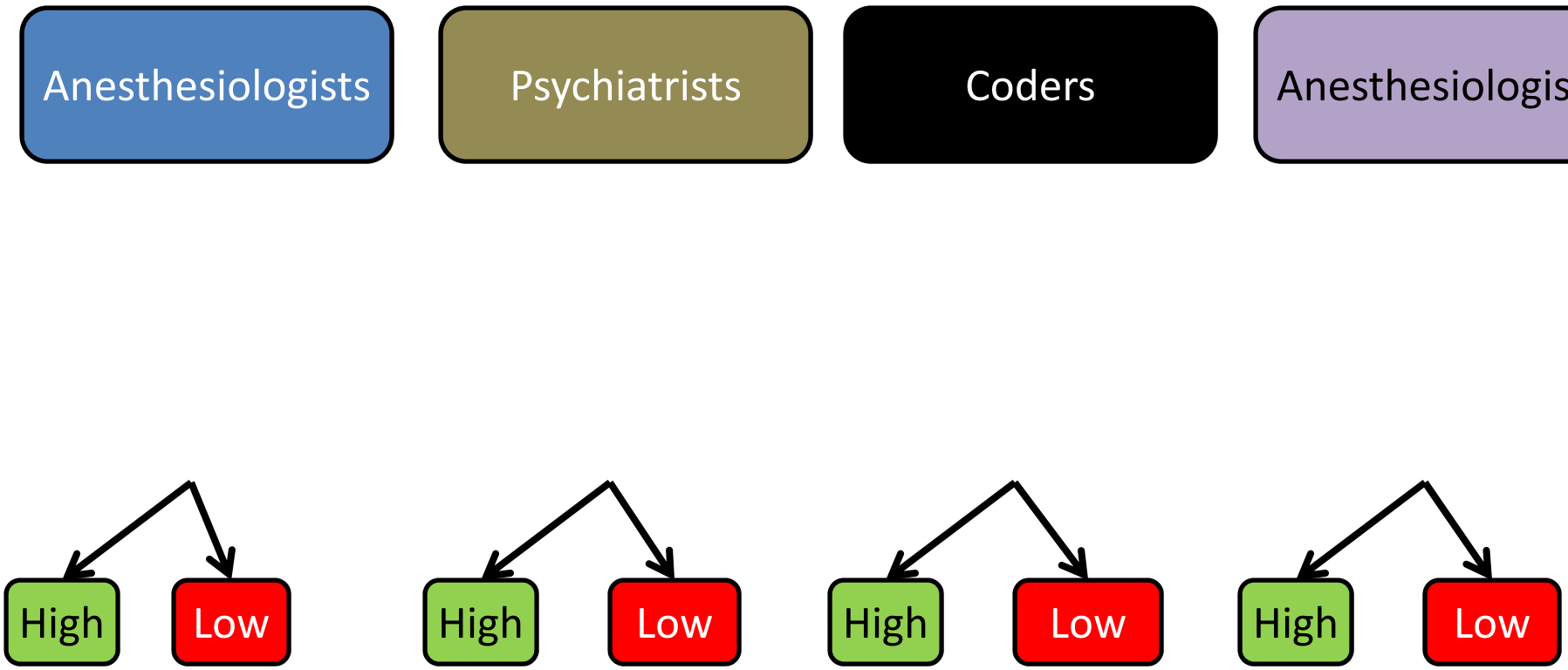
Hypothesis

- 1) Employees can distinguish between high, and low likelihood rules for all HCO areas
- 2) Employees can distinguish between high and low likelihood rules for their own HCO area
- 3) employees can distinguish between high, and low likelihood rules in their own HCO area better than they can in other HCO areas

One respondent has 8 observations
The total number of observations is 8*26=208

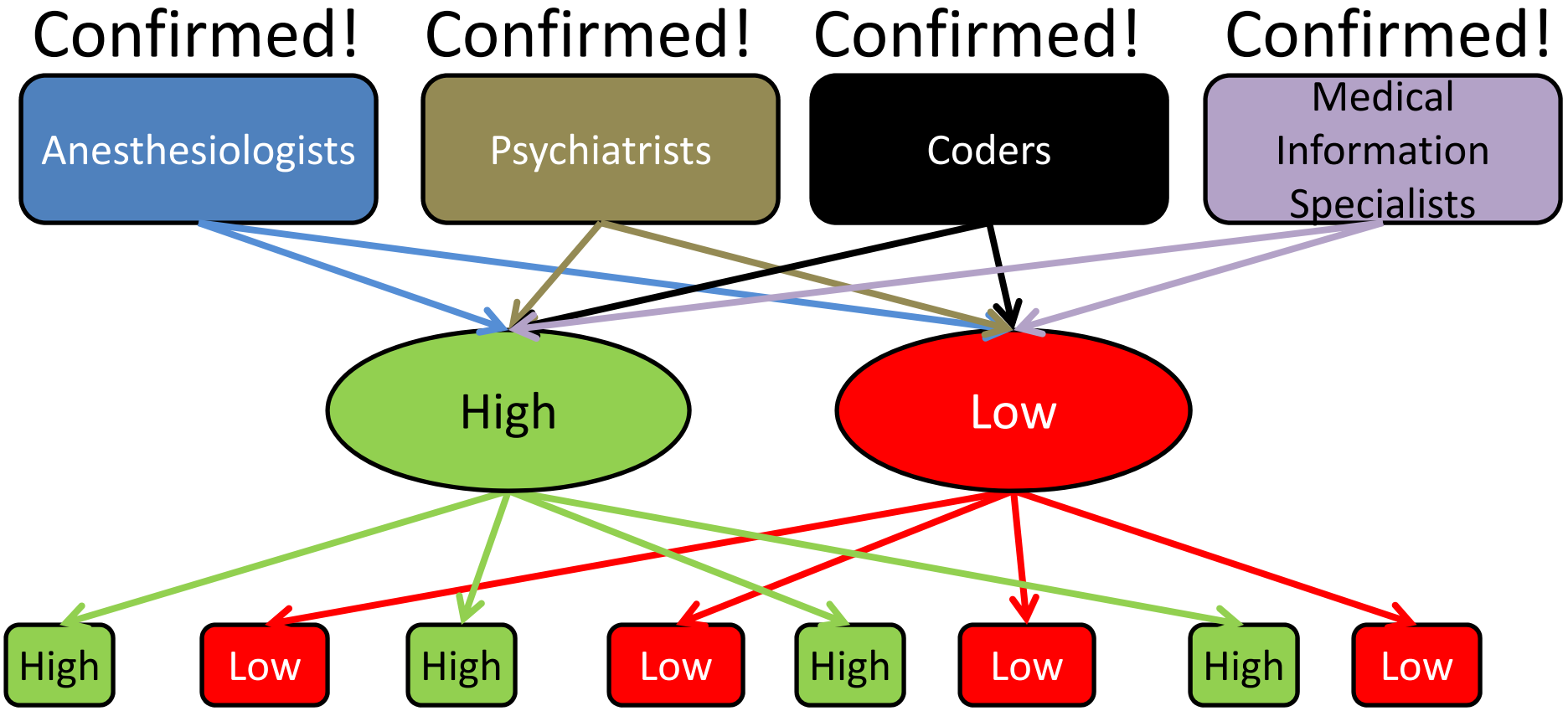
Respondent (ID)	Respondent Type (P)	Rule Type (R)	Rule Class (C)	Average Score of Responses
1	MIS	ANE	High	3
1	MIS	ANE	Low	2
1	MIS	CODE	High	3.3
1	MIS	CODE	Low	2.1
1	MIS	MIS	High	3.1111
1	MIS	MIS	Low	2.125
1	MIS	PSY	High	2.9
1	MIS	PSY	Low	2.05

Hypothesis Test 1 – Rules of All HCO Areas:



Hypothesis Test1 – Rules of All HCO Areas:

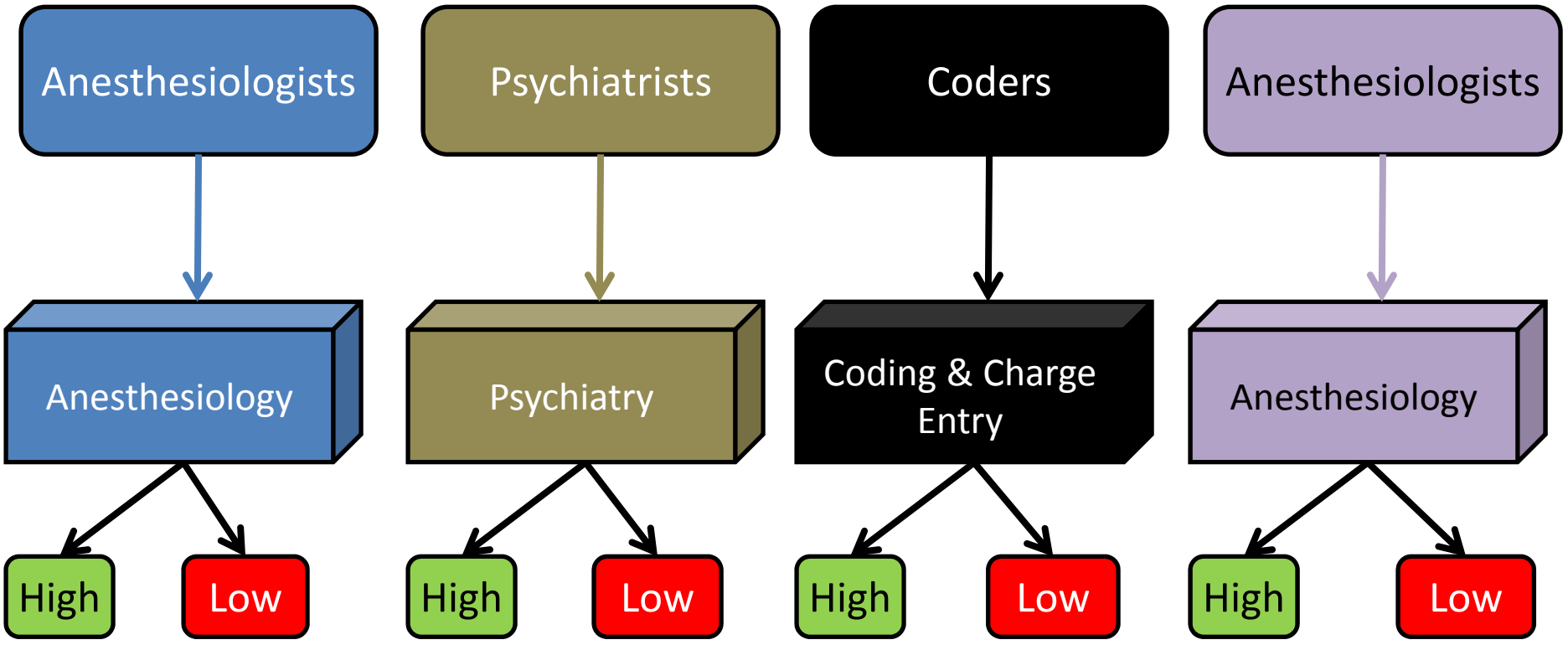
One-sided t-test, 95% confidence



Hypothesis Test 2– Self Assessment:

Linear Mixed Effects Model

One-sided t-test, 95% confidence



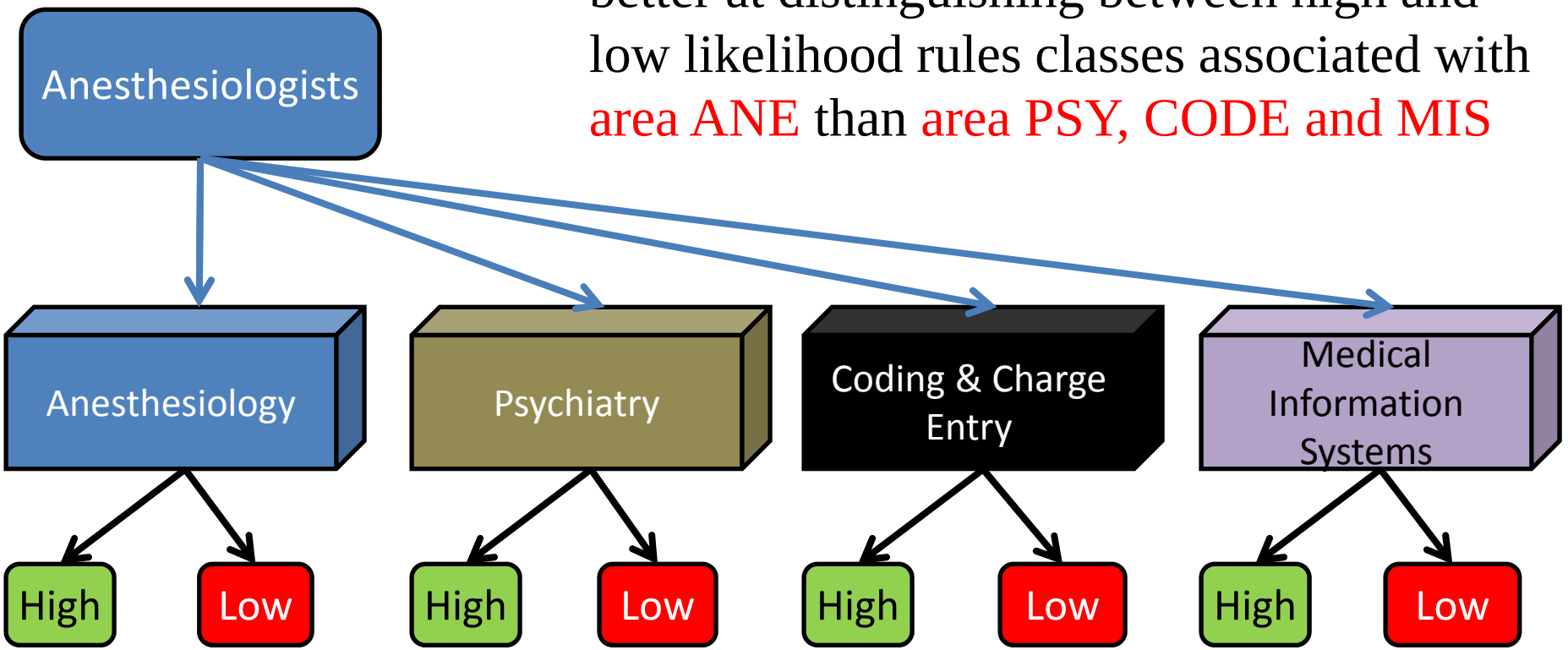
Confirmed! Confirmed! Confirmed! Confirmed!

Hypothesis Test 3– Bias Toward Own Rules

Linear Mixed Effects Model

One-sided t-test, 95% confidence

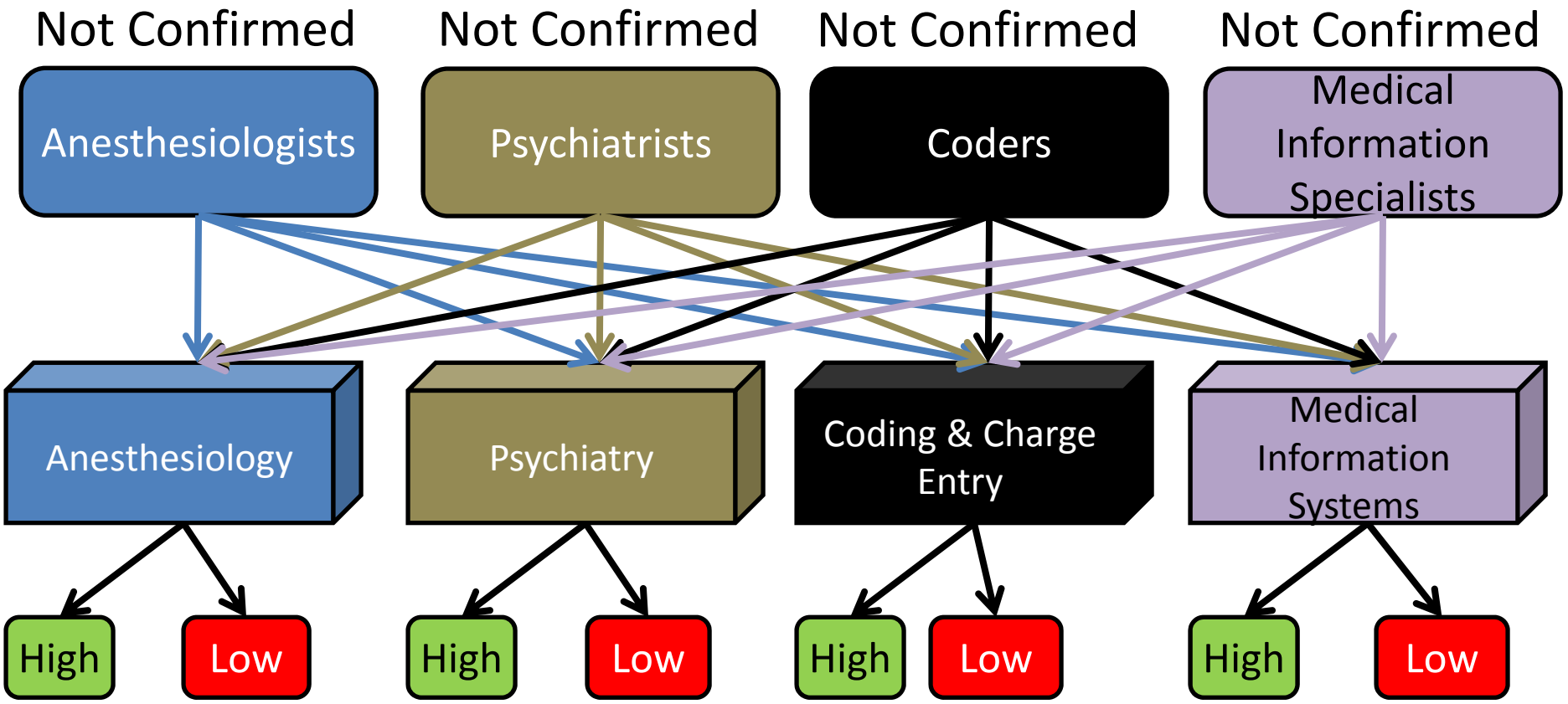
Respondents from HCO area ANE are better at distinguishing between high and low likelihood rules classes associated with area ANE than area PSY, CODE and MIS



Hypothesis Test 3 – Bias Toward Own Rules

Linear Mixed Effects Model

One-sided t-test, 95% confidence



Conclusions

- Healthcare organization employees generally understand what goes on around them…

  … and for other sections of the organization as well!

- Automated healthcare organizational modeling may be possible.

- Anomalies detection through collaborative patterns may be reliable!

Acknowledgements

Vanderbilt

- ***Bradley Malin, Ph.D.***
- ***Jonathan Schildcrout, Ph.D.***
- Dario Giuse, Dr. Ing
- ***Nancy Lorenzi, Ph.D.***
- Steve Nyemba, M.S.

UIUC

- Carl Gunter, Ph.D.

Northwestern

- David Liebovitz, M.D.

Funding

- National Science Foundation
  - CCF-0424422
  - CNS-0964063
- National Institutes of Health
  - R01LM010207

Reference

- Chen & Malin – ACM CODASPY 2011 : **Detection of Anomalous Insiders in Collaborative Environments via Relational Analysis of Access Logs**. ACM Conference on Data and Application Security and Privacy.
- Malin, Nyemba, and Paulett 2011: **Learning Relational Policies from Electronic Health Record Access Logs.** Journal of Biomedical Informatics.
- Chen, Nyemba, & Malin - AMIA 2012 : **Auditing Medical Records Accesses via Healthcare Interaction Networks.** AMIA 2012 Annual Symposium.
- Chen, Nyemba, & Malin – IEEE TDSC 2012 : **Detecting Anomalous Insiders in Collaborative Information Systems.** IEEE Transaction on Dependable and Secure Computing.
- Chen, Lorenzi, Nyemba, Schildcrout & Malin – IJMI 2014 : **We Work with Them? Healthcare Workers Interpretation of Organizational Relations Mined from Electronic Health Records**, International Journal of Medical Informatics.

Q&A

Thanks!