RETROSPECTIVE ON A DECADE OF RESEARCH IN VISUALIZATION FOR CYBERSECURITY

R. Jordan Crouser, **Erina Fukuda & Subashini Sridhar**

SMITH COLLEGE

HUMAN COMPUTATION & VISUALIZATION LABORATORY

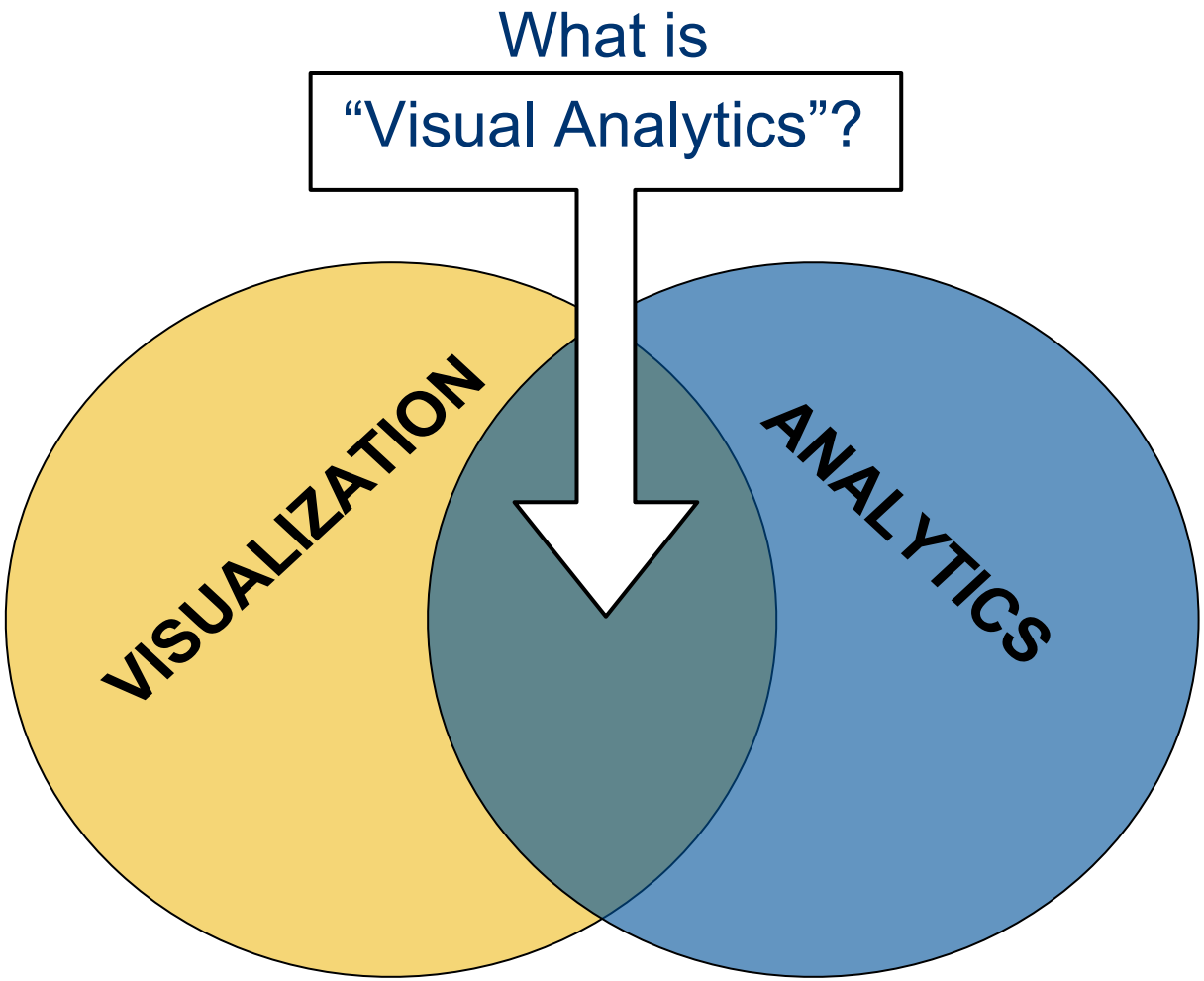
LAS Laboratory for Analytic Sciences

Reflect. Observe. Imagine.

Outline

- 2-minute crash course in **visual analytics**
- Using visualization for cybersecurity
- 2-armed survey
  - Classification using existing frameworks
  - Automated text mining
- Next Steps

Visual Analytics 101



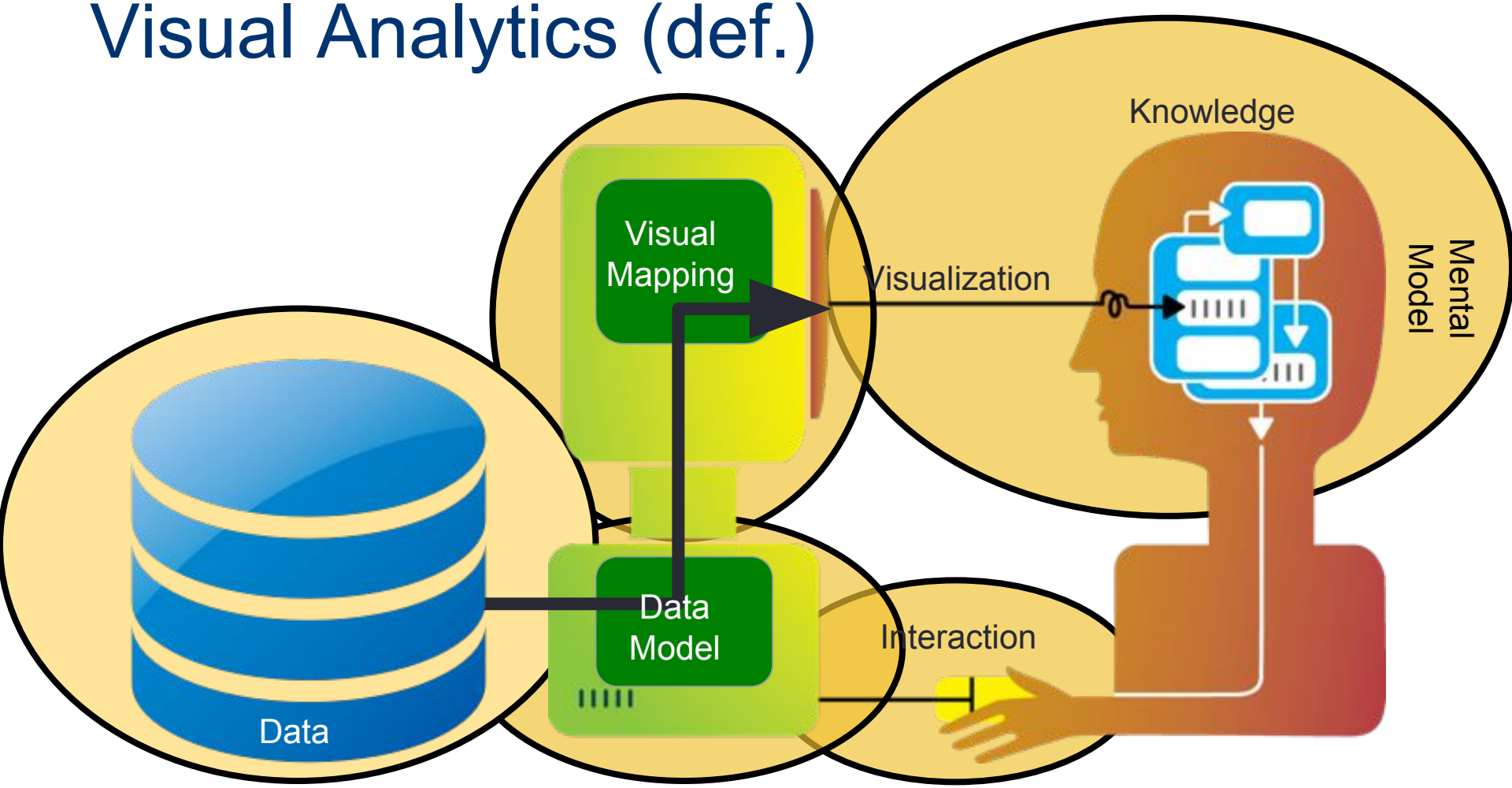
Visualization (def.)

Visual representations of data that reinforce human cognition

Analytics (def.)

Discovery and communication of **meaningful patterns** in data

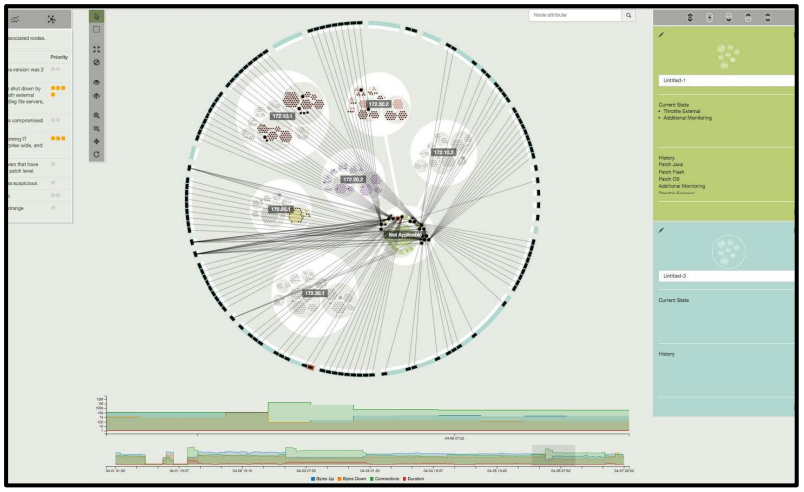
Visual Analytics (def.)



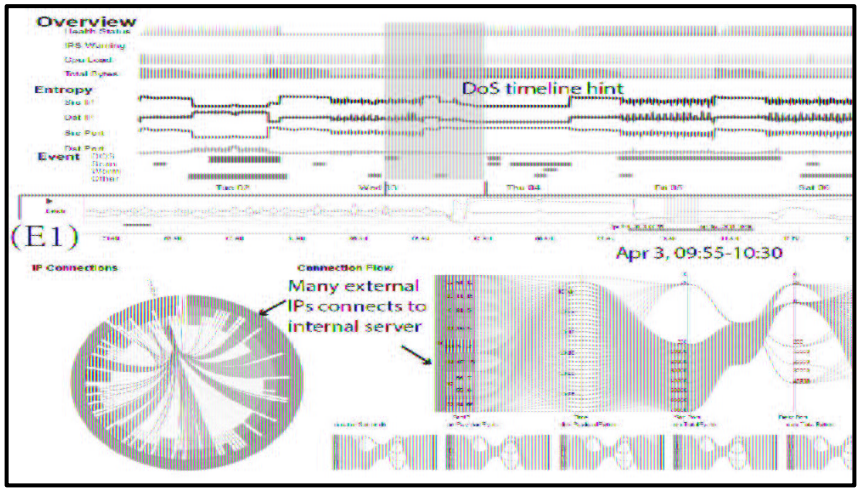
(the science of) **analytical reasoning** facilitated by **interactive visual interfaces**[1]

[1]Thomas, James J., and Kristin A. Cook, eds. *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press, 2005.

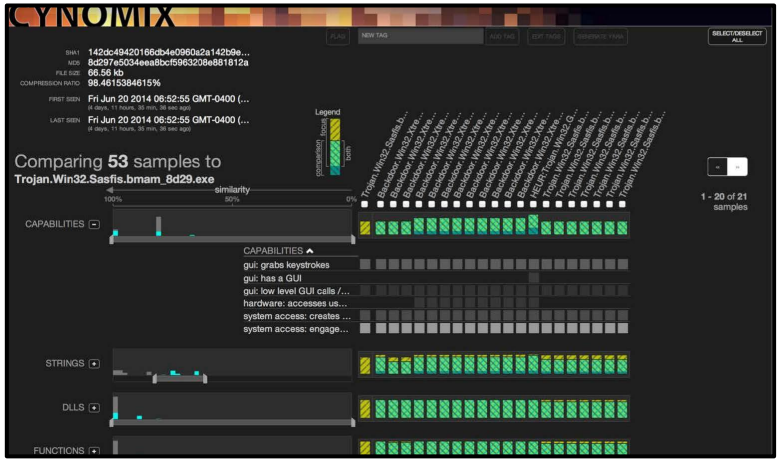
Visualization / VA for cybersecurity



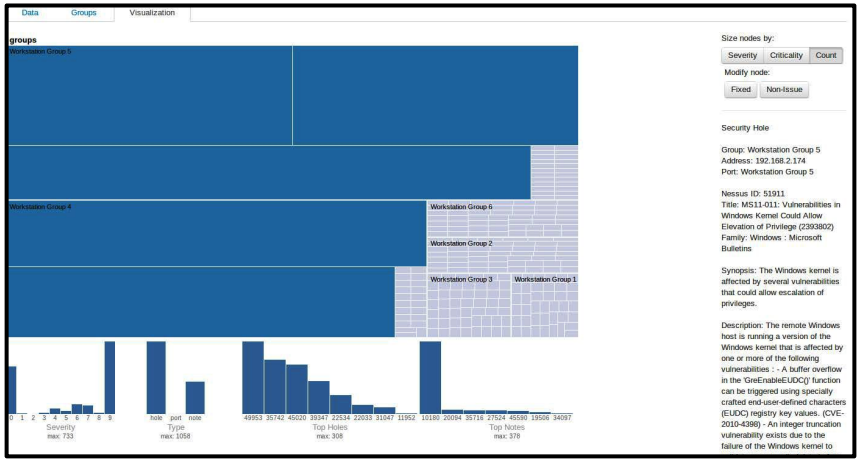
Ocelot (Arendt et al. 2015)



OCEANS (Chen et al. 2014)



SEEM (Gove et al. 2014)

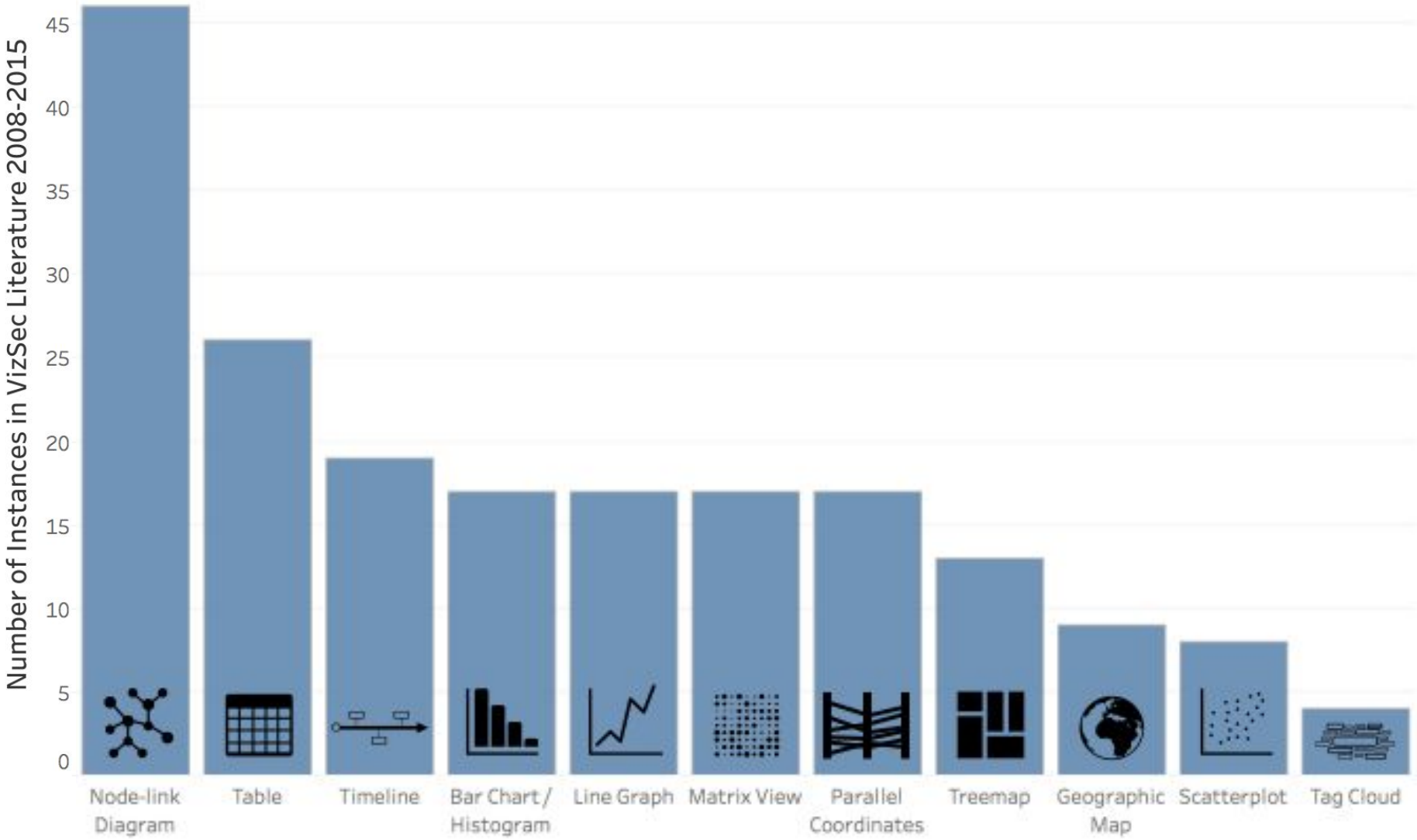


NessusVIS (Harrison et al. 2012)

Methodology pt. 1: traditional lit review

- Corpus of 161 papers published in IEEE Visualization for Cyber Security Conference between 2004 and 2015
- Manually classified papers along following dimensions using existing frameworks from the VIS community
  - Visualization Techniques
  - Interaction Techniques
  - Data Type
  - Analytical Goal
  - Other metadata (authors, date of publication, etc.)

Finding 1: common visual metaphors



Node-link vs. Matrix

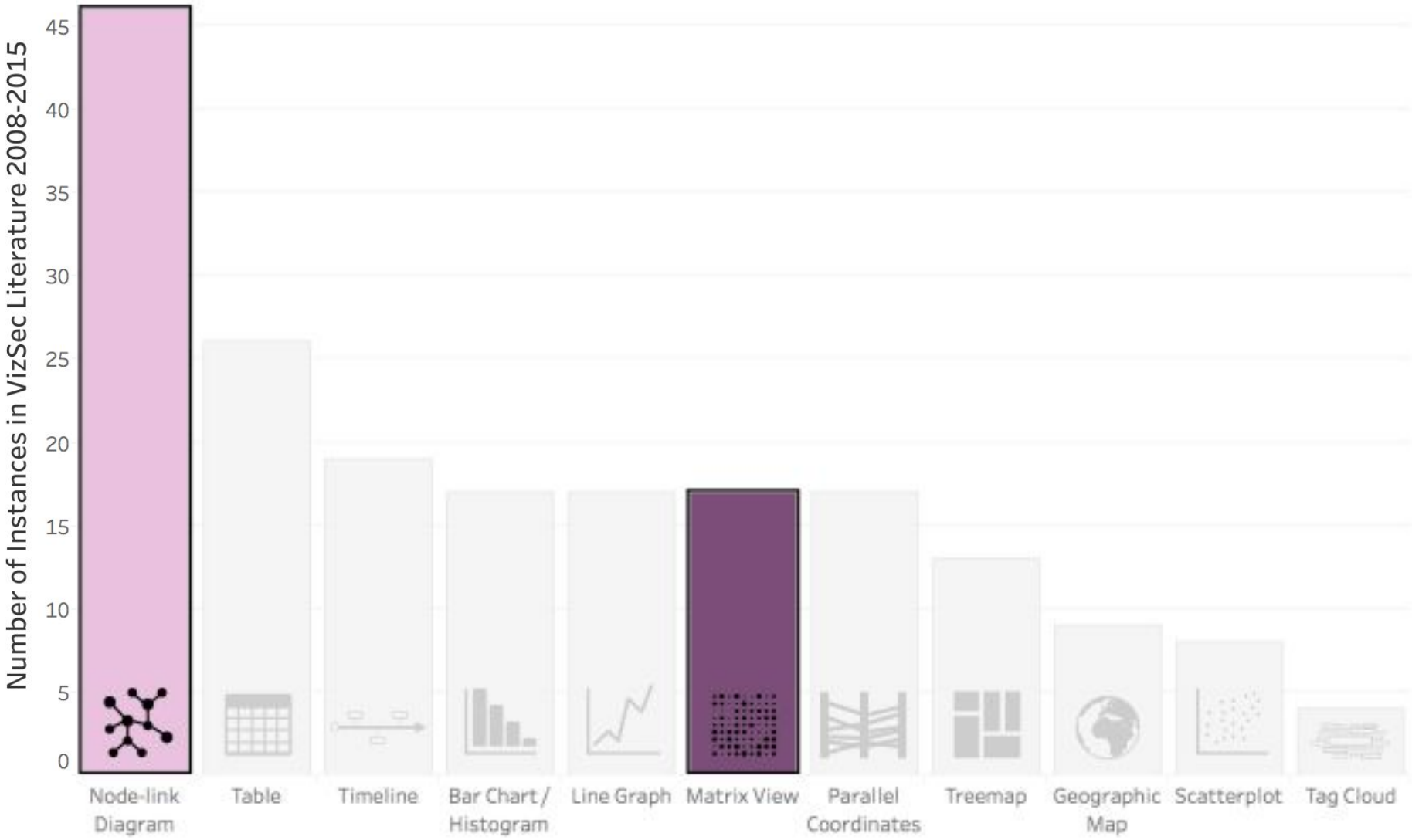
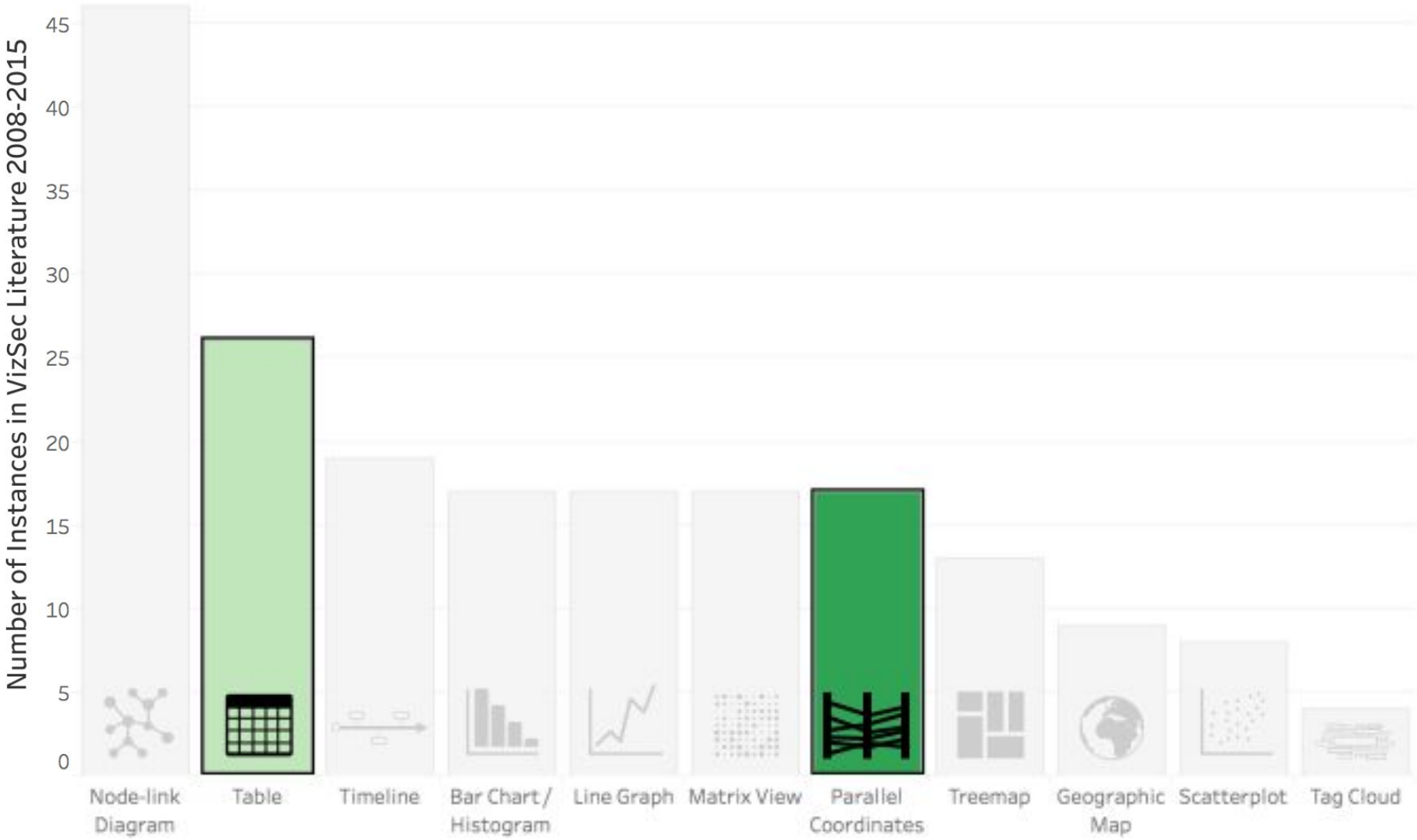
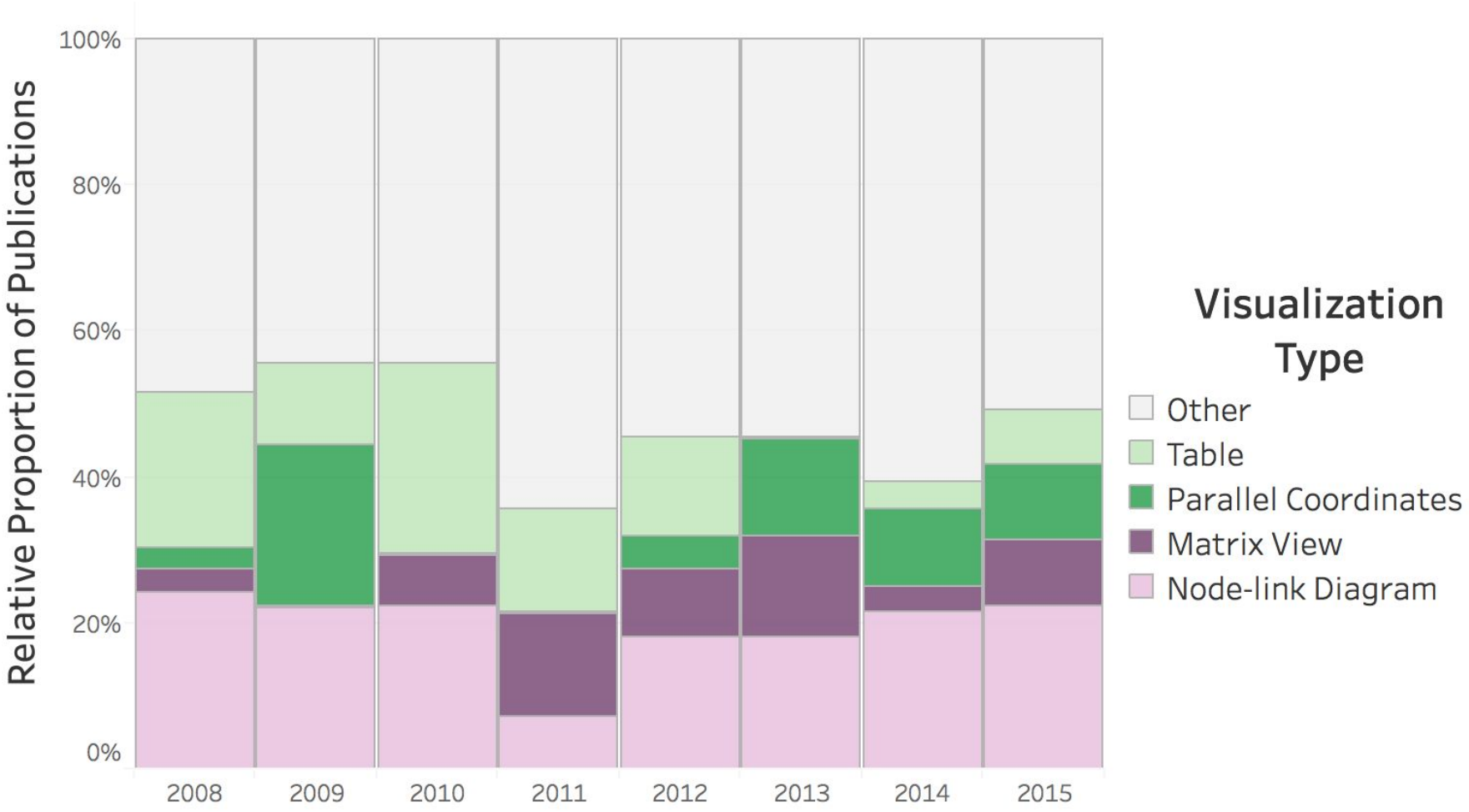


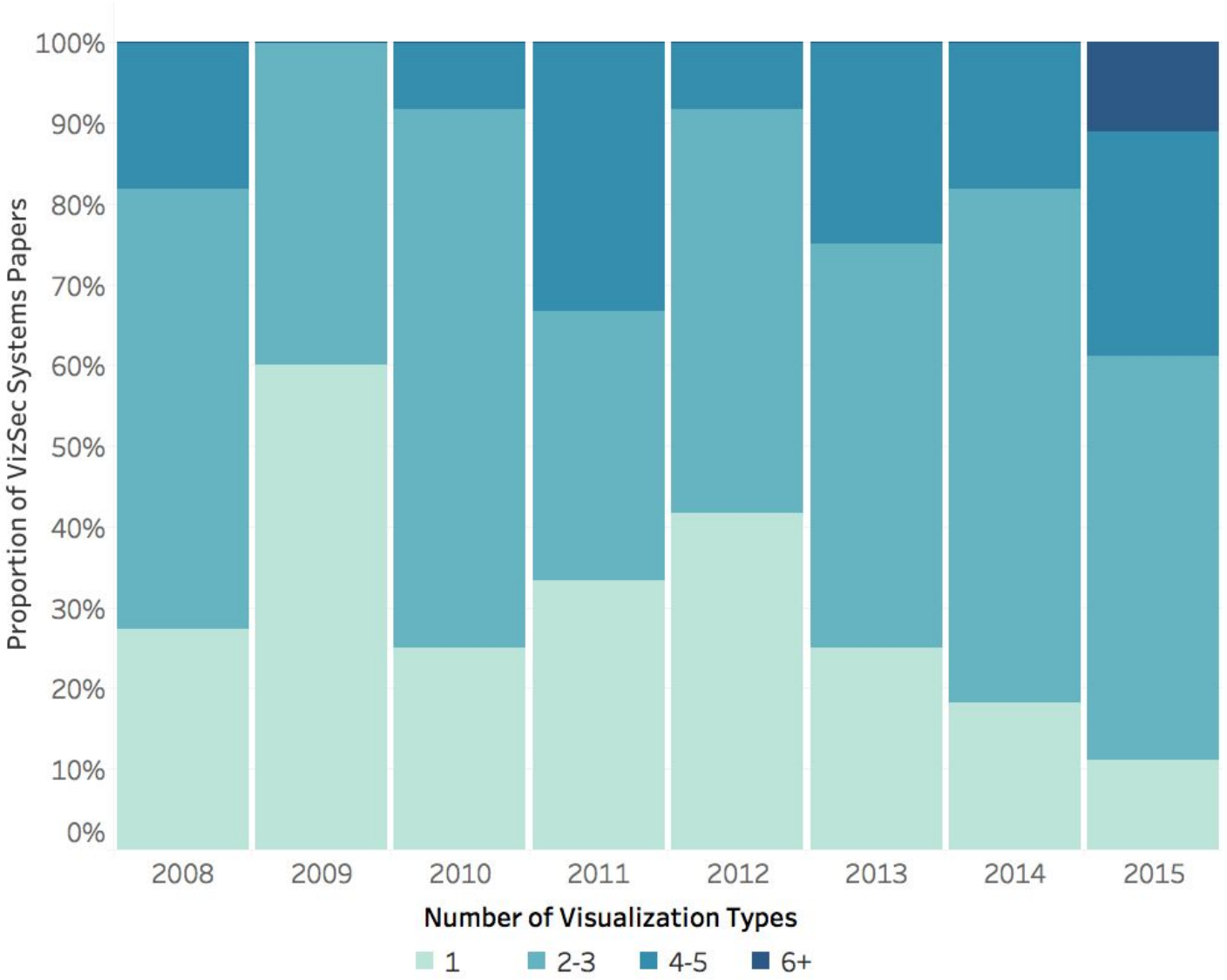
Table vs. Parallel Coordinates



Finding 2: visual metaphors over time

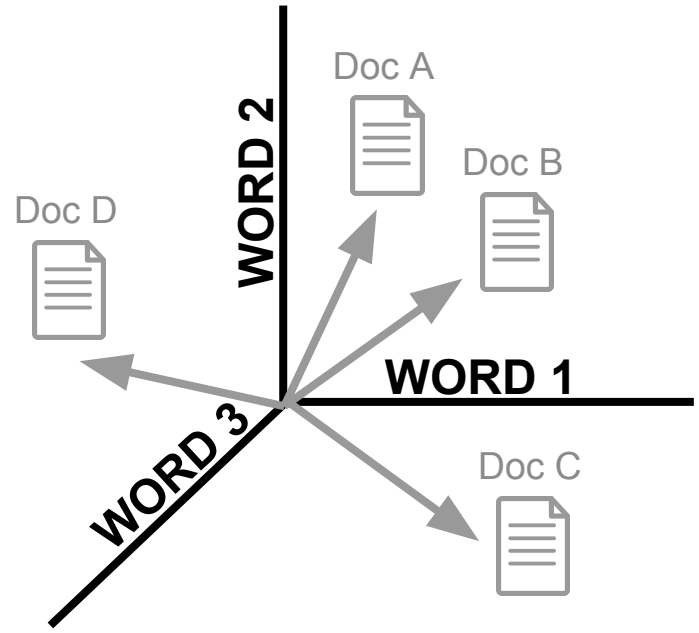


Finding 3: increasing use of multiple vis



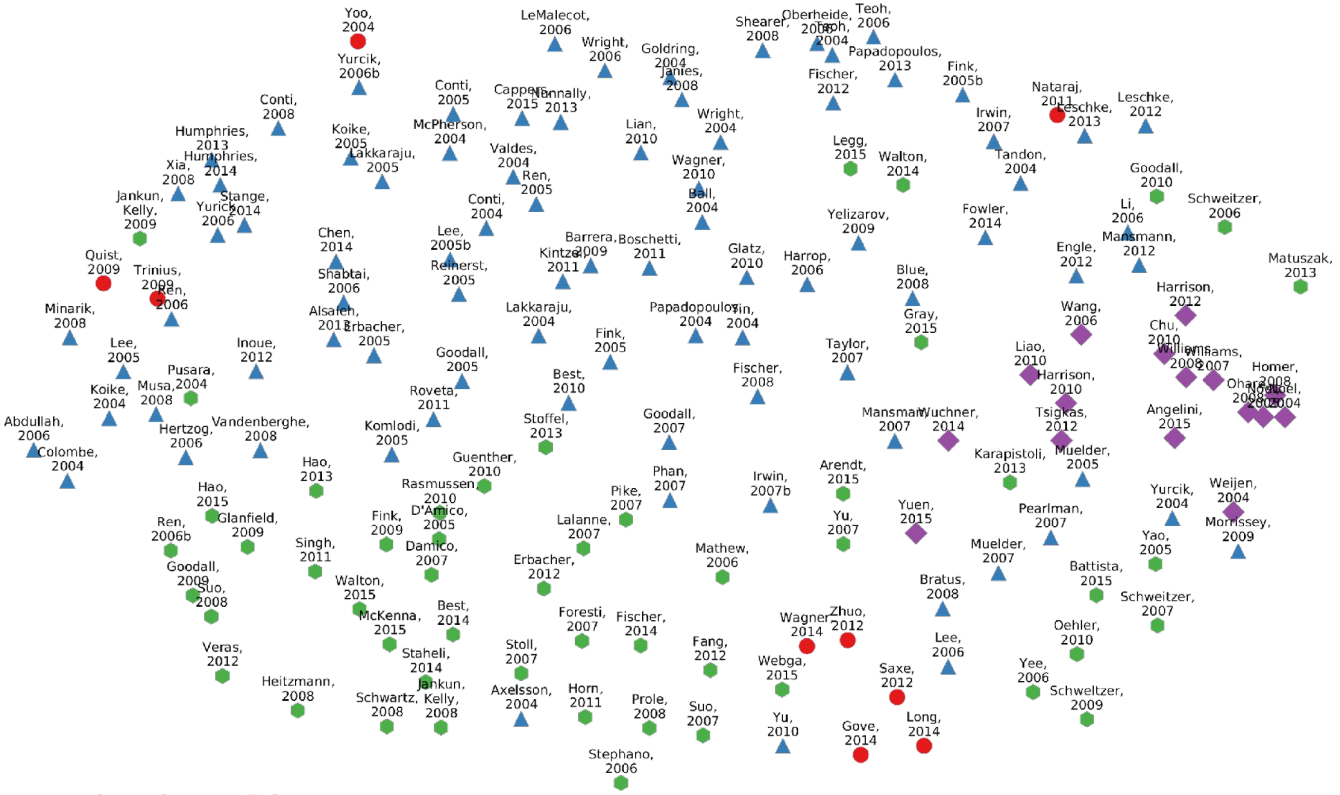
Methodology pt. 2: topic modeling

- Automated text mining:
  - Compute TF-IDF vector for each paper
  - Cluster TF-IDF vectors using k-means clustering to find latent topics



- Project TF-IDF vectors using multidimensional scaling

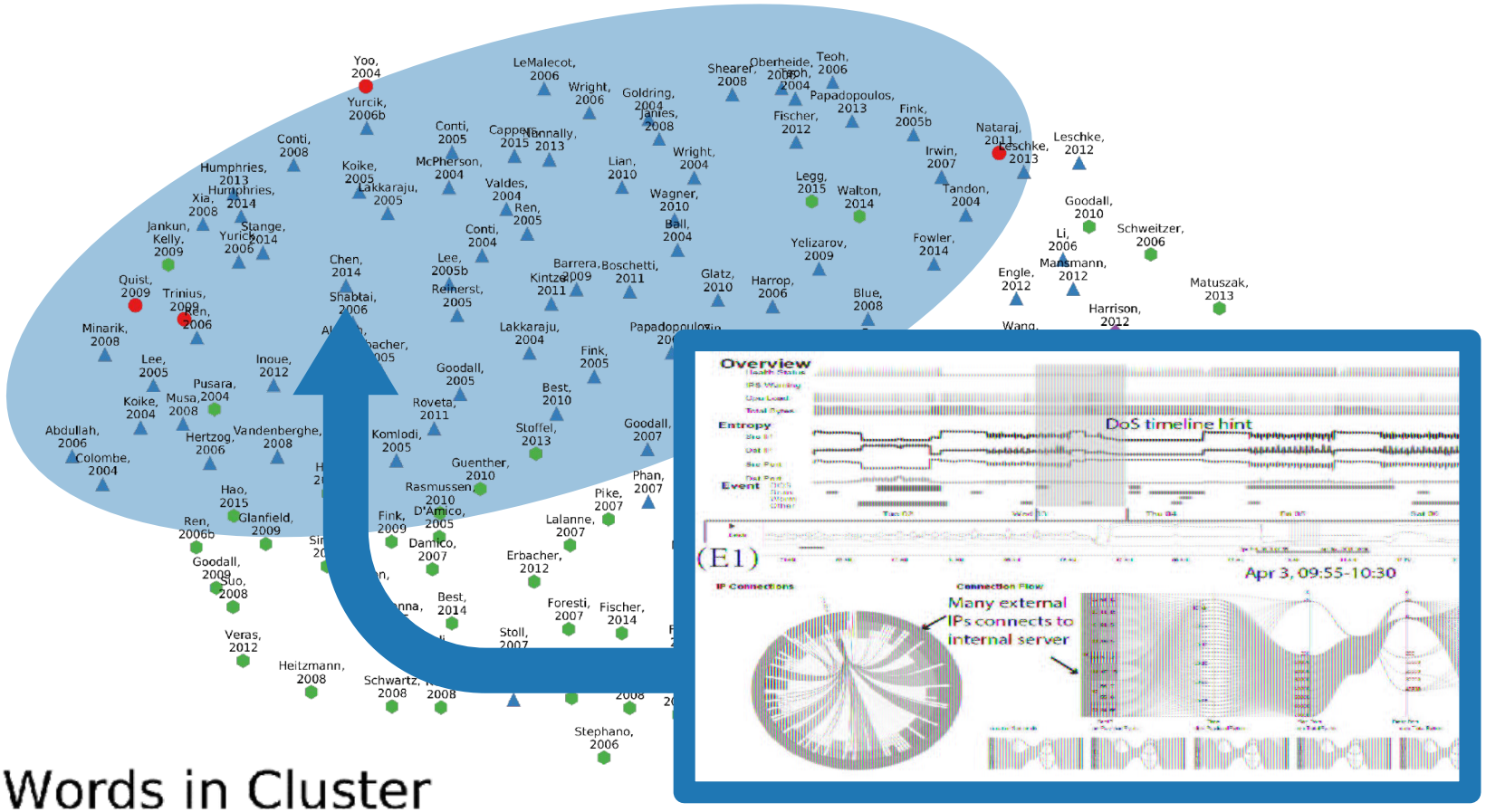
Finding 4: latent topics



Top 5 Words in Cluster

- ip, ported, hosts, traffic, packet
- analysts, task, models, cyber, alerts
- attacks, graph, node, vulnerabilities, exploit
- malware, sample, execution, imaging, virus

Finding 4: latent topics



OCEANS (Chen et al. 2014)

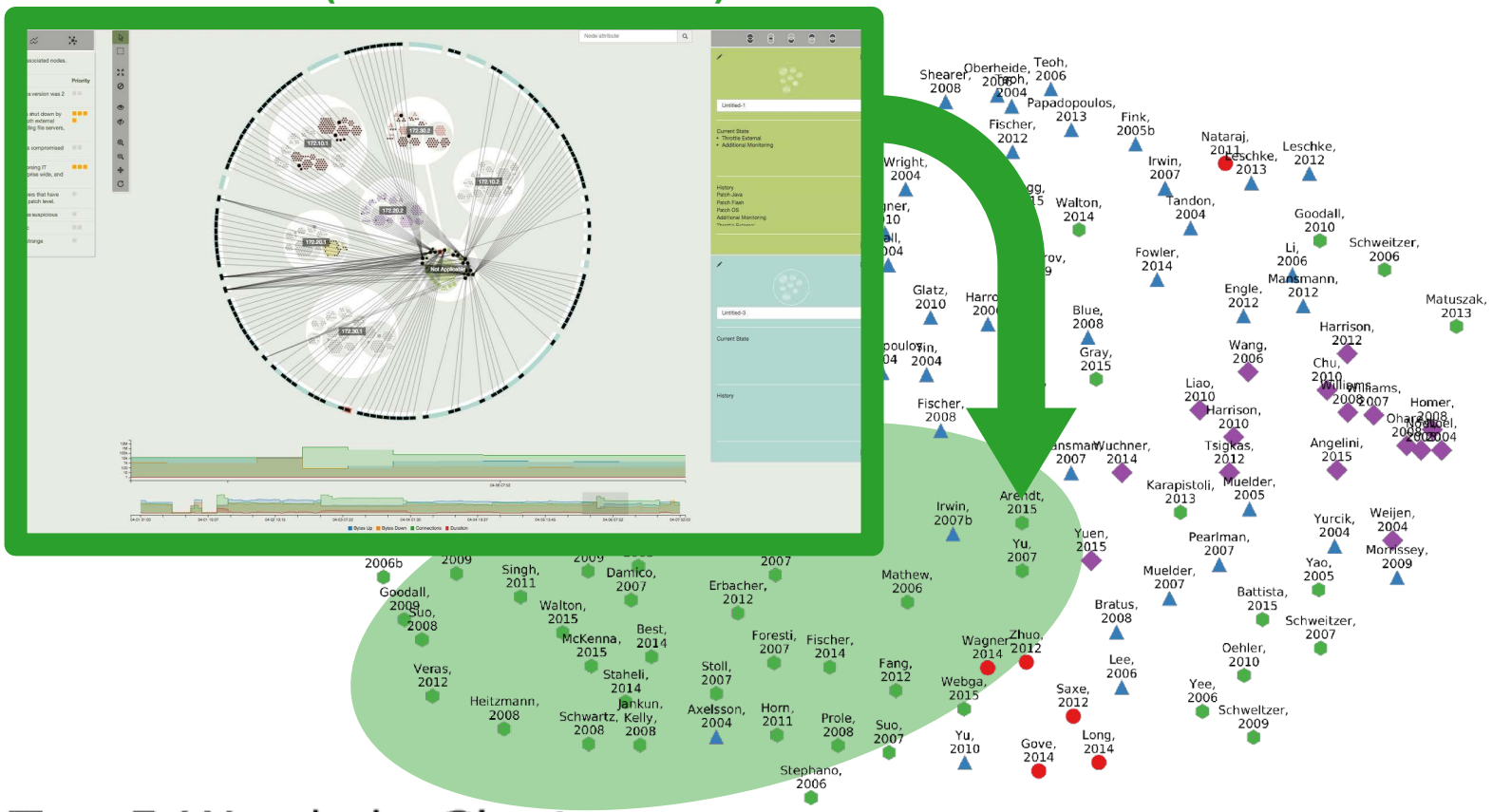
Top 5 Words in Cluster

- ip, ported, hosts, traffic, packet
- analysts, task, models, cyber, alerts
- attacks, graph, node, vulnerabilities, exploit
- malware, sample, execution, imaging, virus

“FORENSIC ANALYSIS”

Finding 4: latent topics

Ocelot (Arendt et al. 2015)



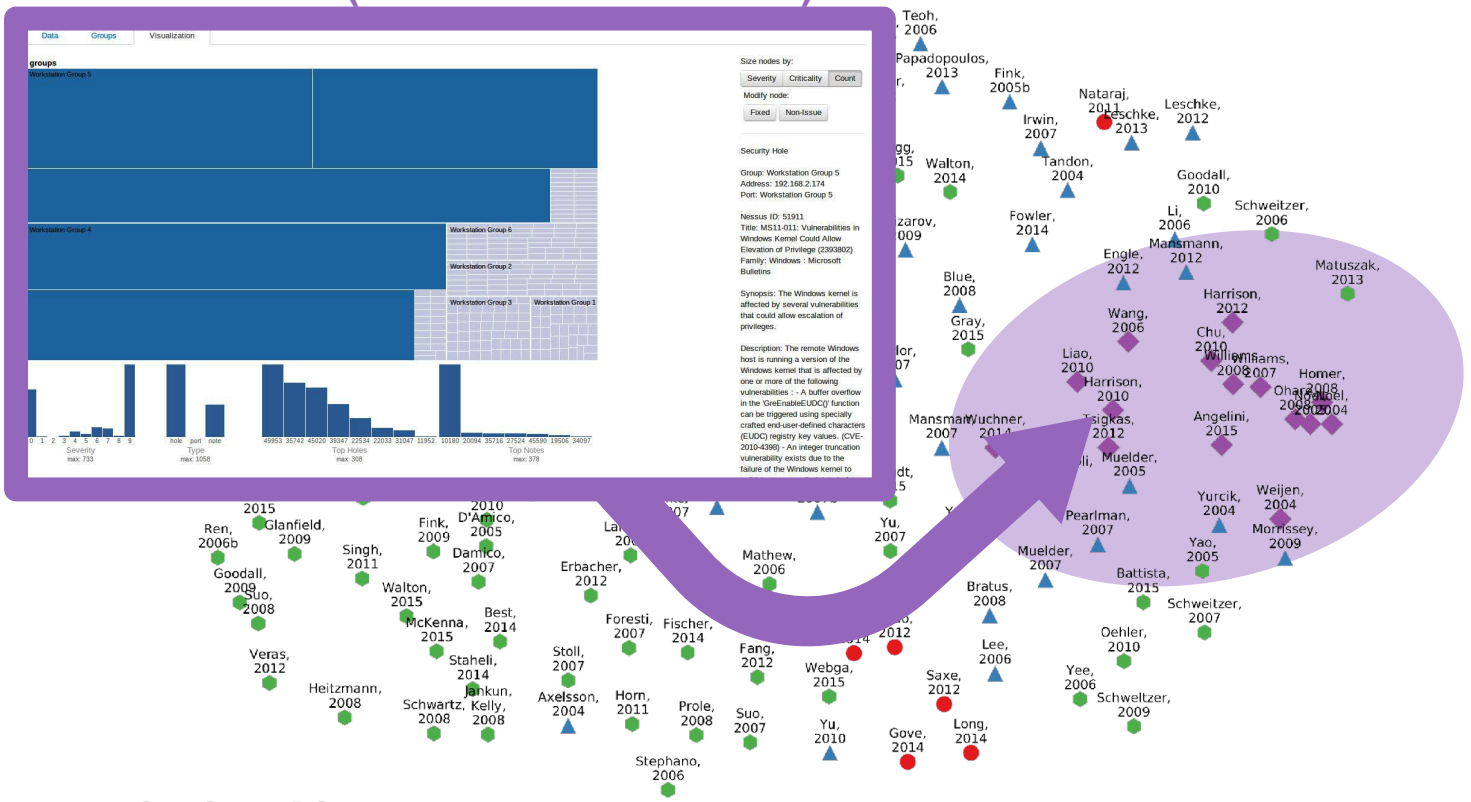
Top 5 Words in Cluster

- ip, ported, hosts, traffic, packet
- analysts, task, models, cyber, alerts
- attacks, graph, node, vulnerabilities, exploit
- malware, sample, execution, imaging, virus

“SITUATIONAL AWARENESS”

Finding 4: latent topics

NessusVIS (Harrison et al. 2012)



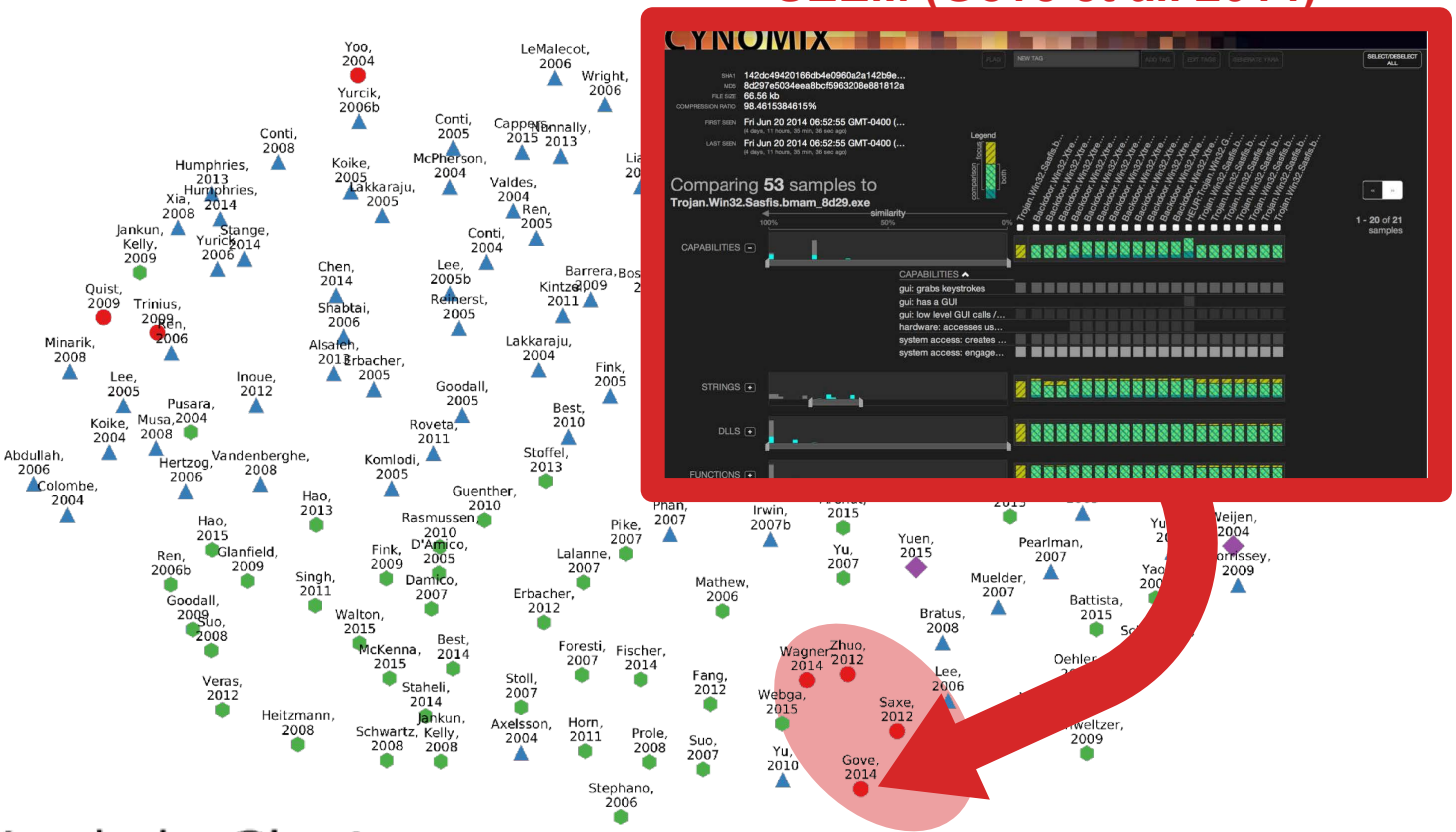
Top 5 Words in Cluster

- ip, ported, hosts, traffic, packet
- analysts, task, models, cyber, alerts
- attacks, graph, node, vulnerabilities, exploit
- malware, sample, execution, imaging, virus

“NETWORK DEFENSE”

Finding 4: latent topics

SEEM (Gove et al. 2014)

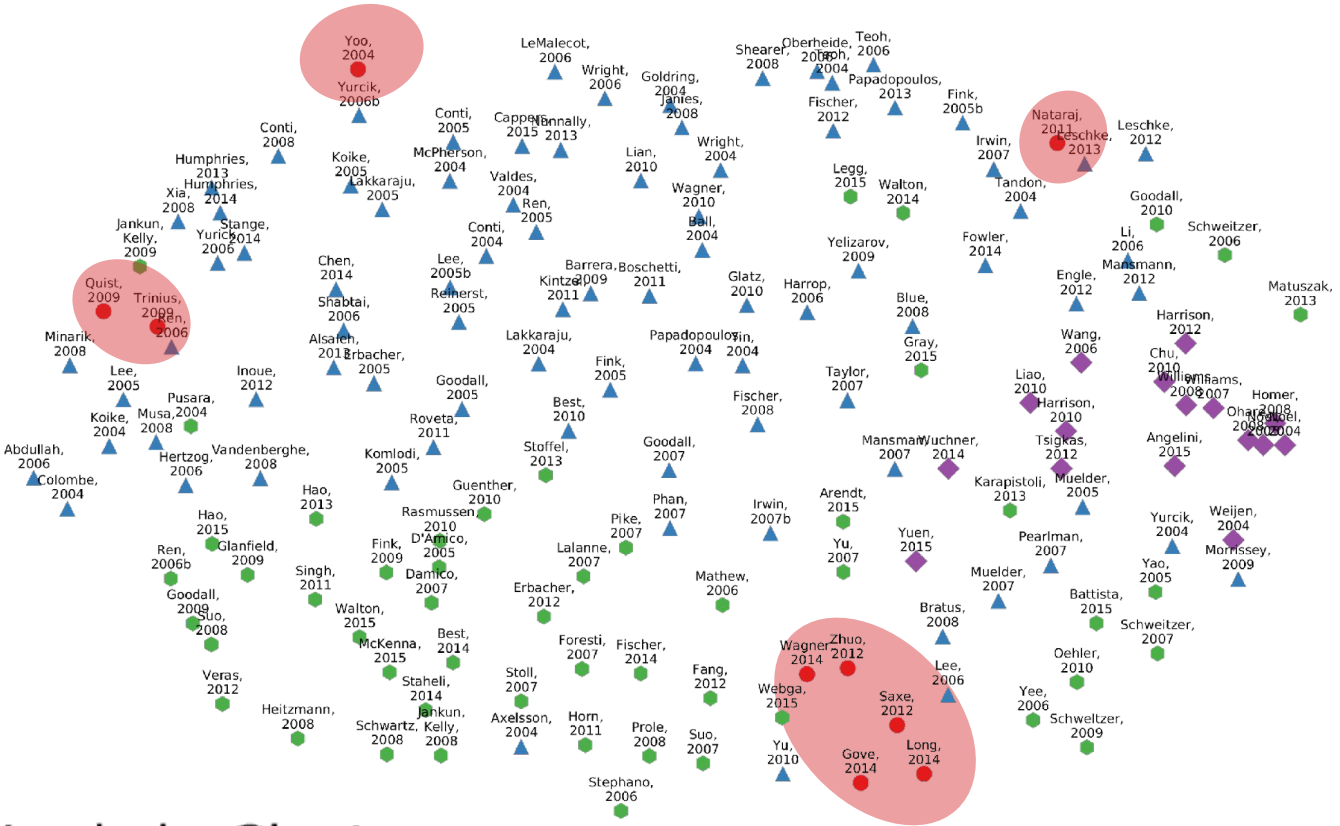


Top 5 Words in Cluster

- ip, ported, hosts, traffic, packet
- analysts, task, models, cyber, alerts
- attacks, graph, node, vulnerabilities, exploit
- malware, sample, execution, imaging, virus

“MALWARE ANALYSIS”

Finding 4: latent topics

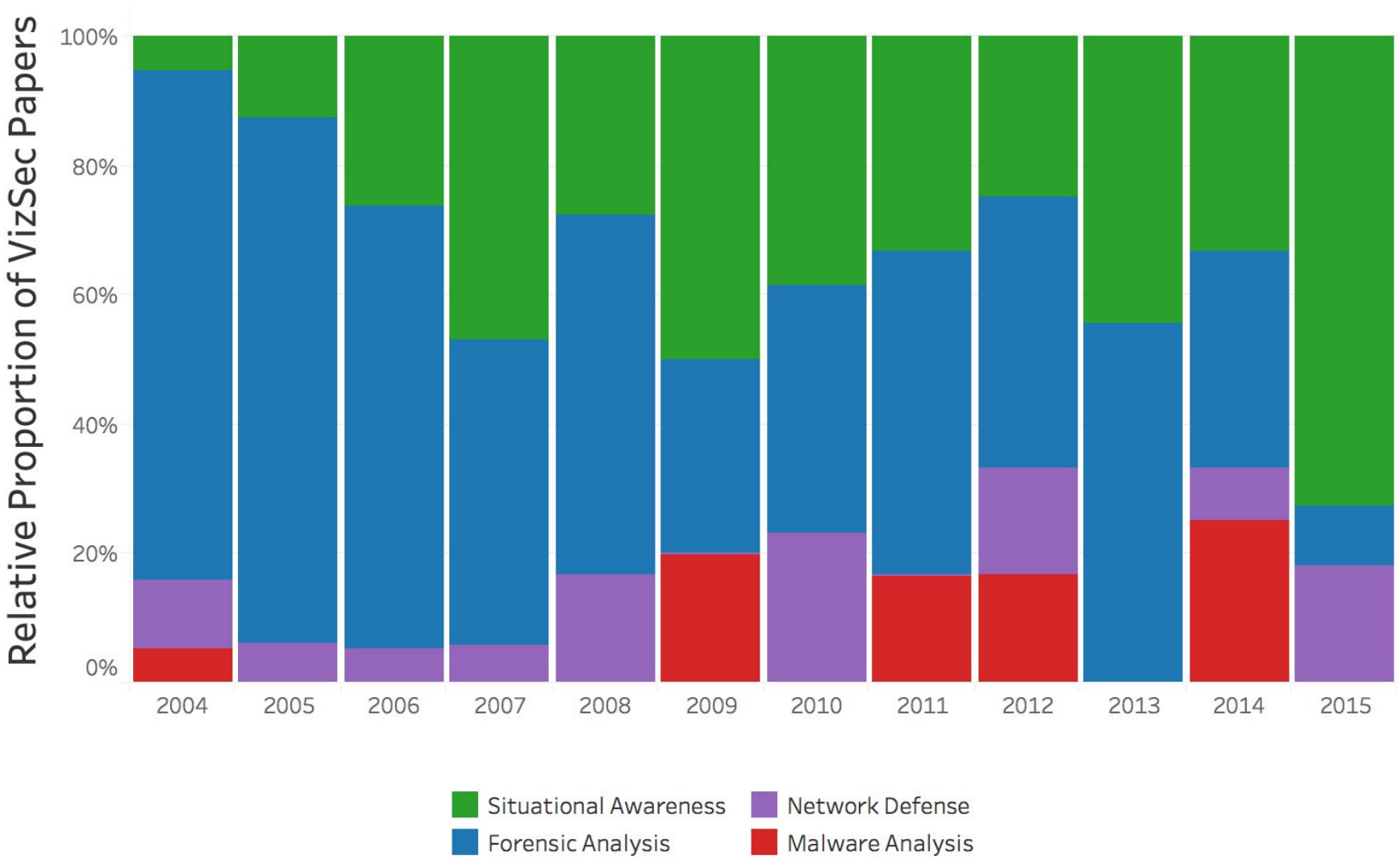


Top 5 Words in Cluster

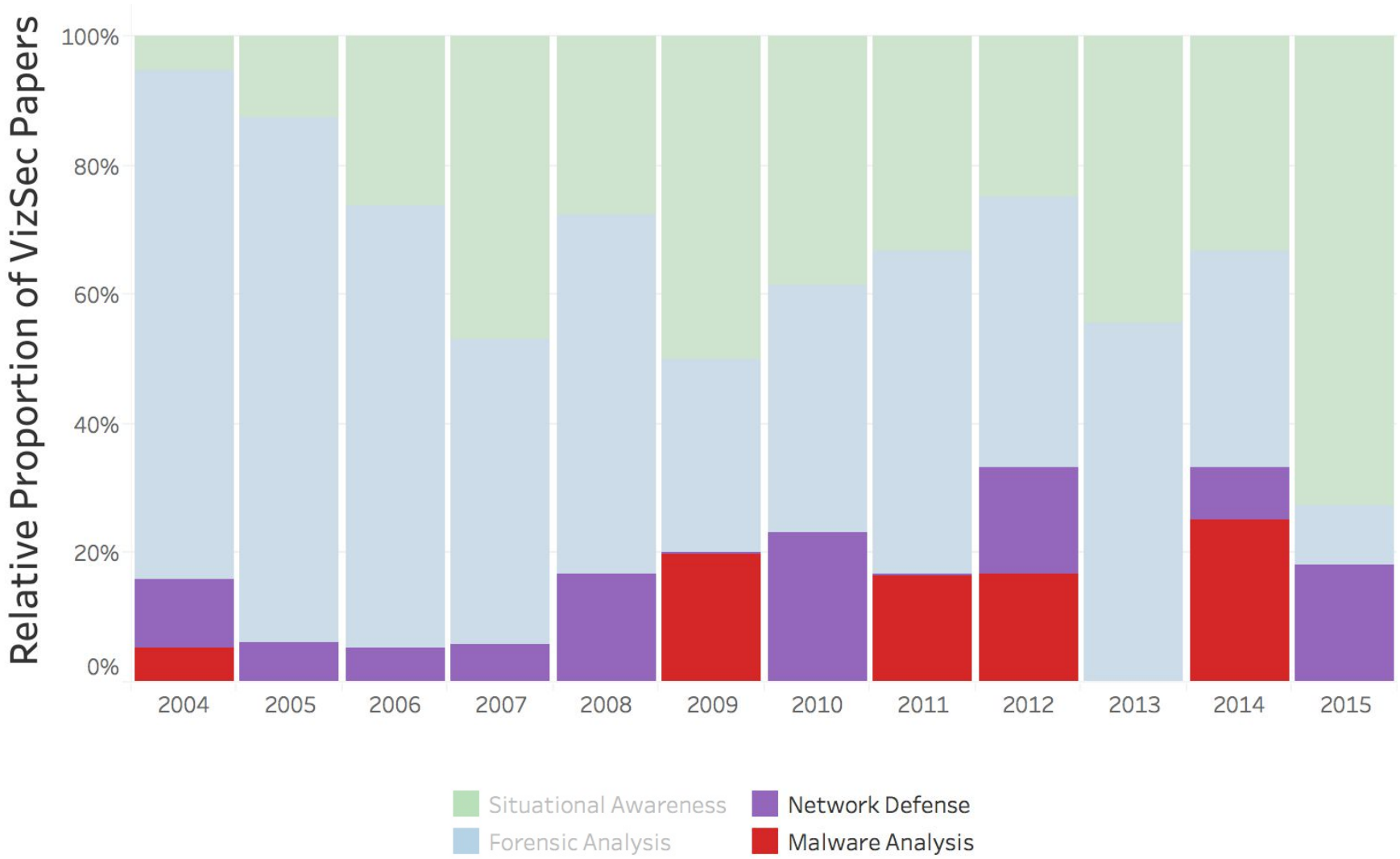
- ip, ported, hosts, traffic, packet
- analysts, task, models, cyber, alerts
- attacks, graph, node, vulnerabilities, exploit
- malware, sample, execution, imaging, virus

“MALWARE ANALYSIS”

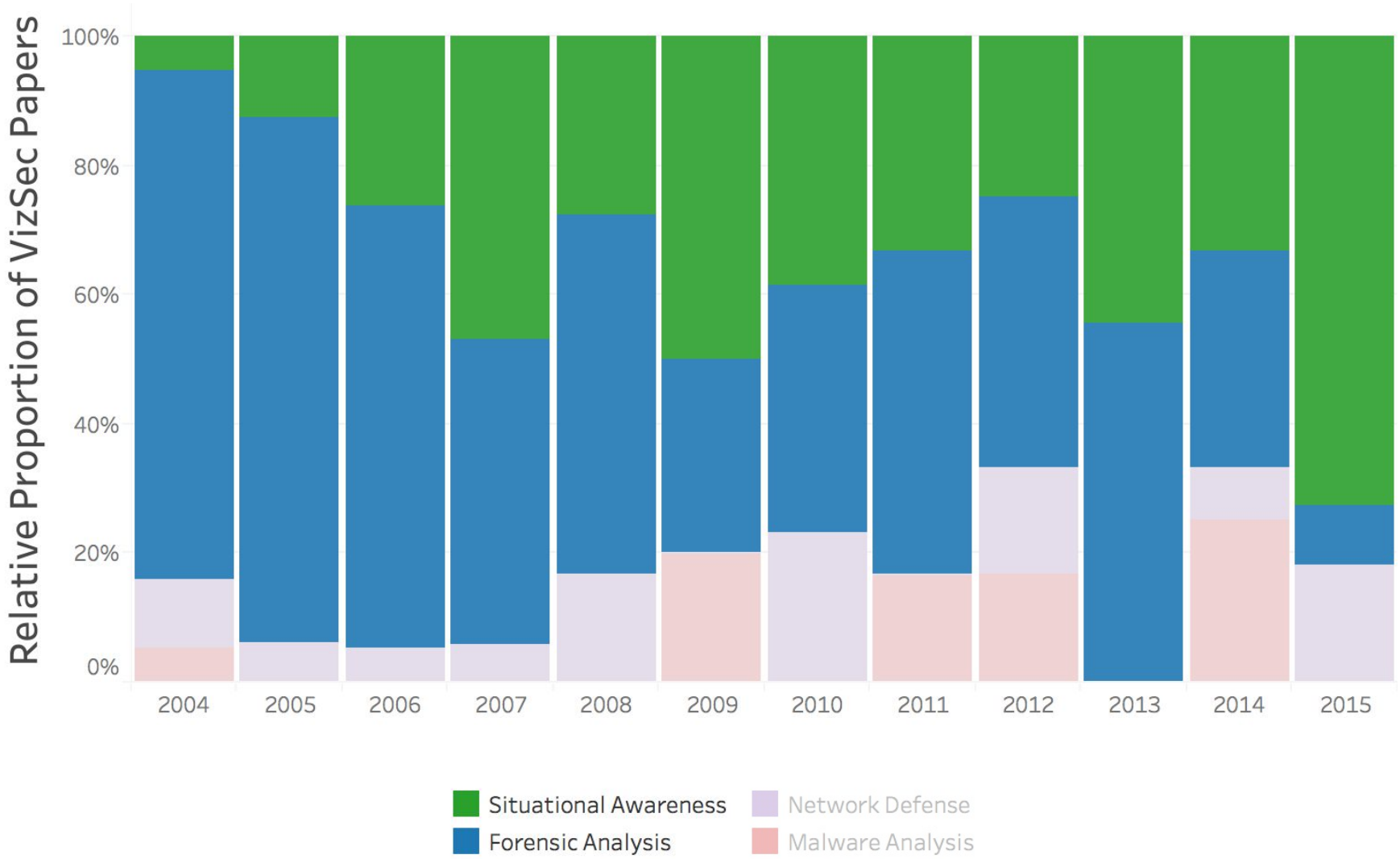
Finding 5: shift in focus over time



Finding 5: shift in focus over time



Finding 5: shift in focus over time



Finding 6: visual metaphors by cluster

	Forensic Analysis	Network Defense	Situational Awareness
Node-link Diagram	14	18	12
Table	8	8	8
Timeline	8	6	3
Bar Chart / Histogram	9	4	3
Line Graph	5	8	4
Matrix View	9	4	4
Parallel Coordinates	8	5	3
Treemap	2	6	4
Geographic	3	3	3
Scatterplot	3	4	1
Tag cloud	3	1	0

Finding 6: visual metaphors by cluster

	Forensic Analysis	Network Defense	Situational Awareness
Node-link Diagram	14	18	12
Table	8	8	8
Timeline	8	6	3
Bar Chart / Histogram	9	4	3
Line Graph	5	8	4
Matrix View	9	4	4
Parallel Coordinates	8	5	3
Treemap	2	6	4
Geographic	3	3	3
Scatterplot	3	4	1
Tag cloud	3	1	0

Finding 6: visual metaphors by cluster

	Forensic Analysis	Network Defense	Situational Awareness
Node-link Diagram	14	18	12
Table	8	8	8
Timeline	8	6	3
Bar Chart / Histogram	9	4	3
Line Graph	5	8	4
Matrix View	9	4	4
Parallel Coordinates	8	5	3
Treemap	2	6	4
Geographic	3	3	3
Scatterplot	3	4	1
Tag cloud	3	1	0

Conclusions and future work

- Identified 4 major subtopics under the VizSec umbrella
  - Different visual metaphors in use in different domains
  - Shift in focus from low- to high-level analytical support over time
- Node link and matrix views are heavily utilized views (esp. in **forensic analysis** and **network defense)**
  - Graph visualization is important, are these the most effective?
  - Opportunities to collaborate with network sciences, etc.
- **Next steps**: platforms to evaluate/disseminate VIS
  - Practitioners lack mechanisms to easily test different visualizations in the context of their analysis
  - Research lack access to domain-specific knowledge necessary to successfully move prototypes from lab to practice

Thank you

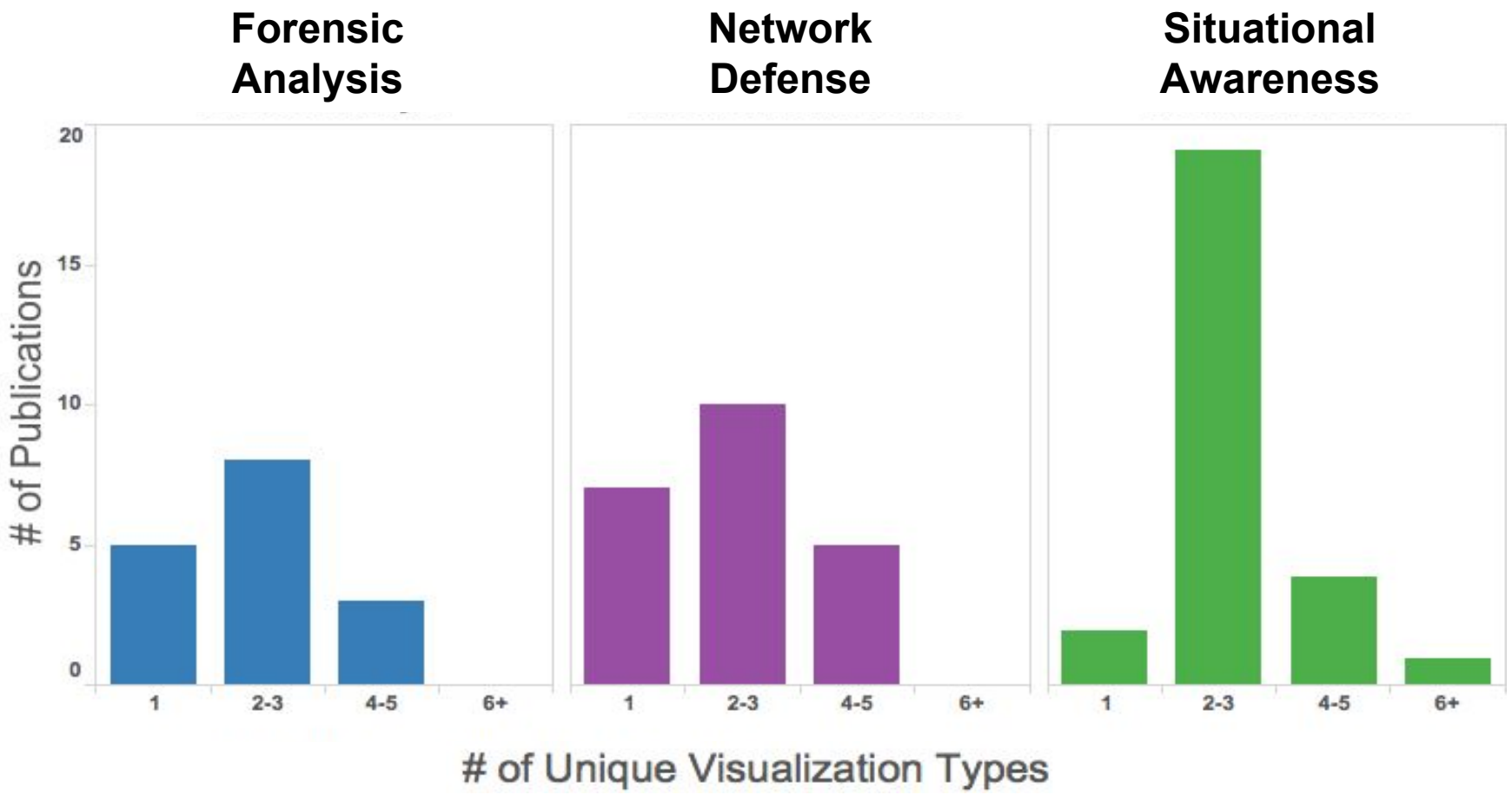
(efukuda) | (ssridhar) | (jcrouser)
@smith.edu

Laboratory for Analytic Sciences
Reflect. Observe. Imagine.

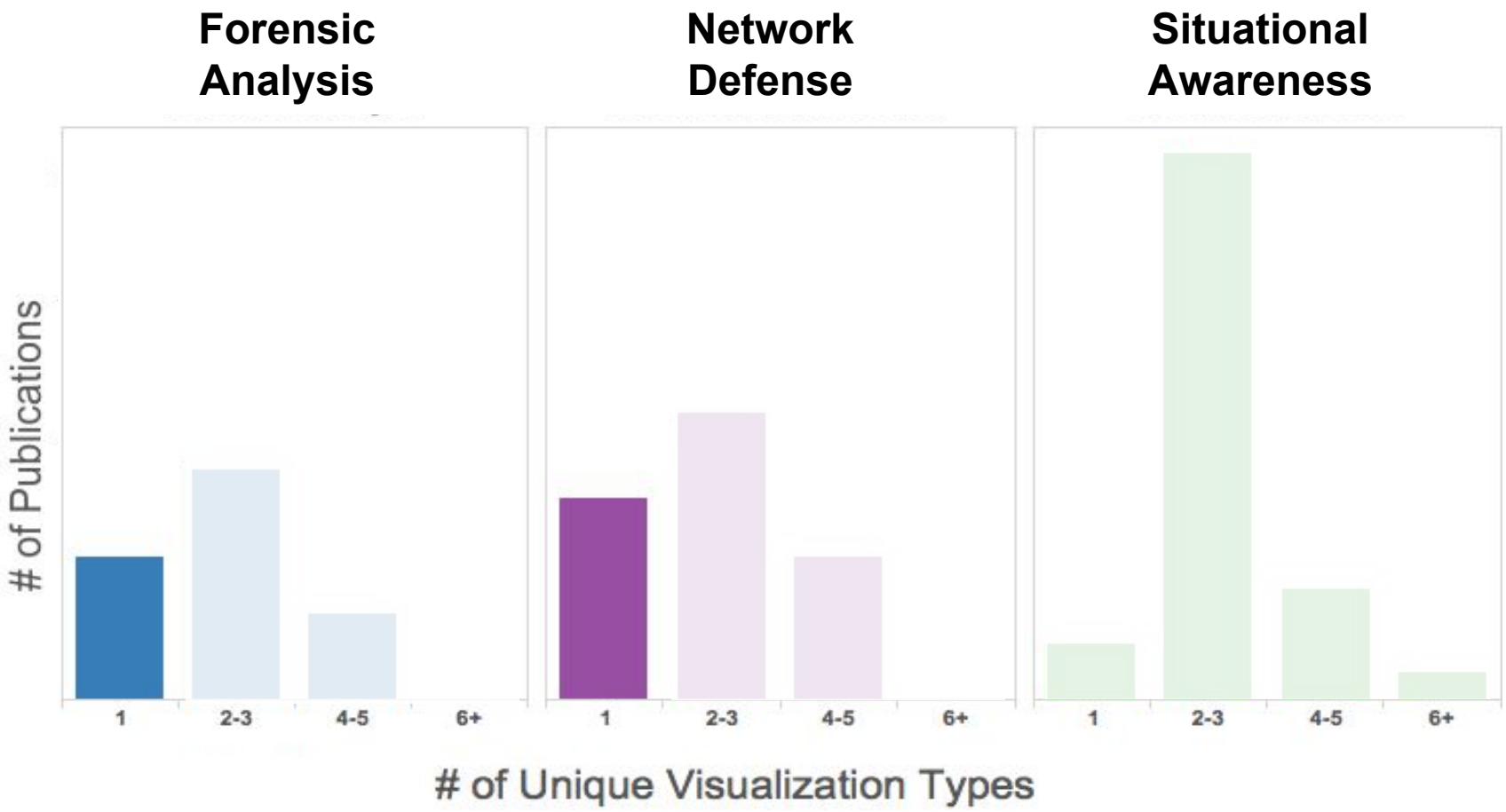
This paper is based upon work supported in part with funding from the Laboratory for Analytical Sciences (LAS). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the LAS and/or any agency or entity of the United States Government.

Additional Slides

Finding 7: use of coordinated views



Finding 7: use of coordinated views



Finding 7: use of coordinated views