TEXAS A&M UNIVERSITY
Engineering

Tackling Cybersecurity Using Graph Mining

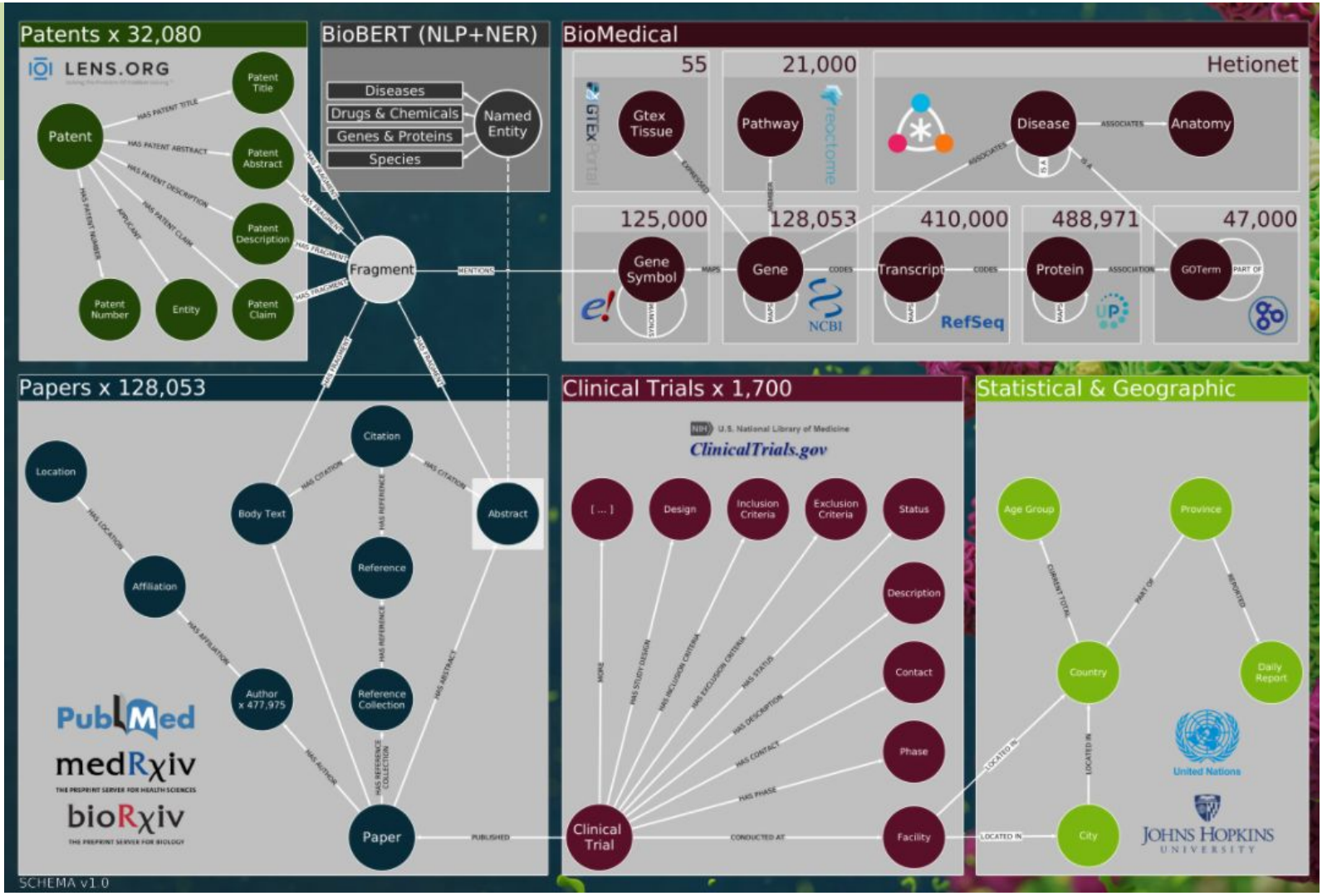
Khanh Nguyen
Assistant Professor
Department of Computer Science and Engineering

Big Graphs are Ubiquitous

TEXAS A&M UNIVERSITY
Engineering

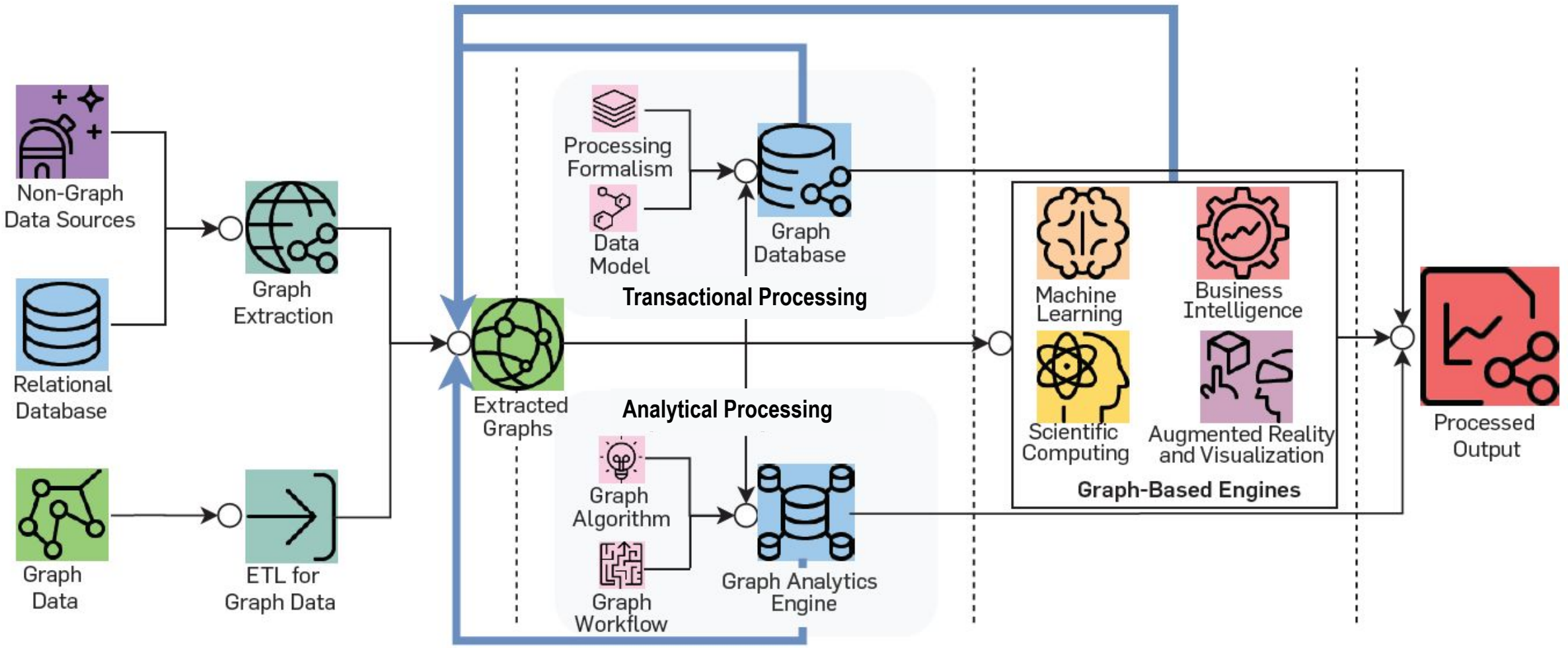
COVID GRAPH

covidgraph.org



The Future Is Big Graphs

TEXAS A&M UNIVERSITY
Engineering

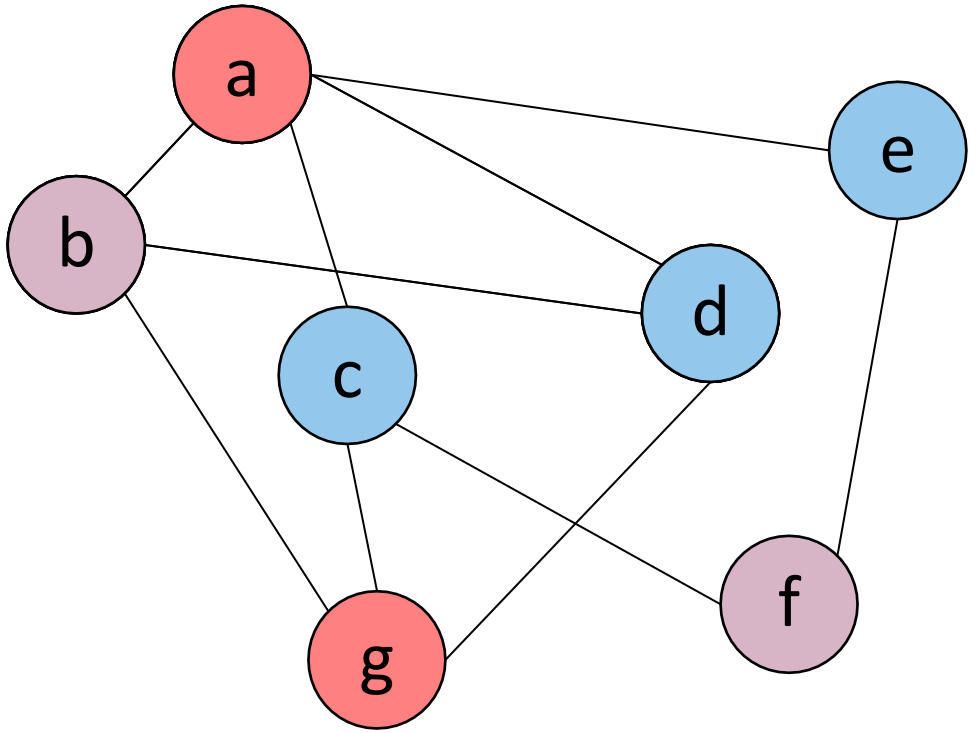


DOI:10.1145/3434642: Communication of ACM 9/2021

TEXAS A&M UNIVERSITY
Engineering

THEME:
GRAPH MINING & CYBERSECURITY

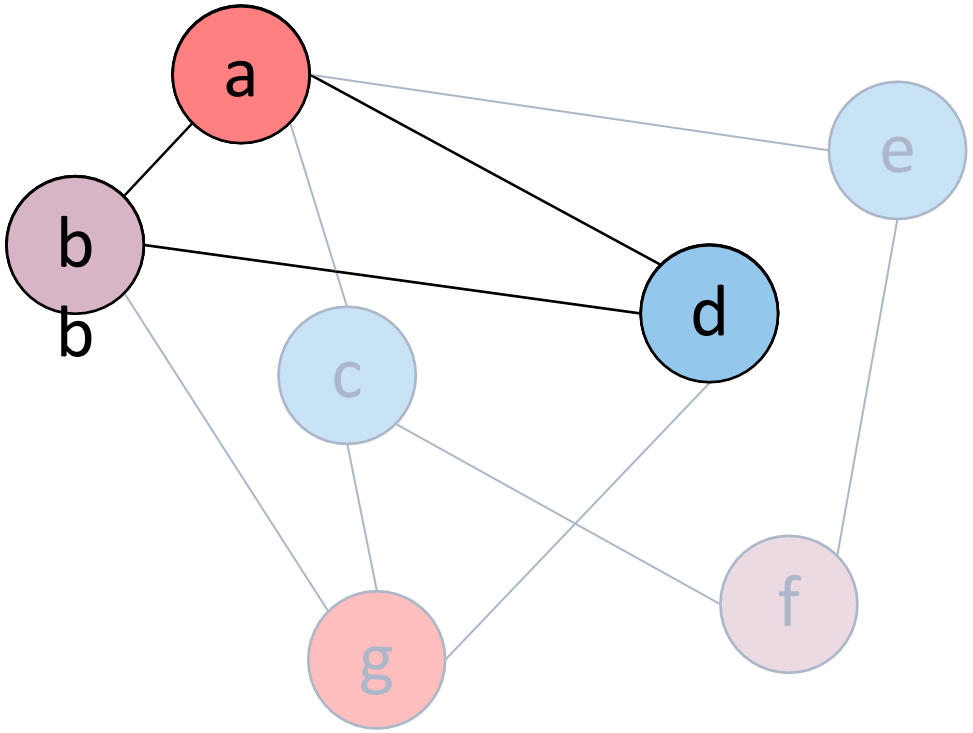
Graph Mining 101



Data Graph

Graph Mining 101

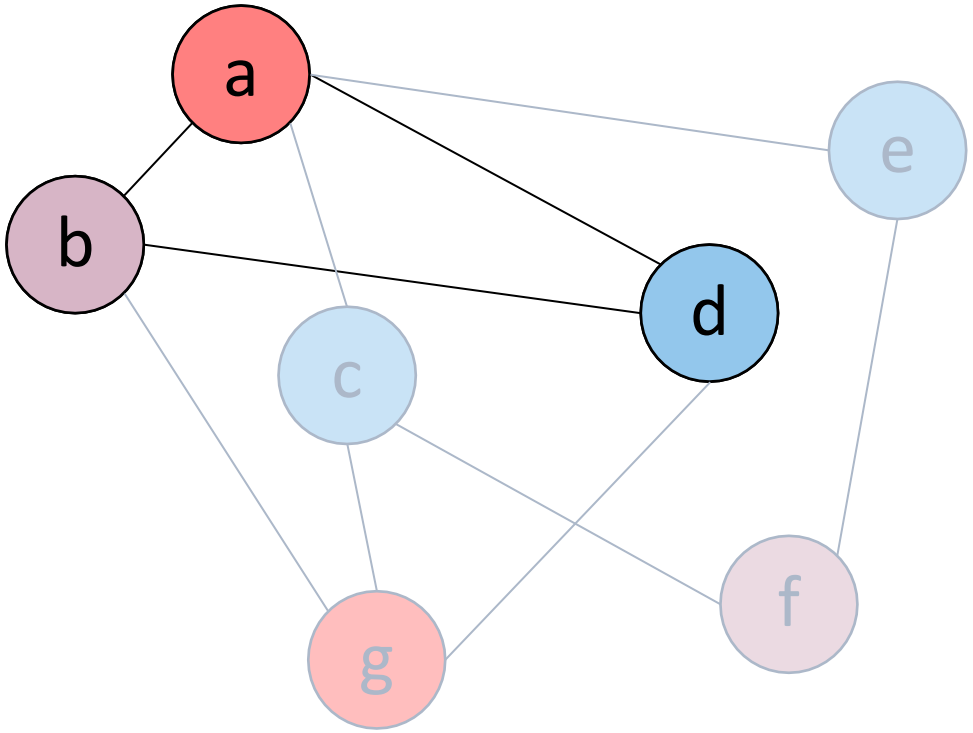
TEXAS A&M UNIVERSITY
Engineering



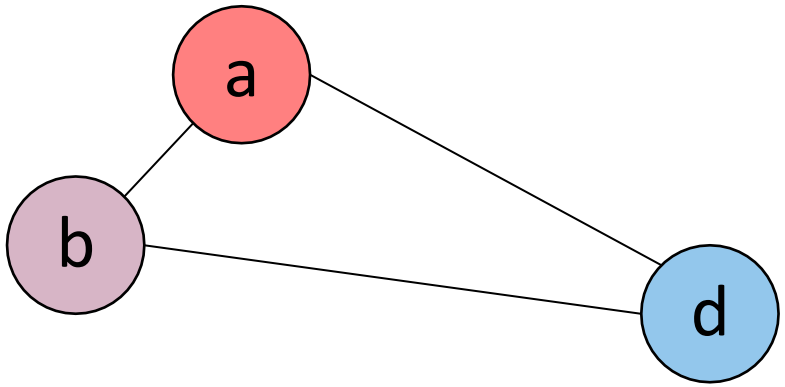
Data Graph

Graph Mining 101

TEXAS A&M UNIVERSITY
Engineering



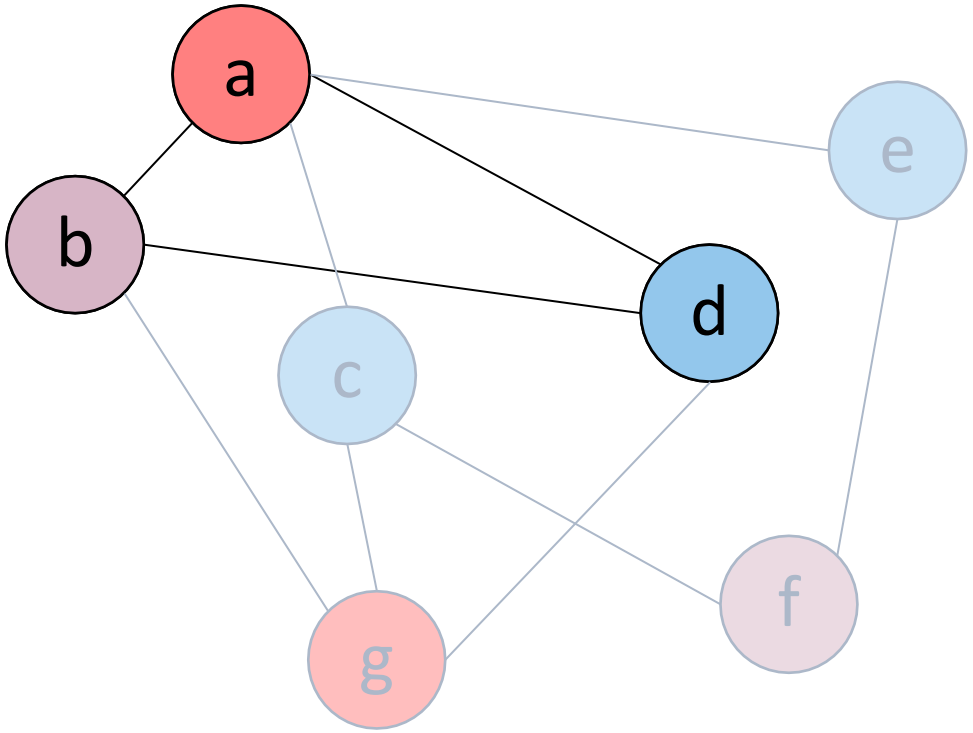
Data Graph



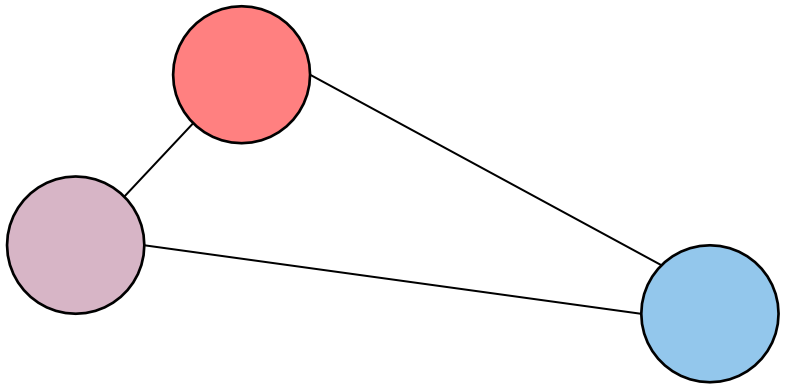
Subgraph

Graph Mining 101

TEXAS A&M UNIVERSITY
Engineering

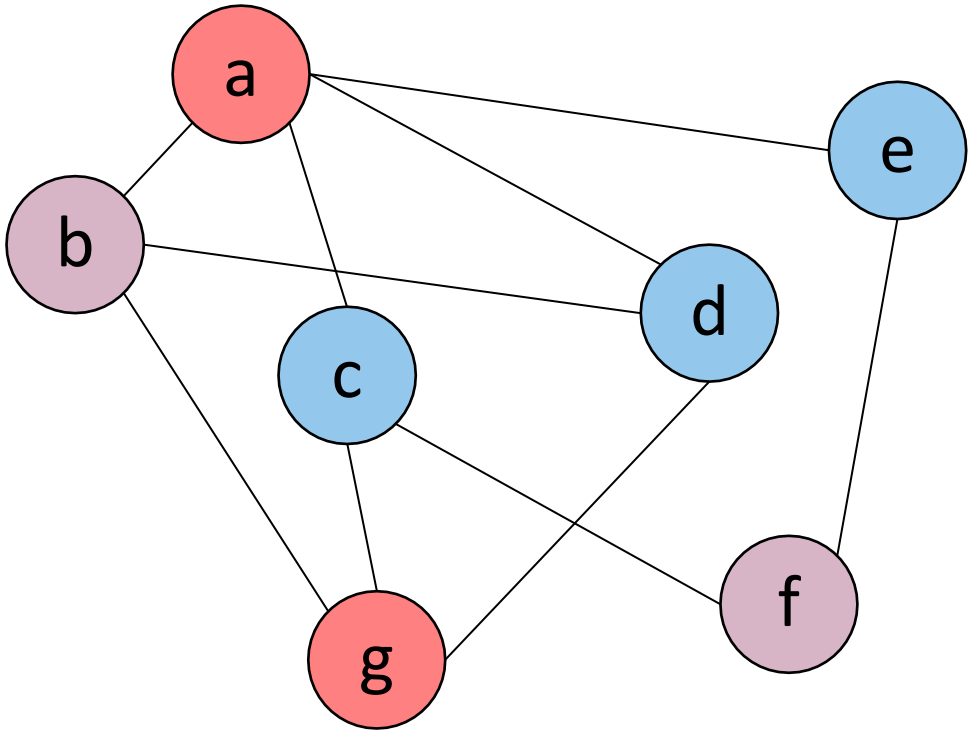


Data Graph

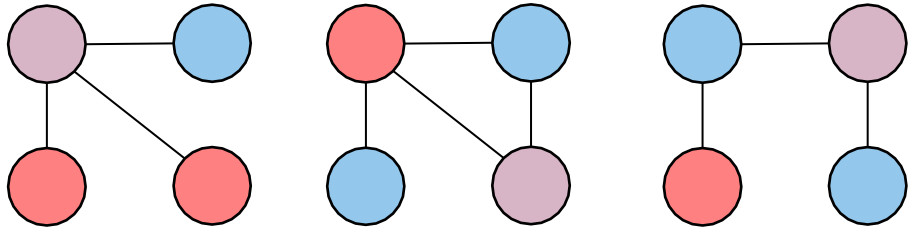


Pattern

FSM: Frequent Subgraph Mining



Data Graph



Frequent Patterns

Research Question #1

TEXAS A&M UNIVERSITY
Engineering

Benign network

Under-attack network

Discriminative pattern?

Research Question #2

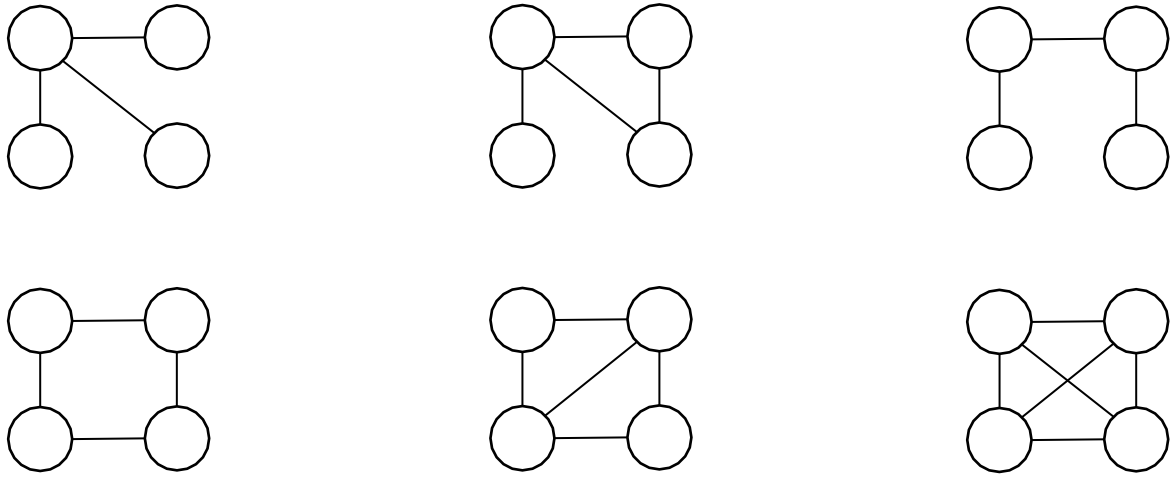
TEXAS A&M UNIVERSITY
Engineering

Statistical distribution of patterns?
(e.g., motif size 2,3,4;
FSM size 3,4,5; etc.)

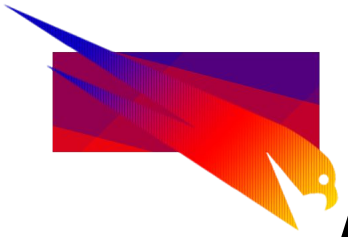
Processing Engines

Scalability Challenge

- 4-motif counting on Orkut graph (|V| = 13M, |E| = 117M)



123,503,340,341,270 subgraphs



PEREGRINE

Jamshidi et. al., EUROSYS'20

A Pattern-Aware Graph Mining System

- ✔ executes 700x faster
- ✔ consumes 100x less memory
- ✔ scales to 100x larger datasets

Unfortunately:

- ○ does not support (multiple) edge labels
- ○ performance suffers with multiple vertex labels: 10x memory overhead and 30x time overhead

https://github.com/pdclab/peregrine

Conclusion

TEXAS A&M UNIVERSITY
Engineering

- Using a state-of-the-art graph mining system to investigate the role of graph patterns in cybersecurity:
  - Online detection
  - Analysis
- Work in progress

khanhtn@tamu.edu