

Finding Patterns in Complex Social Networks

Lennon Ganz
Santa Barbara City College
Computer Science



Mentor: Dr. Yinghui Wu
Faculty Advisor: Dr. Xifeng Yan
Department: Computer Science



Funded by: National Science Foundation

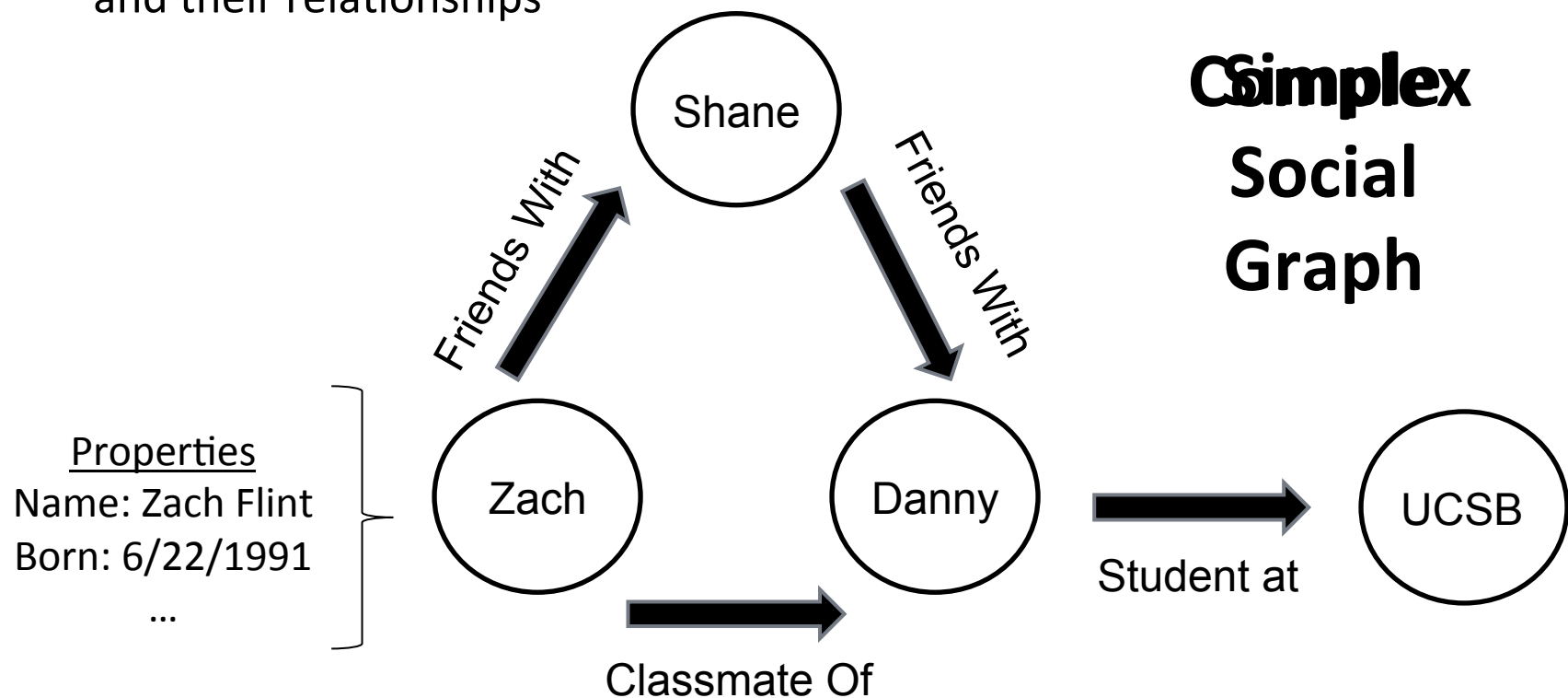
Social Network

“A network of social interactions”

Simple – Contains one type of entity (usually people) and their relationships

Complex – Contains multiple types of entities (people, places, events) and their relationships

**Complex
Social
Graph**



Graph Analytics Systems

Challenges

- Networks are large and complex
- Queries (questions) are difficult to write

My Objective

- Design a simple, intuitive platform to analyze complex networks



Designing a Graph Analysis Platform

Approach

- Understand existing graph databases and systems
- Identify suitable queries and algorithms for social graphs
- Propose and research optimization techniques



Titan

Question

Query -
a question



Algorithm - a set of
instructions for
answering a query

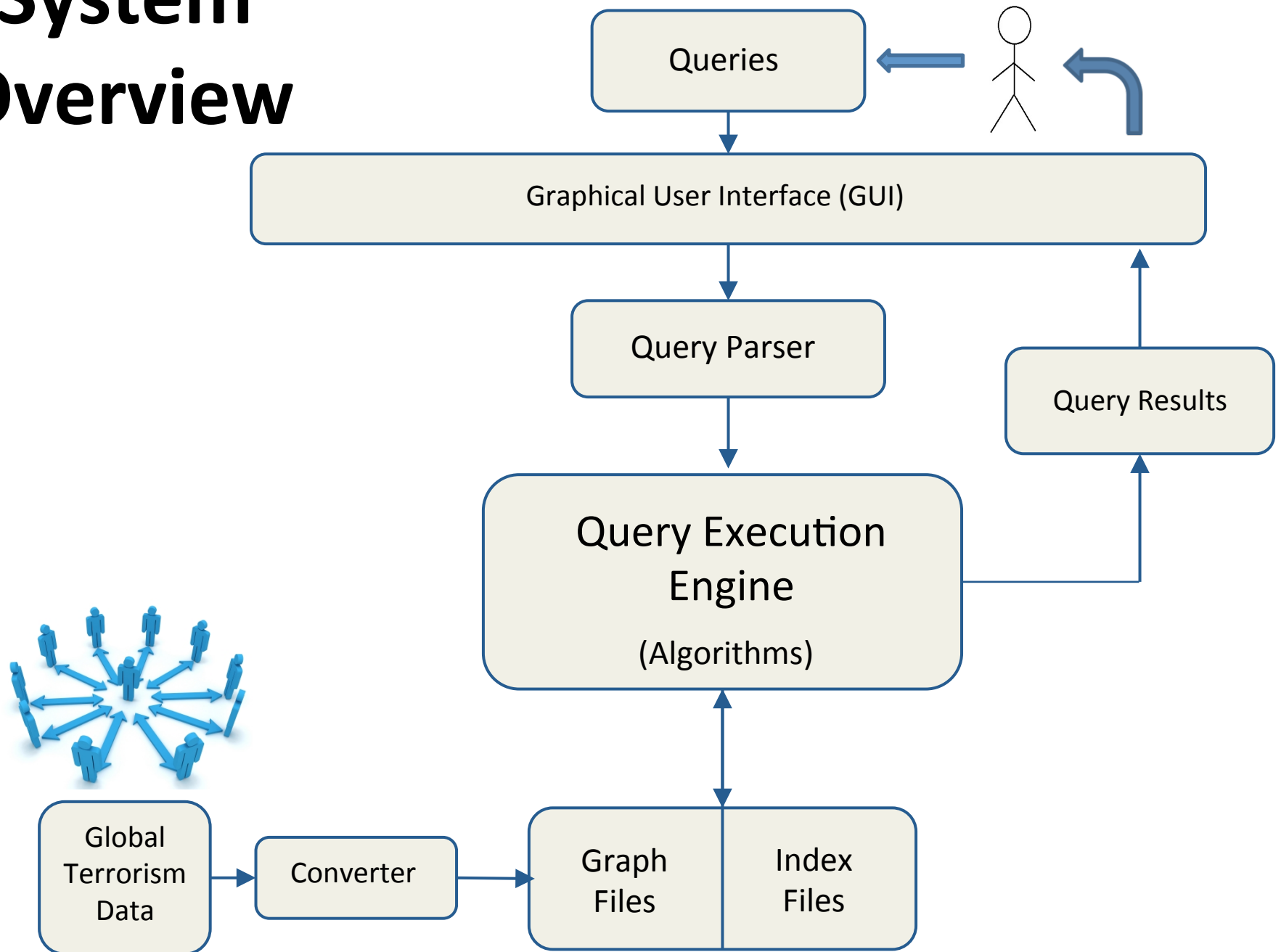
Answer

Types of Queries

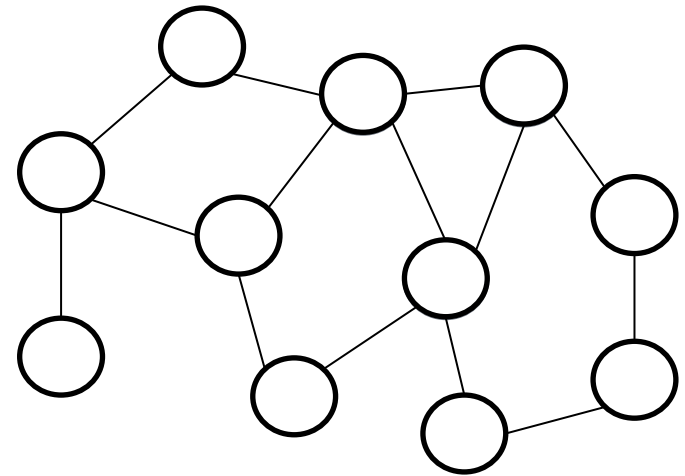
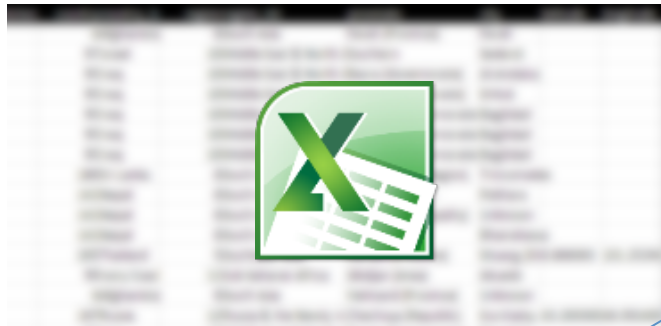
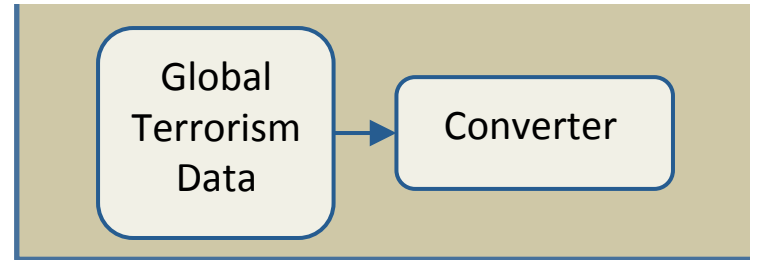


Type	Example
Aggregation	<ul style="list-style-type: none">• “How many attacks were there in 2005?”• “What was the most common type of weapon used?”
Reachability	<ul style="list-style-type: none">• Have these two terrorist groups ever collaborated on an attack?
Pattern Matching	<ul style="list-style-type: none">• “Find instances where a terrorist group targeted a government official who had also been attacked by a different group earlier that year”

System Overview



Building a Graph



Converter

Read data from cells

Add data to new or existing nodes in graph

Connect!

```
curGroup = cell.getStringCellValue();
```

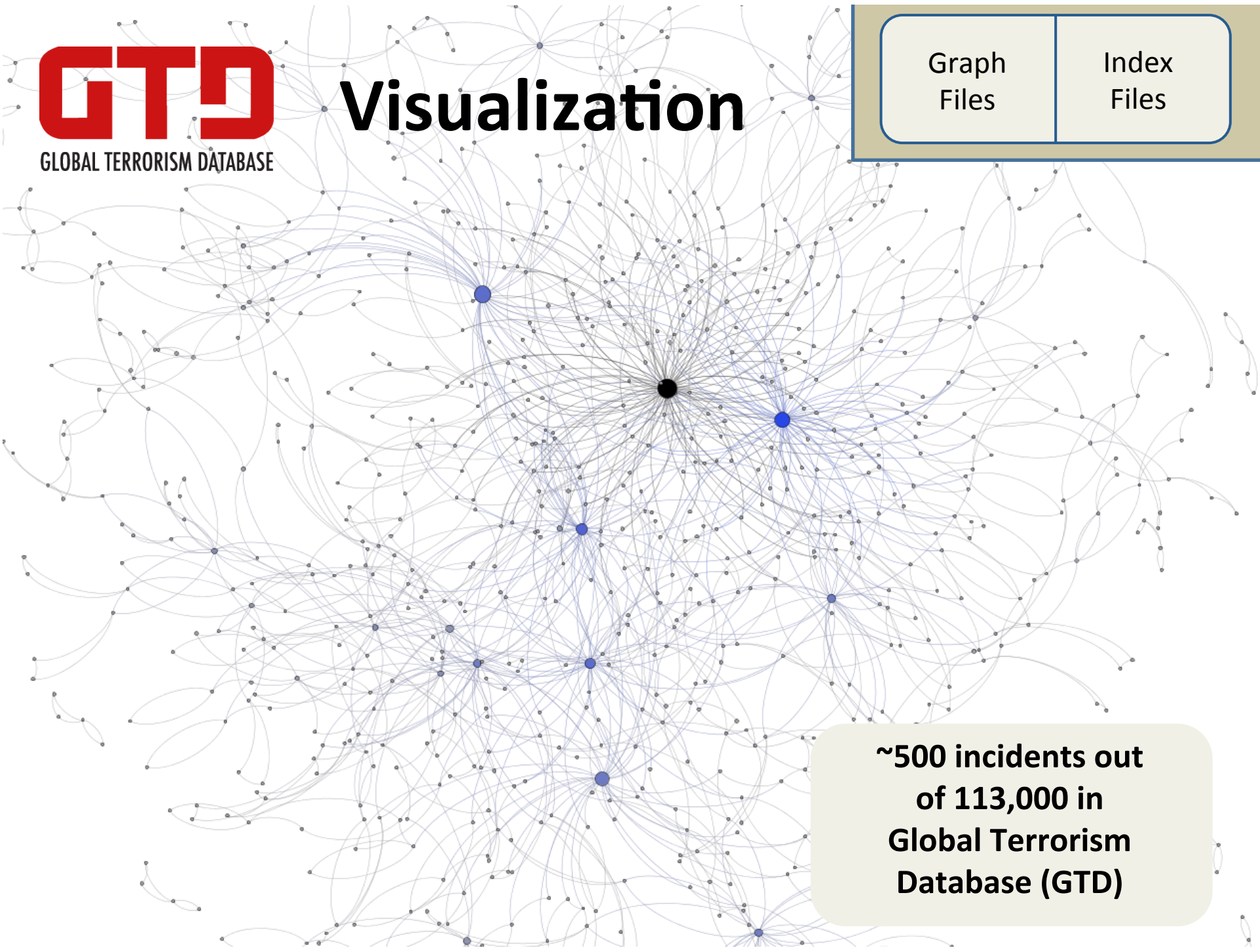
```
v2.addProperty("GROUP_NAME", curGroup);
```

```
e = graph.addEdge(v1, v2);  
e.addLabel("PERPETRATED_BY");
```

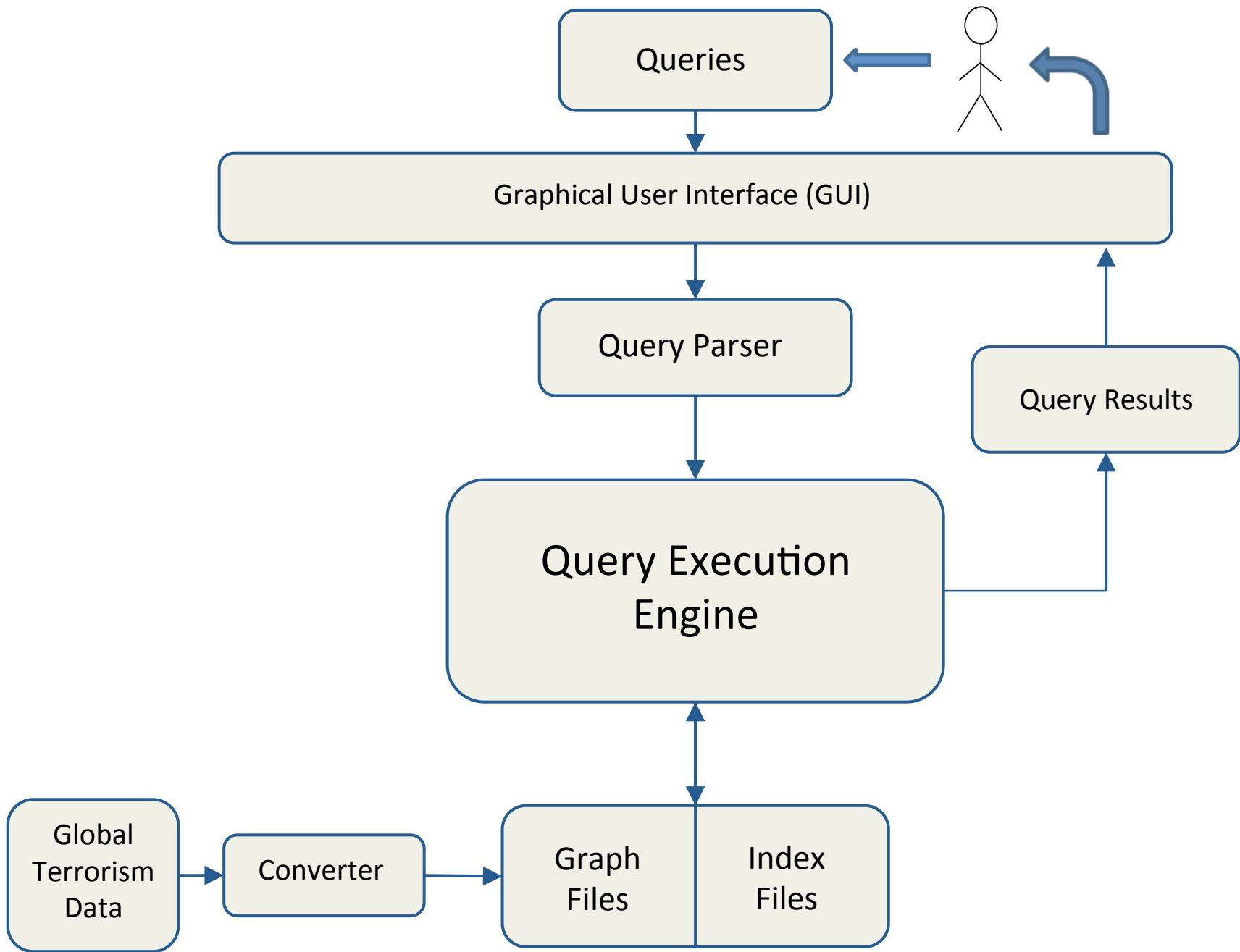


Visualization

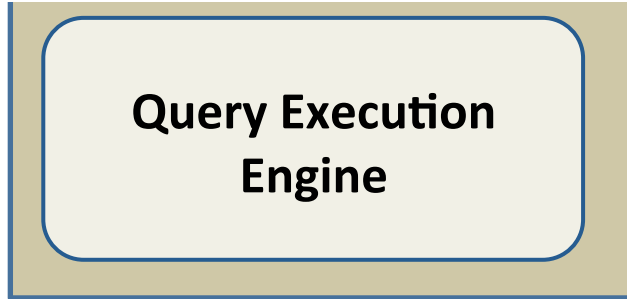
Graph Files	Index Files
-------------	-------------



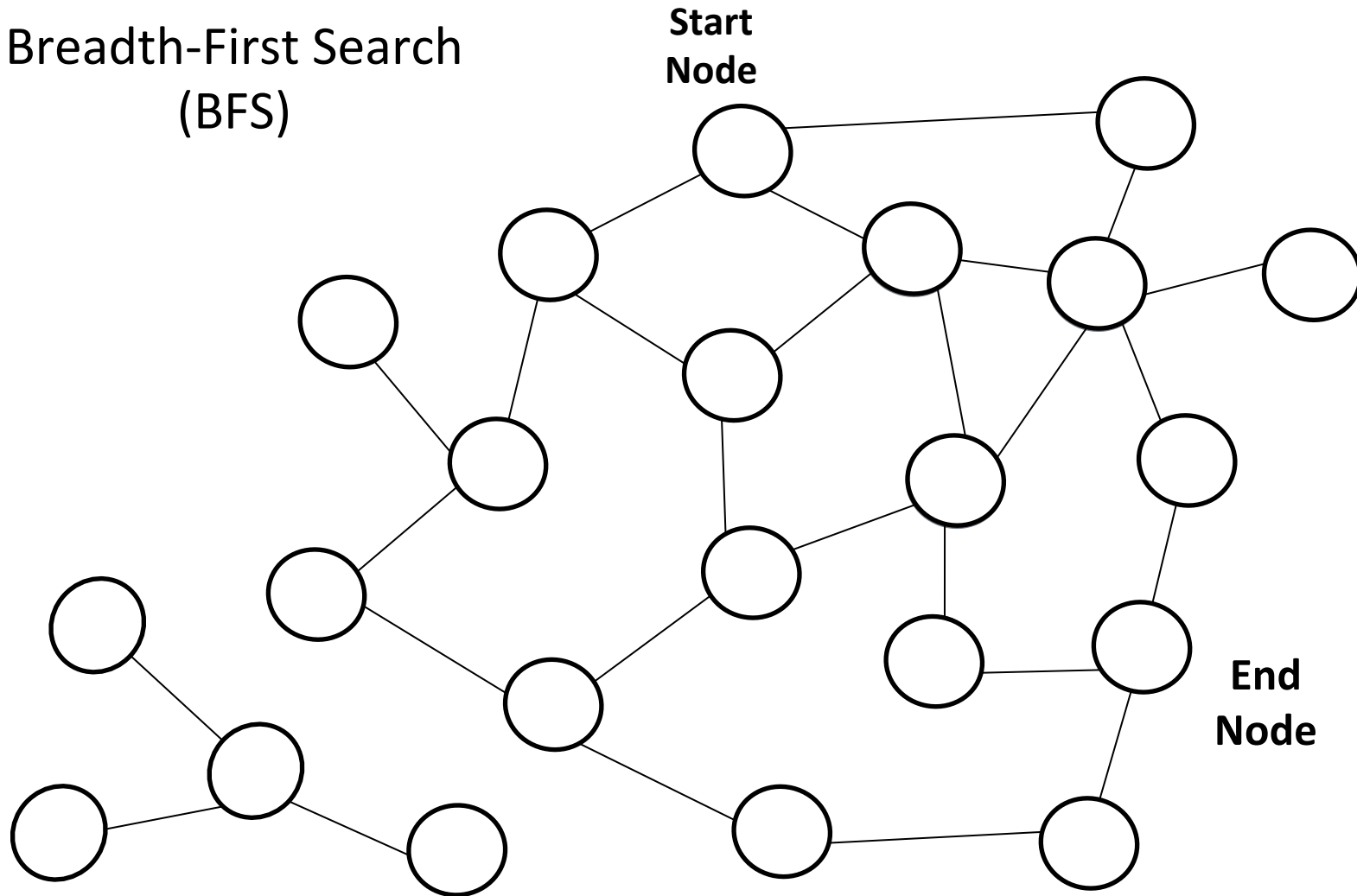
~500 incidents out of 113,000 in Global Terrorism Database (GTD)



Reachability Queries



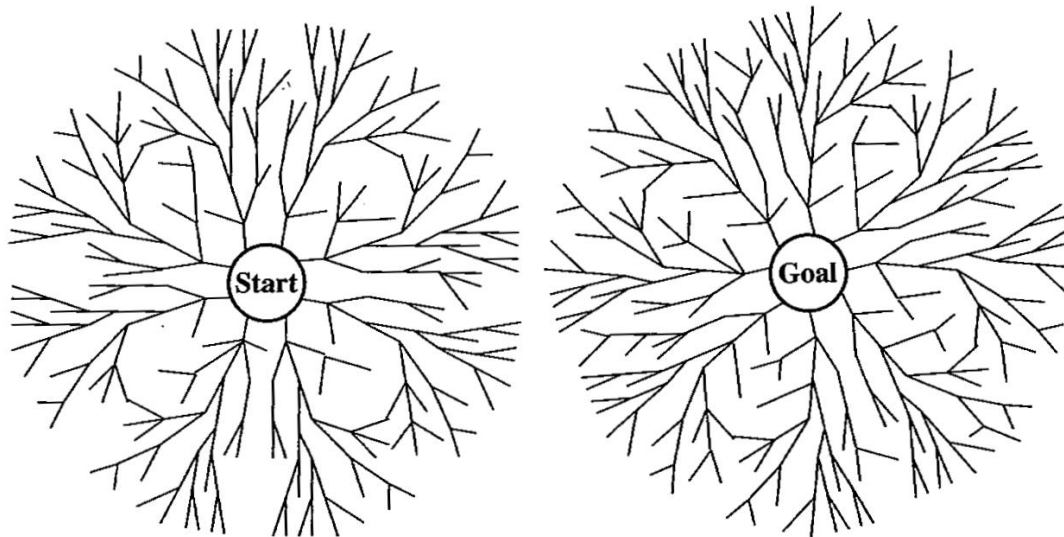
Breadth-First Search
(BFS)



Bidirectional Searching

Query Execution Engine

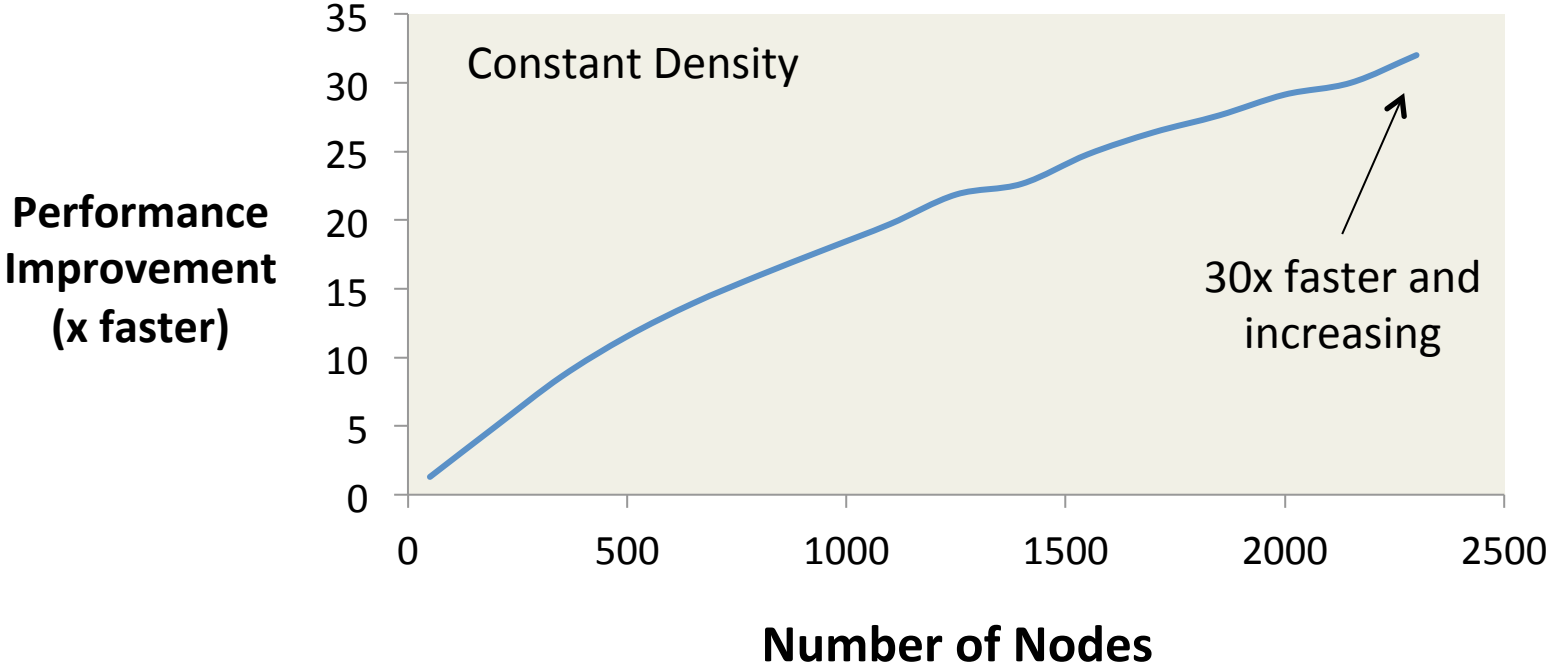
Almost always faster than a normal search



How much faster? 2X? 4X?

Bidirectional Search Performance

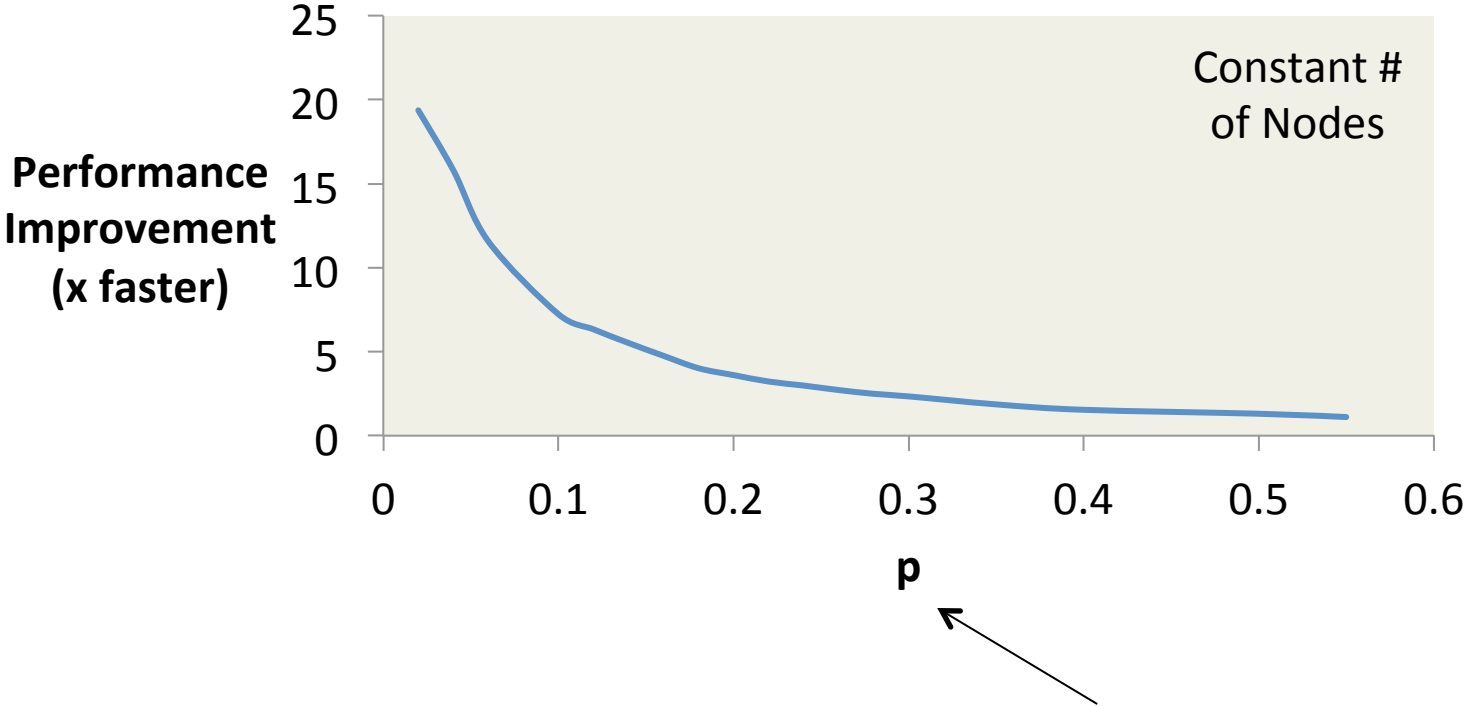
Performance vs. Graph Size



$$Improvement = \frac{BFS \text{ Running Time}}{BIBFS \text{ Running Time}}$$

Bidirectional Search Performance

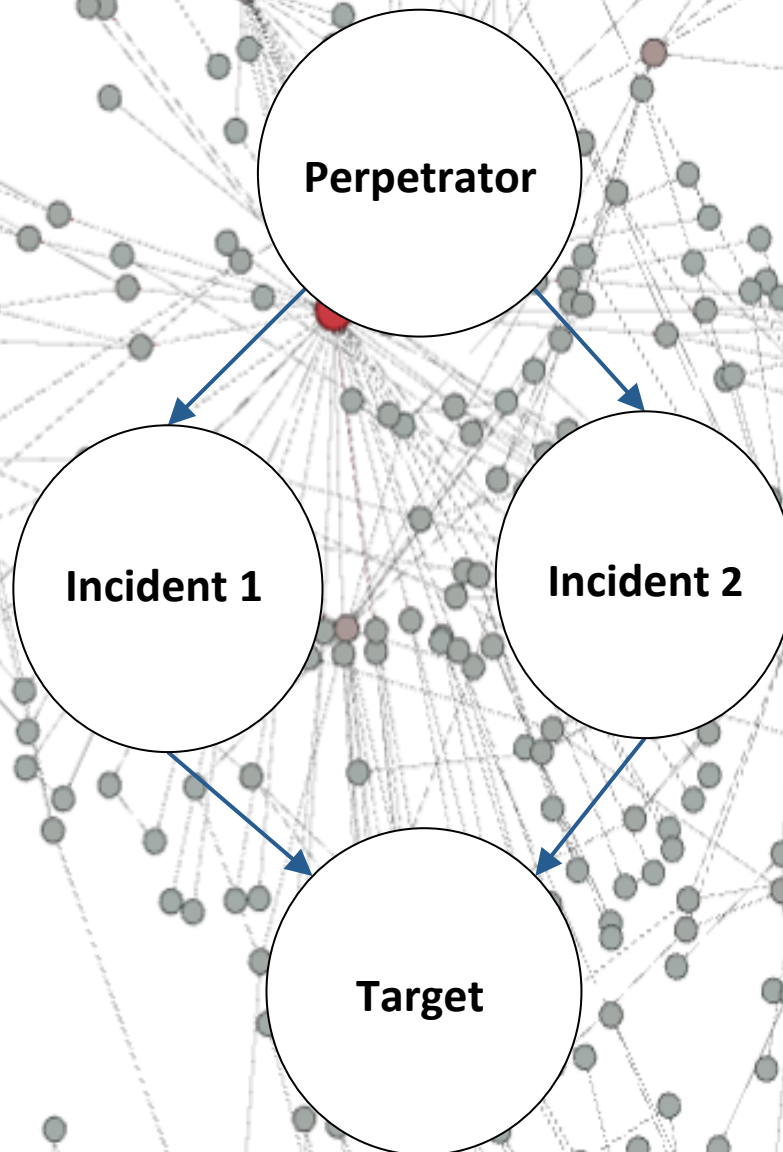
Performance vs. Edge Probability



$Improvement = \frac{BFS \text{ Running Time}}{BIBFS \text{ Running Time}}$ Probability that any two nodes are connected by an edge

Pattern Matching

Query Execution Engine

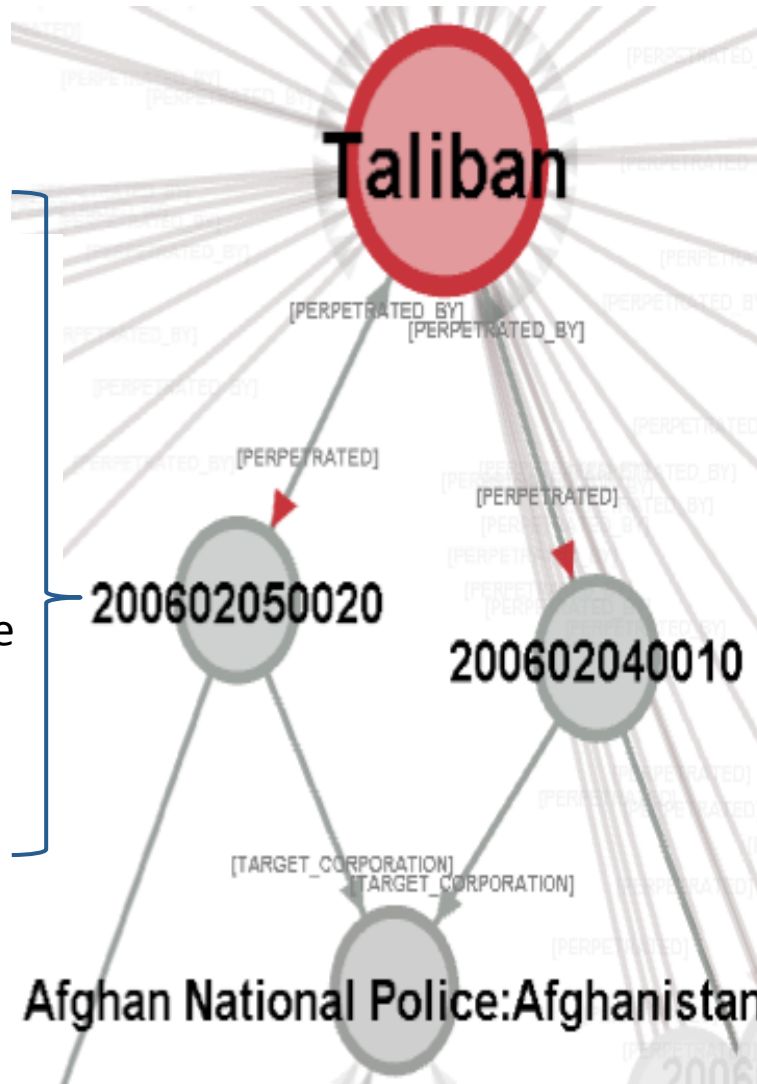


“Search for two distinct incidents that had the same perpetrator and the same target”

Pattern Found

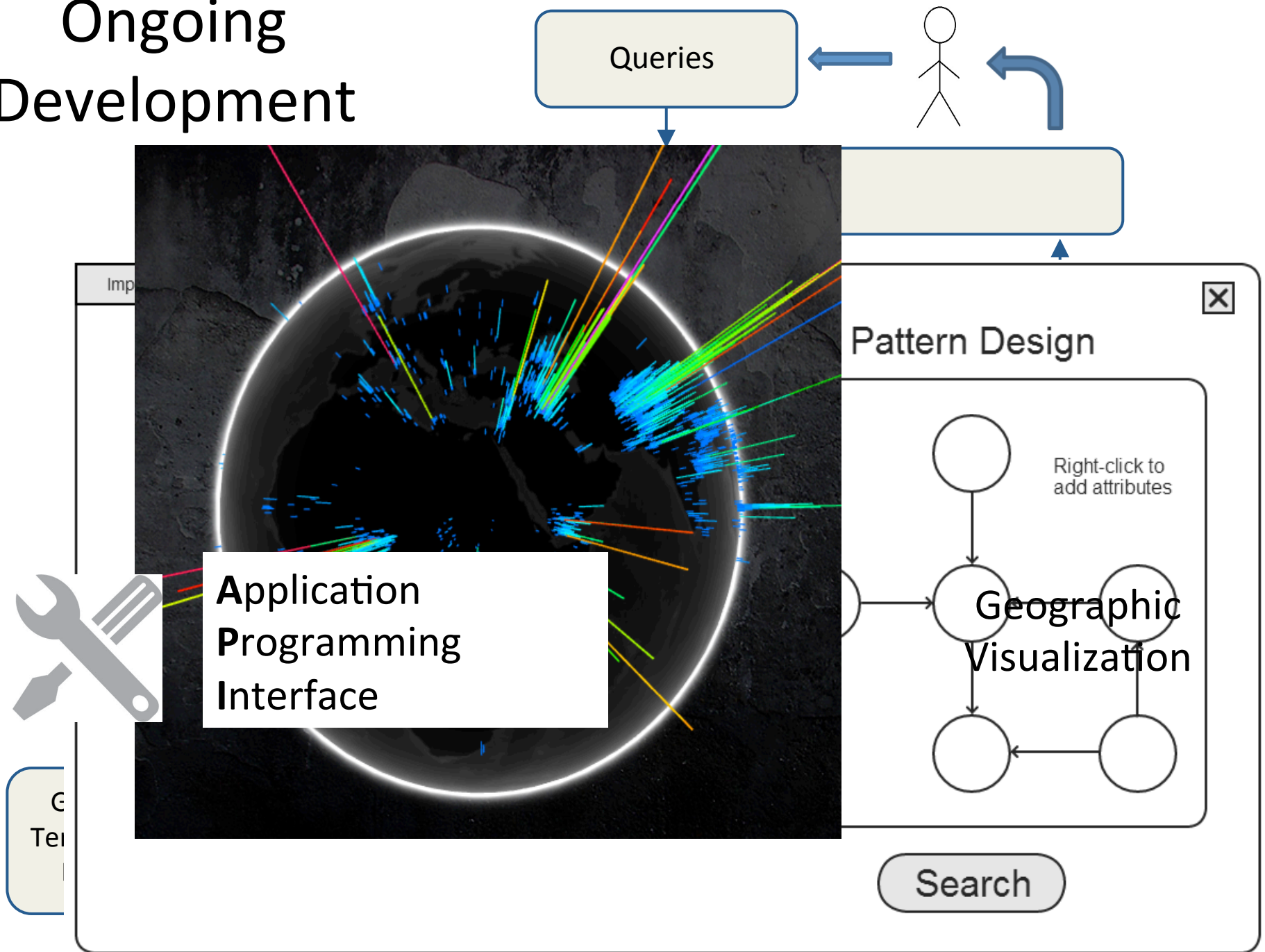
Query Results

200602050020
Target: Afghan National
Police
Date: 2/5/06
Country: Afghanistan
City: Khakrez
Weapon Type: Land Mine
Casualties: 6
Wounded: 2
Summary:



200602040010
Target: Afghan National
Police
Date: 2/4/06
Country: Afghanistan
City: Kandahar
Weapon Type: IED
Casualties: 2
Wounded: 2
Summary:

Ongoing Development



Acknowledgments

Dr. Yinghui Wu



Dr. Xifeng Yan



Zach, Shane, Danny