# Analyzing Data Visualization Tools for Effectiveness of Network Analysis

Min Cheong Kim[1], Nitish Natarajan[2], Yi Shian Ho[3]

[1]Department of Philosophy, [2]Departments of Statistics and Computer Science, [3]Department of Economics, School of Information Sciences, University of Illinois at Urbana-Champaign

ILLINOIS
School of Information Sciences

## INTRODUCTION

- Data visualization tools are used extensively for both academic and non-academic uses to portray different kinds of data.
- Each tool carries different types of visualized data, accepts different inputs and has different limitations and hence, performs different functions.
- They were developed with academic and engineering use cases in mind.

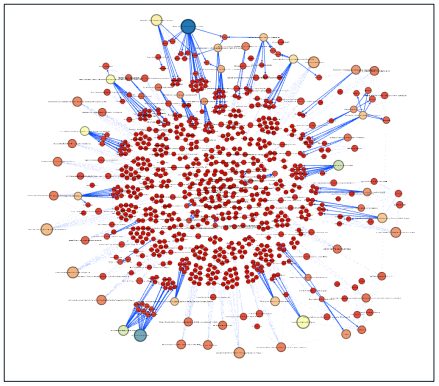
## CRITERIA

- We have chosen 3 data visualization tools for analysis: CiteSpace, Neo4J and Gephi. We have chosen these tools on the basis of "common words" analysis.
- We have chosen these three tools because of their ability to perform academic network analysis.
- Even though all three tools are network analysis tools, we found it useful and important to compare and contrast how they might be used in the network analysis of academic work.

## APPROACH

- We chose a common data set of 1000 academic papers, articles and journals from the WikiProject Zika Corpus for network analysis on the Zika Virus.
- The approach for the tools were as follows:
  - Gephi: Data was obtained as CSV file and loaded to Gephi. A Force Atlas layout was used to distinguish node size.
  - CiteSpace: Metadata was collected from Web of Science database and loaded as TXT files to return author citation network.
  - Neo4J: Data was obtained as a CSV file. Using Cypher, this data was loaded into Neo4J and queried to return the graph.
- We compared the tools and analyzed their effectiveness for different types of usage.

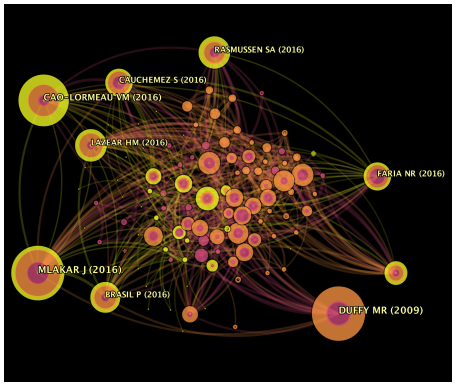
## GEPHI



*Author Co-Citation Analysis of Zika Corpus using Gephi*

- **Input Types:**
  - Spreadsheets (Excel)
  - Graph file formats: GEXF, GraphML, GDF, CSV, PajekNET
- **Output Data:**
  - Social Network Analysis: Mapping community relationships
  - Linked Citation Analysis/ Document Co-Citation Analysis: Use of nodes and edges to visualize and analyze relationships and display graphs in real-time
- **Best Used for:**
  - Analyzing the commonalities between the nodes and easily see which nodes have the most connections

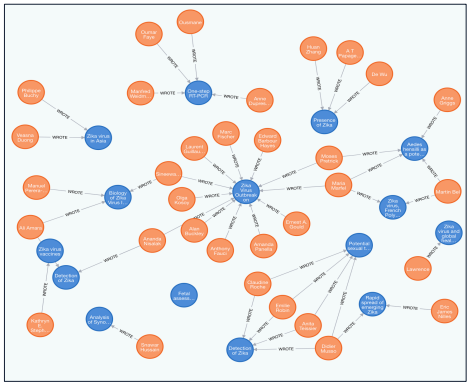
## CITESPACE




*Author Co-Citation Analysis of Zika Corpus using CiteSpace*

- **Input Types**:
  - From Web of Science: TXT format files
  - From Scopus: RIS format files
- **Output Data**:
  - Document Co-Citation Analysis: Analysis of common keywords used
  - Author Co-Citation Analysis: Analysis of how authors cite one another's work
    - Dual-Map Overlays: Analysis on data set based on the present scientific landscape
- **Best Used for**:
  - Author Citation Analysis

## NEO4J



*Author Co-Citation Analysis of Zika Corpus using Neo4J*

- **Input Types:**
  - CSV files
  - JSON files
  - Cypher - Neo4J's querying language
- **Output Data:**
  - Graphical Network Analysis: Graphical representation with nodes, relationships and properties
  - Queried Data Analysis: Queried data in tabular form with the option to be exported as an auto-generated TXT file
- **Best Used for:**
  - Creating graphs through querying data efficiently and easily using Cypher

## CONCLUSION

Each data visualization tools performed a unique action on the Zika Virus Corpus. Based on the strengths of each tool:

- Gephi is best used for visualizing the connections and relationships between each node and mapping out how they are connected to each other.
- CiteSpace is best used for academic paper citation analysis, particularly for analyzing common key words in a specific field of academia and for analyzing how authors are citing one another in their fields.
- Neo4J is best used to analyze vast amounts of information through creating relationships between nodes based on queries.

## ACKNOWLEDGEMENTS

- Supervisor: Professor Jodi Schneider, School of Information Sciences

ILLINOIS