

# Analyzing Biological Networks

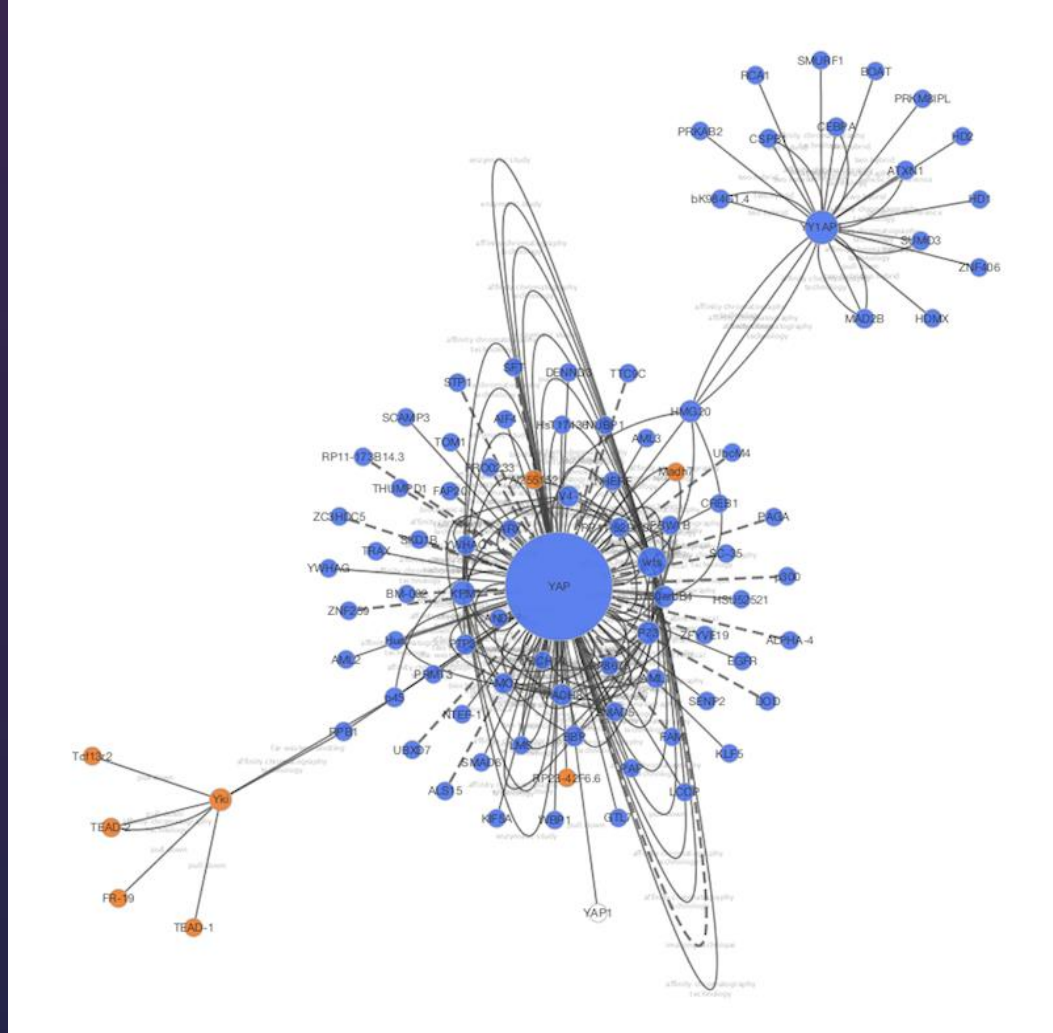
Students: Osmany Pujol, Vyacheslav Lazurenko, Anthony Abello, Sage Pages, Kyle Gareth Brindle

Mentor: Raihanul Bari Tanvir , Md Mezbahul Islam

Instructor/Faculty: Alex Afanasyev, Florida International University

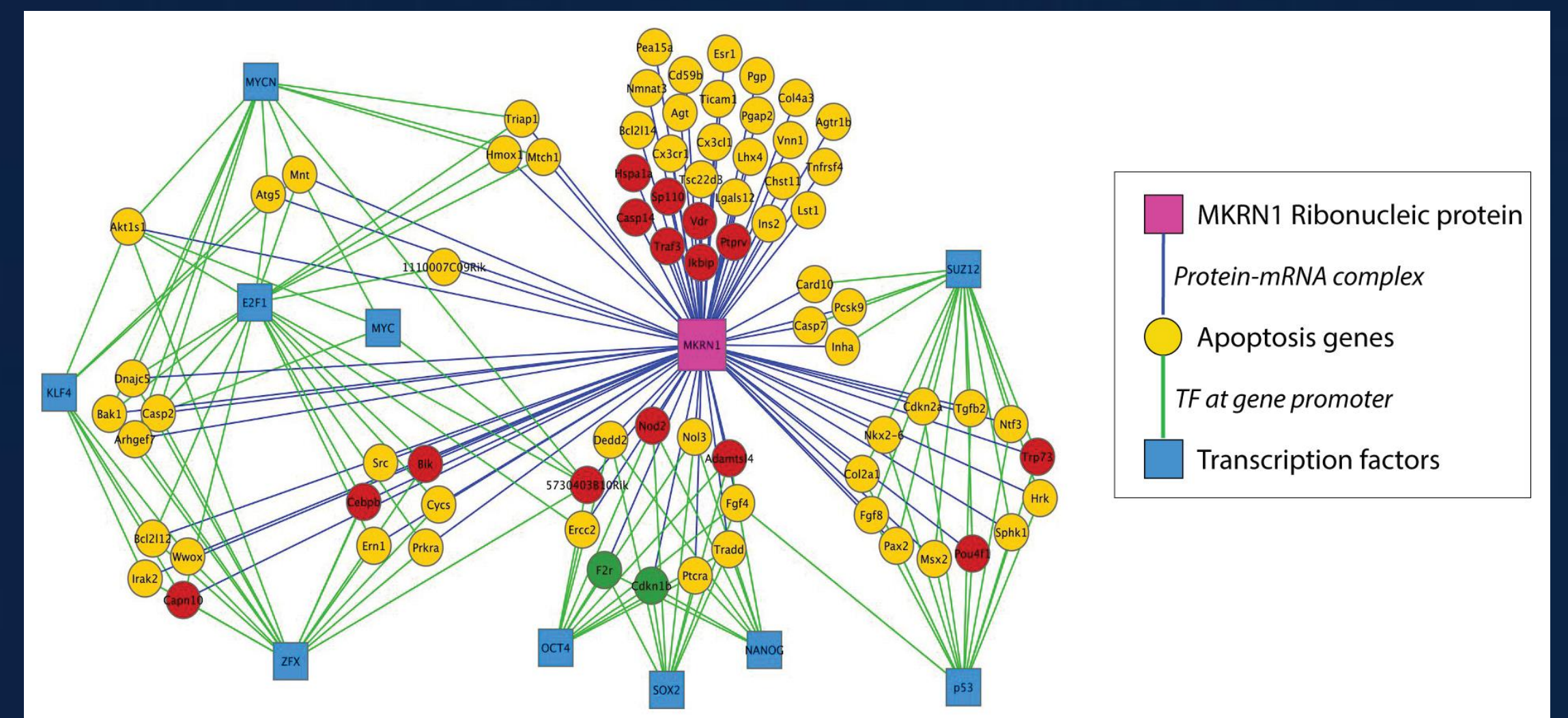
## PROBLEM

When it comes to studying biological networks, scientists and researchers often rely on community detection algorithms. While the open-source tool Cytoscape is commonly used for this purpose, it does not include all available algorithms, especially newer ones.



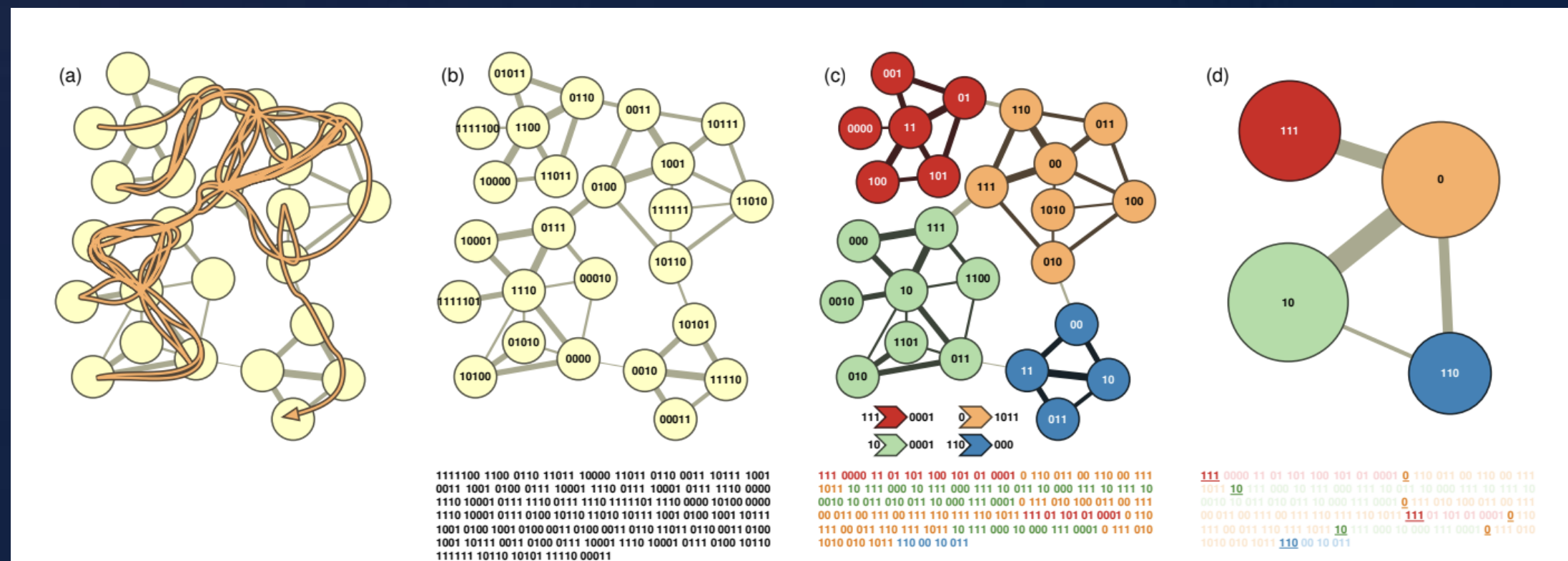
## SOLUTION

For scientists and researchers studying biological networks, community detection algorithms like the Louvain method and Infomap algorithm are popular tools. For our contribution, we focused on adding these implementations to the Cyfinder plugin for Cytoscape to expand its capabilities for detecting communities in undirected graphs.



## IMPLEMENTATION

In order to identify communities within a network, Infomap uses the Map Equation, a calculation developed by researchers to determine the minimum average bit length needed to represent the network (known as MDL). By minimizing the total description of the network according to this principle, the algorithm can generate a community structure that reflects the most significant patterns of connections within the network.



Map Equation for undirected graphs

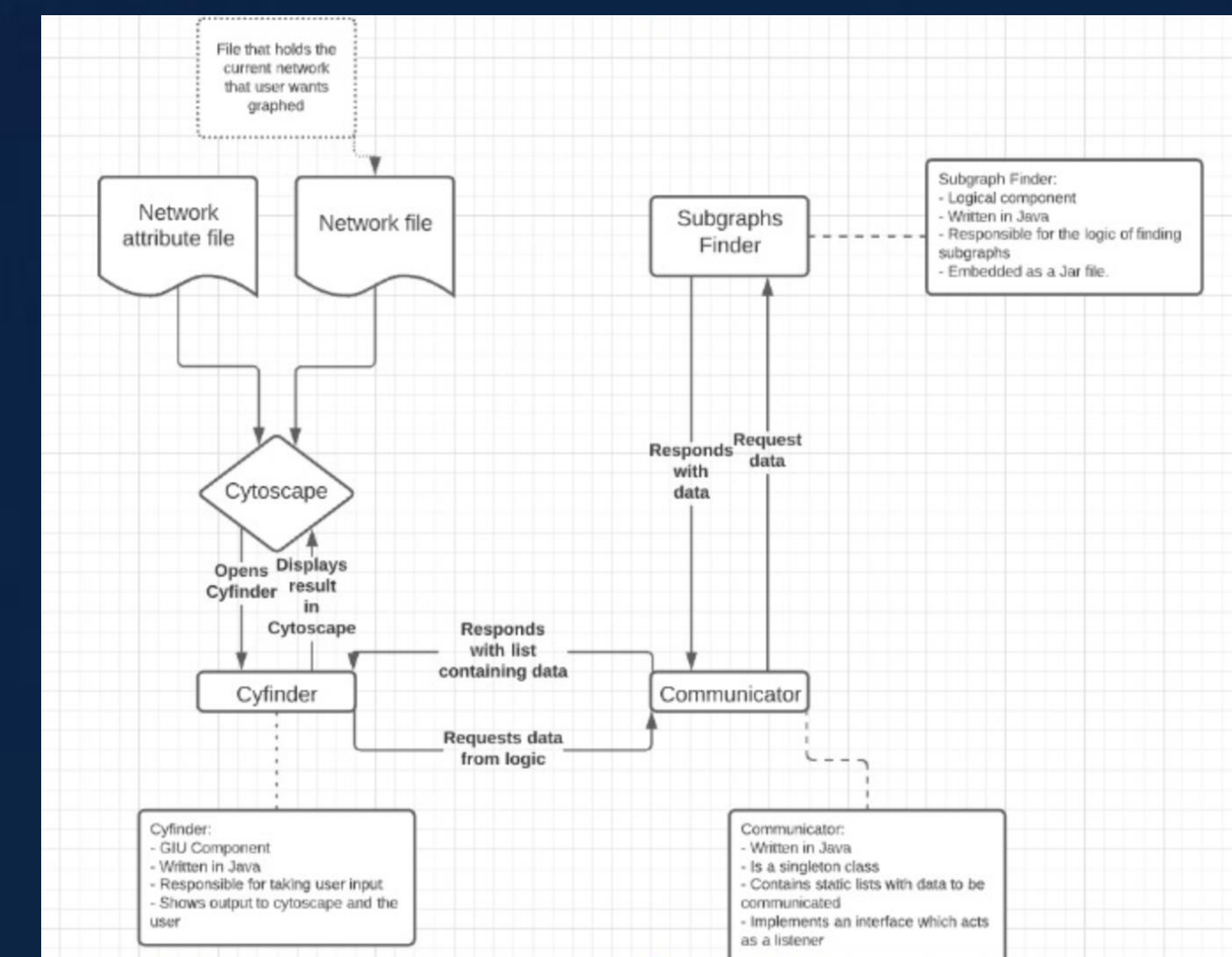
Map Equation Steps:

1. Each node is assigned to its own module (or community) to start with.
2. The algorithm then iteratively optimizes the module assignments by trying to minimize the description length of a random walker's movements on the network graph.
3. The description length is calculated using information theory, and represents the length of the message that would be needed to describe the random walker's movements on the network.
4. During each iteration, the algorithm considers moving nodes between modules to improve the description length of the random walker's movements.
5. The algorithm continues iterating until no further improvements can be made to the description length.
6. The resulting module assignments represent the communities in the network graph.

## REFERENCES

Ono, K. (n.d.). What is Cytoscape? Retrieved November 29, 2022, from [https://cytoscape.org/what\\_is\\_cytoscape.html](https://cytoscape.org/what_is_cytoscape.html) Ono, K. (n.d.). Download Cytoscape. Retrieved November 29, 2022, from <https://cytoscape.org/download.html> Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2010). The map equation. *mapequation*. Retrieved November 26, 2022, from <https://www.mapequation.org/publications.html> Rosvall, M., & Bergstrom, C. T. (2008, January 29). Maps of random walks on complex networks reveal community structure. *pnas*. Retrieved November 27, 2022, from <https://www.pnas.org/doi/10.1073/pnas.0706851105> Infomap - Network community detection using the Map Equation framework. (n.d.-b). *mapequation.org*. <https://www.mapequation.org/infomap/> Rita, L. (2021a, December 14). Infomap Algorithm - Towards Data Science. Medium. <https://towardsdatascience.com/infomap-algorithm-9b68b7e8b86> Wikimedia Foundation. (2022, July 18). *Random Walker algorithm*. Wikipedia. Retrieved November 29, 2022.

## SYSTEM DESIGN



## REQUIREMENTS

|  |                                                                                                          |
|--|----------------------------------------------------------------------------------------------------------|
|  | Java 11.0                                                                                                |
|  | Maven<br>widely used build automation and project management tool for Java-based projects                |
|  | Apache Ant<br>Subgraphs Finder uses ant to build all the logic into a jar file that is used by Cyfinder. |
|  | Cytoscape 3.9.1<br>The software for which we built the community detection algorithm and                 |
|  | Git<br>Version Control and Collaboration                                                                 |

## CONTRIBUTIONS

- Conducted research into various algorithms and research papers.
- Concise understanding of implementation details.
- Removal of Huffman encoding portion of previous implementation.
- Pari programming to conduct tests on various graphs and varying bit lengths.

## SUMMARY

Our team developed the Infomap algorithm for detecting communities in undirected weighted graphs, inspired by a scholarly work. We then integrated the algorithm into the Cytoscape GUI as part of our proposed solution. Although there are ways to optimize the algorithm further, we were unable to include them due to time constraints. The algorithm is adaptable to other types of graphs and offers versatility for community detection.

**ATTENTION:** Subheadings shown above is a recommended structure for a research/scientific poster. Main elements of a poster are Introduction, Research and Conclusion. Researchers can customize the section headers based on their projects' needs. REMOVE THIS TEXT ONCE YOU START WORKING ON YOUR POSTER

## ACKNOWLEDGEMENT

The material presented in this poster is based upon the work supported by XXXXX (name of the project sponsor: federal, state or local government, or corporate partners, where applicable). We/I thank XXXX for his/her/their assistance and mentorship that I/we received throughout the senior design project.