



Knight Foundation
School of Computing
and Information Sciences

Knight Foundation School of Computing and Information Sciences

Spring 2023 Senior Design Project

Analyzing Biological Networks

Students: Vyacheslav Lazurenko

Mentors: Ananda Mondal, Mezbahul Islam, Raihanul Bari Tanvir

Instructor/Faculty: Masoud Sadjadi, Florida International University

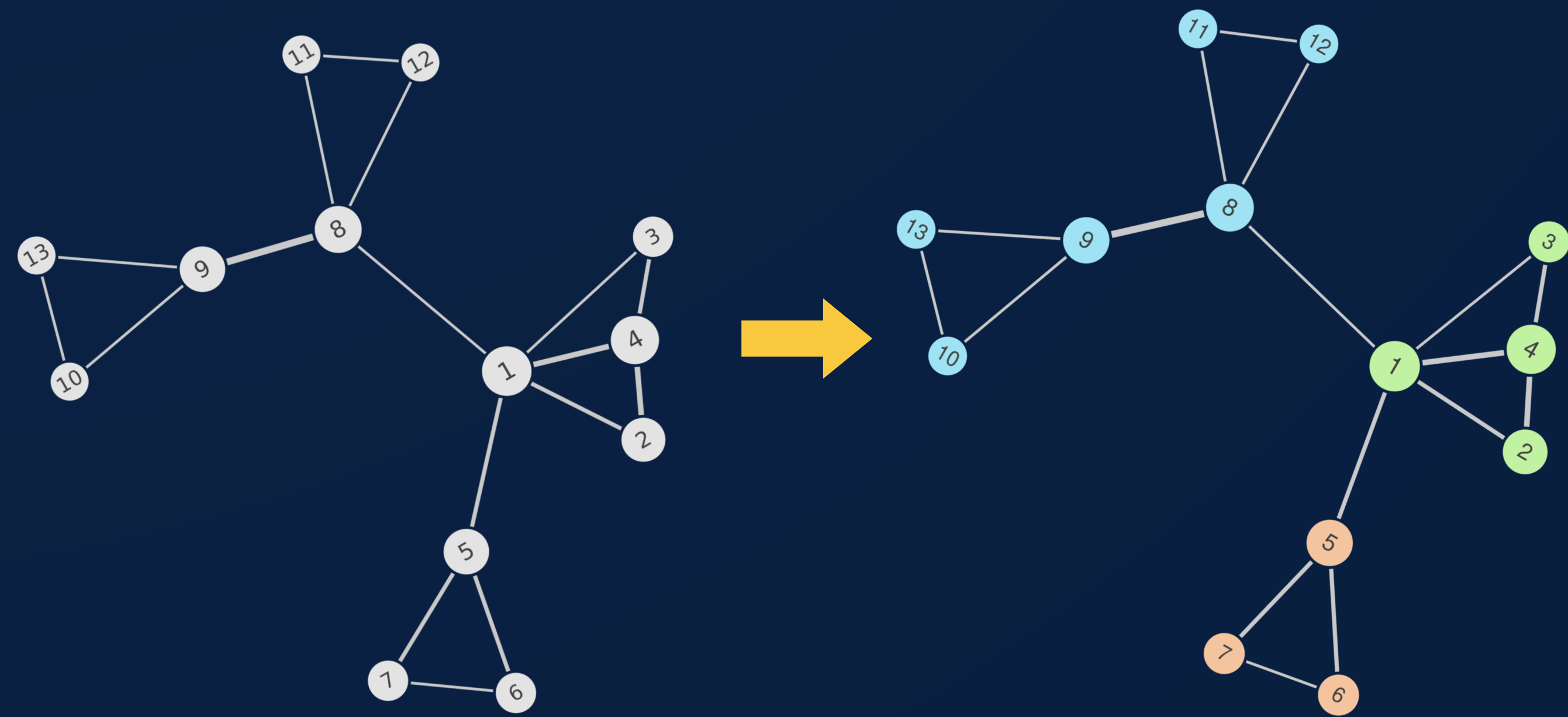
INTRODUCTION

A group of *genes* or *proteins* work together by forming a cluster to accomplish different functions, beneficial or adversarial such as initiation or progression of a disease (e.g., cancer). **Cytoscape** is a bioinformatics software that allows researchers to analyze biological networks in form of graphs. **CyFinder** project is a plugin that gives users the ability to perform many types of clustering algorithms that can be used to detect diseases in early stages.

PROBLEM

Add **Infomap** community detection algorithm to **CyFinder** plugin

The main goal is to identify clusters (communities) in a graph (biological network). Example blow shows input, a graph structure, and output, same graph but where each node is assigned to a community.

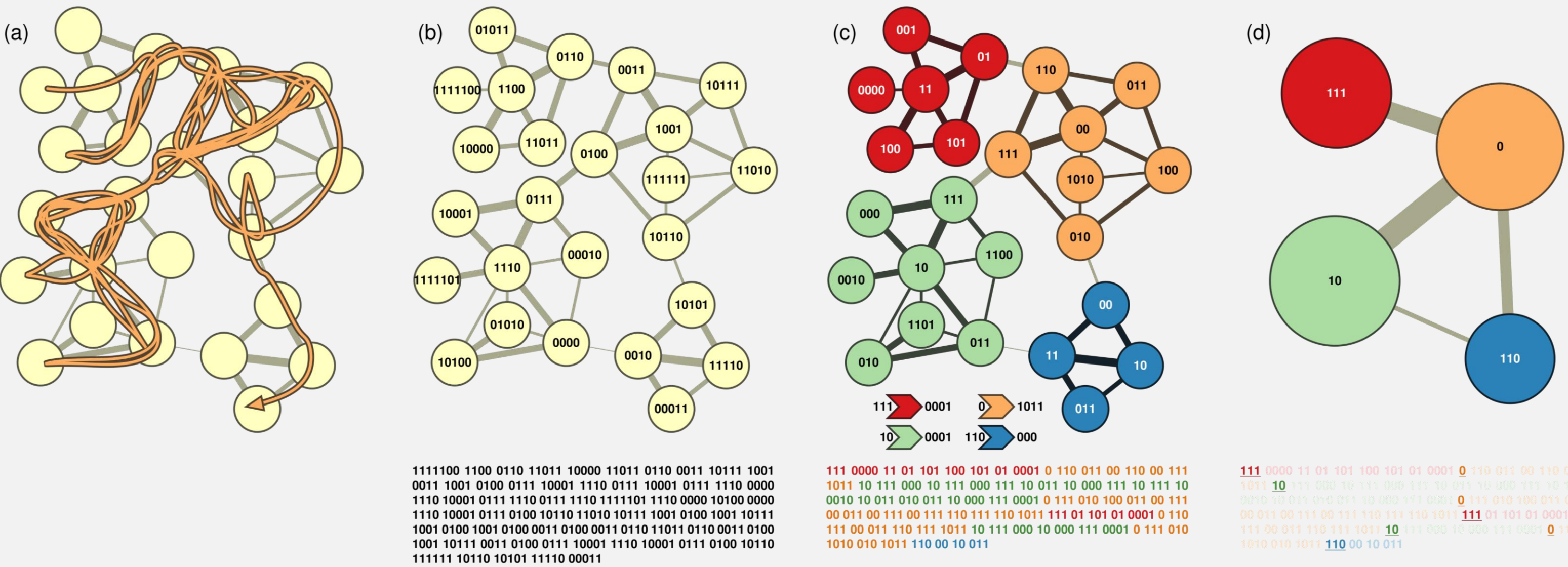


SUMMARY

- Implemented community detection algorithm as a part of plugin for biological network analysis software
- Applied problem solving and analytical thinking skills to interpret and apply complex logic and mathematical concepts behind Infomap cluster detection algorithm
- Tested and verified functionality for the end-user

SOLUTION

Map Equation

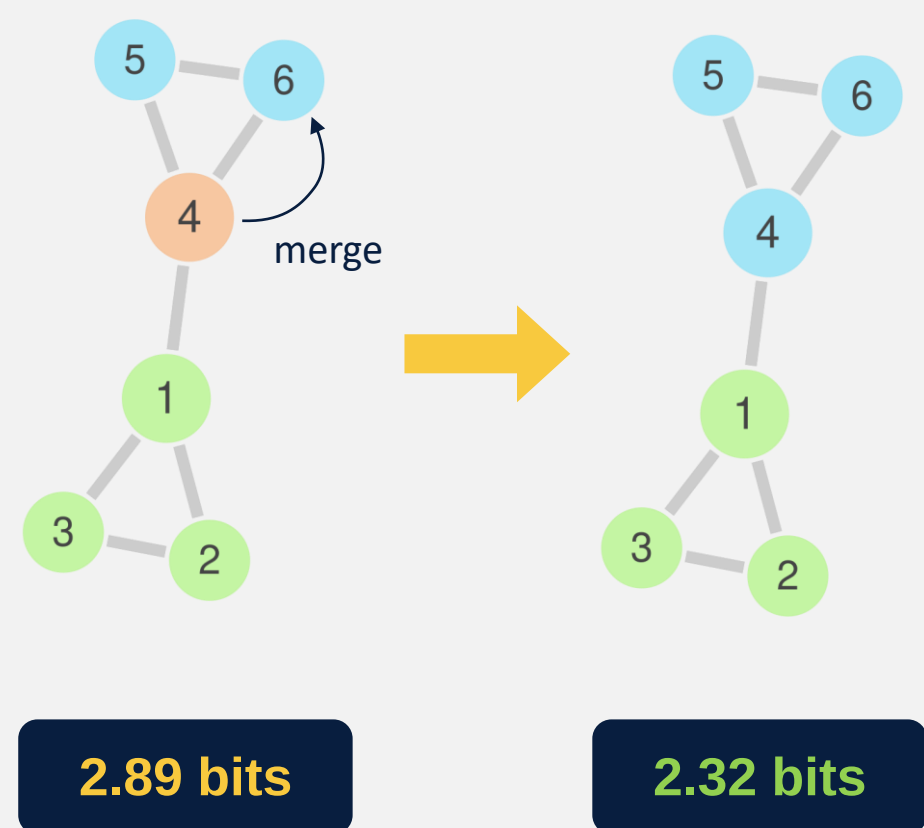


Map Equation is used to evaluate efficiency of graph partition. Logical concept of Map Equation:

- Random walker**: roam the graph and store its path to identify frequency with which each node is visited.
- Each node is assigned a binary code using Huffman Encoding algorithm, that is most frequent nodes will receive shorter codes. These codes are used to describe random walker's path.
- Split the graph into subgraphs (communities), each will have its own Huffman Encoding for nodes, and unique entrance and exit codes. This decreases the total length of random walker's path description.
- Each subgraph represents a node at higher-level graph, process is repeated.

Infomap Algorithm

- Infomap** algorithm uses **Map Equation** to find optimal graph partition
- The goal is to find such community configuration that will **minimize** Map Equation
 - Start with each node in its own module
 - Move node to its neighboring module if it decreases modularity
 - Repeat until no further improvements can be made



MY CONTRIBUTION

- Leading and coordinating my team throughout this project
- Deep understanding of concepts from research papers
- Implementation multi-level algorithm
- Improvement of performance by 50%

TECHNOLOGIES



REFERENCES

- Rosvall, M., Axelsson, D. & Bergstrom, C. The map equation. *Eur. Phys. J. Spec. Top.* 178, 13–23 (2009). <https://doi.org/10.1140/epjst/e2010-01179-1>
- Edler D, Bohlin L, Rosvall M. Mapping Higher-Order Network Flows in Memory and Multilayer Networks with Infomap. *Algorithms*. 2017; 10(4):112. <https://doi.org/10.3390/a10040112>
- Edler, D., Holmgren, A., & Rosvall, M. (n.d.). *Simplify and highlight important structures in complex networks*. MapEquation. Retrieved April 17, 2023, from <https://www.mapequation.org/>

Mathematical Definition of Map Equation

$$L(M) = q_{i \sim} H(Q) + \sum_{i=1}^m p_{i \circ}^i H(P^i)$$

Evaluates **modularity** of a graph partition.
Lower result means better partition.

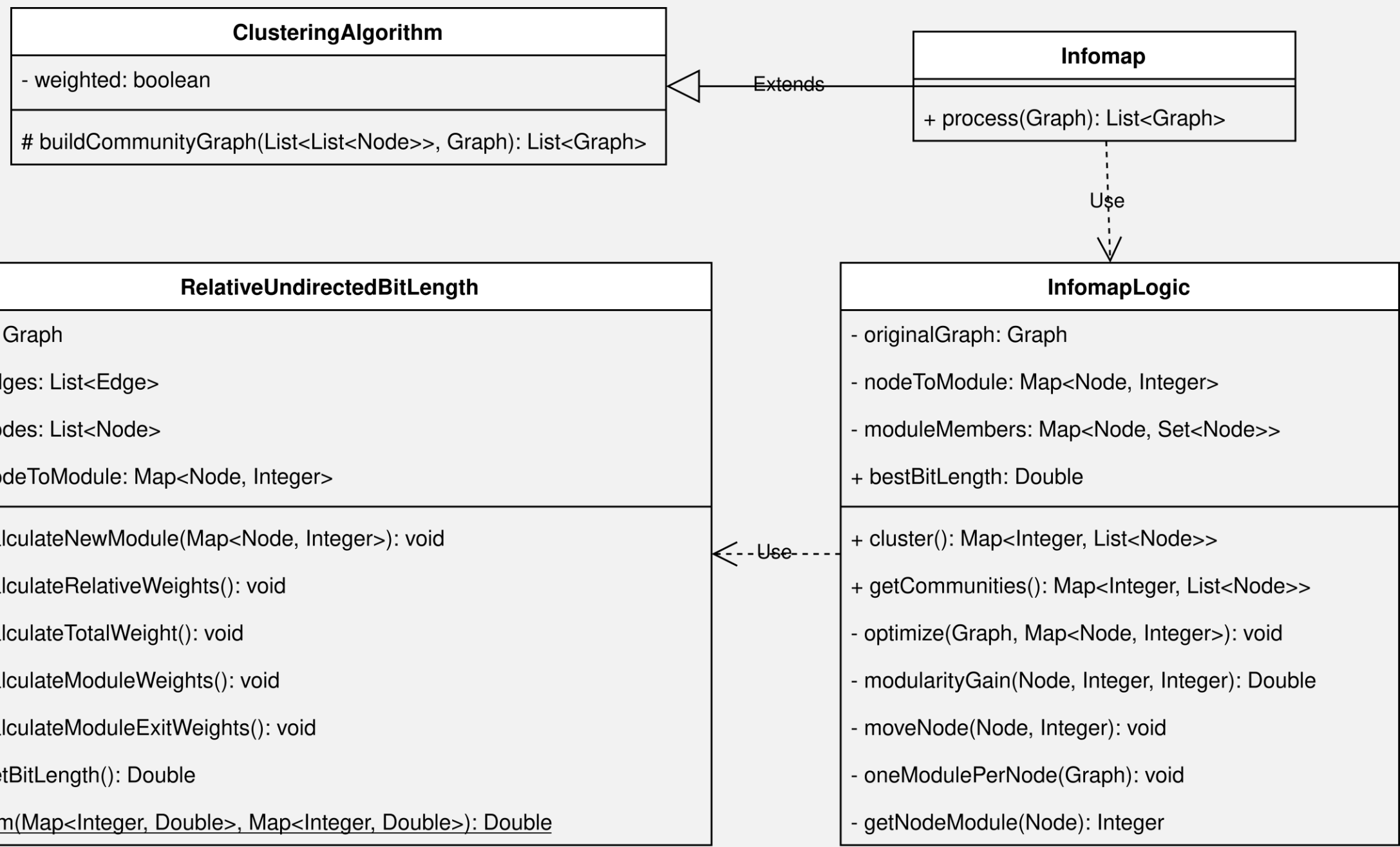
$$H(Q) = - \sum_{i=1}^m \frac{q_{i \sim}}{\sum_{j=1}^m q_{j \sim}} \log \left(\frac{q_{i \sim}}{\sum_{j=1}^m q_{j \sim}} \right)$$

Orange section represents number of bits used to describe movement **between** modules

$$H(P^i) = - \frac{q_{i \sim}}{q_{i \sim} + \sum_{\beta \in i} p_{\beta}} \log \left(\frac{q_{i \sim}}{q_{i \sim} + \sum_{\beta \in i} p_{\beta}} \right) - \sum_{\alpha \in i} \frac{p_{\alpha}}{q_{i \sim} + \sum_{\beta \in i} p_{\beta}} \log \left(\frac{p_{\alpha}}{q_{i \sim} + \sum_{\beta \in i} p_{\beta}} \right)$$

Green section represents number of bits used to describe movement **inside** modules.

SYSTEM DESIGN



Engineering
& Computing

ACKNOWLEDGEMENT

The material presented in this poster is based upon the work supported by Ananda Mondal. I thank Ananda Mondal, Md Mezbahul Islam and Raihanul Bari Tanvir for their assistance and mentorship that I received throughout the senior design project.

FLORIDA INTERNATIONAL UNIVERSITY