

*Florida International University*  
*Knight Foundation School of Computing and Information Sciences*

Software Engineering Focus

# Final Deliverable

Project Title: Analyzing Biological Networks - CyFinder

**Team Members:** Slav Lazurenko, Osmany Pujol, Sage A Pages, Kyle Gareth Brindle, Anthony Abello

**Product Owner(s):** Ananda Mondal

**Mentor(s):** Raihanul Bari Tanvir, Md Mezbahul Islam

**Instructor:** Masoud Sadjadi

The MIT License (MIT)  
Copyright (c) 2016 Florida International University

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

***Abstract***

*This document presents the information necessary to gain a good understanding of the functionality and work done on the Cytoscape plug-in or app called CyFinder. The primary purpose of the CyFinder is to process biological networks that are imported in the form of graphs where each node represents a protein or gene, and each edge represents the interactivity of the objects. This project focuses on the implementation of the Infomap community detection algorithm that will become a part of the rich toolset that CyFinder provides. This algorithm can be used to detect clusters formed by biological entities to detect various diseases like cancer.*

## Table of Contents

|                                                                                                                       |           |
|-----------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Introduction.....</b>                                                                                              | <b>6</b>  |
| Current System.....                                                                                                   | 6         |
| Purpose of New System.....                                                                                            | 7         |
| <b>User Stories.....</b>                                                                                              | <b>8</b>  |
| Implemented User Stories.....                                                                                         | 8         |
| Pending User Stories.....                                                                                             | 8         |
| <b>Project Plan.....</b>                                                                                              | <b>9</b>  |
| Hardware and Software Resources.....                                                                                  | 9         |
| Sprint Plans.....                                                                                                     | 11        |
| Sprint 1.....                                                                                                         | 11        |
| Sprint 2.....                                                                                                         | 12        |
| Sprint 3.....                                                                                                         | 13        |
| Sprint 4.....                                                                                                         | 14        |
| Sprint 5.....                                                                                                         | 15        |
| Sprint 6.....                                                                                                         | 16        |
| Sprint 7.....                                                                                                         | 17        |
| <b>System Design.....</b>                                                                                             | <b>18</b> |
| Architectural Patterns.....                                                                                           | 18        |
| System and Subsystem Decomposition.....                                                                               | 19        |
| Project Dependencies Diagram.....                                                                                     | 20        |
| Design Patterns.....                                                                                                  | 21        |
| <b>System Validation.....</b>                                                                                         | <b>22</b> |
| <b>Validation Image 1 - Comparison of results from infomap website to CyFinder Algorithm...</b>                       | <b>22</b> |
| <b>Appendix.....</b>                                                                                                  | <b>23</b> |
| Appendix A - UML Diagrams.....                                                                                        | 23        |
| Appendix B - User Interface Design.....                                                                               | 23        |
| Appendix C - Sprint Review Reports.....                                                                               | 24        |
| Appendix D - User Manuals, Installation/Maintenance Document, Shortcomings/Wishlist Document and other documents..... | 27        |
| <b>References.....</b>                                                                                                | <b>27</b> |



## INTRODUCTION

Cytoscape is a piece of open-source software that analyzes and provide visual representation of networks. Its applications can range from analyzing molecular networks to gene expression. CySubgraphFinder, also known as CyFinder, is a plugin for Cytoscape that expands upon Cytoscapes pre-existing tools for analyzing and visualizing networks. CyFinder is a Java Archive (JAR) Maven Install File.

Most of CyFinder functionality is found in SubgraphFinder, which contains the logic for all of the algorithms implemented into Cytoscape by CyFinder. This project aimed to implement the map equation, a tool used to evaluate the efficacy of community partitions within a network, and Infomap, an algorithm that makes use of the map equation to detect communities within networks.

## Current System

The current system allows the user to convert directed graphs to undirected graphs using two approaches: additive, which involves the adding another edge between any two connected nodes, or transformative, which changes the given edge object into an object Cyfinder can work with. It also allows user to specify the search conditions (bi-partite, clique or directed clique), select the search algorithm they want to use (Breadth or Depth First Search), if they want the graphs to be ordered (based on edge weight and SUID) and how they want to save the generated subgraphs (add truth table, generate separate subgraphs). Also, it allows the user to sort the graphs based on node count, visualizing a specific k-partite graph based on the user's input and updating the GUI. There is also an option to add k-partite numbers to subgraphs, but it has not been completed. Additional features include enabling users to find the maximum clique and the maximum complete bipartite subgraph within a network; creating a layout option to group the Bipartite Network nodes into two columns; permit users to find collections of clusters of well-connected proteins using six methods: Walktrap, Fast Greedy, Edge Betweenness, Label Propagation, Leading Eigenvector, and Louvain.

## **Purpose of New System**

The purpose of the project is to expand upon the tools CyFinder implements into Cytoscape. The new features include the map equation - a tool used to measure the efficacy of community partitions by predicting the average code length of a network if that network were to be encoded using Huffman encoding. The other primary feature incorporated in this project is the implementation of the Infomap community detection algorithm, which makes use of the map equation to measure its own efficacy. The infomap algorithm begins by finding the local minima of a graph. It then recursively groups these low-level communities by treating them as nodes in a graph, and applying the same algorithm that was applied to the original network. It continues this until no further improvements to the map equation can be found.

## USER STORIES

The following section provides the detailed user stories that were implemented in this iteration of the CyFinder project. These user stories served as the basis for the implementation of the project's features. This section also shows the user stories that are to be considered for future development.

### Implemented User Stories

IMAP 1: Fix Map Equation implementation using Random Walker and Huffman Encoding

IMAP 2: Test Map Equation implementation against mapequation.org solution

IMAP 3: Implement Map Equation average bit length calculation using power iteration method

IMAP 4: Test implementation of power iteration method

IMAP 5: Implement community formation part of Infomap algorithm

IMAP 6: Integrate IMAP 3 into Infomap algorithm as a modularity gain function

IMAP 7: Improve performance of current multi-level Infomap algorithm

IMAP 8: Document history of performance improvements

IMAP 9: Connect current Infomap implementation to Cytoscape

IMAP 10: Adjust Cytoscape GUI dialog for Infomap

### Pending User Stories

IMAP 11: Adding support for directed graphs in map equation

IMAP 12: Implement coarse-tuning (Algorithm 6) from *Mapping Higher-Order Network Flows* Paper

IMAP 13: Implement fine-tuning algorithm (Algorithm 5) described in *Mapping Higher-Order Network Flows* Paper

IMAP 14: Implement repeated supermodule clustering (Algorithm 7) from *Mapping Higher-Order Network Flows* Paper

IMAP 15: Implement repeated submodule clustering (Algorithm 8) from *Mapping Higher-Order Network Flows* Paper



## PROJECT PLAN

This section describes the planning that went into the realization of this project. This project incorporated the agile development techniques and as such required the sprints to be planned. These sprint plannings are detailed in the section. This section also describes the components, both software and hardware, chosen for this project.

## Hardware and Software Resources

The following hardware and software requirements were necessary throughout the development of these CyFinder components.

### Hardware:

- None: there was no need for specific hardware.

### Software:

- JAVA 13: the language that was used to write all of our code for CyFinder this semester.
- Eclipse: the IDE where we implemented all of our features.
  - Apache Maven: used to generate JAR files.
  - Apache Ant: used to build the logic behind CyFinder
- IntelliJ Community edition (any version should work fine): An optional IDE used to implement features.

- Cytoscape 3.9.1: core application where CyFinder is rooted
  - Cytoscape API: powerful graph-based API used by CyFinder
- GitHub: version control software used throughout the project

## Sprint Plans

### *Sprint 1*

#### **Sprint Planning Meeting Minutes:**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kayle Gareth Brindle, Anthony Abello**

Start time: **9:30PM**

End time: **9:45PM**

After discussion, the velocity of the team was estimated to be **20 points (4 per team member)**.

The product owner chose the following user stories to be done during the next sprint. They are ordered based on their priority.

- No specific user stories were provided yet. Main goal is to get familiar with the project, its requirements and work of the previous team.

The team members indicated their willingness to work on the following user stories.

- Every team member will:
  - Learn basics of Cytoscape software program (1 point)
  - Go over documentation and source code from the previous team (2 points)
  - Read other resources provided by the owner (1 point)

*Sprint 2***Sprint Planning Meeting Minutes:**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kayle Gareth Brindle, Anthony Abello**

Start time: **9:30PM**

End time: **9:45PM**

After discussion, the velocity of the team was estimated to be **40 points (8 per team member)**.

The product owner chose the following user stories to be done during the next sprint. They are ordered based on their priority.

- PREP 3: Deepen understanding of Info Map algorithm (3 points)
- PREP 4: Understand current state of Info Map implementation (3 points)
- PLAN 1: Create the roadmap for finishing Info Map implementation (2 points)

The team members indicated their willingness to work on the following user stories.

- Every team member will:
  - PREP 3
  - PREP 4
  - PLAN 1

*Sprint 3***Sprint Planning Meeting Minutes:**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kayle Gareth Brindle, Anthony Abello**

Start time: **9:30PM**

End time: **9:45PM**

After discussion, the velocity of the team was estimated to be **40 points (8 per team member)**.

The product owner chose the following user stories to be done during the next sprint. They are ordered based on their priority.

- FIX 1: Fix the info map implementation from the previous semester (3 points)
- TEST 1: Test the new implementation against the online map equation info map (3 points)
- DOC 1: Write documentation for the implementation of the info map algorithm (2 points)

The team members indicated their willingness to work on the following user stories.

- Every team member will:
  - FIX 1
  - TEST 1
  - DOC 1

*Sprint 4***Sprint Planning Meeting Minutes:**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kayle Gareth Brindle, Anthony Abello**

Start time: **9:30PM**

End time: **9:45PM**

After discussion, the velocity of the team was estimated to be **45 points (9 per team member)**.

The product owner chose the following user stories to be done during the next sprint. They are ordered based on their priority.

- IMAP 4: Test implementation of power iteration method for undirected networks (2 points)
- IMAP 5: Implement community formation part of Infomap algorithm (3 points)
- IMAP 6: Integrate IMAP 3 into Infomap algorithm as a modularity gain function (4 points)

The team members indicated their willingness to work on the following user stories.

- Every team member will:
  - IMAP 4
  - IMAP 5
  - IMAP 6

*Sprint 5***Sprint Planning Meeting Minutes:**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kayle Gareth Brindle, Anthony Abello**

Start time: **9:30PM**

End time: **9:45PM**

After discussion, the velocity of the team was estimated to be **35 points (7 per team member)**.

The product owner chose the following user stories to be done during the next sprint. They are ordered based on their priority.

- IMAP 5: Implement community formation part of Infomap algorithm (3 points)
- IMAP 7: Improve performance of current multi-level Infomap algorithm (4 points)

The team members indicated their willingness to work on the following user stories.

- Every team member will:
  - IMAP 5
  - IMAP 7

*Sprint 6***Sprint 6 Planning Meeting Minutes:**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kyle Gareth Brindle, Anthony Abello**

Start time: **9:30PM**

End time: **9:45PM**

After discussion, the velocity of the team was estimated to be **36 points ( 5 per team member)**.

The product owner chose the following user stories to be done during the next sprint. They are ordered based on their priority.

- IMAP 9: Connect current Infomap implementation to Cytoscape (6 points)

The team members indicated their willingness to work on the following user stories.

- Every team member will:
  - IMAP 9: Connect current Infomap implementation to Cytoscape (6 points)



*Sprint 7***Sprint 7 Planning Meeting Minutes:**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kyle Gareth Brindle, Anthony Abello**

Start time: **9:30PM**

End time: **9:45PM**

After discussion, the velocity of the team was estimated to be **40 points ( 5 per team member)**.

The product owner chose the following user stories to be done during the next sprint. They are ordered based on their priority.

- DOC 1: Prepare the documentation for the entire project(8 points)

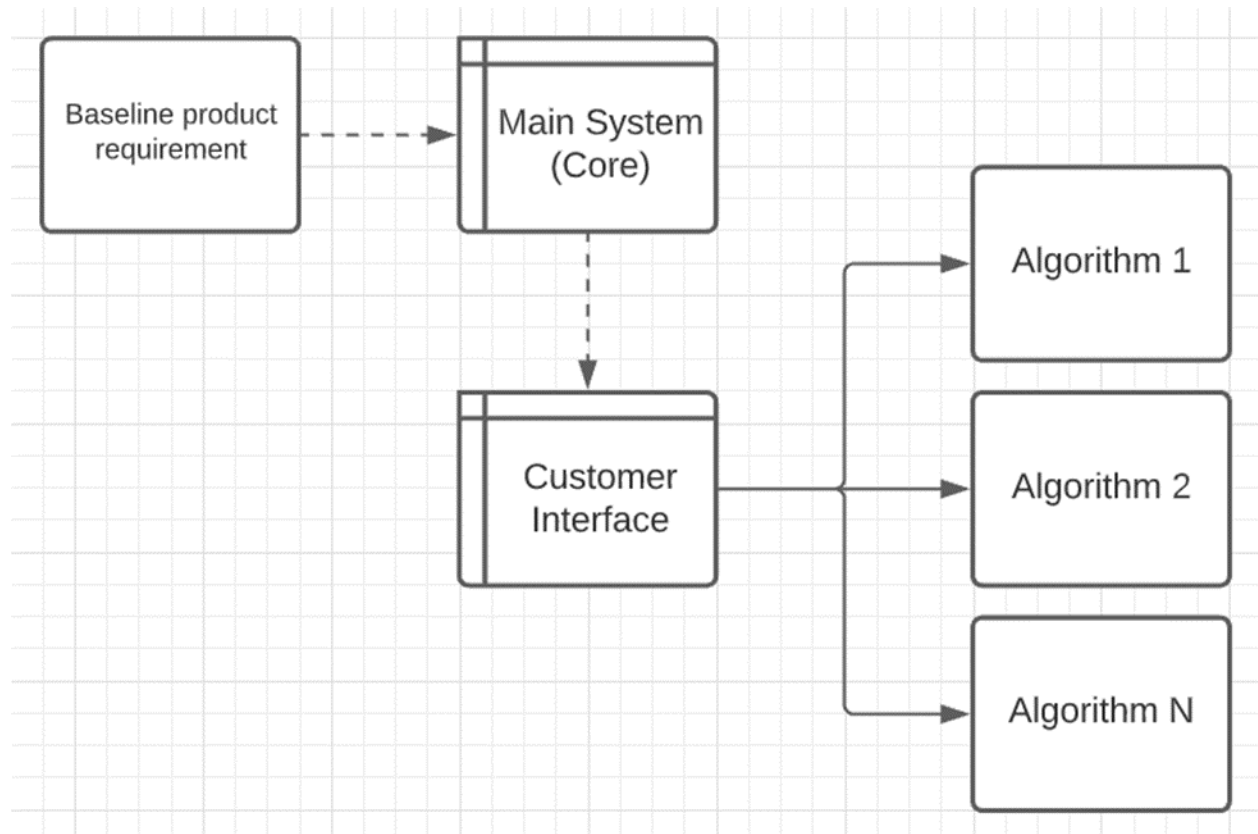
The team members indicated their willingness to work on the following user stories.

- Every team member will:
  - DOC 1: Prepare the documentation for the entire project (8 points)

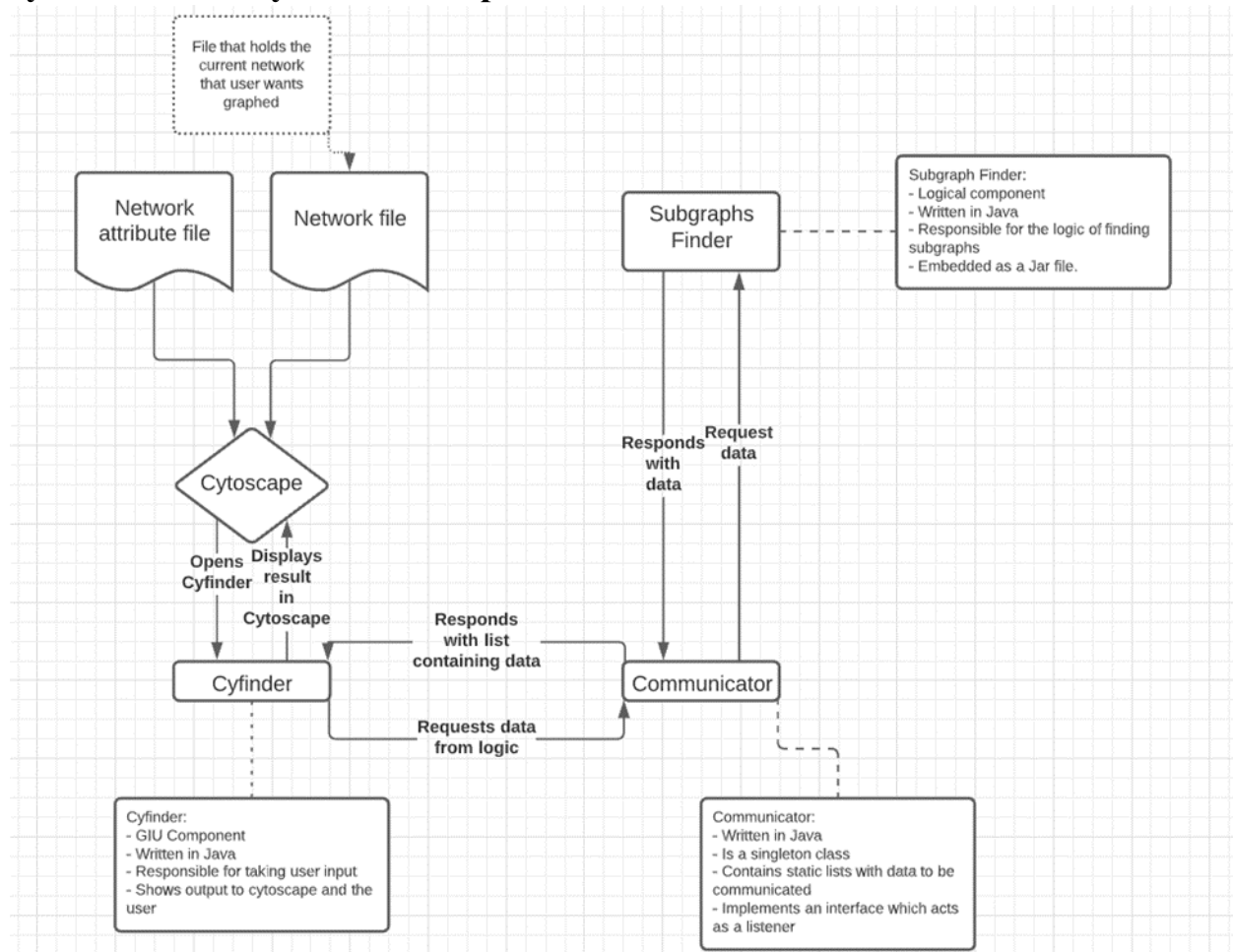
## SYSTEM DESIGN

This section contains information on the design decisions that went into this project. The architecture patterns are outlined and explained. The entire system is shown in a package diagram and the subsystems are explained. Finally, the design patterns used in the project are discussed.

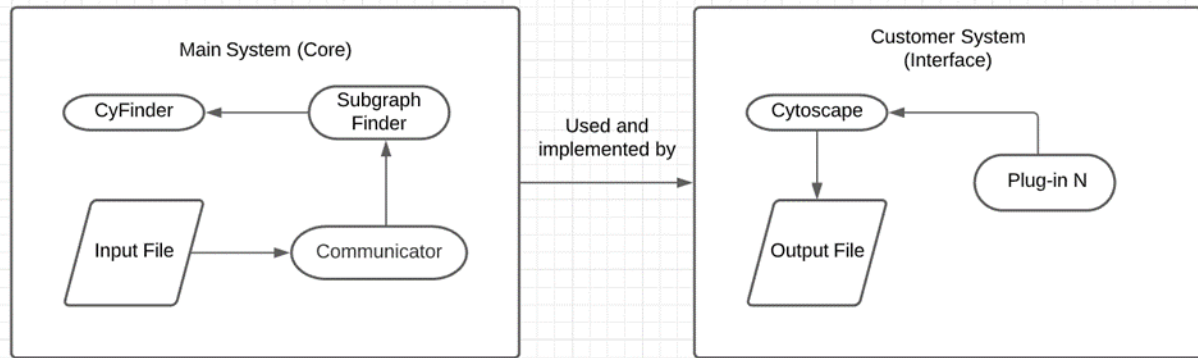
### Architectural Patterns

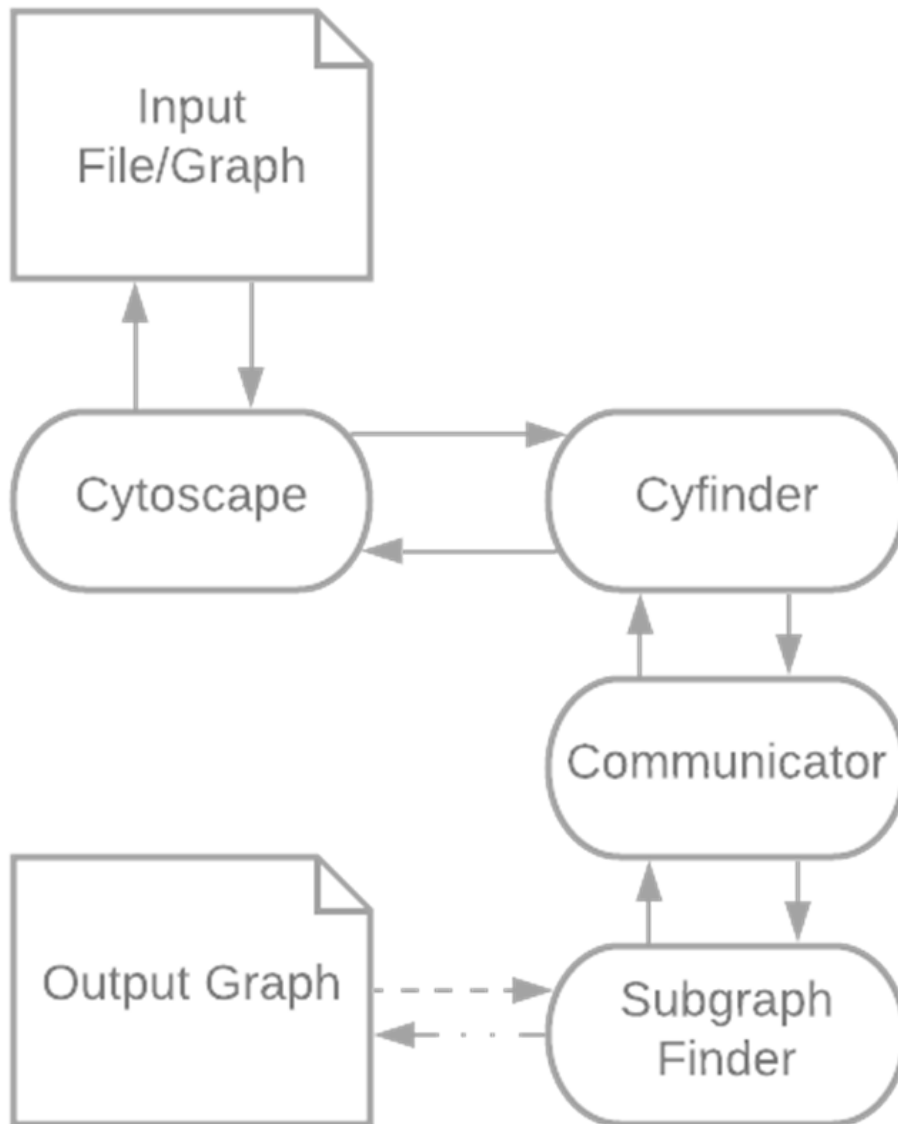


## System and Subsystem Decomposition

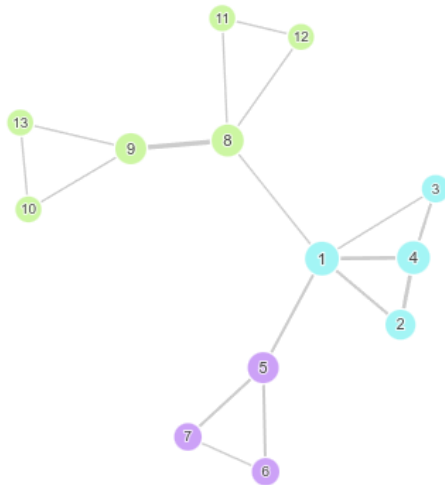


## Project Dependencies Diagram



**Design Patterns**

## SYSTEM VALIDATION



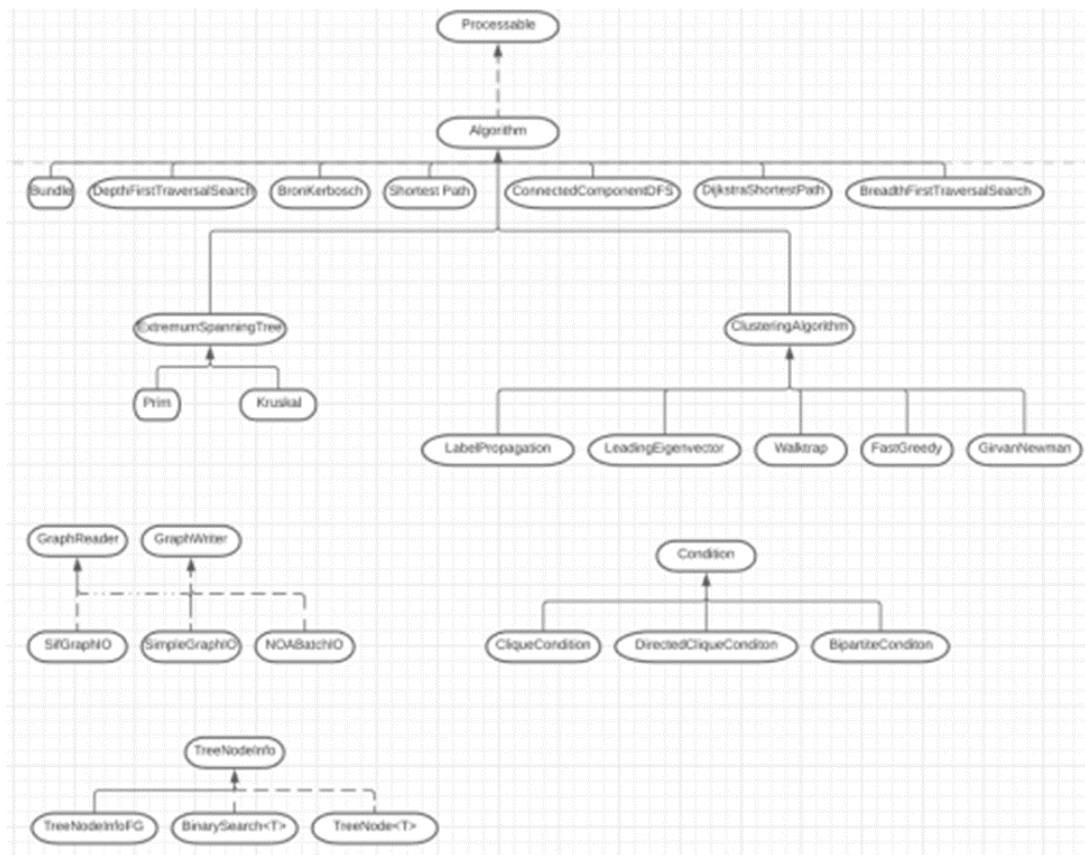
```
# v2.6.1
# ./Infomap network . --clu --ftree
# started at 2023-04-18 23:33:12
# completed in 0.074 s
# partitioned into 2 levels with 3 top
# codelength 2.60244 bits
# relative codelength savings 25.5301
# flow model undirected
# module level 1
# node_id module flow
1 1 0.155172
4 1 0.137931
2 1 0.0862069
3 1 0.0517241
8 2 0.12069
9 2 0.103448
10 2 0.0344828
11 2 0.0344828
12 2 0.0344828
13 2 0.0344828
5 3 0.103448
6 3 0.0517241
7 3 0.0517241
```

```
{0=[1, 2, 3, 4], 1=[5, 6, 7], 2=[8, 9, 10, 11, 12, 13]}
2.6024437910254266
```

## VALIDATION IMAGE 1 - COMPARISON OF RESULTS FROM INFOMAP WEBSITE TO CYFINDER ALGORITHM

## APPENDIX

### Appendix A - UML Diagrams



### Appendix B - User Interface Design

## Appendix C - Sprint Review Reports

### Sprint 1 Review Meeting Minutes

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kyle Gareth Brindle, Anthony Abello**

Start time: **7:30PM**

End time: **8:00PM**

After a show and tell presentation, the implementation of the following user stories was accepted by the product owners:

- Learn basics of Cytoscape software program
- Go over documentation and source code from the previous team
- Read other resources provided by the owner
- PREP 1: Read CyFinder tutorial and run examples
- PREP 2: Go over partial Info Map implementation

The following ones were rejected and moved back to the product backlog to be assigned to a future sprint at a future Sprint Planning meeting.

- N/A

### Sprint 2 Review Meeting Minutes

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kyle Gareth Brindle, Anthony Abello**

Start time: **8:30PM**

End time: **9:00PM**

After a show and tell presentation, the implementation of the following user stories was accepted by the product owners:

- PREP 3: Deepen understanding of Info Map algorithm (3 points)
- PREP 4: Understand current state of Info Map implementation (3 points)
- PLAN 1: Create the roadmap for finishing Info Map implementation (2 points)



The following ones were rejected and moved back to the product backlog to be assigned to a future sprint at a future Sprint Planning meeting.

- IMAP 1: Implement the InfoMap logic (Reason: Rewriting the infomap algorithm to include huffman encoding for module entrance and exits, hoping to make the average bit length accurate to the algorithm from the infomap website)
- IMAP 2: Test InfoMap against online InfoMap implementation

### **Sprint 3 Review Meeting Minutes**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kyle Gareth Brindle, Anthony Abello**

Start time: **8:30PM**

End time: **9:00PM**

After a show and tell presentation, the implementation of the following user stories was accepted by the product owners:

- FIX 1: Fix the info map implementation from the previous semester
- TEST 1: Test the new implementation against the online map equation info map
- DOC 1: Write documentation for the implementation of the info map algorithm
- IMAP 3: Implement map equation average bit length calculation using power iteration method

The following ones were rejected and moved back to the product backlog to be assigned to a future sprint at a future Sprint Planning meeting.

- IMAP 4: Test implementation of power iteration method

### **Sprint 4 Review Meeting Minutes**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kyle Gareth Brindle, Anthony Abello**

Start time: **9:30PM**

End time: **10:00PM**

After a show and tell presentation, the implementation of the following user stories was accepted by the product owners: All.

- IMAP 4: Test implementation of power iteration method for undirected networks (2 points)
- IMAP 6: Integrate IMAP 3 into Infomap algorithm as a modularity gain function (4 points)

The following ones were rejected and moved back to the product backlog to be assigned to a future sprint at a future Sprint Planning meeting.

- IMAP 5: Implement community formation part of Infomap algorithm (3 points)

### **Sprint 5 Review Meeting Minutes**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kyle Gareth Brindle, Anthony Abello**

Start time: **9:30PM**

End time: **10:00PM**

After a show and tell presentation, the implementation of the following user stories was accepted by the product owners: All.

- IMAP 5: Implement community formation part of Infomap algorithm (3 points)
- IMAP 7: Improve performance of current multi-level Infomap algorithm (4 points)
- IMAP 8: Document history of performance improvements (2 points)

The following ones were rejected and moved back to the product backlog to be assigned to a future sprint at a future Sprint Planning meeting.

- IMAP 9: Connect current Infomap implementation to Cytoscape (6 points)

## **Sprint 6 Review Meeting Minutes**

Project: **App Development for Analyzing Biological Networks**

Attendees: **Slav Lazurenko, Osmany Pujol, Sage A Pages, Kyle Gareth Brindle, Anthony Abello**

Start time: **9:30PM**

End time: **10:00PM**

After a show and tell presentation, the implementation of the following user stories was accepted by the product owners: All.

- IMAP 9: Connect current Infomap implementation to Cytoscape (6 points)
- IMAP10: Adjust weight control in Cytoscape for infomap

The following ones were rejected and moved back to the product backlog to be assigned to a future sprint at a future Sprint Planning meeting.

- None

## **Appendix D - User Manuals, Installation/Maintenance Document, Shortcomings/Wishlist Document and other documents**

## **REFERENCES**

Edler, D., Bohlin, L., & Rosvall, M. (2017). Mapping higher-order network flows in memory and multilayer networks with Infomap. *Scientific Reports*, 7(1), 1-13. doi: 10.1038/srep46689

Rosvall, M., & Axelsson, D. (2009). The map equation. *The European Physical Journal Special Topics*, 178(1), 13-23. doi: 10.1140/epjst/e2010-01179-1

D. Edler, A. Holmgren and M. Rosvall, The MapEquation software package, available online at [mapequation.org](http://mapequation.org).