MSF - Modulated Sub-graph Finder

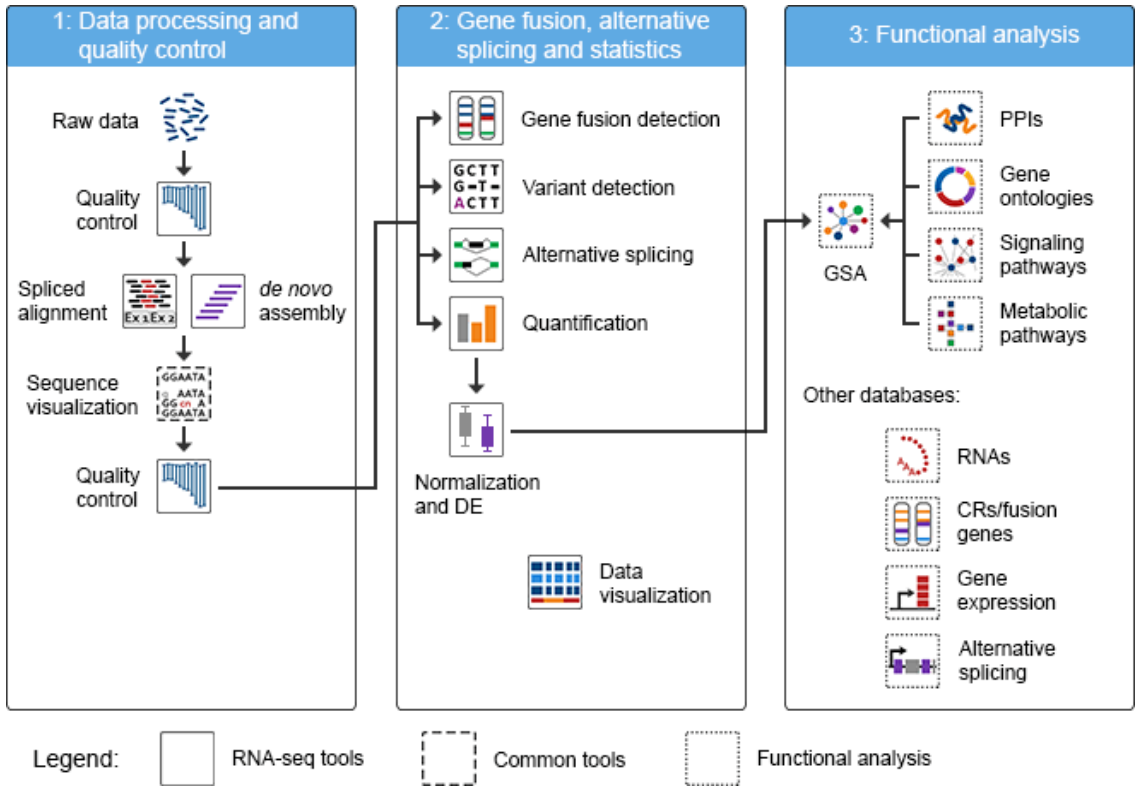
Mariam Farman

University of Vienna, Faculty of Chemistry, Department of Theoretical Chemistry

Ivo L. Hofacker

February 16, 2018

Motivation


1: Data processing and quality control
Raw data
Quality control
Spliced alignment
de novo assembly
Sequence visualization
Quality control
2: Gene fusion, alternative splicing and statistics
Gene fusion detection
Variant detection
Alternative splicing
Quantification
Normalization and DE
Data visualization
3: Functional analysis
GSA
PPIs
Gene ontologies
Signaling pathways
Metabolic pathways
Other databases:
RNAs
CRs/fusion genes
Gene expression
Alternative splicing
Legend:
RNA-seq tools
Common tools
Functional analysis

Limitations of Conventional Techniques

Over Representative Analysis

- User defined cut-off for the log-fold change/p-value
- topology of the genes in the pathways are ignored
- Assumes pathways are independent of each other and ignores the fact that biological pathways cross-talk and overlap

Functional Class Scoring

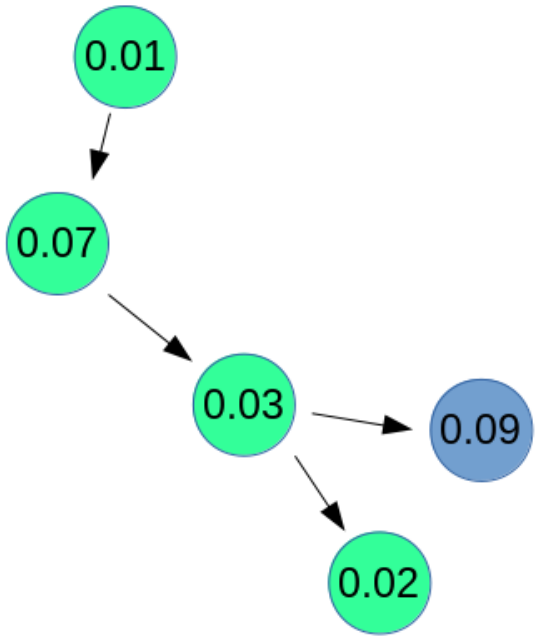
- lacks the topology, cross-talk and overlap of the pathways

Why Use MSF

- Find modulated sub-graphs from whole cell signaling network
- Consider all differentially expressed genes
- Find predefined pathways connected in MSF identified sub-graphs
- Possible sources (targets) of the modulated sub-graphs

How Does MSF Work?

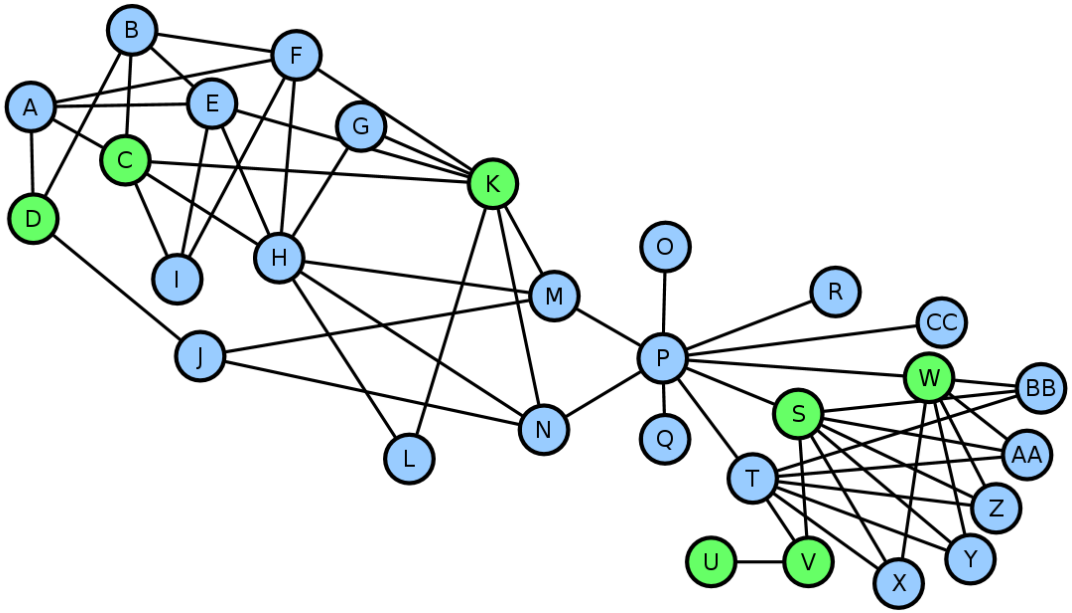
- Heuristic Approach in Java
- Results from DEG analysis as input
- Good quality network (Directed)
- Combining individual gene p-value to generate sub-graph p-value (Hartung)
- Most upstream genes are the sources (Perturbation Points)



Algorithm

Network

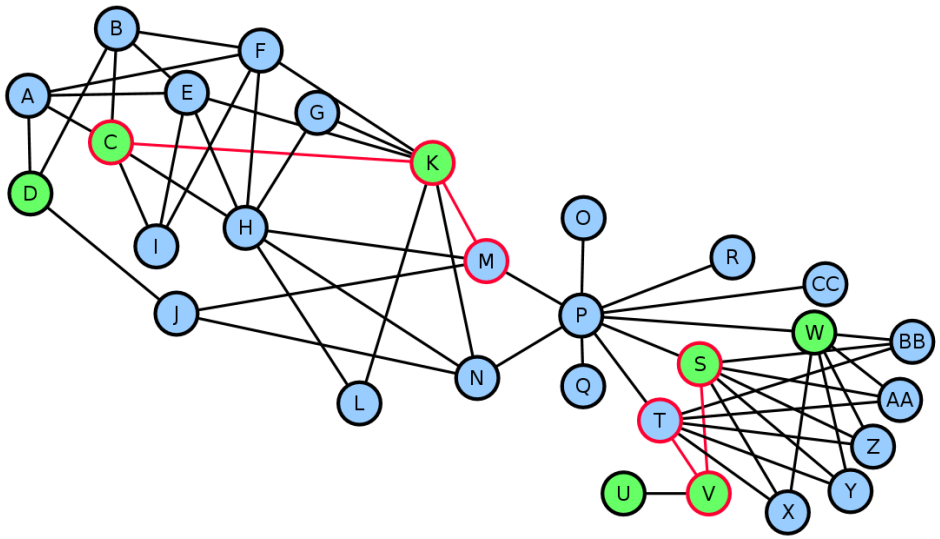
A network with nodes as genes and edges as interactions
Green nodes - Significant p-value genes
Blue nodes - Insignificant p-value genes



Algorithm

Initial

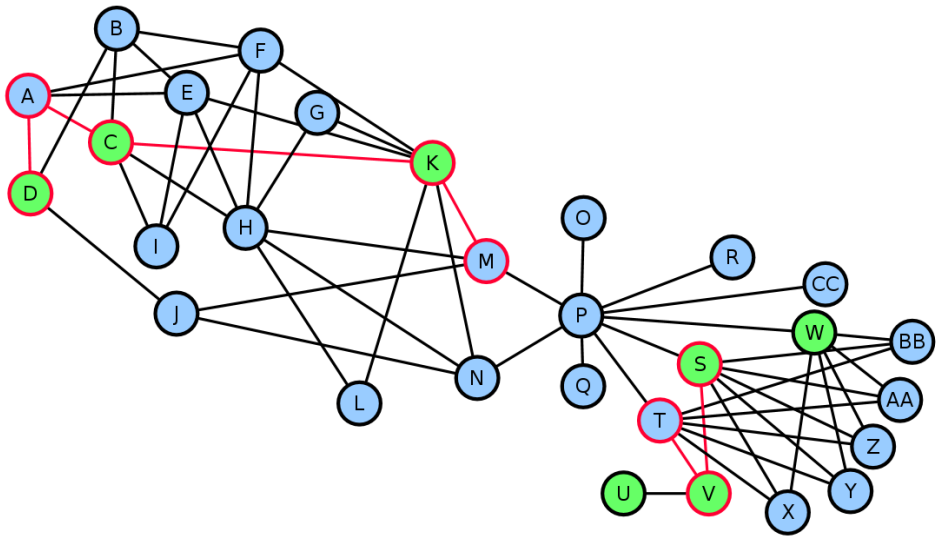
Start with the most significant p-value node - Keep adding next most significant node until new combined p-value $<$ previous combined p-value



Algorithm

Extension

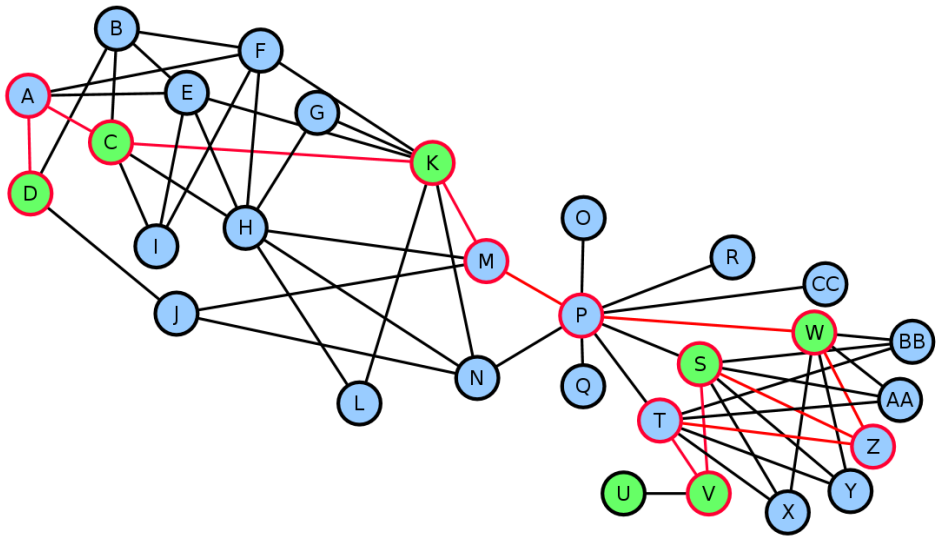
Check if they could further be extended beyond the immediate neighborhood



Algorithm

Merging

Merging modulated sub-graphs by depth first search - new combined p-value $<$ combined p-value sub-graph1 & combined p-value sub-graph2



Case Study

Ebola Virus

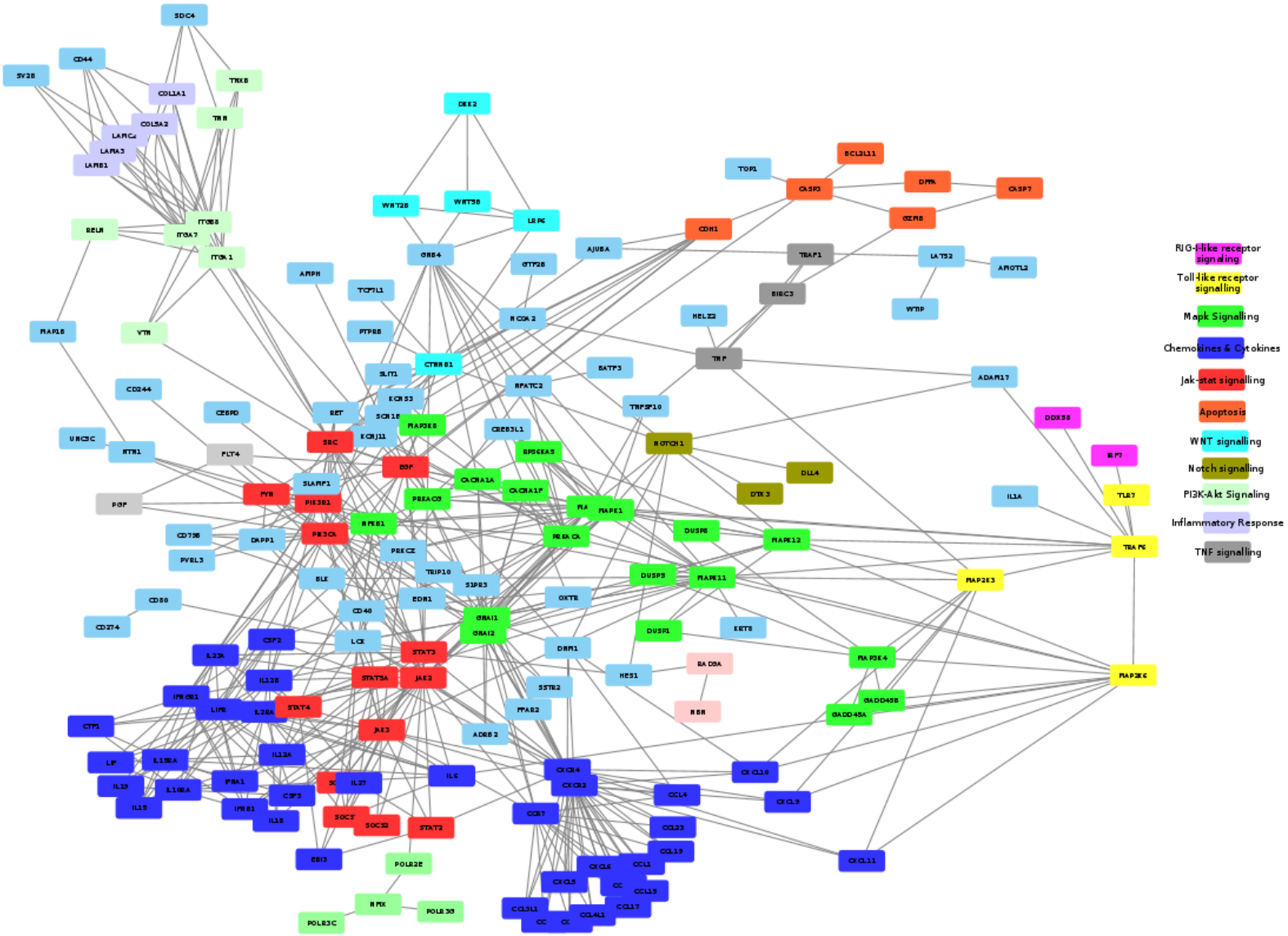
Ebola Virus (EBOV) belongs to the Filoviridea family, that are filamentous, enveloped and single stranded RNA viruses. The initial targets of EBOV are the macrophages and dendritic immune cells. Ebola Virus inhibits the critical innate immune response of the host, which includes the activation of alpha/beta interferon (IFN- α/β)

- The EBOV RNA-seq DEG data published 2017
- The human biological interactions network from Reactome

Case Study

	6hpi	**1dpi**	**2dpi**
IFNA1	yes	yes	yes
IFNB1	yes	no	yes
IFNA1 p-value	0.12	4.16E-13	0.025
IFNB1 p-value	0.24	1.56E-39	6.95E-20

Modulted Sub-graph for 6hpi



Comparison with SPIA & Reactome (6hpi)

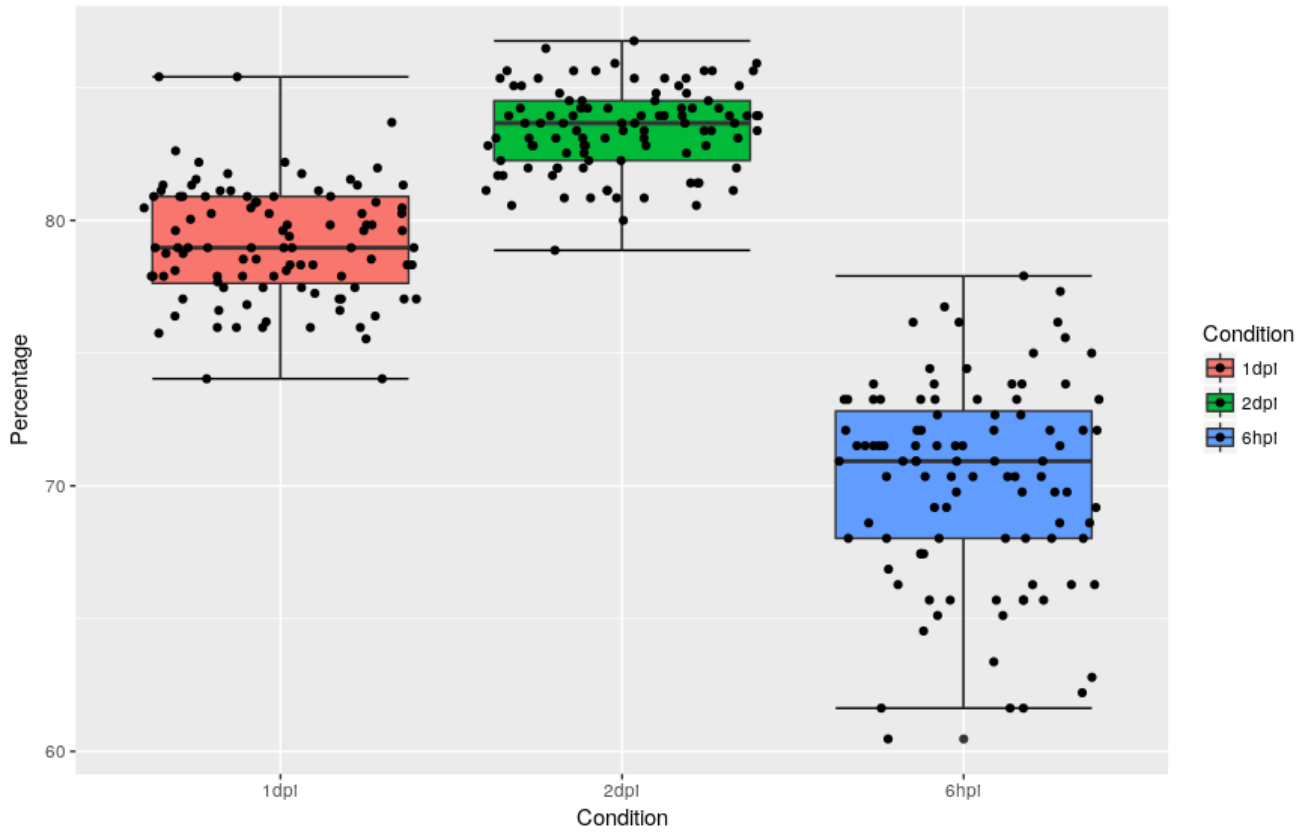
SPIA

- Toll-like receptor signaling
- RIG-I-like receptor signaling
- TNF signaling & IL- signaling

Reactome

- Toll-like receptor signaling

Robustness


Percentage
80
70
60
1dpI
2dpI
6hpI
Condition
Condition
1dpI
2dpI
6hpI

MSF available on github

`https://github.com/MariamFarman/Modulated-SubPath-Finder`

Try it !

Acknowledgment

- Ivo Hofacker
- Fabian Amman