



Implementing Nemo Profile in MASS Bionet

By Drew Andersen



Overview

Problem: Add Nemo Profile to a MASS implementation of network motif finder.

Challenges:

- Locating code base
- Running code across clusters
- Learning core concepts
- Deciding between C++ or Java

Network Motifs

Statistically significant sub-graphs or patterns

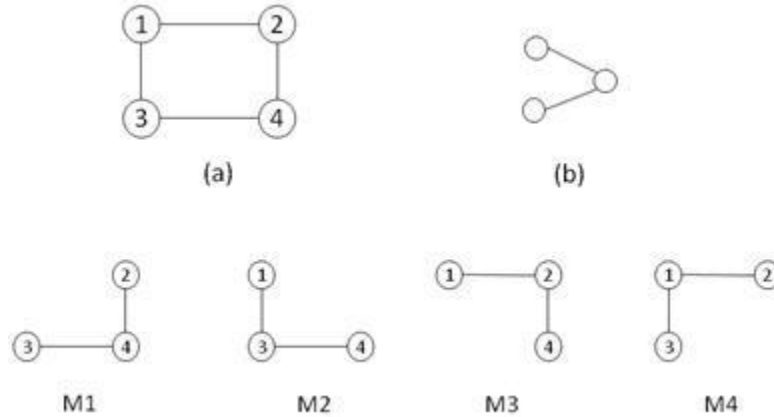


Figure 1

Biological Applications

Protein-protein interactions

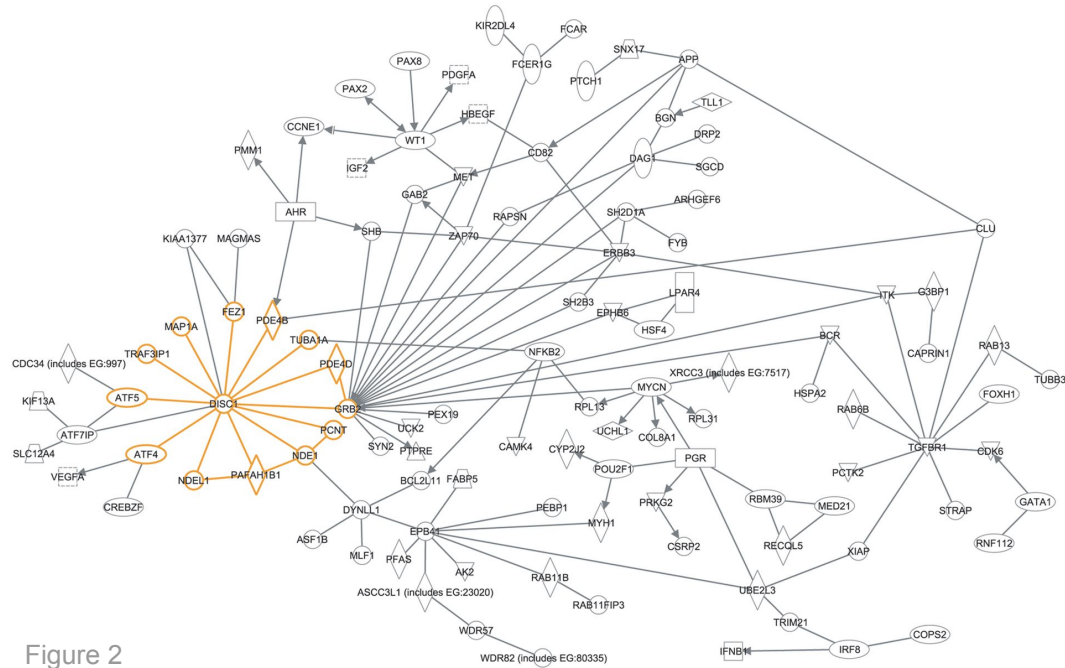
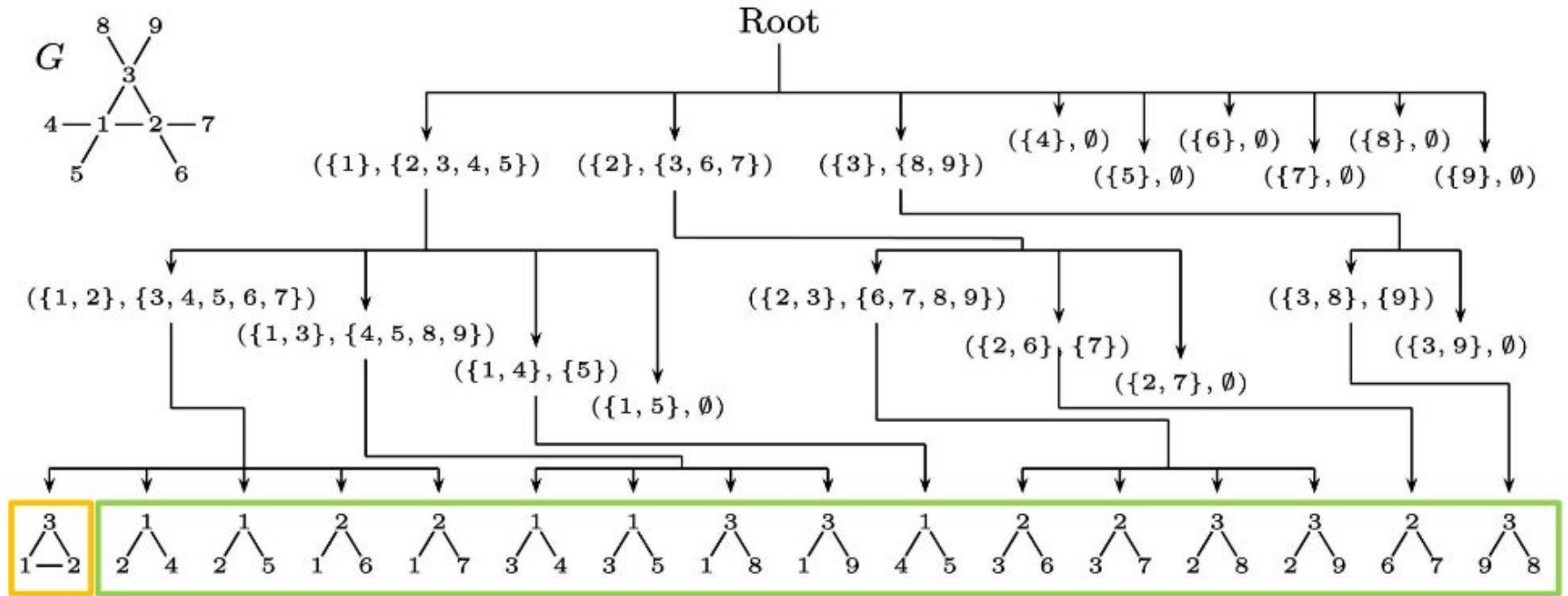


Figure 2

Algorithm: Enumerate Subgraphs (ESU)



Network Motif

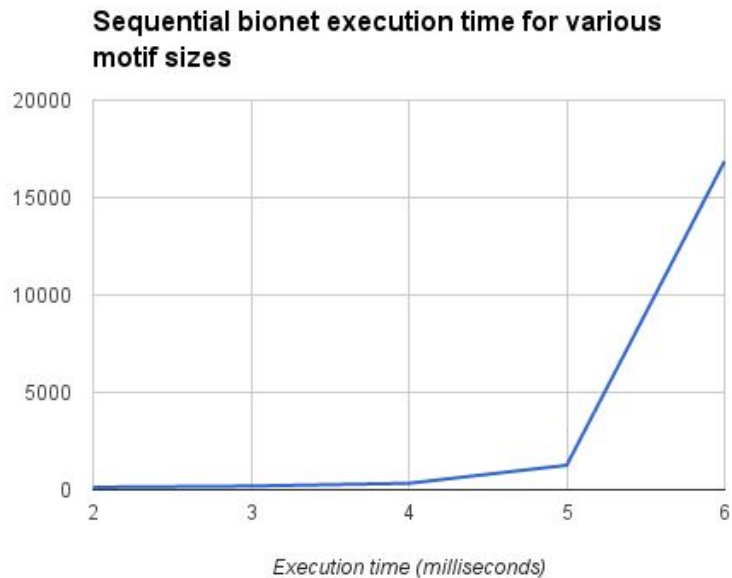
Label	Frequency of Target
C~	110
CF	91848
Cr	1169
CR	59745
C^	606
CN	8148

Bionet

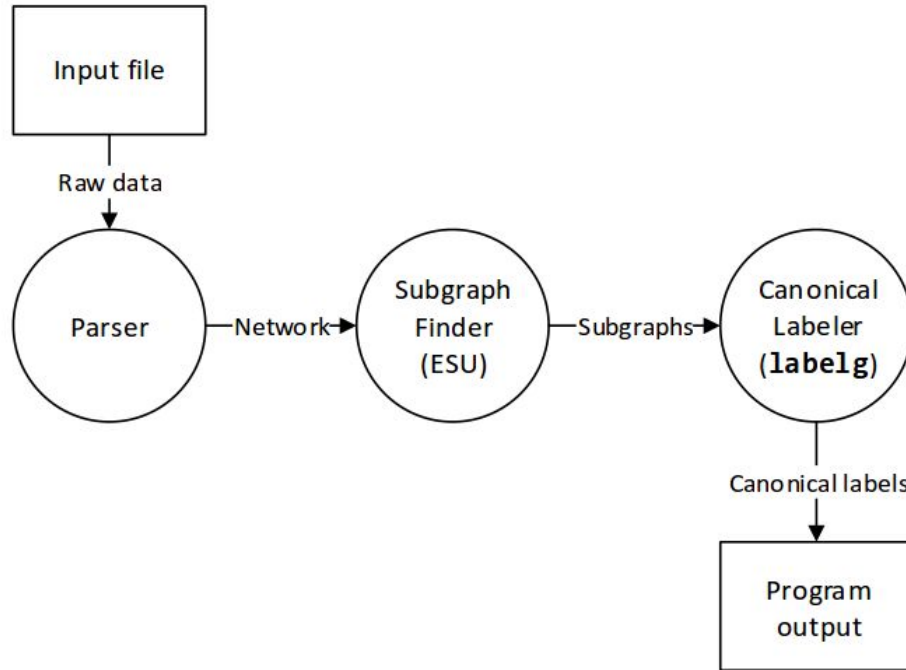
Java implementation of ESU algorithm written by former student Matthew Kipps:

1. Sequential
2. MASS Agent-based
3. MASS Places-based
4. MPI

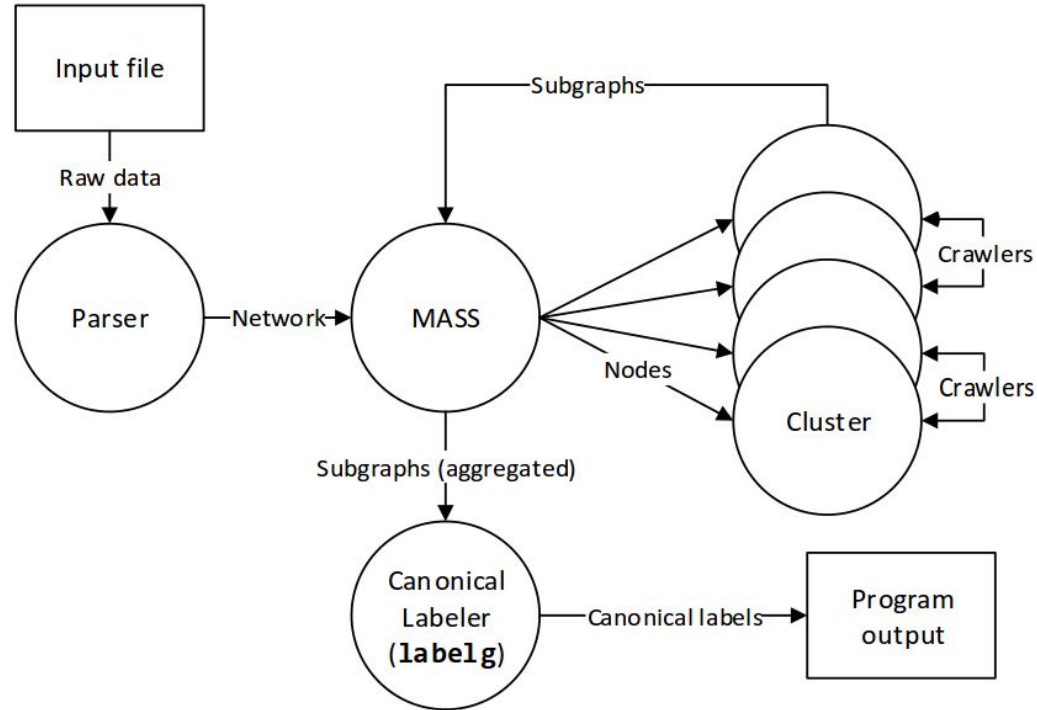
Why parallelize?



Sequential Bionet



MASS Agent-Based Bionet



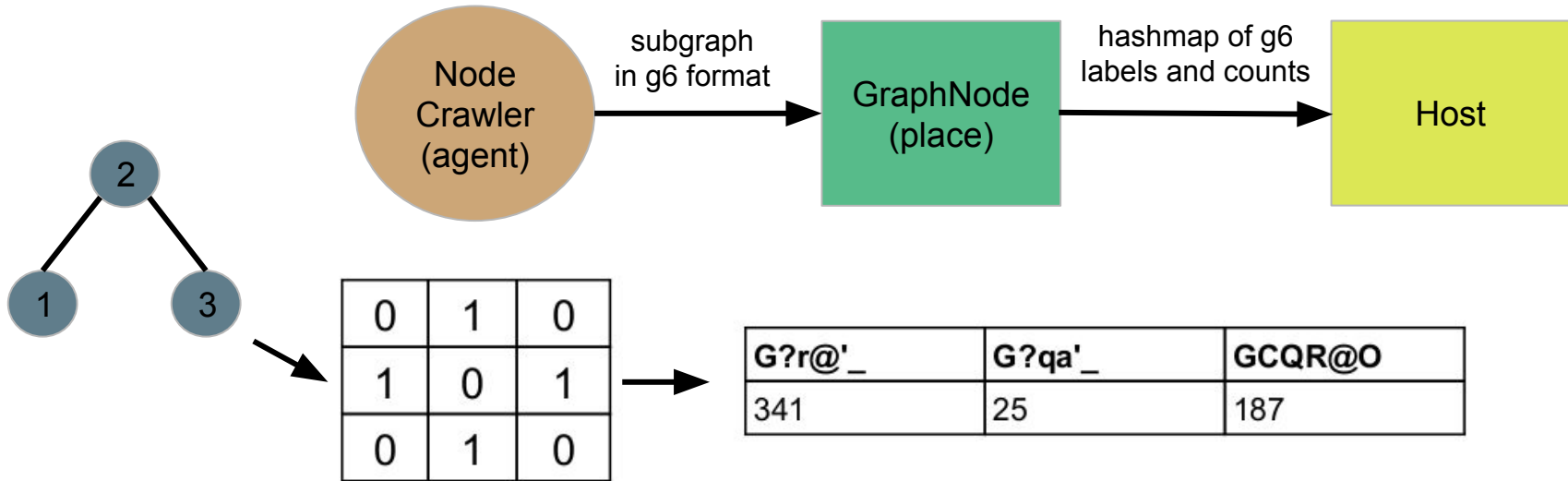
Nemo Profile

What if we want to know to which motifs each node belongs?

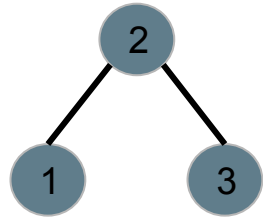
Node	C~	CF	Cr	CR
1	54	2004	2	0
2	74	80	254	95
3	43	1543	84	135
...
n	105	765	90	437

Problem

How to carry node data back to the host in order to implement Nemo Profile?

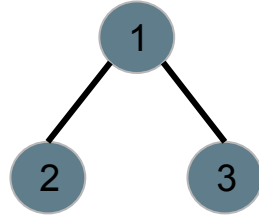


Why convert from g6 format?



0	1	0
1	0	1
0	1	0

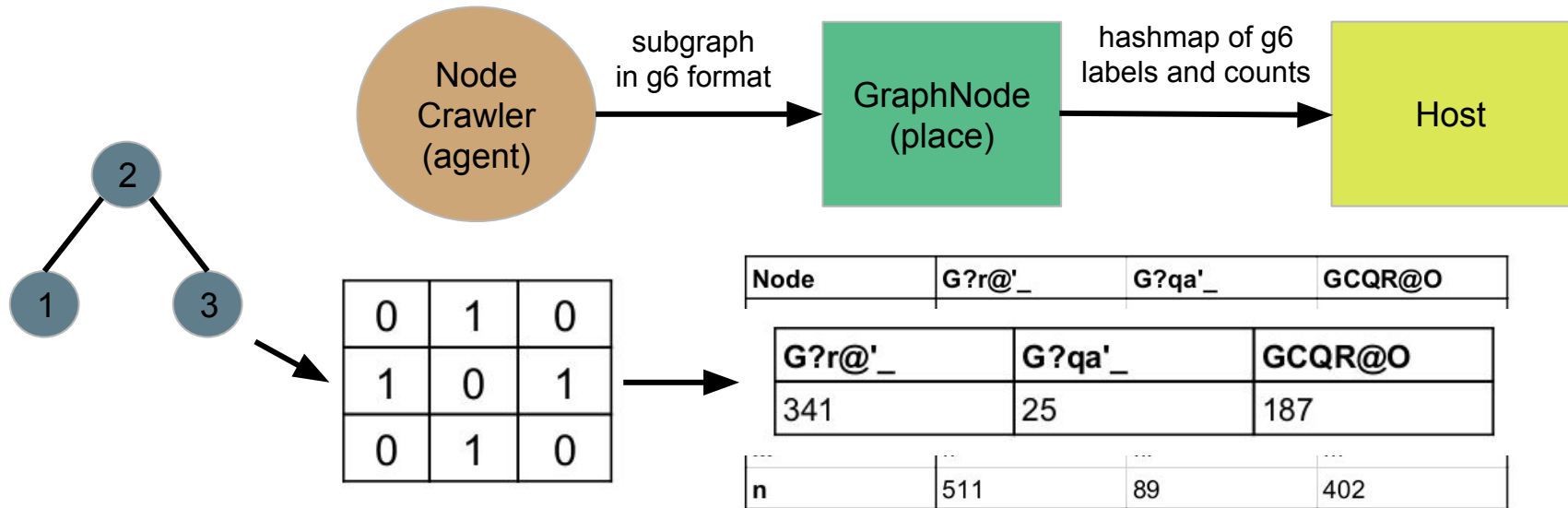
=



0	1	1
1	0	0
1	0	0

Solution

Convert g6 hash map into a hash map of hash maps



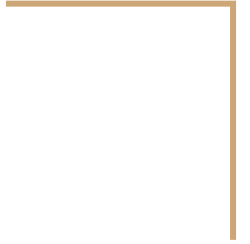
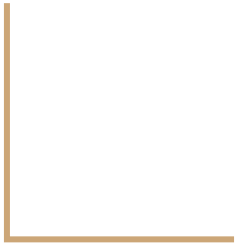
Next steps

1. Implementation (in process)
2. Experiments
3. Research paper / Conference: The 6th IEEE International Conference on Big Data and Cloud Computing (BDCloud 2016)

Long term

1. Reduce agent explosion
2. Add the motif-finding portion of the algorithm

Questions?



References

1. Figure 1: CC BY-SA 3.0, <https://en.wikipedia.org/w/index.php?curid=36822143>
2. Figure 2: By Hennah, Porteous - [The DISC1 Pathway Modulates Expression of Neurodevelopmental, Synptogenic and Sensory Perception Genes <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0004906>], CC BY 2.5, <https://commons.wikimedia.org/w/index.php?curid=7681986>
3. Figure 3: CC BY-SA 3.0, <https://en.wikipedia.org/w/index.php?curid=36822191>
4. Various figures: http://depts.washington.edu/dslab/MASS/reports/MattKipps_report_au14.pdf