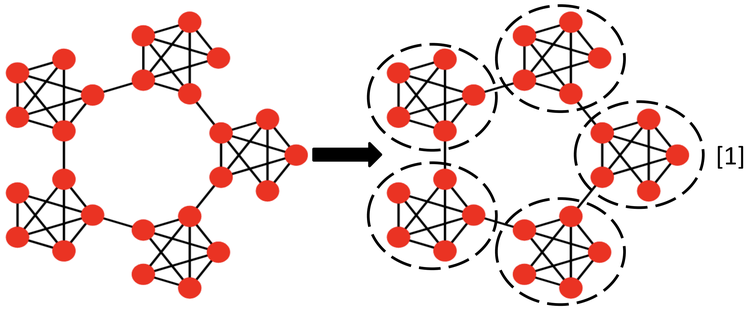
Community Detection: Using a Randomized Approach and Applications for *C. Elegans* Neurons

Jake Jasmer, Tony Ni, Aidan Roessler, Yang Tan; Advised By Layla Oesper

Carleton

The Problem

We want to group nodes together into communities. This grouping should match how the nodes are connected, so that closely connected nodes are in the same community and disconnected nodes are in different communities.



[1]

Figure 1. Communities in a caveman graph

Input: an undirected, graph (weighted or unweighted)

Output: a set of communities in that graph

Quantifying Sets of Communities

We want to compare our set of communities to a random set of communities. We can do this using modularity (Q).

Modularity is a function with output between -1 and 1 for quantifying the value of a set of communities. A random set of communities should result in a value around 0.

Actual weight between nodes i & j

Expected weight between nodes i & j (probability of two half-edges being connected)

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \qquad [2]$$

Total number of half-edges

1 if nodes i & j are in the same community, 0 otherwise

Figure 2. The components of the modularity equation

Types of Algorithms

Divisive (Girvan-Newman):

- Removes edges iteratively one at a time, leaving still connected groups of nodes as communities
- Returns the highest modularity iteration of communities

Agglomerative (Louvain):

- Merges nodes into communities that maximize modularity

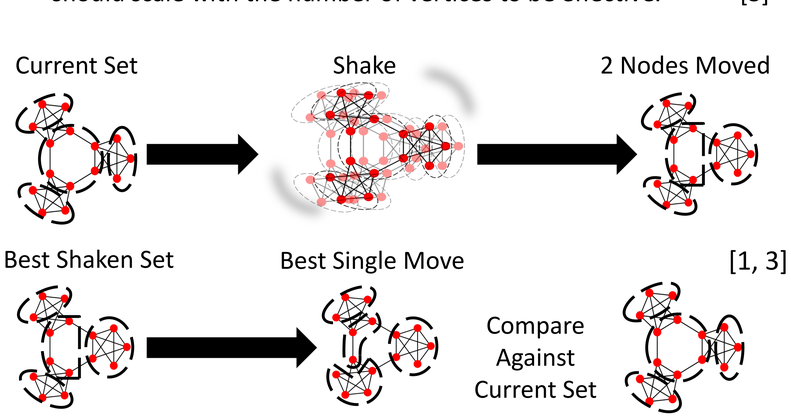
Random (Basic Variable Neighborhood Search or BVNS):

- Uses randomness to continuously generate steps that change the communities
  - Picks the steps that improve modularity

BVNS

Basic Variable Neighborhood Search (BVNS) starts with a random set of communities. It then iterates through the following steps on the current set of communities:

1) Shaking: Move k nodes to a different community. Do this for 1-k_{max} nodes, resulting in k_{max} different shaken sets of communities.
2) Local Search: Take the best shaken set. Move every node to every other possible community and check the modularity. Return the set with the best modularity. This is an exhaustive search at a distance of 1.
3) Comparison: Compare the output of local search to the original solution we had before shaking. The better set of communities becomes the current set of communities.
4) Iterate: Repeat the previous steps with the current solution a given number of times. This is typically around 100, although it should scale with the number of vertices to be effective. [3]



[1, 3]

Figure 3. The steps of BVNS. This figure shows the steps of a single iteration of BVNS

Comparing BVNS to Other Algorithms

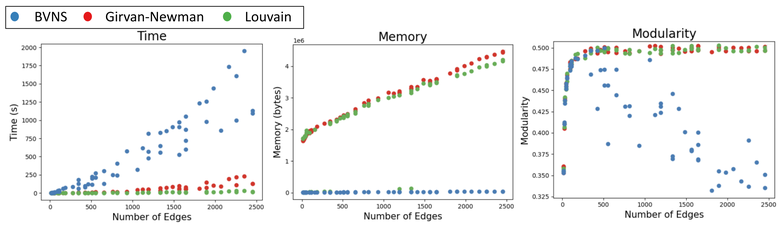
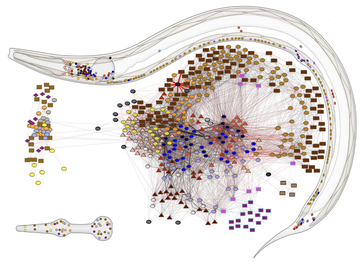


Figure 4. Metrics for each approach on simulated data. These graphs show the time taken, memory used, and modularity for each approach on graphs of different sizes. The number of edges scales with the number of nodes.

- BVNS is slow
- BVNS is memory efficient
- BVNS gives worse modularity as the graph gets larger and the number of iterations doesn't

C. Elegans Neuronal Connection



[4]

Figure 5. Neuronal connections in an adult hermaphrodite *C. Elegans* worm.

The modularity equation changes to account for directed edges.

$$k_i k_j \rightarrow k_i^{\text{out}} k_j^{\text{in}} \qquad [6]$$

- Neurons are nodes and synapses (connections between neurons) are edges
- Communities indicate neuronal circuits with specific functionality
- These feedforward networks are best represented by directed graphs
- Our approaches only work on undirected graphs, so we need to convert neuronal connection graphs to undirected ones

Algorithm	Directed Modularity	Undirected Modularity	ARI	Edit Distance
Directed Approach	0.52	0.49	1.00	0
Girvan-Newman	0.37	0.35	**0.47**	211
Louvain	**0.53**	**0.53**	0.43	**169**
BVNS	0.23	0.25	0.02	406

Table 1. Undirected approaches compared to a directed approach for community detection using a *C. Elegans* dataset. The directed approach was run on the directed version of the graph while the other algorithms were run on the undirected version. Directed modularity is the modularity in the directed graph, and undirected modularity is the same in the undirected graph. ARI measures similarity to the ground truth, in this case directed approach, on a scale from -1 to 1, accounting for randomness. Edit distance is the number of node moves compared to the directed approach solution.

Results For Other Datasets

Dataset	$\|V\|$	$\|E\|$	Algorithm	Communities Found	Time (s)	Peak Memory (mb)	Modularity
Karate Club	34	78	Girvan-Newman	5	0.27	0.07	0.38
			Louvain	**3**	**0.11**	0.07	0.43
			BVNS	5	14.54	**0.03**	**0.44**
College Football	115	613	Girvan-Newman	**10**	28.08	29.19	**0.60**
			Louvain	5	**3.06**	0.20	0.58
			BVNS	15	276.46	**0.04**	0.49
Protein-Protein Interaction	519	679	Girvan-Newman	115	190.33	36.50	**0.95**
			Louvain	162	**59.54**	21.30	0.78
			BVNS	34	19462.02	**19.1**	0.40
Urban Movement	360	2925	Girvan-Newman	5	19592.44	1.40	**0.67**
			Louvain	4	**557.48**	1.35	0.65
			BVNS	26	15581.97	**0.41**	0.27

Table 2. An overview of our results for using each algorithm on choice datasets, including the number of communities found. Modifications were made to transform all datasets into Network X undirected graphs where necessary. The Girvan-Newman algorithm was run with default parameters. The Louvain method was run with two iterations. BVNS was run with 100 iterations and a k_{max} of 3 for all datasets except for Karate Club and College Football where k_{max} was 4.

References

[1] Weisstein, Eric W. *Caveman Graph*. Wolfram Research, Inc., https://mathworld.wolfram.com/CavemanGraph.html. Accessed 26 Feb. 2025.
[2] "Modularity (Networks)." *Wikipedia*, 14 Nov. 2024. *Wikipedia*, https://en.wikipedia.org/w/index.php?title=Modularity_(networks)&oldid=1257345693.
[3] Jovanović, Djordje, et al. "Variable Neighborhood Search Approach to Community Detection Problem." *Numerical Methods and Applications*, edited by Ivan Georgiev et al., Springer Nature Switzerland, 2023, pp. 188–99.
[4] Cook, Steven J., et al. "Whole-Animal Connectomes of Both Caenorhabditis Elegans Sexes." *Nature*, vol. 571, no. 7763, July 2019, pp. 63–71. *www.nature.com*, https://doi.org/10.1038/s41586-019-1352-7.
[5] Emmons, Scott W. "Comprehensive Analysis of the C. Elegans Connectome Reveals Novel Circuits and Functions of Previously Unstudied Neurons." *PLOS Biology*, vol. 22, no. 12, Dec. 2024, p. e3002939. *PLOS Journals*, https://doi.org/10.1371/journal.pbio.3002939.
[6] Leicht, E. A., and M. E. J. Newman. "Community Structure in Directed Networks." *Physical Review Letters*, vol. 100, no. 11, Mar. 2008, p. 118703. *DOI.org (Crossref)*, https://doi.org/10.1103/PhysRevLett.100.118703.