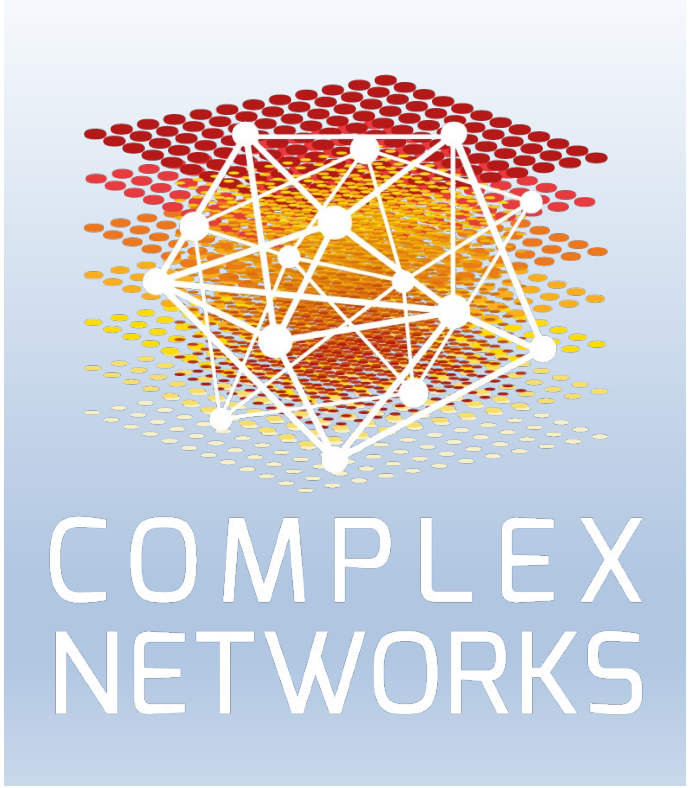
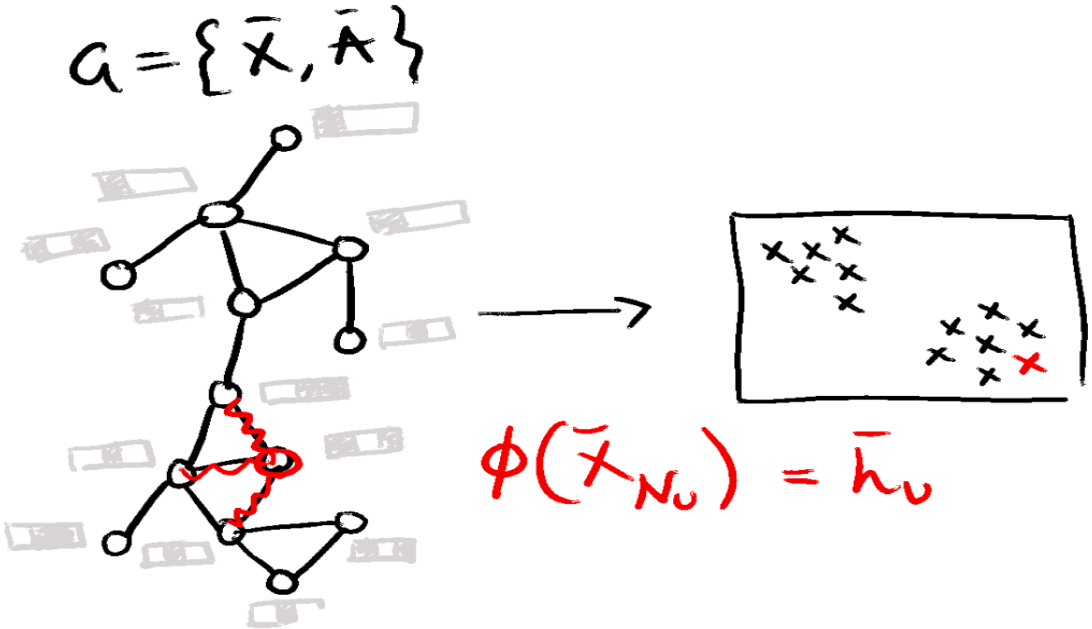
University of BRISTOL



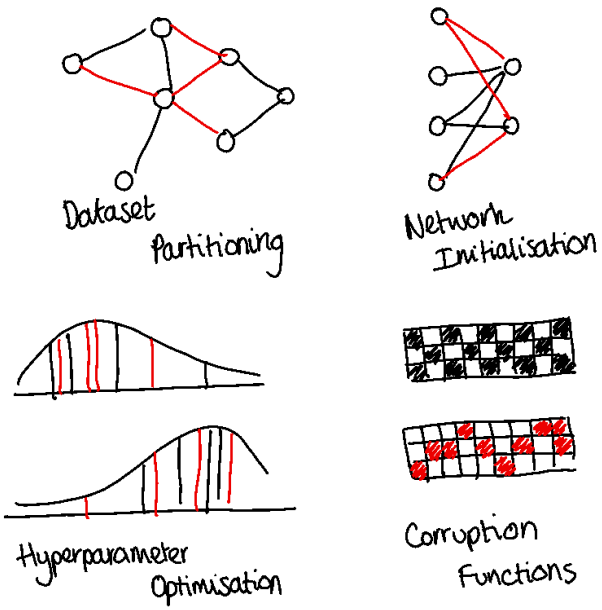
Uncertainty in GNN Learning Evaluations: The Importance of a Consistent Benchmark in Community Detection

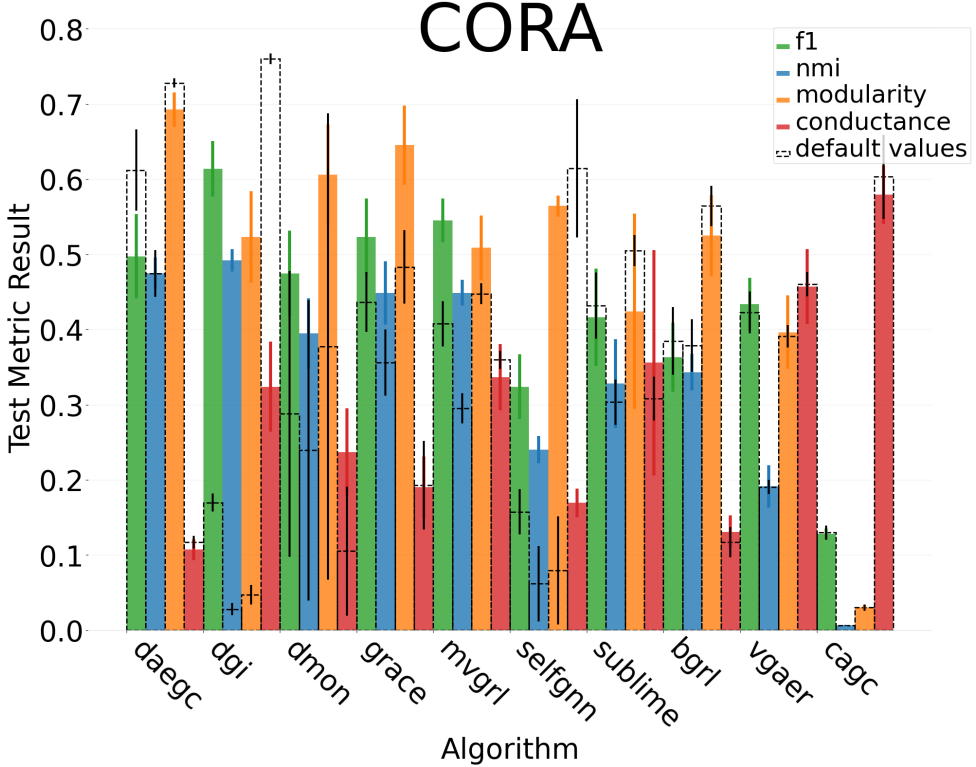
William Leeney and Ryan McConville

Community Detection with GNNs



What is the Problem?





A Solution: W Randomness Coefficient

- $W = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} 1 - \frac{12S}{n^2(a^3 - a)}$
- a: number of algorithms
- n: number of random seeds
- S: sum of squared deviations from mean rank over each random seed.
- $t \in \mathbb{R}^{n \times a}$: test ranking matrix of single dataset on single metric
- $\mathcal{T}$: tests on every dataset and every metric.

```
def w_randomness_coefficient(rankings)
    W = []
    for test in rankings:
        n = test.shape[0] # number of random seeds evaluated on
        a = test.shape[1] # number of algorithms
        S = n*np.var(np.sum(test, axis=0))
        W.append(1 - (12*S)/(n**2*(a**3 - a)))
    return np.mean(W)
```

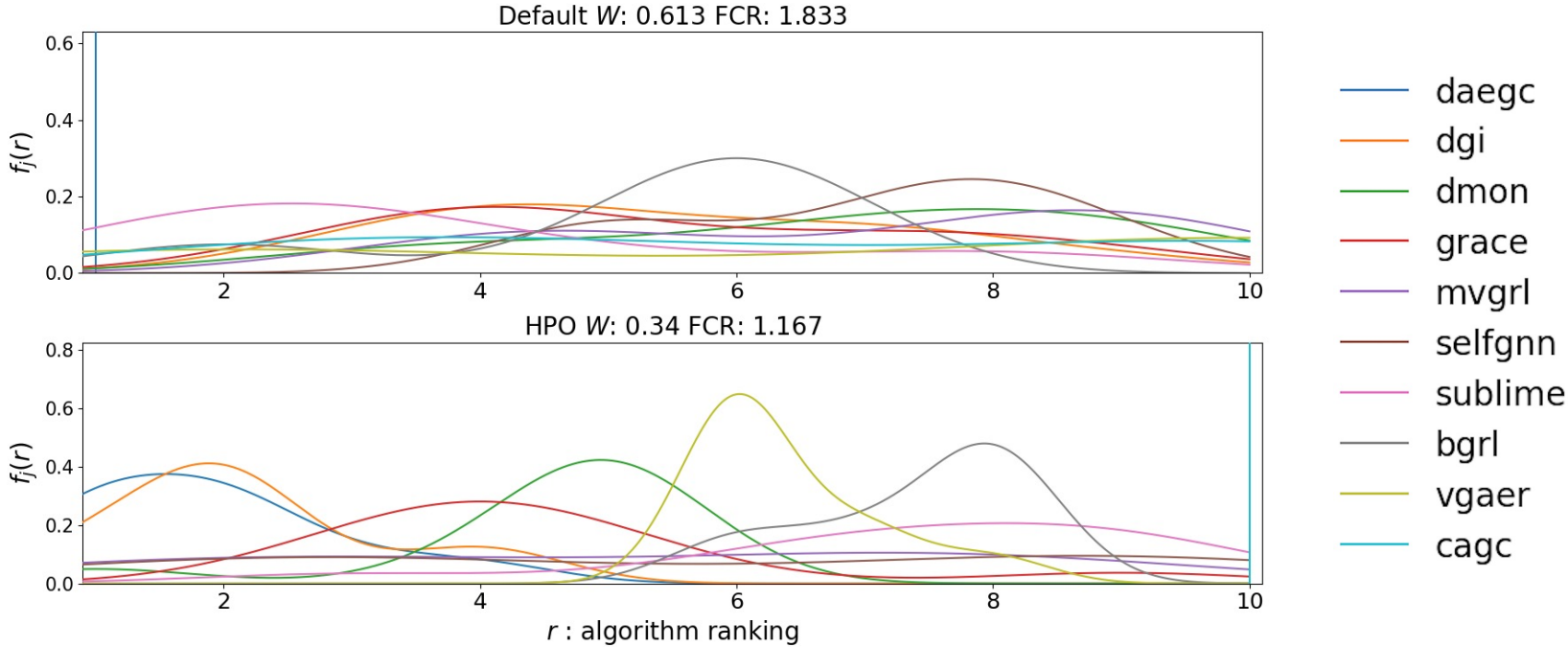
Investigation Framework

- Methods: DAEGC, DMON, DGI, CAGC, GRACE, MVGRL, BGRL, SUBLIME, VGAER
- Metrics: NMI, F1, Modularity, Conductance
- Datasets: AMAC, AMAP, BAT, CITESEER, CORA, UAT, EAT, TEXAS, WISC, CORNELL, DBLP
- Experiment: Default Hyperparameters Vs Hyperparameter Optimisation
- Results: W Randomness Coefficient and Average Framework Comparison Rank

Resource	Allocation
Optimiser	Adam
HPO	MOTPE
Max Epochs	5000
Max HP Trials	300
Training/Validation/Testing Split	0.64/0.16/0.2
Random Seeds	[42, 24, 976, 12345, 98765, 7, 856, 90, 672, 785]
Patience	[25, 100, 500, 1000]
Learning Rate	[0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001]
Decay Rate	[0, 0.05, 0.005, 0.0005]

https://github.com/willleeney/ugle

Algorithm F1 Score Rank Distribution Estimation Comparison on Cora



Results

	Default HPs	HPO
W Randomness Coefficient	0.476	0.489
Framework Comparison Rank	1.829	1.171

Conclusions

- HPO matters: Always carry out a consistent evaluation
- Trustworthy results ≈ Usability in real-world applications.
- W Randomness Coefficient applies to any ML experiment with non-deterministic algorithms

Questions?

- I’m on the Job Market - Thesis Deadline April 2024
- Areas of Expertise: GNNs, Unsupervised Learning, Federated Learning
- Twitter: @WillLeeney