Annealed Heuristic Search for Local Community Detection in Probabilistically Weighted Graphs

Andrei Bardas | Supervisor: Dr. Anna L.D. Latour | EEMCS, Delft University of Technology

1. The Problem

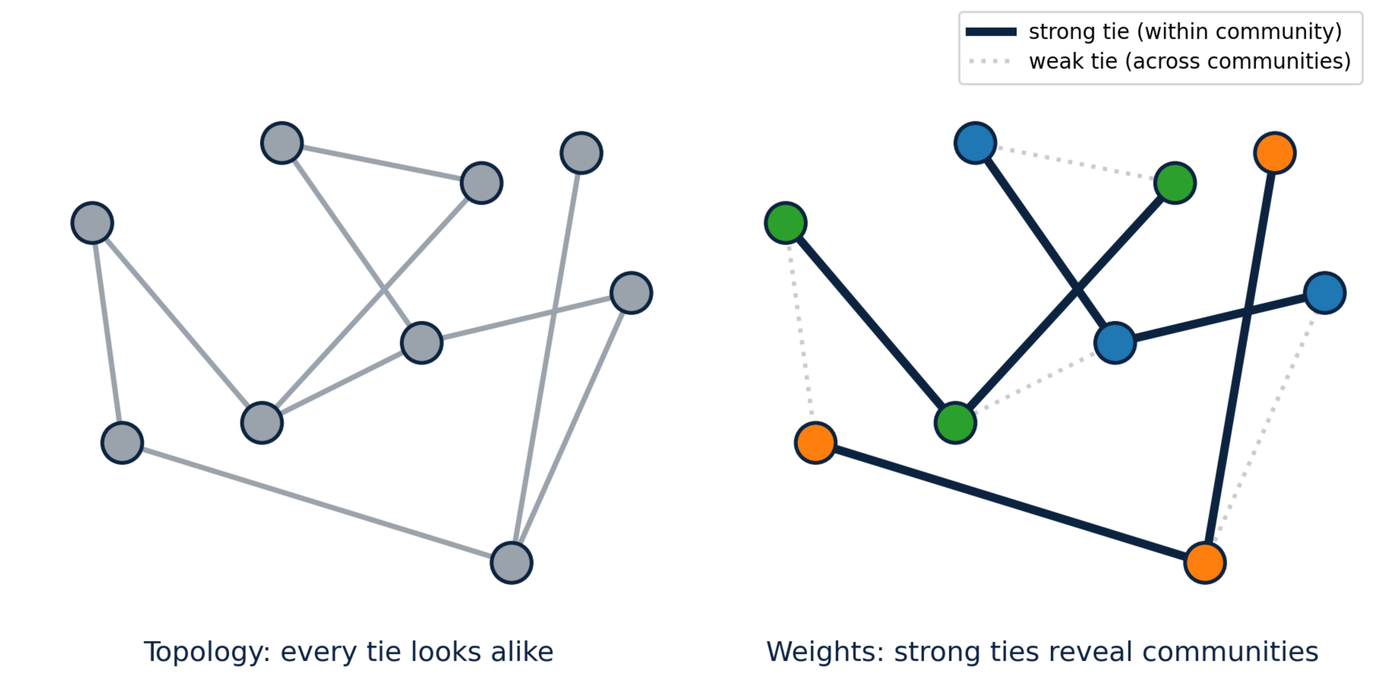
Social Feeds: You are exposed to many accounts, but interact with only a few.

Topology: Dense and weakly structured.

Goal: Grow a community outward from a single seed node.

Edge Weight: The probability that a tie is a genuine within-community connection.

▸ **Membership hides in interaction intensity, not in topology.**



2. Setting & Method

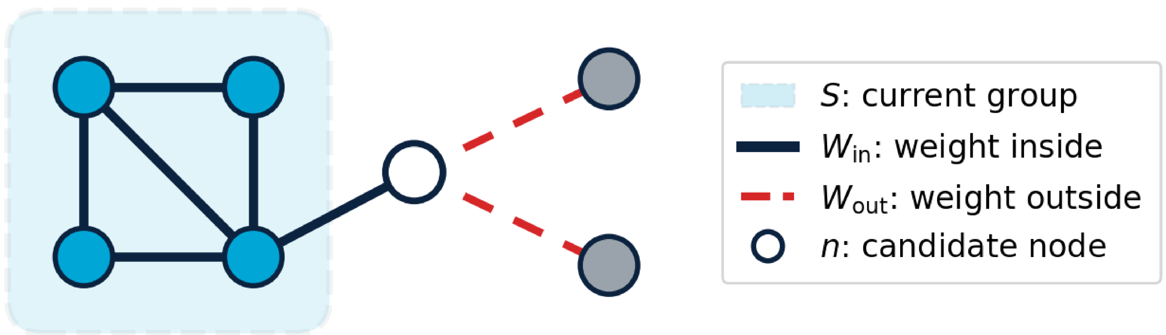
We model the network as a probabilistic graph. A negative-log transform turns edge probabilities into additive costs, so finding the most likely community becomes a **shortest-path search**:

$$C(e) = -\ln p(e)$$

We search with **A***: expand the node of smallest priority, balancing **likelihood** (g, follow strong ties) with **structure** (h, stay tight):

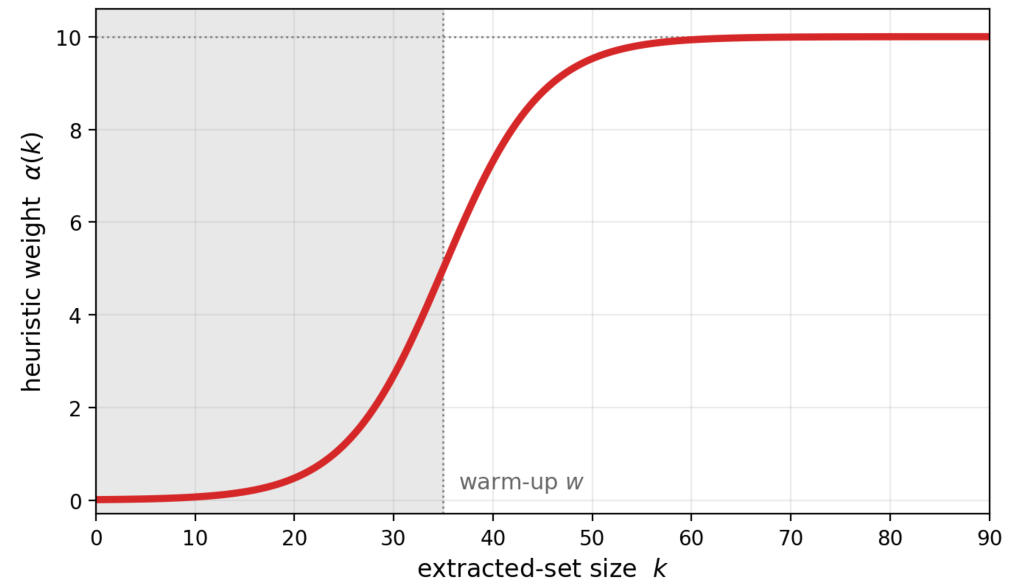
$$f(n) = g(n) + \alpha(k)\, h(n)$$

$h = W_{out}/W_{in}$ penalises a node n whose ties point *outside* the group rather than *inside*.



A **sigmoid warm-up** $\alpha(k)$ suppresses h on small sets, phasing it in past size w.

▸ **Follow the strong ties first; add structure only once the group is large enough to have a stable boundary.**



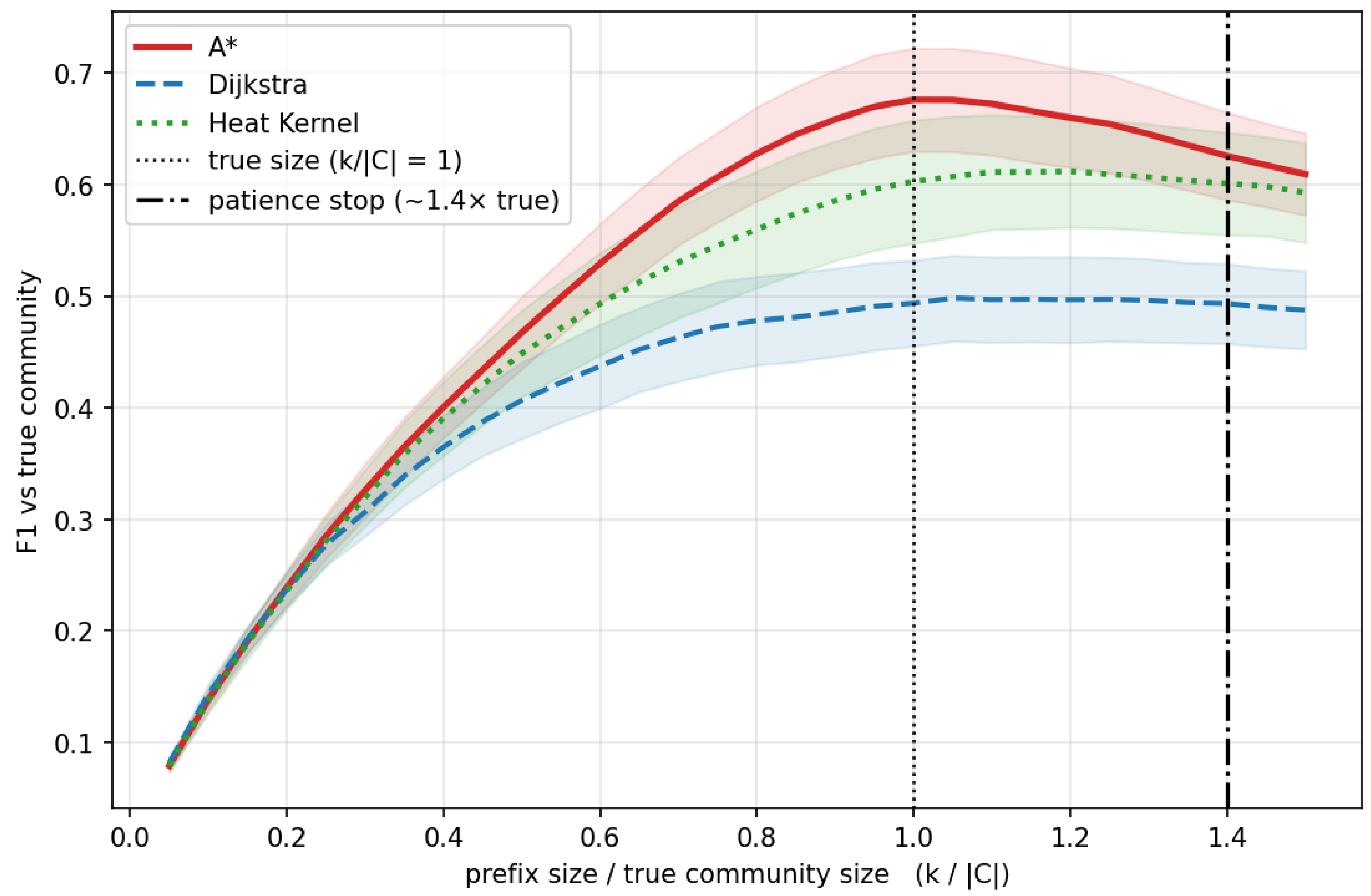
3. Accuracy vs. Baselines

Where it wins, annealed A* beats strong, individually-tuned baselines.

It stays above them at **every** prefix length, not only at the chosen cut.

Patience stop: grow until separation stops improving for several steps, then keep the best-separated point.

▸ **The win is in the ordering, not the stopping rule.**



4. Conditions for the Win

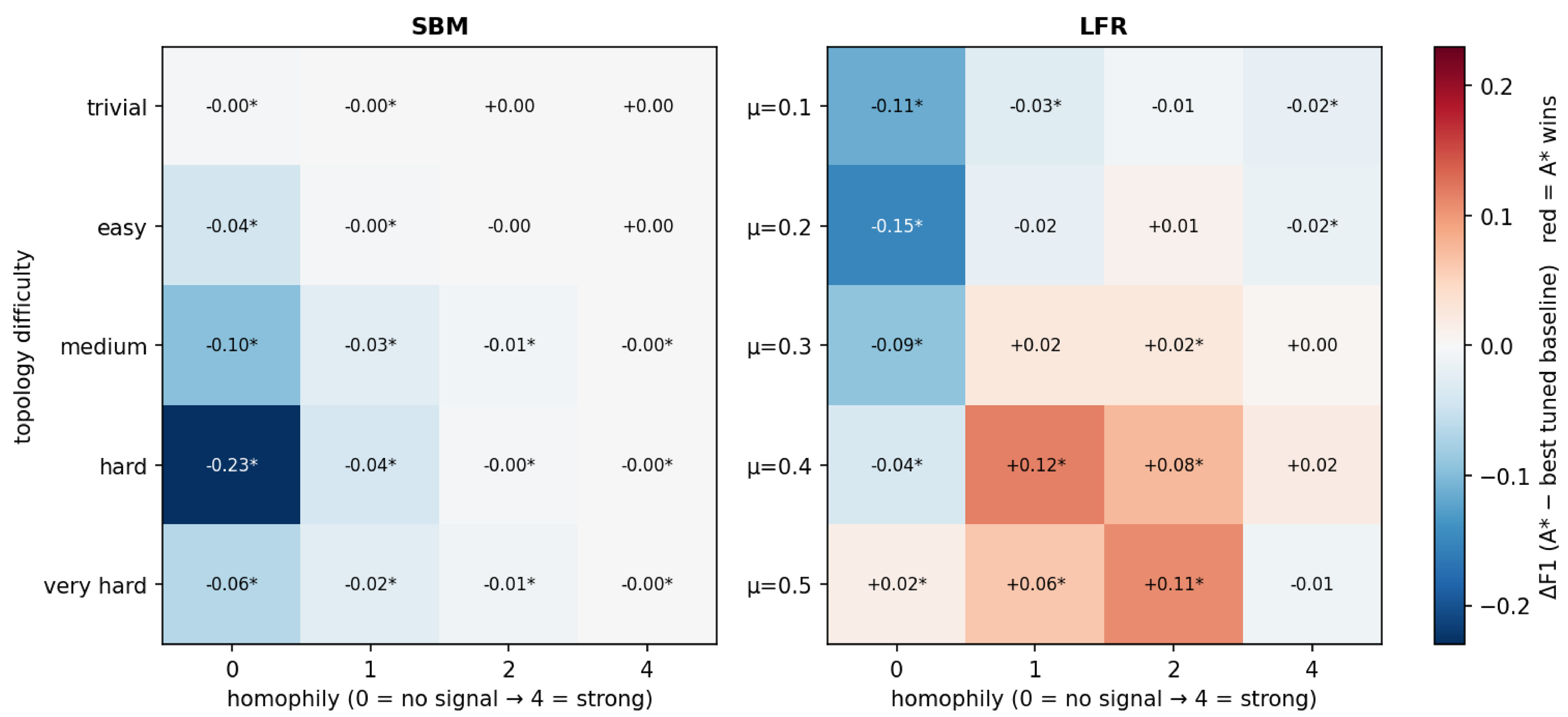
The win concentrates where the topology is weak (mixing μ) and the weights still carry a partial signal (homophily strength η).

Up to **+0.115 F1**, statistically significant on 6 of the hardest LFR cells.

Wins: hardest mixing ($\mu = 0.3$–0.5), as long as the signal is not trivial ($\eta \leq 2$).

Draws / Losses: usually no weight signal ($\eta = 0$), a very strong signal ($\eta = 4$), and all SBM graphs.

▸ **A precise win, not a general one.**



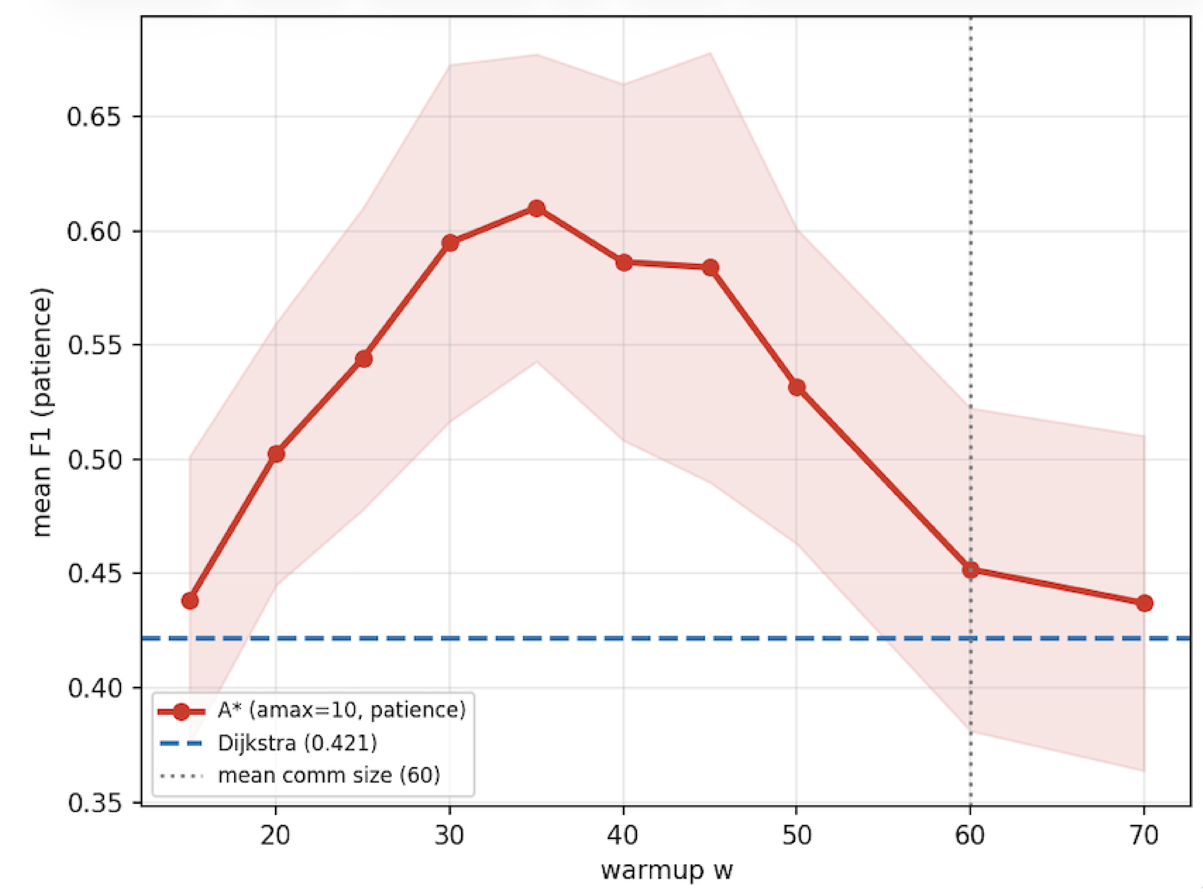
5. Impact of Annealing on Accuracy

Annealed A* clearly beats plain Dijkstra: annealing the structural heuristic into the search is what makes the difference.

Performance peaks when the warm-up size is below the true community size.

The advantage fades back to Dijkstra if the heuristic switches on too late.

▸ **What matters is *when* the heuristic switches on, not just that it does.**



6. Validation on Real Networks

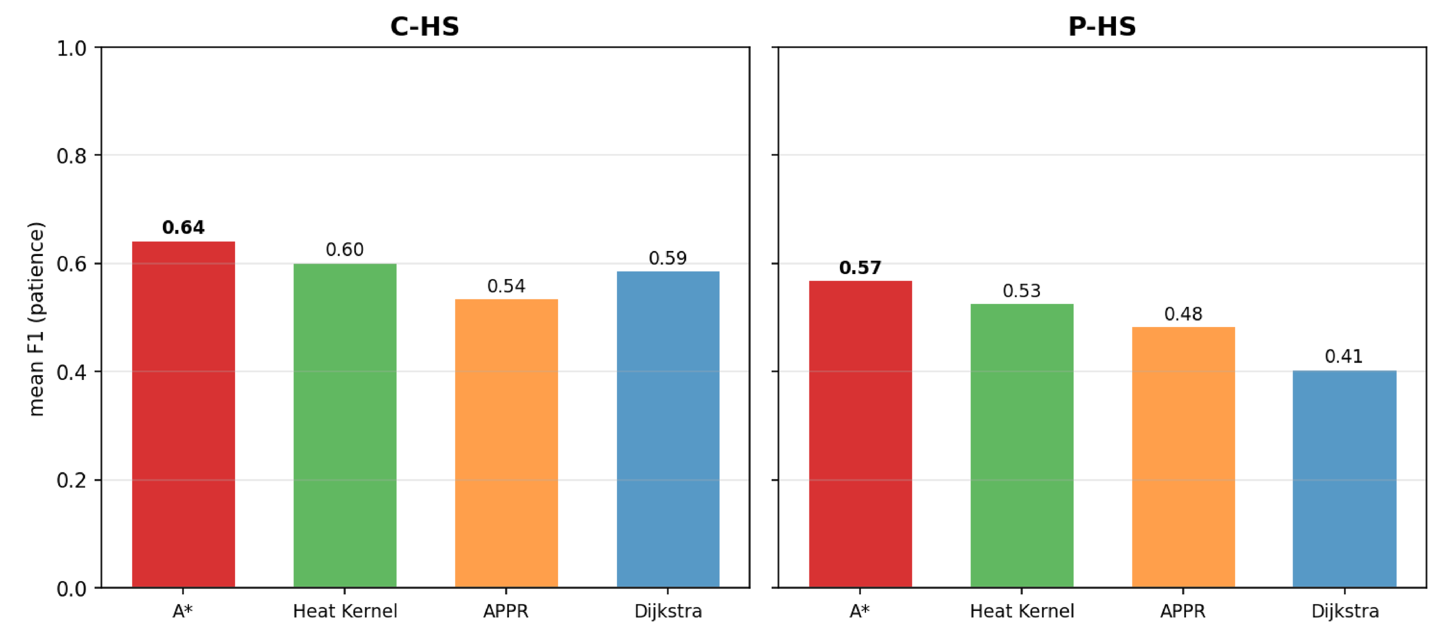
The same effect carries over to real data: two high-school contact networks, where a tie is the time two students spend together.

Communities are school grades.

A* beats the best-tuned baselines (C-HS +0.041, P-HS +0.044).

Indicative only: two graphs, parameters from synthetic data.

▸ **Contact intensity recovers grades that topology hides.**



7. Conclusions

Takeaway: Community membership can hide in interaction intensity rather than topology.

Limitations: Tested on only two real graphs; parameters transferred from synthetic data.

Future Work: Swap contact duration for opinion alignment to analyse polarised online networks; tune directly on real data; extract multiple communities simultaneously.