

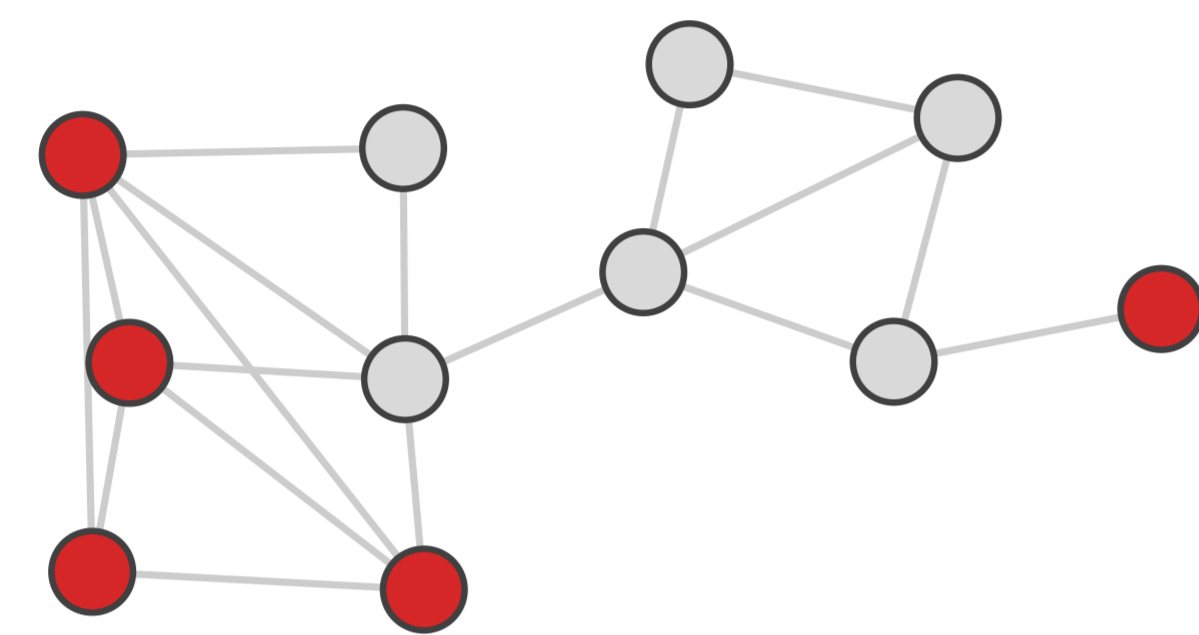
Observation Design for Active Search over Uncertain Graphs

Tudor Şucaliuc | T.Sucaliuc-1@student.tudelft.nl | EEMCS, TU Delft | CSE3000 Research Project

Objectives & Research Question

Active search finds hidden **target** nodes by probing one at a time under a budget. Most methods assume the graph is *known*; here the **topology is hidden too**. We ask:

- ▶ Which **observation model** — *what a probe returns* — lets an agent learn useful structure under joint topology-label uncertainty?
- ▶ What **inference** realises that value, and how is it **acquired under cost**?



● target ● other

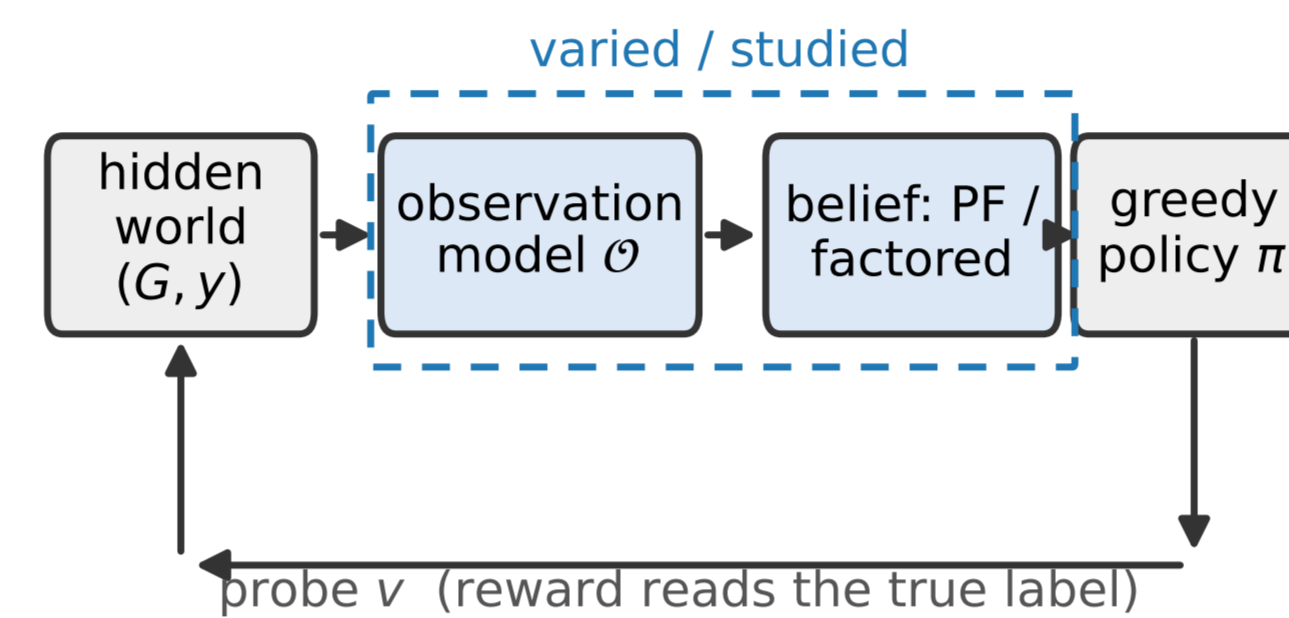
Active search: find the hidden targets (red). They cluster by homophily — but the graph is hidden.

Introduction

Active search. Sequentially probe nodes to find as many targets as possible within a budget — fraud accounts, infected individuals, compromised machines.

The gap. Prior work assumes the graph is given; in reality structure is revealed only *by* probing, so the agent is uncertain about **both labels and edges**.

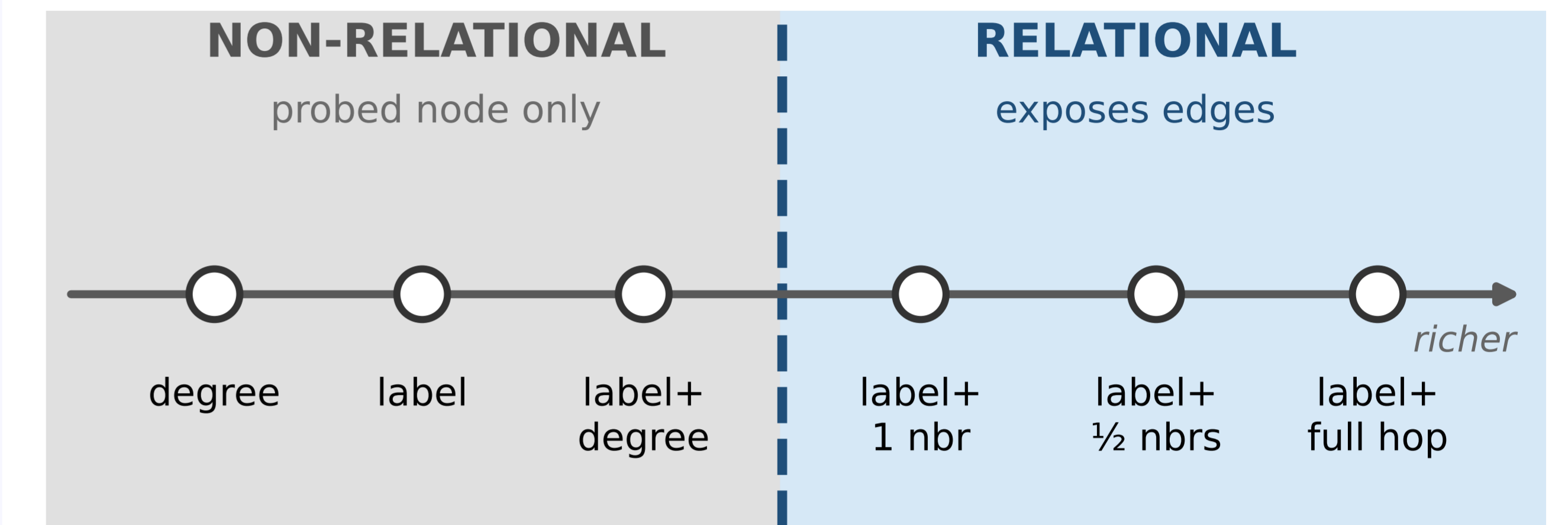
Why it matters. The **observation model** (the map from a probe to what it returns) is usually left implicit. We make it the controlled variable — because Tsugawa & Ohsaki report it can be a “**double-edged sword**”: richer observation helps in some regimes and *hurts* in others.



The agent loop: *probe, observe, update belief, act — repeat*. The **observation model** (and the belief) is what we study; world and greedy policy stay fixed.

Methods

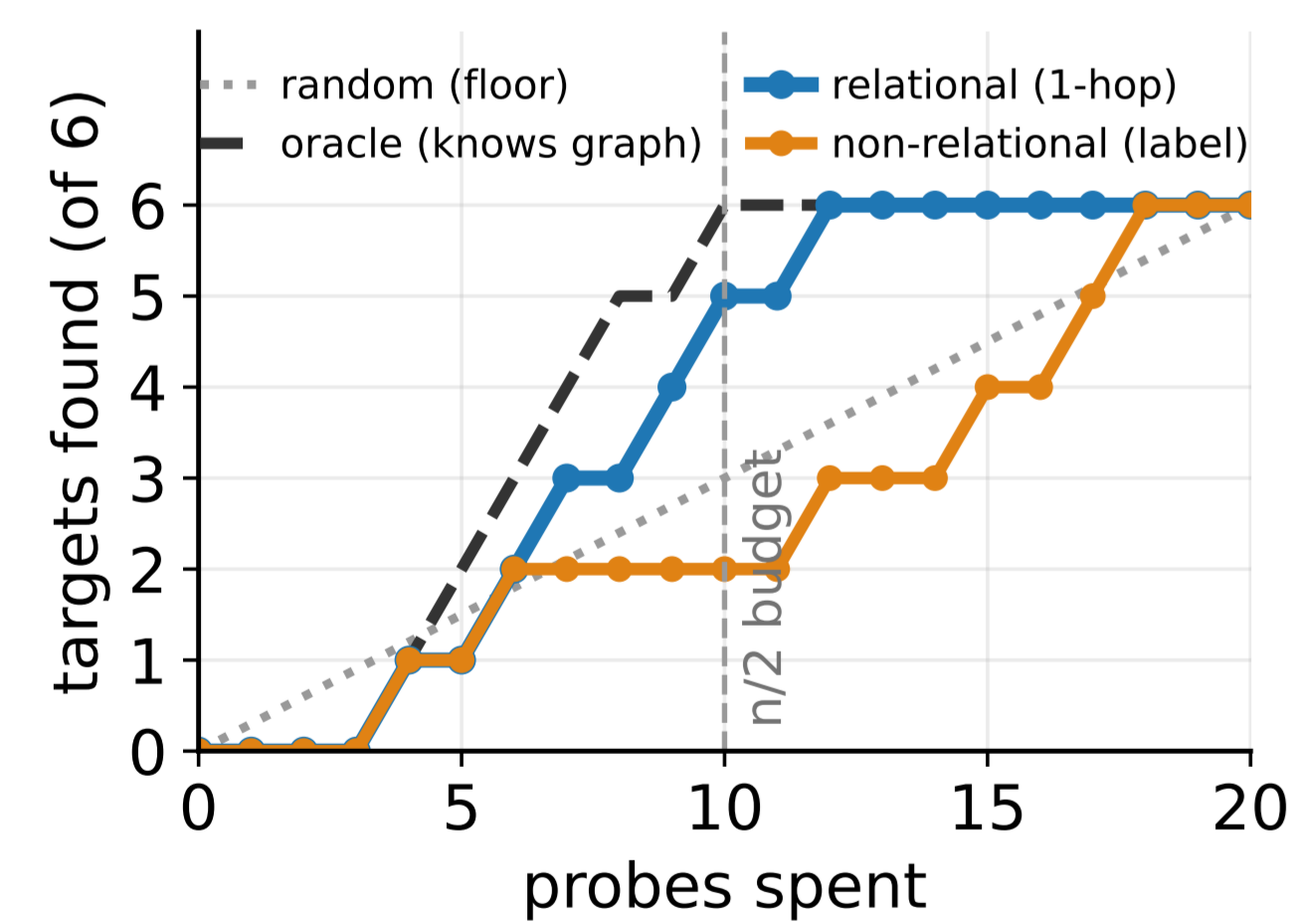
POMDP. A state is a hidden world (graph G , labels y); a probe returns a (noisy) label and a revealed neighbour set; reward +1 per target; metric **recall at budget $n/2$** . We vary *only* the observation, holding world, belief and policy fixed.



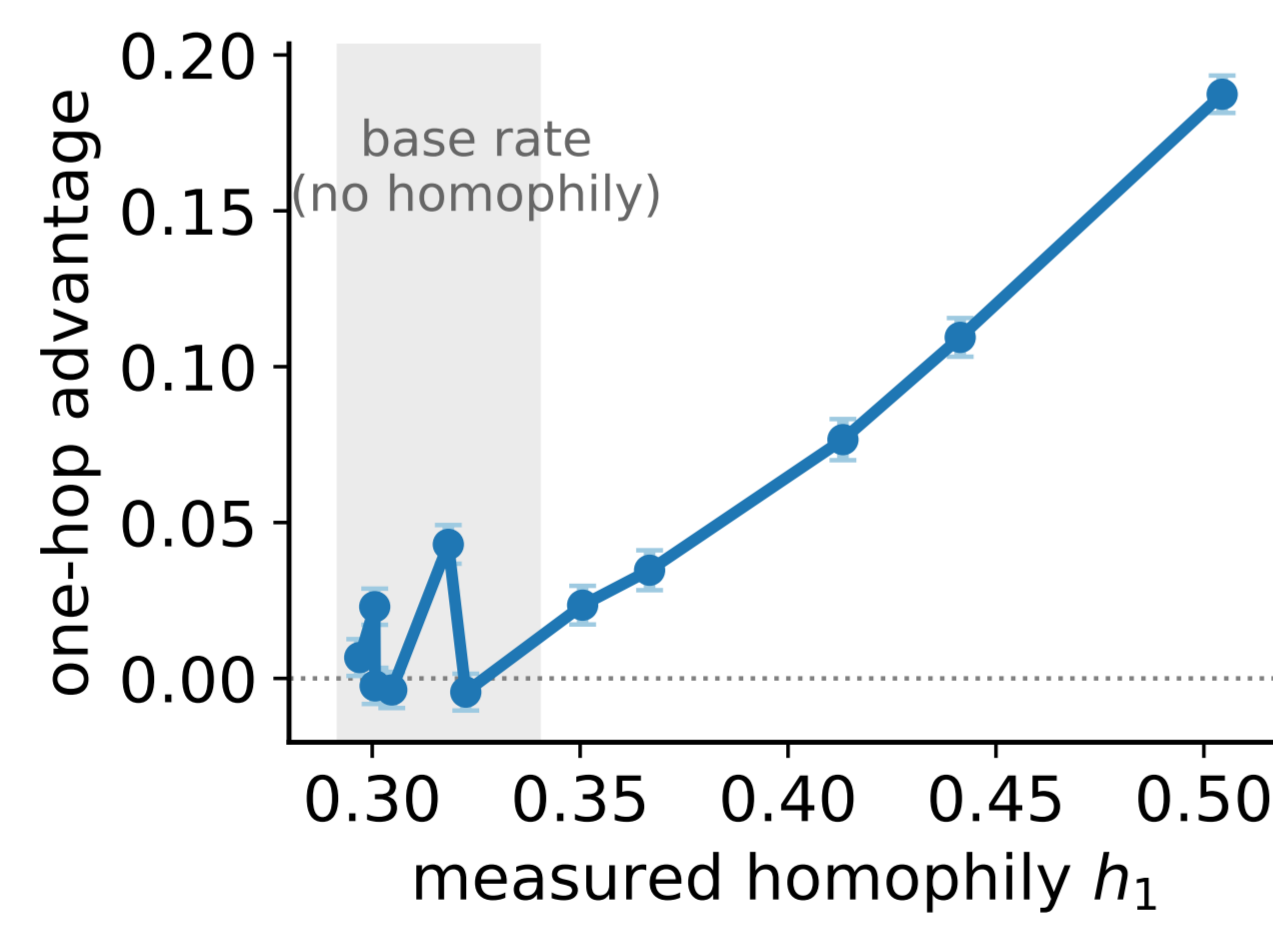
Beliefs. A bootstrap **particle filter** vs a closed-form **factored belief** (per-edge Bayes + a mean-field label score); **greedy** selection. Worlds: ER, community (SBM), epidemic; $n = 20-100$; exact enumeration at $n = 6$.

1 | Value — what is necessary

Under a **node-exchangeable prior** a non-relational observation leaves every unprobed node exchangeable, so *no policy* — even *Bayes-optimal* — *beats random*. One relational hop **nearly matches knowing the whole graph**; its advantage *tracks* the homophily it exposes.



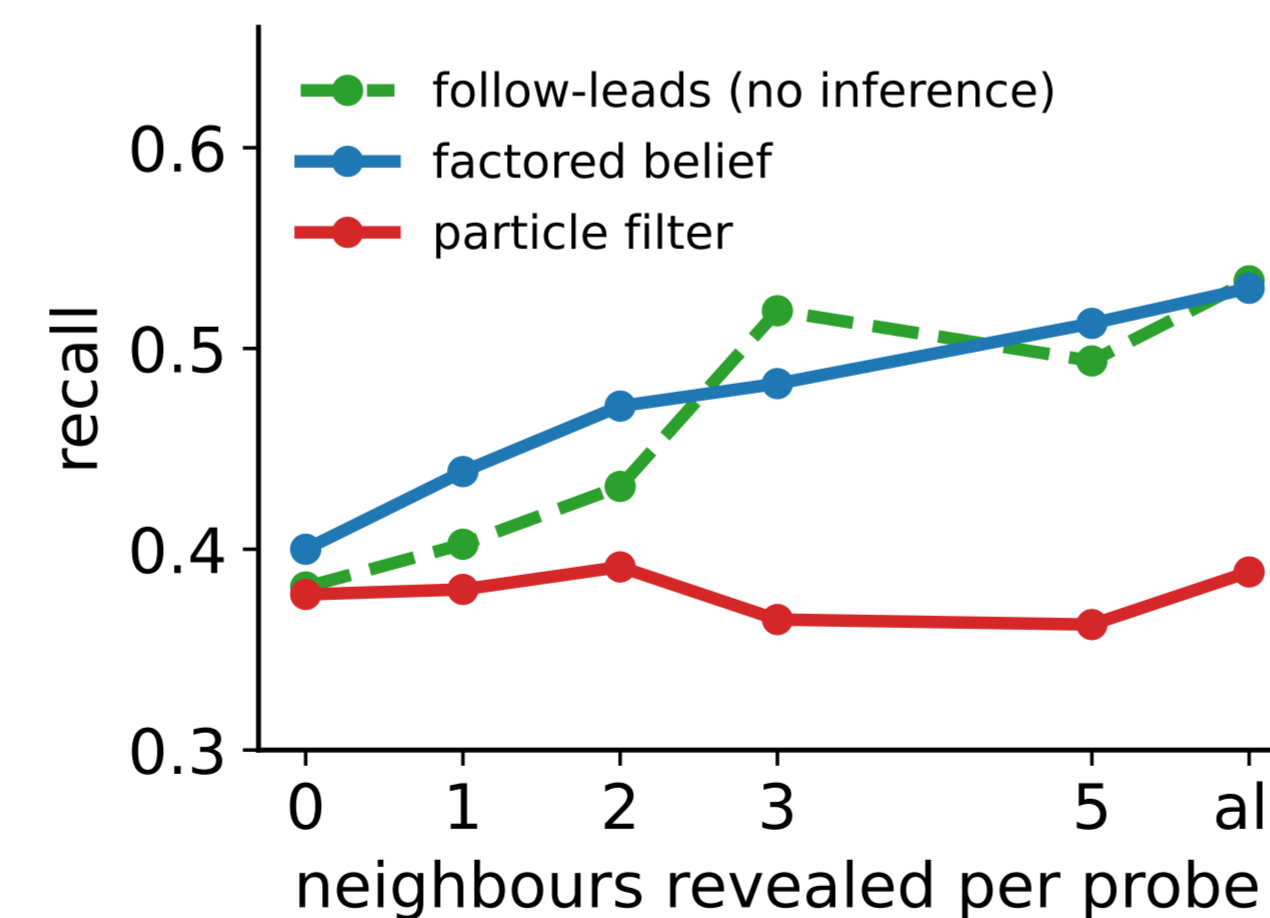
1-hop (blue) rides the oracle; label (orange) stays at random.



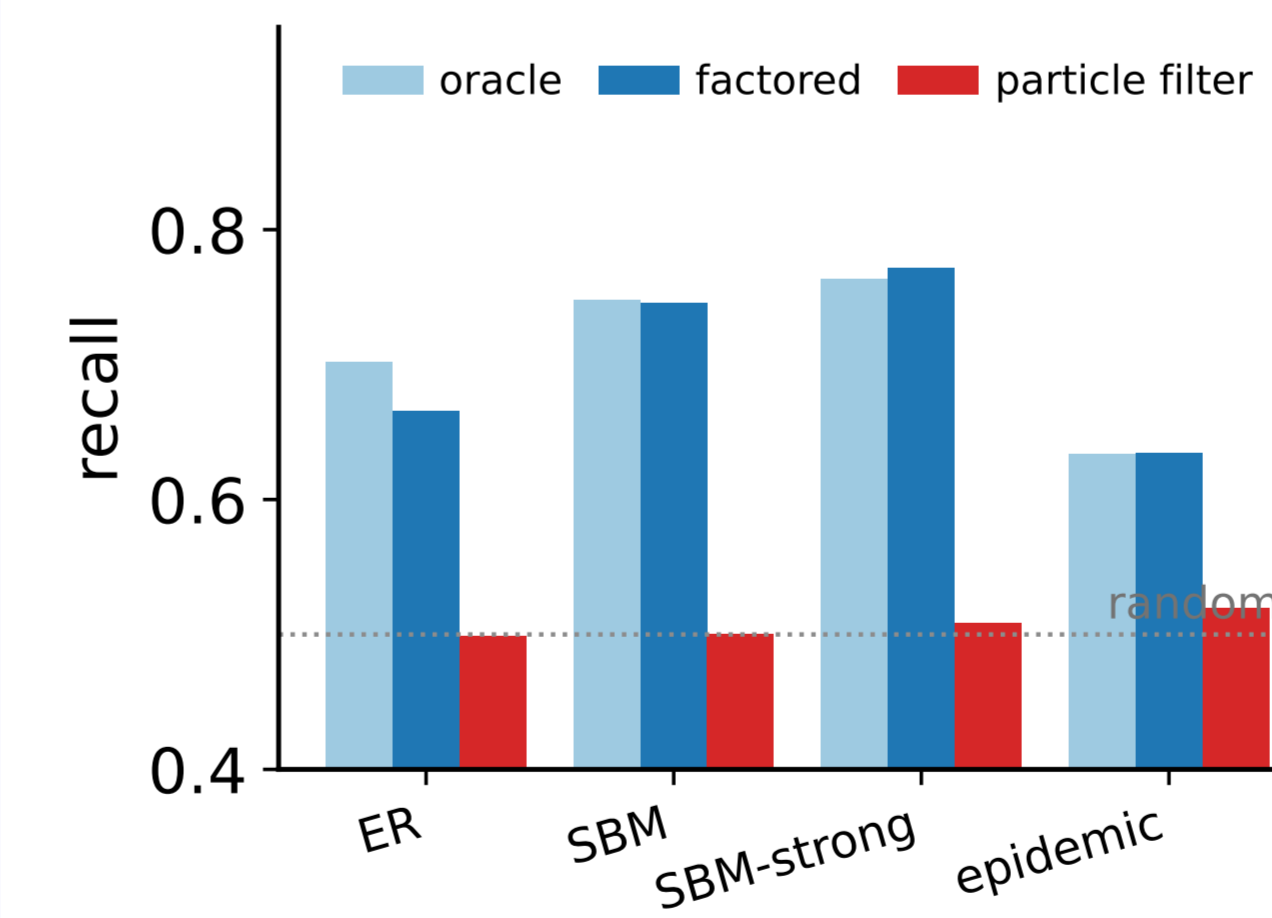
Advantage rises with homophily; ≈ 0 with none.

2 | Inference — what realises it

Two requirements from step 1: the observation must be **relational**, and the belief must **use** it — that is where it breaks. A particle filter **collapses** (effective sample size $2000 \rightarrow 2$ after one probe). The factored belief reaches the **known-graph reference** across every family. The **belief, not the information, is the culprit**: the same observation gives +0.23 recall with a correct-sign belief, -0.16 with a wrong-sign one.



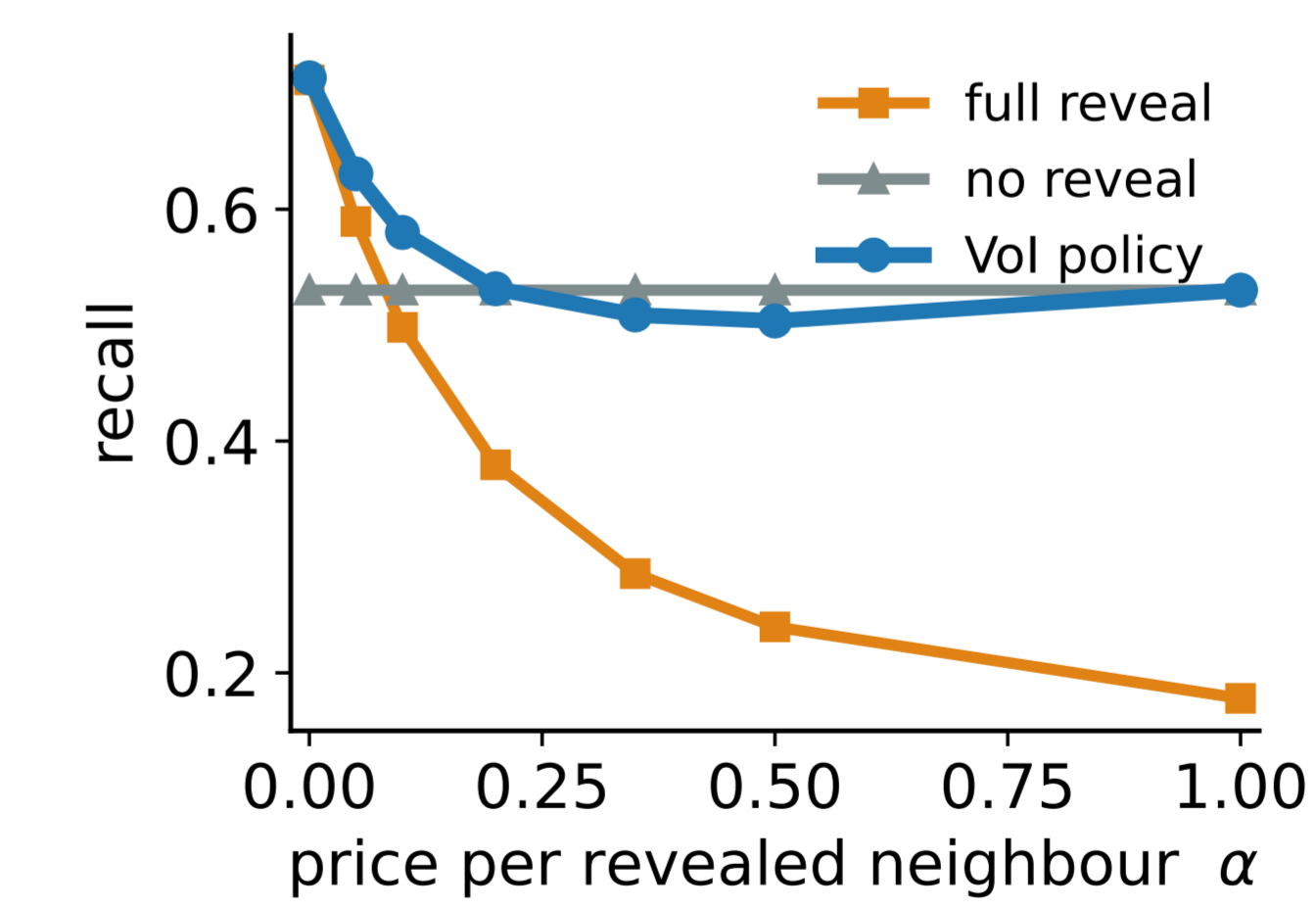
Factored matches follow-leads; PF stays flat.



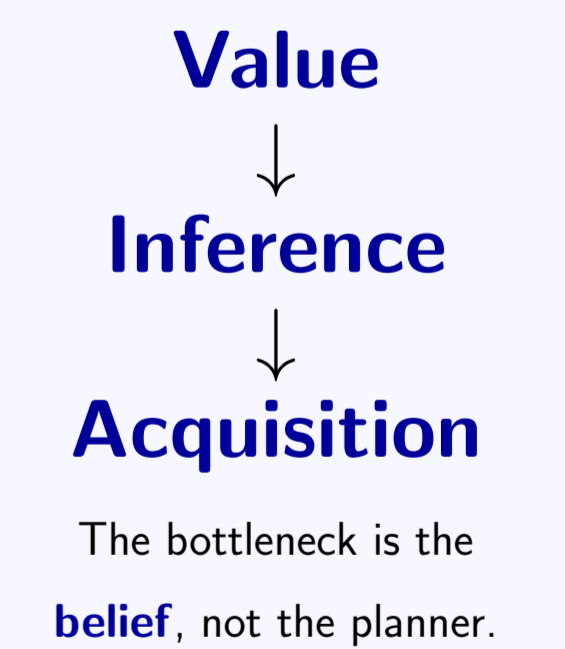
Factored reaches the oracle; PF does not.

3 | Acquisition — under cost

Price each revealed neighbour at α . A **value-of-information** rule estimates each lead's worth *online* and reveals only while it beats the price — acquiring **only the topology that pays**, with no cost model given.



Vol tracks the better of full / no reveal at every price.



Discussion

The harm in the “double-edged sword” traces to the **belief**, not the information — only a per-neighbour *price* makes a richer observation a clear liability (economic, not informational). The findings fit standard theory: **belief-state sufficiency**, the **naive-Bayes phenomenon** (a factored model ranks correctly even with a wrong joint), and greedy **matching the exact optimum** despite violated adaptive submodularity.

Limitations: synthetic graphs; greedy node selection; a node-exchangeable prior behind the proof; a strong-baseline (not proven-optimal) reference.

Conclusion

For joint structure-and-label search under topology uncertainty: relational observation is **necessary** and one hop nearly suffices; the bottleneck is the **belief, not the planner** — *factorise the belief, don't sample it*; and value-of-information acquires topology efficiently under cost.