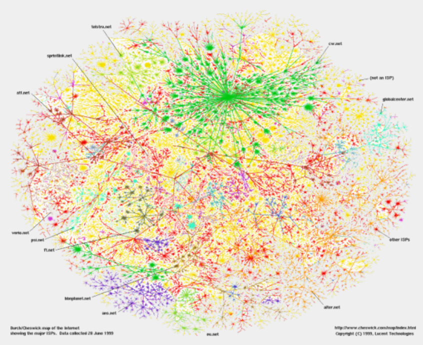
SparRL: Graph Sparsification via Deep Reinforcement Learning

Ryan Wickman
University of Memphis

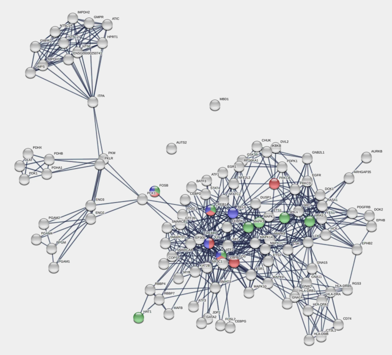
1. Motivation

- Graph size often dominates the efficiency of graph analytic workloads!
- Graphs are ***ubiquitous*** and ***huge*** in size in various domains

Social Network



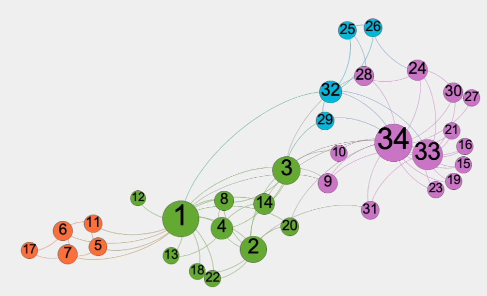
Internet Topology



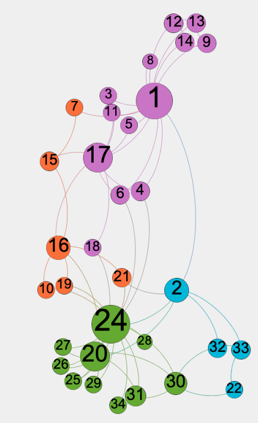
Protein-Protein Interaction Network

2. Problem

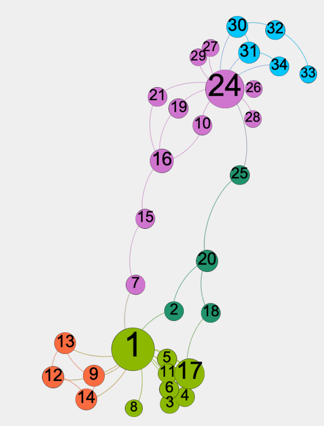
- **Graph sparsification** is a data reduction technique where an edge-reduced graph of similar structure is preferred.
- Derive $G' \subseteq G$ such that $F(G') \approx F(G)$



Karate Graph



Karate Graph (20% edge pruned)



Karate Graph (40% edge pruned)

3. Contributions

We propose a deep reinforcement learning algorithm for objective invariant graph sparsification

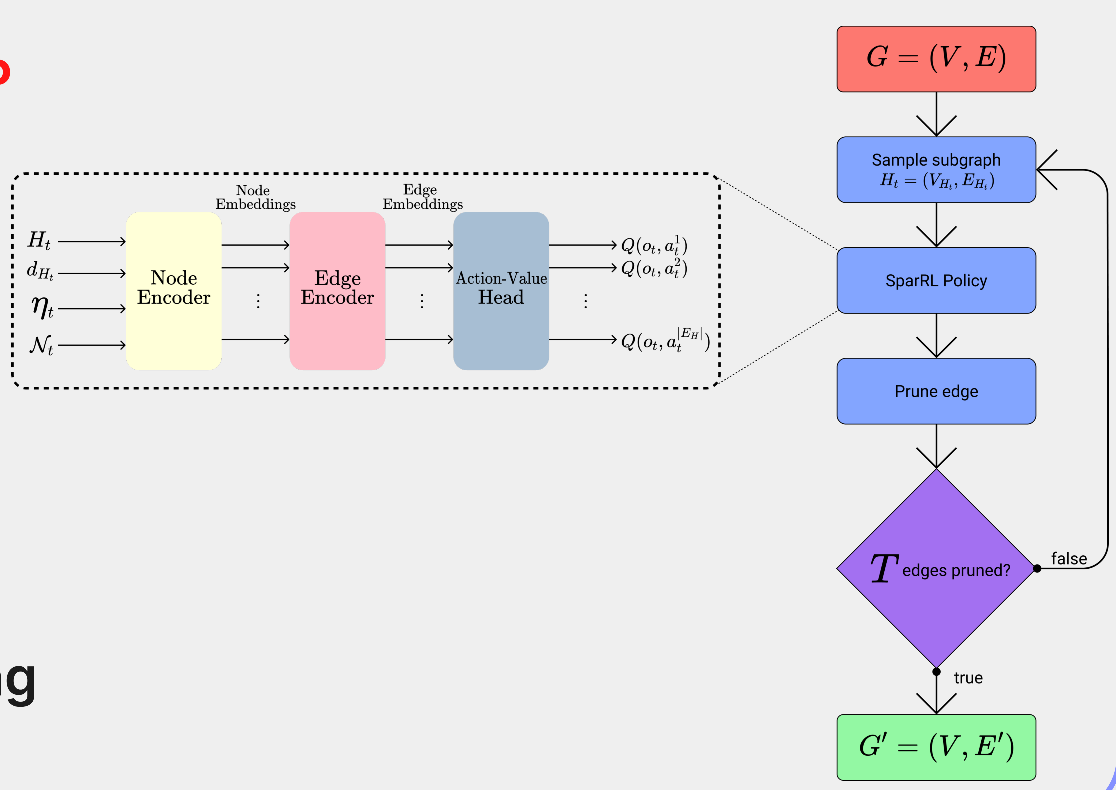
- **Highly configurable through reward function**
  Any scalar objective that can be modeled as a function of the graph can be optimized!
- **Shown to outperform all other baselines on all graphs and objectives**

4. Solution

- **We model graph sparisification as a POMDP**
  $(\mathcal{S}, \mathcal{A}, P, R, \Omega, \mathcal{O}, \gamma)$
- **Solved using Double DQN**
  The policy outputs a value for each edge and we prune the edge with the highest value

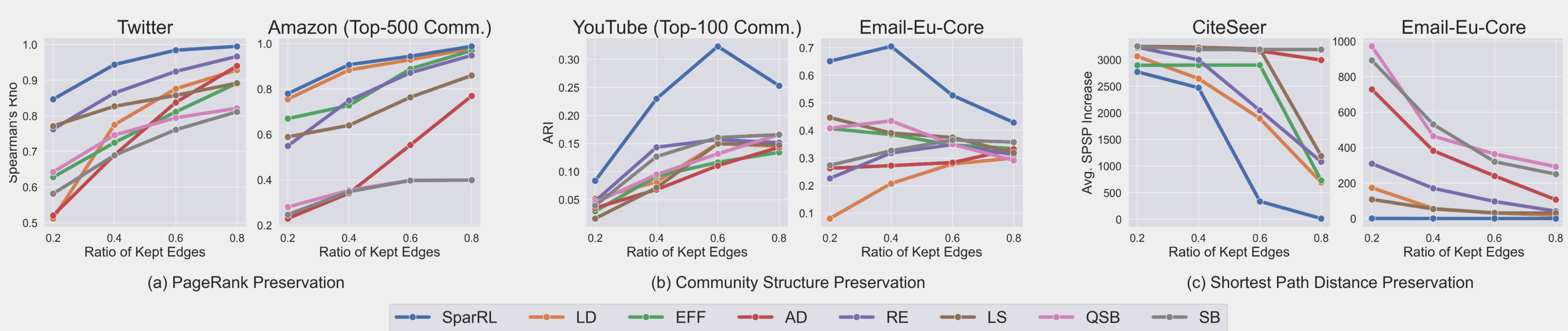
$$a_t = \arg\max_{a \in H_t} Q(o_t, a)$$

- **Time Complexity of pruning T edges is $O(|E_H|T)$**



5. Experiments & Conclusion

- **Outperforms on all tested benchmarks:**
  PageRank, single-pair shortest path, community detection
- **Outperforms all tested baseline methods:**
  Random Edge (RE), Local Degree (LD), Edge Forest Fire (EFF), Algebraic Distance (AD), L-Spar (LS), Simmelian Backbone (SB), Quadrilateral Simmelian Backbone (QSB)



(a) PageRank Preservation

(b) Community Structure Preservation

(c) Shortest Path Distance Preservation

- **SparRL is the first task-adaptive and effective reinforcement learning-based framework for graph sparsification**
  Generality evident by its performance on multiple objectives on a variety of graphs
- **In the future, we plan to extend SparRL**
  Test in a parallel setting, repurpose for graph learning tasks (e.g., link prediction, label classification etc.), and test on a dynamic graph setting

Table 1: SparRL compared against t-spanner for various stretch values t over CiteSeer. (x%: edge kept ratio)

Method	t=3 (99.65%)	t=4 (99.63%)	t=8 (97.82%)	t=16 (93.74%)	t=32 (90.78%)
t-spanner	0.0082	0.0054	0.0405	0.1187	0.1911
SparRL	**0.0031**	**0.0043**	**0.0350**	**0.0974**	**0.1820**