CRExplainer: Global Graph Counterfactual Explainer through Common Recourse

Gregoire Fournier, Sourav Medya

University of Illinois at Chicago

Goal

Find a set of recourse on graphs, $\mathbb{F}$, that provides the best counterfactual explanation coverage:

$$max_{\mathbb{F}} \text{ coverage}(\mathbb{F}) \text{ s.t. size}(\mathbb{F}) = N$$

Introduction

- Given a **GNN** trained on binary classification into 'accept' and 'reject' classes, a **global counterfactual explanation** explanation consists in generating a small set of 'accept' graphs relevant to all of the input 'reject' graphs.
- In the context of **common recourse**, we want to find a small set of **recourse**, or transformation on graphs, which turn a 'reject' graph into an 'accept' graph.
- Although counterfactual explanation has been studied [1], the problem of finding common recourse for global counterfactual explanation is new.
- **Common recourse counterfactual explanation** has been applied to credit application or alternative sanctions problems, where fairness is central. It can also be used to provide high level insights in computational biology or computer security.

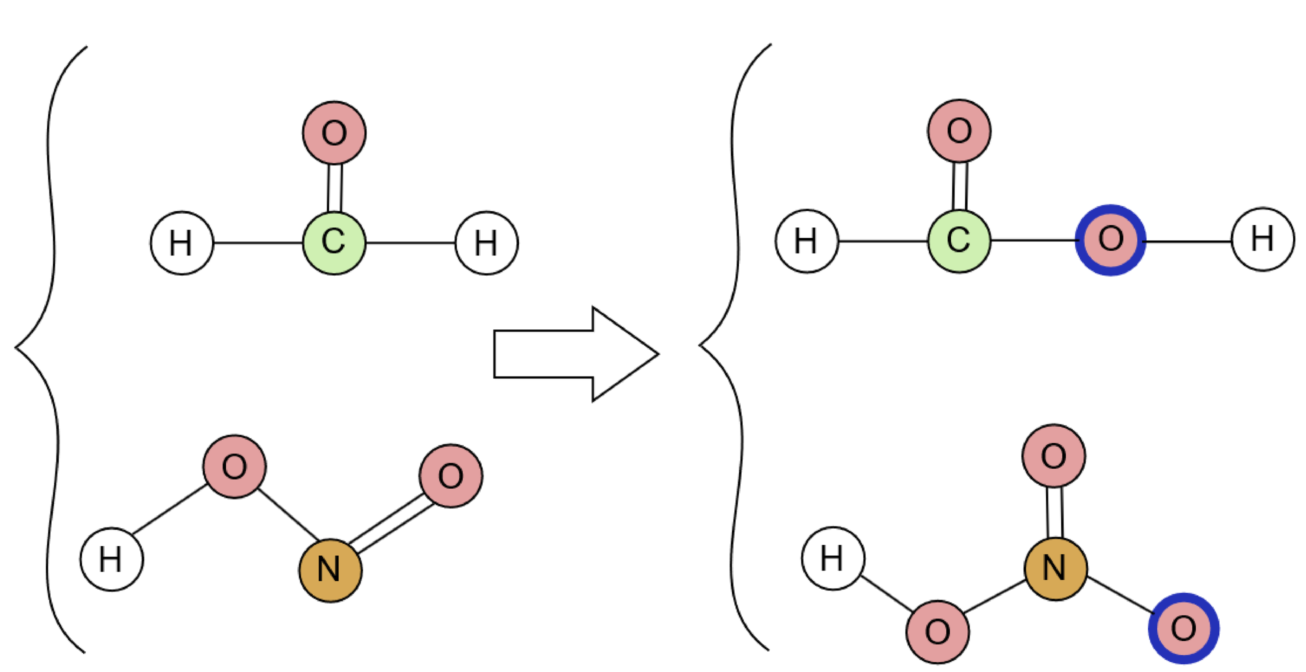

Figure 1:Common recourse for counterfactual explanation

Problem analysis

- The problem is NP-hard, which can be seen by reduction from max coverage.

Common Recourse Explainer

To find good common counterfactual recourse, CRExplainer combines:

- The **graph embedding algorithm** GREED [2]: to evaluate the recourse and the distances between graphs reached through the random walk.
- A **multi-head vertex reinforced random walk** in the graph-edit space:
  - CRExplainer finds counterfactuals through a vertex reinforced random walk (**VRRW**) [3] with the following transition rule:
  $$p(u, v) \propto p_\phi(v)N(v) \tag{1}$$
  - The teleportation is optimized for **coverage**. Define $g(G)$ be the number of close counterfactuals covering an input graph G. Then the probability to teleport to G is:
  $$p_\tau(G) = \frac{\exp(-g(G))}{\Sigma_{G' \in \mathbb{G}} \exp(-g(G'))} \tag{2}$$
  - The random walk uses **multiple heads**: at each step, we randomly select one of the heads, and move it toward a counterfactual. The other heads follow to according to the best a common recourse neighbor available.
- A **clustering algorithm** DBSCan: to generate common recourse by form clusters on recourse to find common ones.

Metrics

For $\mathbb{F}$ a set of recourse we define:

- coverage($\mathbb{F}$): the number of input graphs for which at least one counterfactual is obtained through one of the recourse in $\mathbb{F}$.
- cost($\mathbb{F}$): the total distance from the covered input graphs to their attained counterfactual through $\mathbb{F}$.

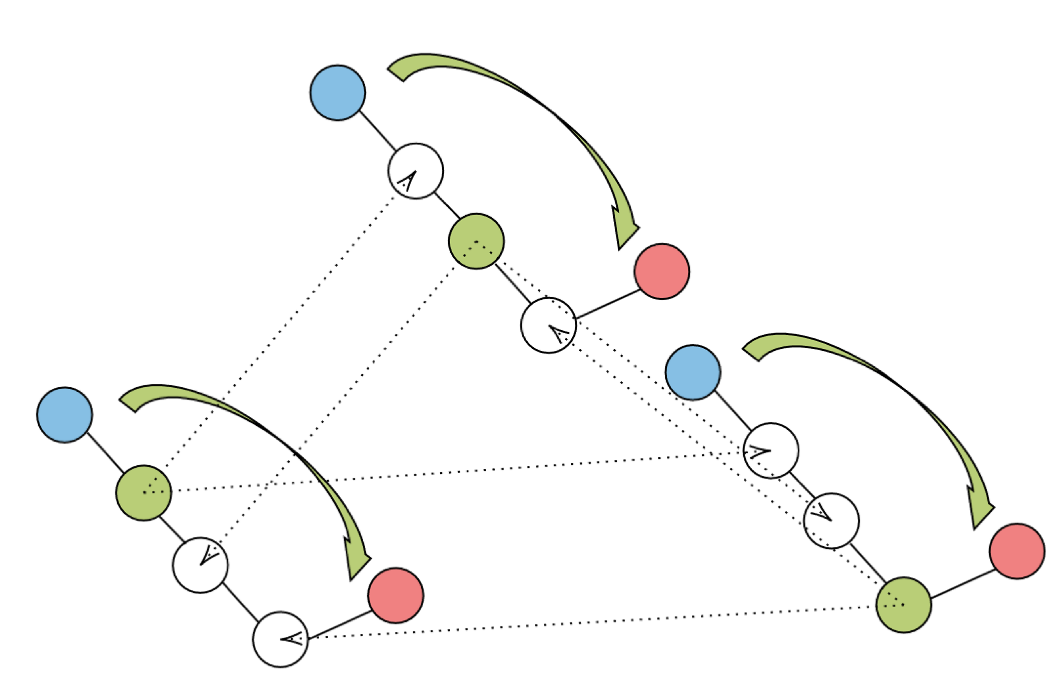

Figure 2:Our proposed multi-head vertex reinforced random walk

Experimental Results

We benchmark CRExplainer on real world datasets against popular explainers:

- CRExplainer produces better quality global common recourse.
- CRExplainer's common recourse are significantly less costly than the other explainers tested.

	NCI1		Mutagenicity		Proteins	
	Coverage	Cost	Coverage	Cost	Coverage	Cost
Ground-Truth	5.6%	161	31.6%	158	95.9%	31.1
RCExplainer	8.6%	161	31.5%	157	x	x
GCFExplainer	15.6%	128	29.3%	131	92.1%	27.7
CRExplainer	**18.3%**	**112**	**32.3%**	**115**	**97.3%**	**26.4**

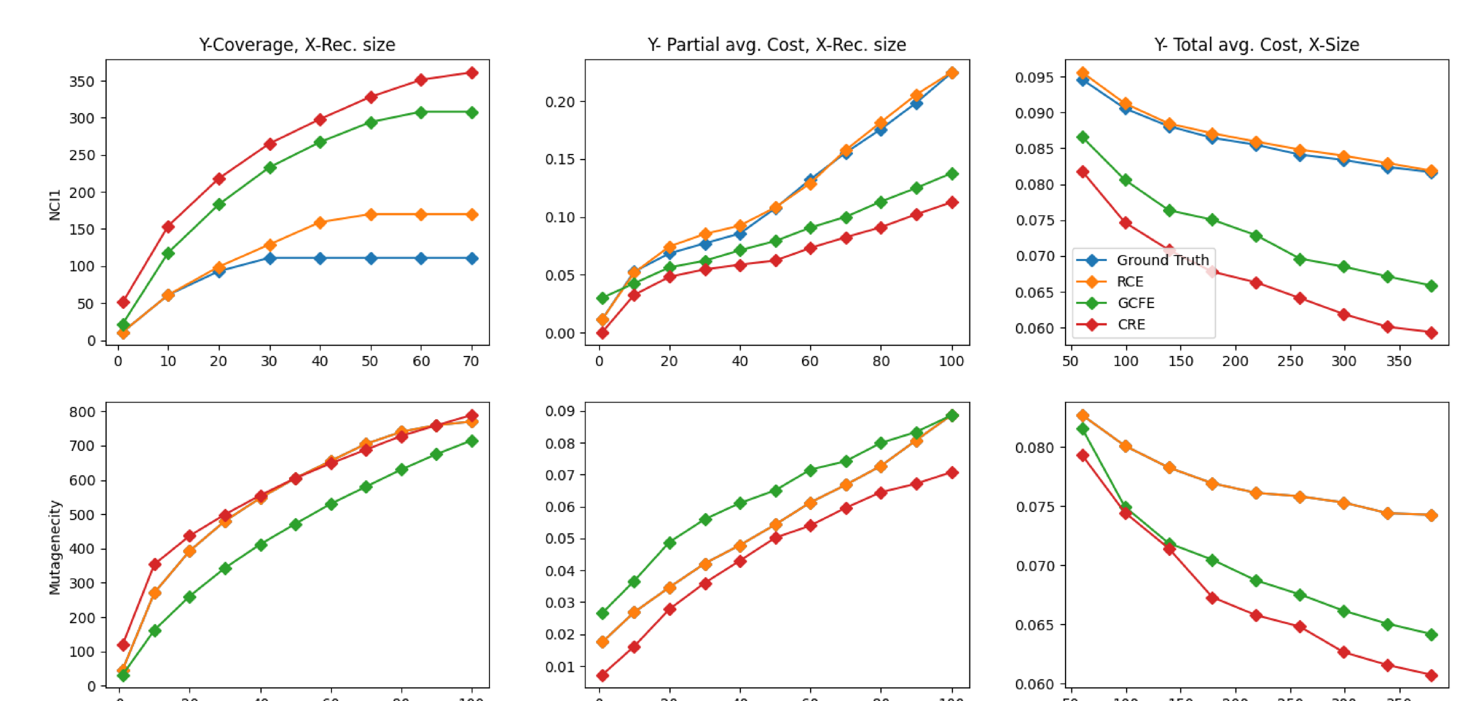

Figure 3:Coverage and cost of different common recourse counterfactual explanation

References

[1] Mert Kosan, Zexi Huang, Sourav Medya, Sayan Ranu, and Ambuj Singh. Global counterfactual explainer for graph neural networks, 2022.

[2] Rishabh Ranjan, Siddharth Grover, Sourav Medya, Venkatesan Chakaravarthy, Yogish Sabharwal, and Sayan Ranu. Greed: A neural framework for learning graph distance functions, 2023.

[3] Robin Pemantle. Vertex-reinforced random walk, 2004.