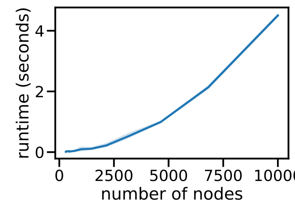
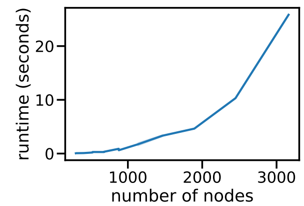
Certifiable Robustness to Graph Perturbations: Author Response

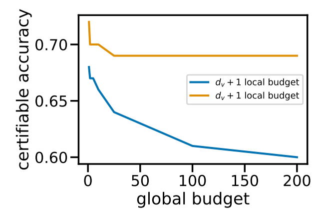
R1/R2/R3: Limited Focus. As suggested, we will clarify in the paper that our focus is on certifying PPNP and label/feature propagation; and not every possible GNN. Certifying any of these approaches is highly relevant: e.g. label propagation is quite popular in practice (often as part of more complicated pipelines in industry), and the strong empirical performance of PPNP has already been independently verified [1]. We can also trivially extend our approach to certify a recently proposed model termed Simple Graph Convolution (SGC) [2] which is equivalent to feature propagation. Certifying SGC is useful since it is one of the few GNNs that demonstrates scalability to graphs with millions of nodes. In future work, we can extend our approach to GCNs by using a similar analysis to Xu et al. [3] (Theorem 1) which shows that the influence between nodes in a k-layer GCN is proportional to a k-step random walk distribution by e.g. bounding the influence with (truncated) PageRank to obtain a certificate.



(a) Runtime: local (VI).



(b) Runtime: global (RLT).



(c) Bound on certifiable accuracy.

R2: SDP relaxation. Let $(y_1, y_2, \dots)$ be the variables corresponding to β^0_{ij}, x^0_{ij}, etc. The SDP relaxation replaces the product terms $y_i y_j$ by an element Y_{ij} of an $n \times n$ matrix Y and adds the constraint $Y - yy^T \succeq 0$. Since in the original $QCLP$ there are no terms of the form $y_i y_i$ corresponding to the elements on the diagonal, we can make the diagonal elements Y_{ii} arbitrarily high to make the matrix $Y - yy^T$ positive semidefinite and trivially satisfy the constraint.

R2: NP-hard proof. We provide a proof sketch that adding the global budget makes the problem NP-hard by constructing a polynomial reduction from the 1-IN-3SAT problem which is NP-complete. The problem: Given a boolean 3-CNF formula s.t. the clauses contain only un-negated atoms, does there exist a truth assignment s.t. in each clause, exactly one literal is true. First, add a single node t, and one node for each literal $l_1, \dots, l_n$ and each clause $c_1, \dots, c_m$. Let $\mathcal{E}_f$ (non-fragile set) contain: one edge from each node to t, one edge from t to each clause c_j, and one edge from each clause c_j to its three literals ($3m$ in total). Let $\mathcal{F}$ (fragile set) contain $3m$ edges, one from each literal l_i to its clauses, and let $\mathcal{E} = \mathcal{E}_f \cup \mathcal{F}$. Set the global budget $B = 2m$, and the teleport vector and reward vector as $\mathbf{z} = \mathbf{r} = \mathbf{e}_t$. Such reward means that we are maximizing the PageRank $\pi(\mathbf{z})_t$ of the single node t, or equivalently minimizing the expected first hitting time h_t to t. Intuitively, for this graph removing any fragile edge decreases h_t, which means we can always improve the objective by removing more edges, up to the budget $B = 2m$. Thus, there are exactly m fragile edges left (i.e. $2m$ removed) in the optimal configuration $\mathcal{O}^*$. Let f_j be the number of fragile edges in $\mathcal{O}^*$ pointing to clause c_j. Claim: 1-IN-3SAT is satisfiable iff in the *optimal* solution each $f_j = 1$. First note that for any optimal solution, if one edge from some literal is in $\mathcal{O}^*$ then all edges from that literal are in $\mathcal{O}^*$ (up to the budget). The reason is that adding an additional edge from a literal already in $\mathcal{O}^*$ to some clause leads to a smaller h_t increase than adding an edge from a literal not yet in $\mathcal{O}^*$ to some clause. Given this, the right-to-left direction of the claim above is trivial: Since each $f_j = 1$, every clause has exactly one literal set to true. It follows: 1-IN-3SAT is satisfiable. Left-to-right: Given that 1-IN-3SAT is satisfiable. Assume that the optimal configuration $\mathcal{O}'$ contains some clause c_1 with $f_1 = 2$. Since $|\mathcal{O}'| = m$, there must be a clause $c_2 \neq c_1$ with $f_2 = 0$. Now c_1 forms 2 cycles with its literals which increases h_t, but having $f_2 = 0$ decreases h_t. The former increase is always larger then the later decrease, thus a configuration where some c_j's have $f_j = 2$ always has a larger h_t compared to any configuration where all $f_j = 1$. Since such a configuration exists (satisfiability holds), $\mathcal{O}'$ cannot be optimal. Similarly, this holds if some $f_j = 3$. Thus, it follows that if 1-IN-3SAT is satisfiable $\mathcal{O}^*$ recovers the truth assignment and all $f_j = 1$.

R2: Only global budget. Our approach is not designed to handle only global budget since the proposed upper bounds explicitly depend on having local budget. Deriving tight upper bounds for the "global only" case is left for future work.

R2/R3: Runtime. To show how the runtime scales with number of nodes we randomly generate SBM graphs of increasing size. In Fig. 1a we see the mean runtime for local budget (VI algorithm). Even for graphs with more than 10K nodes the certificate runs in a few seconds. Similarly, Fig. 1b shows the runtime for global budget (RLT relaxation). The runtime can be easily reduced by: (i) stopping early whenever the worst-case margin becomes negative, (ii) using Gurobi's distributed optimization capabilities to reduce solve times, and (iii) having single preprocessing for all nodes.

R3: Overall accuracy. Notice that the ratio of nodes that are both certifiably robust and at the same time have a correct prediction is a lower bound on the overall classification accuracy since the worst-case perturbation can be different for each node. We plot this ratio in Fig. 1c for Citeseer. We will include this finding in the updated paper.

[1] Fey, M. and Lenssen, J. E. Fast graph representation learning with pytorch geometric. *arXiv:1903.02428*, 2019.
[2] Wu et al. Simplifying graph convolutional networks. In *ICML 2019*.
[3] Xu, K. et al. Representation learning on graphs with jumping knowledge networks. In *ICML 2018*.