Shortest Paths and Distances with Differential Privacy

Adam Sealfon

MIT

Privacy Tools for Sharing Research Data

A National Science Foundation Secure and Trustworthy Cyberspace Project

with additional support from the Sloan Foundation and Google, Inc.

ABSTRACT

We introduce a new model for differentially private analysis of weighted graphs in which the graph topology (V, E) is assumed to be public and the private information consists only of the edge weights w. This can express hiding congestion patterns in a known system of roads.

Differential privacy requires that the output of an algorithm provides little advantage, measured by privacy parameters ε and δ, for distinguishing between neighboring inputs, which are thought of as inputs that differ on the contribution of one individual. In our model, two weight functions w and w' are considered to be neighboring if they have l_1 distance at most one. We study the problems of privately releasing a short path between a pair of vertices and of privately releasing approximate distances between all pairs of vertices. We are concerned with the *approximation error*, the difference between the length of the released path or released distance and the length of the shortest path or actual distance.

For the problem of privately releasing a short path between a pair of vertices, we prove a lower bound of Ω(V) on the additive approximation error for fixed privacy parameters ε and δ. We provide a differentially private algorithm that matches this error bound up to a logarithmic factor and releases paths between all pairs of vertices, not just a single pair. The approximation error achieved by our algorithm can be bounded by the number of edges on the shortest path, so we achieve better accuracy than the worst-case bound for pairs of vertices that are connected by a low-weight path consisting of o(V) vertices.

For the problem of releasing all-pairs distances, we show that for bounded-weight graphs with edge weights in [0,M] we can release all distances with approximation error roughly $O((VM)^{1/2})$ for fixed ε, δ > 0. For trees we show that we can release all-pairs distances with approximation error $O(\log^{2.5}(V))$.

THE MODEL

Today's navigation systems have access to current traffic data and use it to direct drivers. Data about traffic may be based on user reports or GPS locations. We would like to be able to route drivers while protecting this private user information.

In this setting, the network connections are publicly known, since the road map is available to all users. The private information consists of weights on the network edges.

We define two weightings of the same graph to be *neighboring* if they differ by one unit in the weight of a single edge, or more generally if they differ by at most one in l_1 norm. This privacy notion hides the effect of any individual or group which affects the edge weights by no more than one unit.

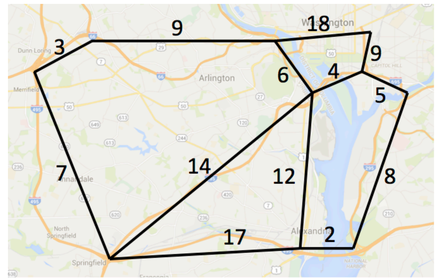
Formally, in our setting, an algorithm A is *(ε, δ)-differentially private* on a graph G = (V, E) if for all pairs of edge weight functions w and w' such that $\|w - w'\|_1 = \Sigma_{e \in E} |w(e) - w'(e)| \leq 1$ and for all sets of outputs S, we have that

$$\Pr[A(w) \in S] \leq e^{\varepsilon} \Pr[A(w') \in S] + \delta.$$

OBJECTIVES

We study two main questions concerning routing in this model.

- Given a weighted graph and a pair of vertices s,t, release some path between s and t whose length is as short as possible while preserving the privacy of the edge weights.
- Given a weighted graph, release approximate distances between all pairs of vertices under differential privacy.



RESULTS

Shortest paths:

- Any differentially private shortest path algorithm must (on some inputs) in expectation release a path which is Ω (n) longer than the shortest path, where n = |V| is the number of vertices.
- We give an algorithm which achieves this error up to a logarithmic factor. If the true shortest path consists of only k edges, our algorithm releases a path with expected error O(k log n).

All-pairs distance estimation:

- We can privately release all distances with error O(n log n) on each distance.
- For bounded-weight graphs with all weights in [0,M], we can release all distances with error $O((nM)^{1/2})$.
- On trees, we can release all distances with error $O(\log^{2.5} n)$.

Additional problems:

- Using similar techniques, we can also provide bounds for releasing low-weight spanning trees and low-cost matchings under differential privacy.

BASIC INGREDIENT: THE LAPLACE MECHANISM

Let f be any real-valued function, and let the sensitivity Δf be the maximum change in f between two neighboring inputs. The Laplace mechanism on database x samples noise Y ~ Lap(1/ε) according to the Laplace distribution with parameter 1/ε, and releases f(x)+Y.

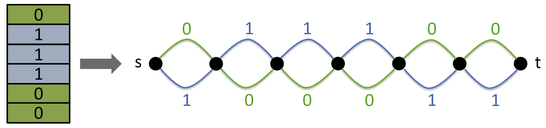
The Laplace mechanism is (ε,0)-differentially private. [Dwork *et al.*, 2006]

SHORTEST PATHS

Lower bound:

The lower bound is based on a reduction from the problem of reconstructing a large fraction of the rows of a database, which is impossible under differential privacy.

Theorem: Any (ε, δ)-differentially private algorithm A for the approximate shortest path problem must on some input release a path whose length is Ω(n) longer than the shortest path.



We obtain an almost-matching upper bound by applying the Laplace mechanism to release a noisy version of each edge weight in the graph. The algorithm releases short paths for all pairs of vertices with the same accuracy that our lower bound shows is necessary for releasing a path between a single pair of vertices. With a small adjustment to the algorithm, we obtain improved accuracy for pairs of vertices with a shortest path consisting of o(n) edges, allowing us to beat our lower bound on these inputs.

Theorem: There is an (ε, δ)-differentially private algorithm A which releases paths between all pairs of vertices s, t ∈ V such that with probability 1 – γ, if there is a path of weight W consisting of at most k vertices, the path released has weight at most $W + (2k / \varepsilon) \log(n^2 / \gamma)$.

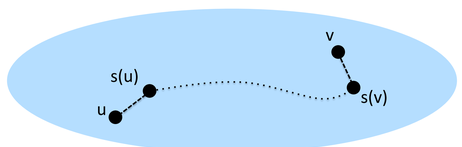
ALL-PAIRS DISTANCES

On general graphs, a straightforward application of the Laplace mechanism combined with composition theorems achieves error roughly O(n/ε) by either releasing noisy weights for each edge or directly adding Laplace noise to each of the n^2 distance queries. We obtain better error for two special types of graphs, trees and graphs with edges of bounded weight.

Bounded-weight graphs:

On bounded-weight graphs a simple combinatorial algorithm allows us to release all-pairs distances with error roughly $O(n^{1/2})$.

Theorem: Let G = (V, E) and Mε ∈ (1/n, n). Then there is a (ε, δ)-differentially private algorithm A that given edge weights w in [0, M] outputs with probability 1 – γ all-pairs distances with additive error roughly $O((nM/\varepsilon)^{1/2})$.



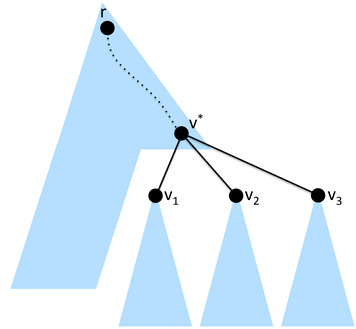
Proof idea: A k-covering of an unweighted graph is a subset S⊆V of the vertices such that any vertex v has distance at most k from some vertex s(v) ∈ S. We will find a k-covering of size $(nM\varepsilon)^{1/2}$ for $k = (n / M\varepsilon)^{1/2}$. Releasing approximate distances between only the vertices in S will allow us to estimate the distance between every pair of vertices in the graph.

Trees:

For trees we can release all-pairs distances with error only $\log^{2.5} n$. The result can be viewed as a generalization of a known result about privately answering threshold queries for a totally ordered domain, which corresponds to releasing distances on the path graph.

Theorem: There is an (ε, δ)-differentially private algorithm A which given a tree T releases with probability 1 – γ approximate distances between all pairs of vertices with error $O(\log^{2.5}(n) \log(1 / \gamma)) / \varepsilon$.

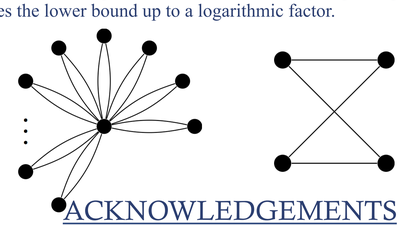
Proof idea: For trees there is a single path between any pair of vertices, so the problem of releasing all-pairs distances reduces to the problem of releasing single-source distances for an arbitrary root node.



We will use a recursive algorithm to solve the single-source problem. Given a root vertex r, we can always find a vertex v^* such that the subtree rooted at v^* has more than n / 2 children but the subtree rooted at each child $v_1, \ldots, v_t$ of v^* has at most n / 2 children. Using the Laplace mechanism, we will release noisy distances between r and v^* and between v^* and each of its children. We will then recursively apply the same procedure to the subtree rooted at each child v_i and to the remainder of the graph.

SPANNING TREES AND MATCHING

Using techniques similar to those we use for approximate shortest paths, we also obtain upper and lower bounds for the problems of finding almost-minimum spanning trees and low-cost matchings in our model. For both problems we obtain a lower bound of Ω(n) on the approximation error, and show that the algorithm which releases noisy edge weights almost matches the lower bound up to a logarithmic factor.



ACKNOWLEDGEMENTS

I am grateful to Salil Vadhan and Shafi Goldwasser for their guidance and mentorship, and to the Department of Energy Computational Science Graduate Fellowship for generously supporting my research.

Contact: asealfon@mit.edu