Differentially Private Graph Data Release: Inefficiencies & Unfairness

Ferdinando Fioretto
University of Virginia

Diptangshu Sen
Georgia Institute of Technology

Juba Ziani
Georgia Institute of Technology

Abstract

Networks in sectors like telecommunications and transportation often contain sensitive user data, requiring privacy enhancing technologies during data release to ensure privacy. While Differential Privacy (DP) is recognized as the leading standard for privacy preservation, its use comes with new challenges, as the noise added for privacy introduces inaccuracies or biases. DP techniques have also been found to *distribute these biases disproportionately across different populations, inducing fairness issues.* This paper investigates the effects of DP on bias and fairness when releasing network edge weights. We specifically examine how these privacy measures affect decision-making tasks, such as computing shortest paths, which are crucial for routing in transportation and communications networks, and provide both theoretical insights and empirical evidence on the inherent trade-offs between privacy, accuracy, and fairness for network data release.

1 INTRODUCTION

Networks underlie many important application domains, such as telecommunications, social networks, energy grids, and transportation. Therefore, it is necessary to publish network information to better understand their properties and serve a multitude of purposes like routing (transportation and computer networks), understanding (mis-)information propagation (social networks), for research and development (e.g., energy grids), or to inform public policy.

However network data often contains sensitive information and its release thus poses a key challenge. For example, releasing energy data can provide malicious entities insights into system vulnerabilities; data from social network and telecommunication can expose personal information about individuals' preferences, social interactions, and activities; transportation data can inadvertently reveal sensitive personal details like home addresses, healthcare-related visits, and other personal information (NYT, 2018, 2019).

Therefore, when releasing network data, it is crucial to protect potentially sensitive information. To this end, *Differential Privacy* (DP) (Dwork et al., 2006) has emerged as the leading paradigm for preserving individual privacy in aggregate-level data release. Notably, this privacy framework has been adopted in various deployments, including the 2020 U.S. Census (Bureau, 2023), Apple's device data collection and federated learning frameworks (Apple, 2017), and Google's location data and maps services (Google, 2024).

In a nutshell, DP relies on calibrated noise addition on the outputs of a computation to provide strong privacy guarantees. However, while this process ensures that the amount of sensitive information that can be "leaked" remain bounded, the added noise can introduce biases, potentially impacting the reliability of the data. While bias is a natural consequence of any private method, a concerning issue with DP is that it can distribute errors and biases *unevenly* across different groups, leading to concerns about fairness.

This work investigates the implications of DP on bias and fairness in network data release, focusing on routing recommendations. This constitutes a departure from previous research that primarily centered on the release of population histograms (e.g., in the U.S. Census) absent such network structure. Specifically, we examine the common scenario where the network structure is known but the edge weights need to be released privately. Our analysis shows how these perturbations influence tasks such as computing the shortest path and recommending optimal routes. Figure 1 presents an overview of our privacy model and data release, which we introduce in more detail in Section 3.

Proceedings of the 28[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258.

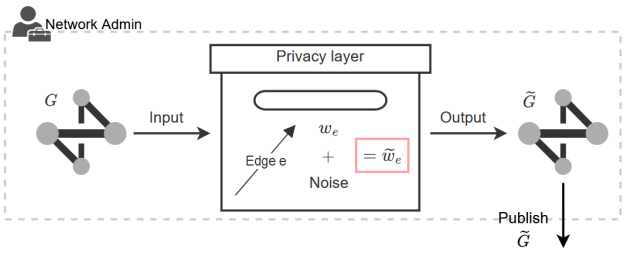


Figure 1: Schematic of privacy model: The network administrator privatizes graph G by adding calibrated noise to each edge weight w_e and publishes privatized graph $\widetilde{G}$, which is used to run downstream tasks.

Summary of contributions. In this work, we: **(1)** propose a model for differentially private network data release, assuming common knowledge of graph topology but requiring protection of sensitive edge weights through calibrated noise addition. This setup is detailed in Section 3. **(2)** investigate the bias and unfairness effects of using private (noisy) graph data to solve downstream optimization problems – particularly, the problem of computing shortest paths on the graph and recommending best routes to users. To the best of our knowledge, we are the first who seek to understand the tradeoff between privacy and fairness in the context of private graph data release. **(3)** develop a theoretical framework explaining how DP-induced biases could disproportionately affect certain groups, particularly through the mechanics of noise accumulation over different path lengths and the availability of alternate routes (Section 4). **(4)** finally, through extensive simulations on diverse network topologies, we demonstrate how privacy-related disruptions can vary by network type (Section 5). Importantly, this analysis also identifies network structures that are inherently more resilient to privacy-induced biases.

Literature Review. Observations that algorithms can mimic and amplify data biases have led to a new research area focusing on defining, analyzing, and mitigating unfairness (see Barocas et al. (2023); Mehrabi et al. (2021a); Pessach and Shmueli (2022)). The source of unfairness is often attributed to either data properties or model properties. For example, group size imbalance can create performance disparities (Mehrabi et al., 2021b). Additionally, constraining the model's hypothesis space to satisfy privacy (Bagdasaryan et al., 2019; Tran et al., 2021a), sparsity (Hooker et al., 2019, 2020; Tran et al., 2022), or robustness (Xu et al., 2021; Nanda et al., 2021; Tran et al., 2024) can also lead to disparate outcomes.

Particularly relevant is the study of disparate impacts caused by privacy-preserving algorithms, which has seen important developments (Fioretto et al., 2022). Much of this research, like ours, focuses on *differential privacy* (Dwork et al., 2006, 2014) as the formal notion leading to unfairness.

In particular, in the context of private data release (which involves revealing a full, privatized version of a dataset as opposed to simply releasing targeted statistics), Pujol et al. (2020) empirically showed that decision tasks made using DP datasets may disproportionately affect some groups of individuals over others. They noticed that the use of DP census data to allocate funds to school district produces unbalanced allocation errors, with some school districts systematically receiving more (or less) than what warranted. Later, Tran et al. (2021b) theoretically attributed these observations to two main factors: (1) the "shape" of the decision problem and (2) the presence of non-negativity constraints in post-processing steps Zhu et al. (2021, 2022).

To the best of our knowledge, no other work has studied the tension between privacy and fairness in downstream tasks on differentially-private *network* data. Related works like Sealfon (2016); Chen et al. (2023) do study DP computation of shortest paths but focus on releasing shortest path statistics, not the entire network, and do not address bias and fairness. Our paper builds on the intersection of privacy and fairness, providing an analysis of unfairness in a new context involving complex network structures.

2 PRELIMINARIES: DIFFERENTIAL PRIVACY

Differential Privacy (DP) (Dwork et al., 2006, 2014) provides a framework to safeguard individual data privacy by ensuring that the inclusion or exclusion of a single individual's data does not significantly affect the outcome of any analysis. Thus, an adversary cannot reliably determine whether an individual's data is part of the dataset based on the output of a computation.

Consider a mechanism $\mathcal{M}$ operating on a dataset $x = (x_1, \ldots, x_n)$, where each x_i represents an individual's data. Two datasets x and x' are called *neighboring* if they differ in exactly one individual's data: formally, if $\exists j \in [n]$ such that $x_j \neq x'_j$, and $x_i = x'_i$ for all $i \neq j$.

Formal definition. A randomized mechanism $\mathcal{M}$ satisfies (ε, δ)-DP if, for all neighboring datasets x and x', and for all subsets of outputs $O \in Range(\mathcal{M})$:

$$\Pr\left[\mathcal{M}(x) \in O\right] \leq \exp(\varepsilon) \Pr\left[\mathcal{M}(x') \in O\right] + \delta.$$

The privacy parameter ε controls the level of privacy: smaller ε implies stronger privacy but may reduce utility due to increased noise. As $\varepsilon \to 0$, we approach

perfect privacy as the output becomes independent of any single data point.

Let f be a query or computation applied to the data. DP mechanisms add noise to the computation based on the *query sensitivity* $\Delta f = \max_{\text{neighbors } x,x'} \|f(x) - f(x')\|$, where Δf quantifies the maximum potential change in the function's output across two neighboring databases. Lower sensitivity indicates small changes between outputs for neighboring databases, and thus require less noise to achieve privacy.

Numerical queries (which output real numbers) can be made differentially private by adding calibrated Gaussian noise to their true outputs. Given query f,

Lemma (Gaussian mechanism)**.** *The mechanism* $\mathcal{M}(f, x, \varepsilon, \delta) = f(x) + Z$ *where* $Z \sim \mathcal{N}\left(0, \sqrt{2\ln(1.25/\delta)} \cdot \Delta f/\varepsilon\right)$*, satisfies* (ε, δ)*-DP.*

In practice, there is a trade-off between privacy and utility: stronger privacy (smaller ε) requires adding more noise, which can reduce the accuracy of the output. *This paper focuses on understanding how this reduction in utility may be disproportionately distributed among different populations in the context of networks.*

Post-processing invariance. Differential privacy satisfies several key properties Dwork et al. (2014). In particular, it is immune to post-processing:

Theorem (Dwork et al. (2014))**.** *Let* $\mathcal{M}$ *be a* (ε, δ)*-DP mechanism and* f *any randomized function. Then the composition* $f \circ \mathcal{M}$ *also satisfies* (ε, δ)*-DP.*

3 MODEL: SETTINGS & GOALS

We consider the problem of *differentially private graph data release.* Formally, let $G = (V, E, \boldsymbol{w})$ be a weighted graph with vertex set V, edge set E, and weights $\boldsymbol{w} : E \to \mathbb{R}_{\geq 0}$. For each edge $e \in E$, $w(e)$ is used to denote the its weight, here used to represent the "time" or "cost" it takes to traverse it. Without loss of generality, we consider connected graphs G in which any two nodes are reachable from each other. Importantly, in this work we consider weights $\boldsymbol{w}$ that are functions of sensitive user data and whose values must be protected. For instance, the weights might represent traffic congestion based on commuter locations or the strength of private social relationships in a network. We write $w(e) = f_e(x_1, \ldots, x_n)$ where the x_i denotes the sensitive users information, $i \in [n]$.

Graph release model under DP. Consider a network administrator releasing a weighted graph G to a third party while preserving data privacy. To achieve this, a modified graph $\widetilde{G} = (V, E, \tilde{\boldsymbol{w}})$ is produced, keeping the nodes and edges the same but altering the edge weights $\tilde{w}$ to ensure differential privacy. This privatized graph retains the public network topology but safeguards sensitive weight information.

The administrator uses the Gaussian mechanism to release the weights $\tilde{\boldsymbol{w}}$. For each edge $e \in E$, produces a private weight: $\tilde{w}(e) = \max(0, w(e) + Z(e))$, where $Z(e) \sim \mathcal{N}(0, \sigma^2)$ is a centered Gaussian random variable[1]. The max function ensures all weights remain non-negative, note that this step retains differential privacy due to post-processing guarantees. If the sensitivity Δf of the function $f_e(\cdot)$ is bounded for all $e \in E$, the released graph with $\sigma = \sqrt{2\ln(1.25/\delta)} \cdot \Delta f/\varepsilon$ guarantees (ε, δ)-differential privacy. A higher σ provides stronger privacy guarantees. In this paper, we focus on σ as the main parameter controlling noise and privacy.

Impact of DP on bias and fairness. Adding noise for privacy and the subsequent post-processing (to ensure non-negativity of edge weights) can introduce bias in outcomes of tasks performed on the privatized graph $\widetilde{G}$. This paper aims to **(1)** characterize such bias both theoretically and experimentally, and **(2)** understand how different segments of the network may be *disproportionately affected*, leading to unfairness.

Our primary task is the computation of the *shortest path.* For any two vertices $i, j \in V$, let $\mathcal{P}_{ij}$ be the set of all paths between them. The *length* of a path $P \in \mathcal{P}_{ij}$ is $w_G(P) = \sum_{e \in P} w(e)$. The *shortest path* in the original graph G is:

$$P^*_{ij} = \arg\min_{P \in \mathcal{P}_{ij}} w_G(P) = \arg\min_{P \in \mathcal{P}_{ij}} \sum_{e \in P} w(e).$$

Our goal is to evaluate the extent to which DP mechanisms, when applied to graph G to produce graph $\widetilde{G}$, impact this computation. In the privatized graph $\widetilde{G}$, the *perceived* shortest path is computed as:

$$\tilde{P}_{ij} = \arg\min_{P \in \mathcal{P}_{ij}} w_{\widetilde{G}}(P) = \arg\min_{P \in \mathcal{P}_{ij}} \sum_{e \in P} \tilde{w}(e).$$

Although users compute paths using $\widetilde{G}$, the actual cost they incur corresponds to the original weights in G. Therefore, our evaluation metric is based on $w_G(\tilde{P}_{ij}) = \sum_{e \in \tilde{P}_{ij}} w(e)$. The *realized bias* or *error* is:

$$B_{ij}(\tilde{P}_{ij}) = \sum_{e \in \tilde{P}_{ij}} w(e) - \sum_{e \in P^*_{ij}} w(e).$$

[1]Weights on adjacent edges may be correlated due to uneven distribution of an individual's data along their path. While adding correlated noise could be an option, we avoid this for a key reason: it requires assumptions about the functional form of edge weights or their correlation. If these assumptions are wrong, the privacy guarantee can fail. It is standard for DP guarantees to hold in the *worst-case* with *minimal assumptions*, which is the approach we adopt here.

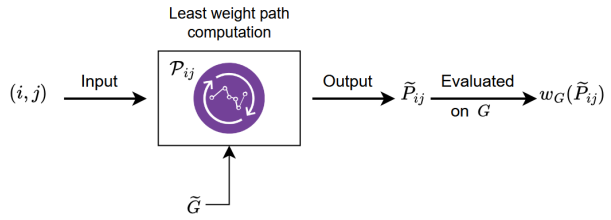


Figure 2: Evaluation Framework: Given node pair (i,j) and privatized graph $\widetilde{G}$, a user computes the shortest path on the set $\mathcal{P}_{ij}$. Computation returns $\tilde{P}_{ij}$ as the perceived shortest path, but the user decision is evaluated on original graph G incurring a cost of $w_G(\tilde{P}_{ij})$ and realizing bias $B_{ij} = w_G(\tilde{P}_{ij}) - w_G(P^*_{ij})$.

Given the stochastic nature of $\tilde{\boldsymbol{w}}$, the *perceived* shortest path $\widetilde{P}_{ij}$ varies. We thus define *expected bias* as:

$$\mathbb{E}[B_{ij}] = \mathbb{E}_{\tilde{\boldsymbol{w}}}\left[\sum_{e\in\tilde{P}_{ij}} w(e) - \sum_{e\in P^*_{ij}} w(e)\right]. \qquad (1)$$

In the numerical section, we will often work with relative errors or bias, defined as

$$R_{ij} = \frac{\mathbb{E}[B_{ij}]}{\sum_{e\in P^*_{ij}} w(e)}, \qquad (2)$$

which represents the percentage change in the length of the recommended path compared to the true shortest path. Figure 2 summarizes the evaluation framework.

4 BIAS: A THEORETICAL PERSPECTIVE

This section presents the main theoretical insights of our work. Our primary contribution is characterizing the bias of the shortest path computation due to privacy noise and understanding how it drives unfair outcomes across different types of source-destination pairs on graphs. We introduce our first result in Claim 4.1 which provides insights about the sign or direction of the bias (proof in Appendix Section A.1).

Claim 4.1. *The realized bias of the shortest path computation due to privacy noise is always greater than or equal to zero.*

A direct consequence of the above claim is that the *expected bias* and *expected relative bias* are non-negative. Note that all our numerical results in Section 5 plot empirical probabilities for incurring different levels of expected relative bias.

When it comes to fairness impacts of privacy, there are two main competing effects that drive which groups of node pairs will *unfairly* face more disruptions (on average) due to privacy:

1. The first of those is the *effective relative noise effect* which is explored in Section 4.1: when the number of path alternatives is fixed, we show that node pairs which are farther apart have a lower likelihood of being affected by privacy noise.
2. On the other hand, we also demonstrate the *path cardinality effect* in Section 4.2, i.e., the higher the number of different paths available to travel between the source and destination, the higher is the likelihood of shifting to a *worse* path due to privacy noise and incurring a large bias. This effect favors node pairs which are closer because they usually have a smaller number of alternate path options.

The trade-off between these two effects explains most of our observations in the numerical experiments section. We also provide a dual interpretation of our main theorem in Section 4.2 which helps us to derive high probability bounds on the realized bias of any shortest path computation.

Before we present our main results, we need to introduce some additional notation for ease of exposition. From now on, we drop the subscript "ij" whenever it is clear from context to simplify notations.

Definition 4.1. *For any two paths P_1 and P_2 in $\mathcal{P}_{ij}$, we define $S_{P_1,P_2} \subset E$ as follows:*

$$S_{P_1,P_2} := \{e \in E :\ e \in (P_1 \setminus P_2) \cup (P_2 \setminus P_1)\},$$

i.e., S_{P_1,P_2} is the set of those edges which belong in exactly one of the two paths P_1 and P_2.

If n_{P_1} and n_{P_2} denote the number of edges in paths P_1 and P_2 respectively, then $|S_{P_1,P_2}| \le n_{P_1} + n_{P_2}$ with equality when P_1 and P_2 have no overlapping edges.

4.1 Effective Relative Noise Effect

In this segment, we are interested in understanding the disparate impacts that privacy noise has on node pairs which are close by versus node pairs which are far apart, when the number of alternate path options is kept fixed for each pair. We measure the impact of noise by estimating the probability that for any two given paths, the *worse* one is perceived to be *better* when computations are done using privatized graph $\widetilde{G}$. Higher the value of this probability, higher is the impact of noise. We make the following conjecture:

Conjecture 4.2. *Node pairs which are closer incur, on average, larger levels of relative noise and hence are more impacted by privacy as opposed to node pairs which are farther apart.*

In order to gain intuition about why the above conjecture may be true, we will start by presenting the following technical result. Let P^* be the true shortest path between nodes i and j and $P' \neq P^*$ be any other alternate path. Define the **gap** α_{P',P^*} as $\alpha_{P',P^*} = w_G(P') - w_G(P^*)$. We assume that $\alpha_{P',P^*} > 0$ which means that P^* is strictly better than P'. Then,

Lemma 4.3. *The probability that path P' is perceived to be* shorter *than the true best path P^* on a privatized graph $\widetilde{G}$, i.e., $\mathbb{P}\left[w_{\widetilde{G}}(P') < w_{\widetilde{G}}(P^*)\right]$, is given by:*

$$q = \Phi^c\left(\frac{\alpha_{P',P^*}}{\sigma\sqrt{|S_{P',P^*}|}}\right),$$

where $\Phi^c(\cdot)$ is the complementary CDF of a standard normal random variable. We call "q" the ***path deviation probability****.*

Proof. The proof is in Appendix A.2. □

Intuition about Conjecture 4.2: We can obtain valuable insights about our earlier conjecture from Lemma 4.3. Suppose for a given pair of nodes, there are exactly 2 paths which have $|S|$ distinct edges between them and they differ in weight by amount α. This implies that the gap α is contributed by exactly $|S|$ edges on which the effective privacy noise has standard deviation $\sigma\sqrt{|S|}$. Therefore, the ratio $\frac{\sigma\sqrt{|S|}}{\alpha}$ represents the *effective relative noise* (effective noise relative to the weight gap between paths). Now, suppose we scale the number of edges by a factor of $M > 1$ to represent a node pair which are farther apart than the first pair. Assuming that all edge weights are i.i.d. samples from some distribution $\mathcal{D}$ and this new pair of nodes also have exactly 2 paths, the path gap between them should also scale by M in expectation. In this case, the *effective relative noise* is $\frac{1}{\sqrt{M}} \cdot \frac{\sigma\sqrt{|S|}}{\alpha}$. Because of the additional $\frac{1}{\sqrt{M}}$ factor, the effective relative noise is *smaller* on average for the pair of nodes farther apart. Therefore by Lemma 4.3, node pairs which are farther apart have on average, a lower likelihood of picking the worse path and hence are less affected by privacy noise.

Other observations from Lemma 4.3: Recall that the standard deviation of the privacy noise σ depends on the privacy parameter ε and the sensitivity of the weight function Δf. The dependence is of the following form: $\sigma \propto \frac{\Delta f}{\varepsilon}$. This implies that at higher levels of privacy (smaller ε), the probability q would be larger. This is intuitive: stronger privacy requires more perturbation to the edge weights and therefore there is a higher chance that the order is flipped, i.e., a previously longer path is perceived to be shorter. We can argue similarly for the case where the sensitivity of $f(\cdot)$ is high. Higher sensitivity of $f(\cdot)$ implies we need more noise to achieve the same level of privacy. This leads to higher q. We plot these dependencies in Figure 8 in Appendix Section C.

4.2 Path Cardinality Effect

In this segment, we are interested in understanding the disparate impacts that privacy noise has on node pairs which have many alternate path choices as opposed to node pairs which have fewer paths. We call this effect the *path cardinality* effect. In this case, we measure the impact of noise by estimating the probability of realizing bias at least as large as β, given some $\beta > 0$. Again, a higher probability indicates a higher impact of noise. We now make the following conjecture:

Conjecture 4.4. *Node pairs which have a large path cardinality are, on average, more impacted by privacy noise as opposed to node pairs which have fewer alternate path options.*

We present our main technical result that supports our conjecture in Theorem 4.5. Before stating the theorem, we need to introduce the following definition and set notations:

Definition 4.2. *(β-worse paths) Any path $P \in \mathcal{P}_{ij}$ is said to be β-worse, if:*

$$w_G(P) \geq w_G(P^*) + \beta,$$

where P^* is the least weight path between nodes i and j on graph G.

Therefore, given $\beta > 0$, we can partition set $\mathcal{P}_{ij}$ into two sets $\mathcal{P}_{ij}^{\geq\beta}$ and $\mathcal{P}_{ij}^{<\beta}$:

$$\mathcal{P}_{ij}^{\geq\beta} := \{P \in \mathcal{P}_{ij} :\ w_G(P) \geq w_G(P^*) + \beta\}$$

$$\mathcal{P}_{ij}^{<\beta} := \{P \in \mathcal{P}_{ij} :\ w_G(P) < w_G(P^*) + \beta\}$$

We are now ready to present our theorem:

Theorem 4.5 (Upper Bound on Bias Probability)**.** *Let q_β be the probability that the realized bias of shortest path computation using a privatized graph $\widetilde{G}$ is at least β. Then q_β is upper bounded as follows:*

$$q_\beta \leq \sum_{P\in\mathcal{P}_{ij}^{\geq\beta}} \Phi^c\left(\frac{\alpha_{P,P^*}}{\sigma\sqrt{|S_{P,P^*}|}}\right) \leq \left|\mathcal{P}_{ij}^{\geq\beta}\right| \cdot \Phi^c\left(\frac{\beta}{\sigma\sqrt{S_{max}}}\right),$$

where $S_{max} = \max_{P\in\mathcal{P}_{ij}^{\geq\beta}} |S_{P,P^*}|$.

Proof. The proof is in Appendix A.3. □

Observations from Theorem 4.5: We can derive useful insights from the expression of the upper bound. It is immediate that it depends on the cardinality of the set $\mathcal{P}_{ij}^{\geq\beta}$. I.e., the higher is the number of β-worse candidate paths, higher the probability that the shortest path changes to one such path which is exactly the intuition for Conjecture 4.4. The dependence on β is actually two-fold: firstly, as β increases, the term $\Phi^c\left(\frac{\beta}{\sigma\sqrt{S_{max}}}\right)$ decreases. Additionally, a higher β decreases the cardinality of $\mathcal{P}_{ij}^{\geq\beta}$. Essentially, this means that if β is large, the probability that we end up shifting to a β-worse path decreases very quickly (refer to Figure 9 in the Appendix). This idea will be explored in greater depth in Corollary 4.7.

Remark 4.1. *Note that the upper bound is tight when $|\mathcal{P}_{ij}^{<\beta}| = 1$ and $|\mathcal{P}_{ij}^{\geq\beta}| = 1$. In this case, we recover the exact expression we derived in Lemma 4.3, implying that our results are consistent.*

Theorem 4.6 (Lower Bound on Bias Probability)**.** *Fix any $\beta > 0$ and any $|P_i j^{\geq\beta}|$. There exists a graph instance G and a pair of nodes i, j where i) all paths between i and j are comprised of exactly k edges and ii) there are $|P_{ij}^{\geq\beta}|$ paths $P \neq P_{ij}^*$ with $w_G(P) = w_G(P_{ij}^*) + \beta$ such that for any $\delta > 0$:*

$$q_\beta \geq (1-\delta)\left(1 - \left(1 - \Phi^c\left(\frac{\beta - g(\delta)}{\sigma\sqrt{k}}\right)\right)^{|\mathcal{P}_{ij}^{\geq\beta}|}\right),$$

where g is a function of δ.

Proof. The proof is in Appendix A.5. □

When β grows large, $\Phi^c\left(\frac{\beta-g(\delta)}{\sigma\sqrt{k}}\right)$ grows to 0. In that case, we can use the binomial approximation of $(1-x)^n$ as $x \to 0$ to note that

$$q_\beta \geq (1-\delta)\left|\mathcal{P}_{ij}^{\geq\beta}\right| \cdot \Phi^c\left(\frac{\beta - g(\delta)}{\sigma\sqrt{k}}\right).$$

This recovers the dependency of Theorem 4.5 up to a $1-\delta$ multiplicative factor and a small linear shift in the Φ^c term, showing that our upper bound is essentially tight. We note that we recover a linear dependency in the number of alternative paths that are at least β-worse, as in Theorem 4.5.

Note: There is an alternate interpretation of Theorem 4.5 in terms of high-probability bounds on the realized bias, which leads to the following corollary:

Corollary 4.7. *Suppose, B_{ij} is the realized bias while computing the shortest path between nodes i and j using a privatized graph $\widehat{G}$. Then,*

$$\mathbb{P}\left[B_{ij} < \sqrt{2}\left(\sigma z^* \sqrt{S}\right)\right] \geq 1 - \gamma,$$

where $z^* = z_{1-\frac{\gamma}{|\mathcal{P}_{ij}|}}$ is the value at which the standard normal CDF evaluates to $1 - \frac{\gamma}{|\mathcal{P}_{ij}|}$ and S denotes the maximum number of edges in any path in $\mathcal{P}_{ij}$.

Proof. The proof can be found in the Appendix Section A.4 and follows directly from Theorem 4.5. □

Theorem 4.5 showed that as β increases, the probability of incurring a bias at least as large as β decreases sharply. This implies that the probability of incurring a large bias is very "small". This is exactly what Corollary 4.7 claims. Thus, Theorem 4.5 and Corollary 4.7 are duals of each other.

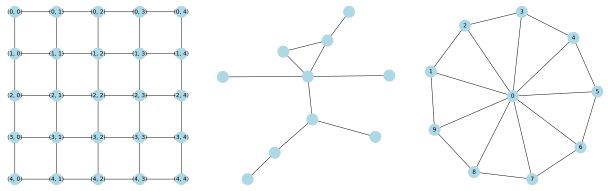
5 EXPERIMENTS: BIAS AND UNFAIRNESS

Next, we provide experimental results that extend and empirically validate our theoretical findings. The goal is to simulate the behavior of a DP release task on graphs that closely mimic real-world networks focusing on the impact of privacy on bias and fairness. We present the experimental setup in Section 5.1, and present a flavor of the results on 2 classes of graphs—grid graphs and scale-free graphs in Section 5.2. We provide additional experiments in Appendix B on a third class of graphs called "wheel graphs".

5.1 Experimental setup

The experiments investigate **three** classes of graphs: **i)** 2-dimensional grid graphs, **ii)** scale-free graphs, and **iii)** wheel graphs (defined and studied in Appendix B). While 2-D grids and wheel graphs closely emulate transportation networks in the real world (for example, Chicago and New York City have road networks laid out in an orthogonal grid pattern), scale-free graphs are often used to model other widely prevalent networks like social networks, the world wide web, etc. Thus, these graph classes cover a large variety of real-world networks. We use the following sets of parameters to generate synthetic networks for each graph class (Figure 3):

- **2-D grid graphs**: A grid graph of size N has N^2 nodes and $2N^2 + N$ edges.
- **Scale-free graphs**: These graphs have a degree distribution following a power law and are parametrized by their size (number of nodes N) and the exponent of the power law (γ). A higher γ indicates very few high-degree nodes, characteristic of many real-world networks like social networks. Unlike grid graphs, scale-free graphs are random, meaning that even with the same parameters, graph topologies may vary from one instance to another.



(a) $N = 5$ (b) $N = 10, \gamma = 2$ (c) $N = 10$

Figure 3: Schematics of different graph classes

Given a ground truth graph, we generate 100 private realizations by independently adding standard Gaussian noise to each edge (noise variance depends on privacy parameters ϵ and δ), followed by the post-processing step (Section 3). We report results that are averaged over all private realizations.

5.2 Results & Insights

5.2.1 Metrics

Given graph G, we aim to empirically estimate the probability that a randomly chosen node-pair (i, j) experiences a certain level of relative bias in its shortest path computation under privacy noise.

Path classification through relative bias. We consider the following levels of relative bias: i) 0% (indicating the shortest path remains unchanged), ii) $0 - 10\%$, iii) $10 - 20\%$, iv) $20 - 40\%$, v) $40 - 60\%$, vi) $60 - 100\%$, and vii) $> 100\%$. We classify node-pairs by first computing the shortest path weight between all pairs of distinct nodes on G and constructing the weight distribution of these paths. Each node-pair is then categorized based on the quartile of the weight distribution in which its true shortest path weight lies. We will index these categories 1 through 4. Category 1includes node-pairs whose shortest path weight lies in the first quartile (nodes are very close), while Category 4 includes those in the last quartile (nodes are very far apart). This categorization allows us to investigate whether privacy noise impacts node pairs differently based on their distance. When presenting our observations, we often compare Category 1 and Category 4 pairs because they represent the two extremes of the spectrum and are expected to have the maximum amount of disparity.

Impact of sparsity. We also study a variant of $2D$ grid graphs parametrized by a *sparsity factor* (Sp), which is the percentage of edges with a ground-truth weight of zero.[2] The motivation is that real-life transportation networks often have a significant proportion of edges with zero or near-zero traffic.

5.2.2 2D grid graphs

First, we study the $2-$D grid graph. For each ground truth graph instance, the edge weights are drawn independently from a $Uniform[0, 1]$ distribution. We generate results for different grid sizes ($N = 10, 40$) and different levels of noise (level of noise measured as the standard deviation relative to mean edge weight)(Figure 4). We make the following observations:

As the level of noise increases, node-pairs across all categories are more likely to incur a strictly positive relative bias. This follows directly from Lemma 4.3: for any node pair (i, j) and any path P, a higher noise level leads to a higher probability that $w_{\widetilde{G}}(P) < w_{\widetilde{G}}(P^*)$. Aggregating over all paths in $\mathcal{P}_{ij}$, the overall probability of a strictly positive relative bias increases.

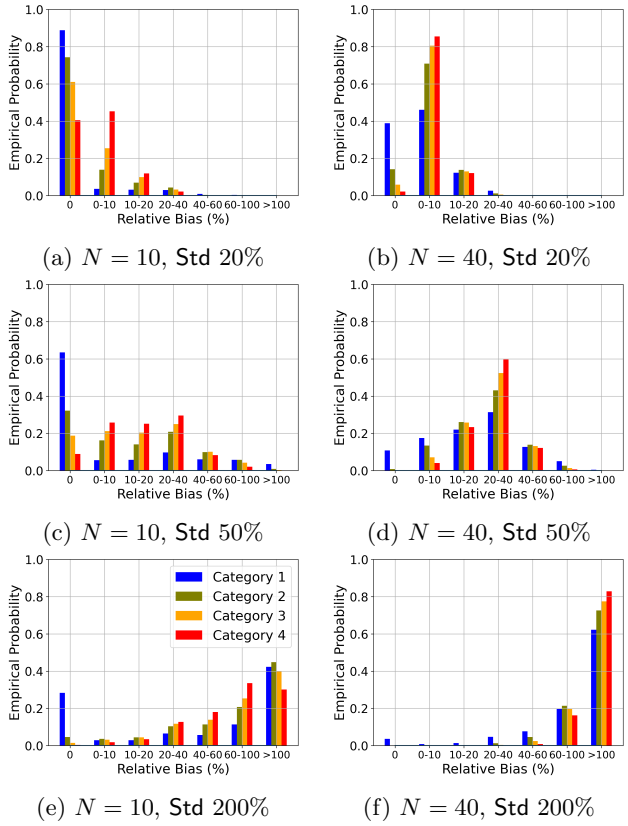
However, there is a clear disparity between the source-destination pairs in Category 1 and those in Category 4. At any noise level, Category 1 pairs are much more likely to remain unaffected compared to Category 4 pairs. Category 4 pairs usually represent nodes that are very far apart. On 2-D grid graphs, pairs of nodes that are farther apart have a larger set of alternative paths (higher $|\mathcal{P}_{ij}|$) and a higher number of edges on these paths (higher S_{max}), thus facing a higher risk of being affected by privacy noise. Here, the *path cardinality effect* explained in Section 4.1 overtakes the *effective relative noise* effect, in favor of shorter paths.

These trends are consistent across graph sizes N. However, as the grid size increases, the bar plots become increasingly right-heavy. This indicates that for the same noise level, a larger graph is more likely to induce higher magnitudes of relative bias across all categories of node pairs. This is again a consequence of the *path cardinality effect* which is amplified on large graphs.

Sparsity analysis. To further shed light on the disparities introduced by privacy, we present results on a grid graph of size $N = 20$ for different sparsity factors and at different levels of noise. (Figure 5). Here, sparsity introduces two interesting effects that are in tension with each other:

Impact on the number of bad paths: As Sp increases, most paths have low total weight. Also, there are fewer bad paths whose weight is significantly worse than that of the best path, which makes it *less likely across all categories of node pairs to switch to a worse-off path*; for example, in the extreme case where the sparsity

[2]Note that even at a sparsity of 0, there may be a significant amount of edges with a low ground truth weight, albeit not zero. For example, about 5 percent of edges are expected to have weight < 0.05 under a uniform distribution.



(a) $N = 10$, Std 20% (b) $N = 40$, Std 20%

(c) $N = 10$, Std 50% (d) $N = 40$, Std 50%

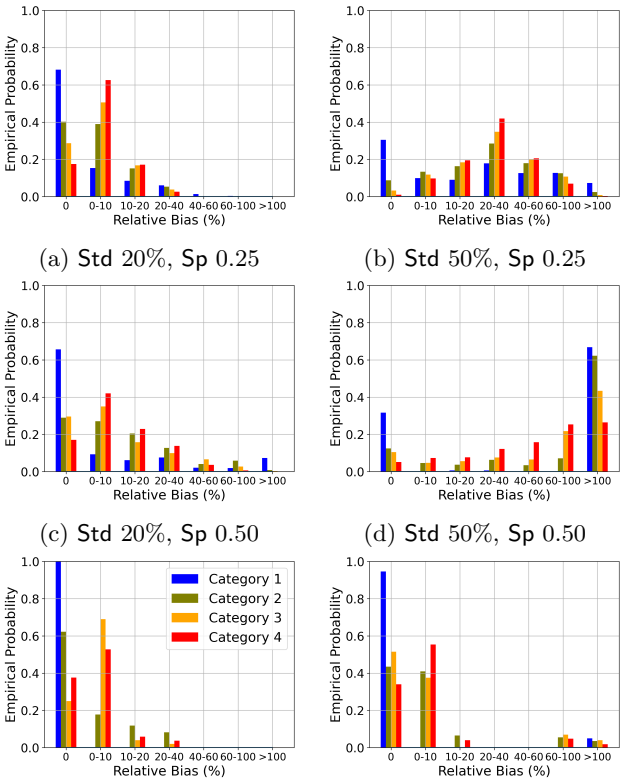
(e) $N = 10$, Std 200% (f) $N = 40$, Std 200%

Figure 4: 2D grid graphs - Empirical probability estimates of incurring different levels of relative bias on shortest path computation.

factor is 1, all paths have weight 0 and are equivalent. Further, *longer paths are disproportionately affected and more likely to switch to a worse path than shorter paths*: this is because node pairs which are farther apart are more likely to have a short alternative due to sparsity.

Impact on path weight estimation bias: The post-processing step means that noisy weights, if negative, are rounded up to 0—this introduces positive bias on edge weights. However, this bias affects edges disproportionately. In particular, edges whose weights are closer to 0 experience more positive bias (as these edges have a high probability of needing to be rounded up after noise addition). This means that paths with fewer edges are disproportionately more likely to be overestimated compared to paths with more edges.

Figure 5 shows the tension between these two effects. For a noise level of 20 %, the first effect dominates, leading to less overall relative bias, and this bias seems to affect Category 4 node pairs more than Category 1 pairs. As the noise level increases to 50 %, the second effect starts becoming important. However, at very high levels of sparsity (Sp 0.75), the first effect again seems to take over with Category 1 node pairs becoming extremely robust to privacy noise and Category 4 pairs being more affected.



(a) Std 20%, Sp 0.25 (b) Std 50%, Sp 0.25

(c) Std 20%, Sp 0.50 (d) Std 50%, Sp 0.50

(e) Std 20%, Sp 0.75 (f) Std 50%, Sp 0.75

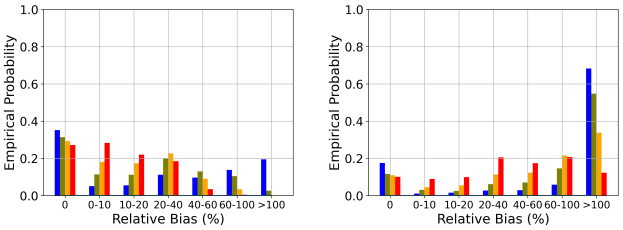
Figure 5: 2D grids graphs: Effects on privacy noise on path change statistics when graphs are sparse

5.2.3 Scale-free graphs

We conclude this section with a study of scale-free graphs where the parameter of interest is the power γ of the underlying degree distribution. Note that scale-free graphs can often have multiple disconnected components (including many singleton nodes of degree zero). However, for our simulation, we always pick its largest connected component. All ground truth edge weights are drawn independently from $Uniform[0, 1]$. The main observations from Figure 6 are:

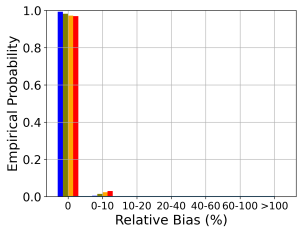
Similar to earlier results, higher levels of noise lead to a higher likelihood of incurring large relative bias across all categories of node pairs. At low levels of noise, Category 1 node pairs still continue to be more robust to noise compared to their Category 4 counterparts (a consequence of Theorem 4.5).

A striking observation is that at low values of γ ($\gamma \leq 2$), Category 1 node pairs are much more likely to incur significant amounts of relative bias (> 100 %) compared to Category 4 pairs at moderate to high levels of

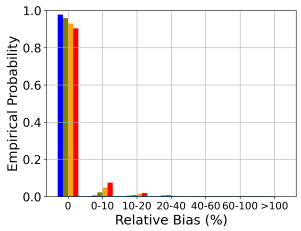


(a) Std 20%, $\gamma = 1.5$

(b) Std 50%, $\gamma = 1.5$



(c) Std 20%, $\gamma = 3$



(d) Std 50%, $\gamma = 3$

Figure 6: Results for scale-free graphs with $N = 100$.

noise. This is in sharp contrast with the results in 2-D grid graphs where, typically, Category 4 pairs were *worse-off* due to privacy. This is largely because of graph topology. When $\gamma \leq 2$, the graph has multiple densely connected centers that branch off into tree-like sub-graphs. A large proportion of Category 1 pairs are located close to the centres and therefore have a large number of path alternatives. The *path cardinality effect* increases their likelihood of incurring high bias. Further, Category 4 pairs are predominantly located on either side of connected centres–this means that they have, on average, the same number of path alternatives as their Category 1 counterparts, but those paths have a high degree of overlap and only diverge near the centre. This causes Category 4 pairs to incur the same levels of absolute bias as the Category 1 pairs, but they incur much smaller levels of relative bias because their paths are longer on average.

This trend becomes less significant for $\gamma > 2$ due to change in the graph topology. As γ increases, the number of nodes of high degree decrease significantly and the graph becomes less dense and more tree-like. As a result, for most node-pairs, there exists a unique path to go from source to destination which explains the low levels of bias incurred across all node categories, i.e., increased robustness to privacy noise.

Acknowledgements

This research was supported by the US National Science Foundation (NSF) under grants SaTC-2133169, CAREER 2401285 and CAREER 2336236. Any opinions and findings expressed in this material are those of the authors and do not reflect the views of their funding agencies.

References

Apple (2017). Learning with privacy at scale. `https://machinelearning.apple.com/research/learning-with-privacy-at-scale`.

Bagdasaryan, E., Poursaeed, O., and Shmatikov, V. (2019). Differential privacy has disparate impact on model accuracy. *Advances in neural information processing systems*, 32.

Barocas, S., Hardt, M., and Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT press.

Bureau, C. (2023). Why the census bureau chose differential privacy. `https://www.census.gov/library/publications/2023/decennial/c2020br-03.html`.

Chen, J. Y., Ghazi, B., Kumar, R., Manurangsi, P., Narayanan, S., Nelson, J., and Xu, Y. (2023). Differentially private all-pairs shortest path distances: Improved algorithms and lower bounds. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 5040–5067. SIAM.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer.

Dwork, C., Roth, A., et al. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407.

Fioretto, F., Tran, C., Hentenryck, P. V., and Zhu, K. (2022). Differential privacy and fairness in decisions and learning tasks: A survey. In *International Joint Conference on Artificial Intelligence*, pages 5470–5477. ijcai.org.

Google (2024). Get information about busy areas from google maps. `https://support.google.com/maps/answer/11323117?hl=en`.

Hooker, S., Dauphin, Y., Courville, A., and Frome, A. (2019). Selective brain damage: Measuring the disparate impact of model pruning.

Hooker, S., Moorosi, N., Clark, G., Bengio, S., and Denton, E. L. (2020). Characterising bias in compressed models. *ArXiv*, abs/2010.03058.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021a). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6):1–35.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021b). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6):1–35.

Nanda, V., Dooley, S., Singla, S., Feizi, S., and Dickerson, J. P. (2021). Fairness through robustness: Investigating robustness disparity in deep learning. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 466–477.

NYT (2018). Your apps know where you were last night, and they're not keeping it secret. `https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html`.

NYT (2019). Twelve million phones, one dataset, zero privacy. `https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html`.

Pessach, D. and Shmueli, E. (2022). A review on fairness in machine learning. *ACM Computing Surveys (CSUR)*, 55(3):1–44.

Pujol, D., McKenna, R., Kuppam, S., Hay, M., Machanavajjhala, A., and Miklau, G. (2020). Fair decision making using privacy-protected data. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 189–199.

Sealfon, A. (2016). Shortest paths and distances with differential privacy. In *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, pages 29–41.

Tran, C., Dinh, M., and Fioretto, F. (2021a). Differentially private empirical risk minimization under the fairness lens. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 27555–27565. Curran Associates, Inc.

Tran, C., Fioretto, F., Kim, J.-E., and Naidu, R. (2022). Pruning has a disparate impact on model accuracy. In *Advances in Neural Information Processing Systems*, volume 35. Curran Associates, Inc.

Tran, C., Fioretto, F., Van Hentenryck, P., and Yao, Z. (2021b). Decision making with differential privacy under a fairness lens. In *International Joint Conference on Artificial Intelligence*, pages 560–566. ijcai.org.

Tran, C., Zhu, K., Hentenryck, P. V., and Fioretto, F. (2024). Fairness increases adversarial vulnerability. In *International Joint Conference on Artificial Intelligence*, page TBA. ijcai.org.

Xu, H., Liu, X., Li, Y., Jain, A. K., and Tang, J. (2021). To be robust or to be fair: Towards fairness in adversarial training.

Zhu, K., Fioretto, F., and Hentenryck, P. V. (2022). Post-processing of differentially private data: A fairness perspective. In *International Joint Conference on Artificial Intelligence*, pages 4029–4035. ijcai.org.

Zhu, K., Hentenryck, P. V., and Fioretto, F. (2021). Bias and variance of post-processing in differential privacy. In *AAAI Conference on Artificial Intelligence*, pages 11177–11184. AAAI Press.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Not Applicable]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

(d) Information about consent from data providers/curators. [Not Applicable]

(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

(a) The full text of instructions given to participants and screenshots. [Not Applicable]

(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

A Missing Proofs

A.1 Proof of Claim 4.1

Proof. Suppose, some path $P \in \mathcal{P}_{ij}$ is the new *perceived* shortest path on privatized graph $\widetilde{G}$ instead of the true shortest path P^*_{ij} on G. In this case, the realized bias $B_{ij}(P)$ is given by:

$$B_{ij}(P) = \sum_{e \in P} w(e) - \sum_{e \in P^*_{ij}} w(e) = w_G(P) - w_G(P^*_{ij}).$$

Now, since P^*_{ij} is the *true* shortest path on G, by definition, it must be that:

$$w_G(P) \geq w_G(P^*_{ij}) \quad \forall\ P \in \mathcal{P}_{ij},$$

which directly implies that $B_{ij}(P) \geq 0$. Since the above holds for any general path $P \in \mathcal{P}_{ij}$, this concludes the proof of the claim. □

A.2 Proof of Lemma 4.3

Proof. Recall that $Z(e)$ is the amount of noise added to edge $e \in E$. We know that $Z(e)$'s are i.i.d. normal mean-zero random variables with variance σ^2. The proof idea is to express the event of choosing the wrong shortest path equivalently as an event when a certain linear inequality condition on $Z(e)$'s is satisfied. Then we can exploit the normality and independence properties of $Z(e)$'s to reason about the probability. The complete proof is presented below.

Note that the wrong path P' can be chosen if and only if $w_{\widetilde{G}}(P') < w_{\widetilde{G}}(P^*)$. Therefore,

$$\begin{aligned}
q &= \mathbb{P}\left[w_{\widetilde{G}}(P') < w_{\widetilde{G}}(P^*)\right] \\
&= \mathbb{P}\left[w_G(P') + \sum_{e \in P'} Z(e) < w_G(P^*) + \sum_{e \in P^*} Z(e)\right] \\
&= \mathbb{P}\left[\sum_{e \in P' \setminus P^*} Z(e) - \sum_{e \in P^* \setminus P'} Z(e) < w_G(P^*) - w_G(P')\right] \\
&= \mathbb{P}\left[\sum_{e \in P' \setminus P^*} Z(e) - \sum_{e \in P^* \setminus P'} Z(e) < -\alpha_{P',P^*}\right] \\
&= \mathbb{P}\left[\sum_{e \in P' \setminus P^*} Z(e) + \sum_{e \in P^* \setminus P'} Y(e) < -\alpha_{P',P^*}\right].
\end{aligned}$$

In the last step above, we substitute $Y(e) = -Z(e)$ for all $e \in P^* \setminus P'$. Note that $Y(e)$ and $Z(e)$ are identically distributed (because mean-zero Gaussian random variables are symmetric). Since each $Z(e), Y(e) \sim N(0, \sigma^2)$ and they are independent of each other, $\sum_{e \in P' \setminus P^*} Z(e) + \sum_{e \in P^* \setminus P'} Y(e) \sim N(0, |S_{P',P^*}|\,\sigma^2)$. This implies:

$$\begin{aligned}
q &= \mathbb{P}\left[\frac{\sum_{e \in P' \setminus P^*} Z(e) + \sum_{e \in P^* \setminus P'} Y(e)}{\sigma\sqrt{|S_{P',P^*}|}} < \frac{-\alpha_{P',P^*}}{\sigma\sqrt{|S_{P',P^*}|}}\right] \\
&= \Phi\left(\frac{-\alpha_{P',P^*}}{\sigma\sqrt{|S_{P',P^*}|}}\right) = \Phi^c\left(\frac{\alpha_{P',P^*}}{\sigma\sqrt{|S_{P',P^*}|}}\right).
\end{aligned}$$

The last step invokes the symmetry of a standard normal variable which allows, for any $a > 0$, $\Phi(-a) = \Phi^c(a)$. This concludes the proof of the lemma. □

A.3 Proof of Theorem 4.5

Proof. The proof idea is as follows: we can express q_β as the probability of the event that there exists a path in $\mathcal{P}_{ij}^{\geq \beta}$ which has the lowest weight on privatized graph $\widetilde{G}$. Since only one path can be the shortest path on any

realization of $\widetilde{G}$, the above event decomposes into a union of disjoint sub-events (a specific path in $\mathcal{P}_{ij}^{\geq\beta}$ is the new shortest path on $\widetilde{G}$). The technical parts of the proof deal with upper bounding the probability of each of these sub-events for which we use Lemma 4.3.

We can express q_β as follows:

$$\begin{aligned}
q_\beta &= \mathbb{P}\left[\text{shortest path on } \widetilde{G} \text{ is } \beta\text{-worse}\right] \\
&= \mathbb{P}\left[\exists\ P \in \mathcal{P}_{ij}^{\geq\beta} : w_{\widetilde{G}}(P) < w_{\widetilde{G}}(R)\ \forall\ R \in \mathcal{P}_{ij} \setminus P\right] \\
&\overset{(i)}{=} \sum_{P \in \mathcal{P}_{ij}^{\geq\beta}} \mathbb{P}\left[w_{\widetilde{G}}(P) < w_{\widetilde{G}}(R)\ \forall\ R \in \mathcal{P}_{ij} \setminus P\right] \\
&= \sum_{P \in \mathcal{P}_{ij}^{\geq\beta}} \mathbb{P}\left[\bigcap_{R \in \mathcal{P}_{ij} \setminus P} \{w_{\widetilde{G}}(P) < w_{\widetilde{G}}(R)\}\right].
\end{aligned}$$

The equality in step (i) above follows from the fact that events of the type $\{w_{\widetilde{G}}(P) < w_{\widetilde{G}}(R)\ \forall\ R \in \mathcal{P}_{ij} \setminus P\}$ are disjoint since two different paths cannot the best simultaneously (the event that two continuous random variables are equal, occurs with probability 0). Now, for each $P \in \mathcal{P}_{ij}^{\geq\beta}$, note that $P^* \in \mathcal{P}_{ij} \setminus P$. Therefore, we have:

$$\begin{aligned}
&\mathbb{P}\left[\bigcap_{R \in \mathcal{P}_{ij} \setminus P} \{w_{\widetilde{G}}(P) < w_{\widetilde{G}}(R)\}\right] \\
&\leq \mathbb{P}\left[w_{\widetilde{G}}(P) < w_{\widetilde{G}}(P^*)\right] = \Phi^c\left(\frac{\alpha_{P,P^*}}{\sigma\sqrt{|S_{P,P^*}|}}\right),
\end{aligned}$$

where the last equality follows from Lemma 4.3. It is important to note that we cannot compute the probability of the intersection event in closed form because the individual events are not mutually independent (two paths may have overlapping edges). Summing over all $P \in \mathcal{P}_{ij}^{\geq\beta}$, we derive the following upper bound:

$$q_\beta \leq \sum_{P \in \mathcal{P}_{ij}^{\geq\beta}} \Phi^c\left(\frac{\alpha_{P,P^*}}{\sigma\sqrt{|S_{P,P^*}|}}\right).$$

Finally, noting that $\alpha_{P,P^*} \geq \beta$ for all $P \in \mathcal{P}_{ij}^{\geq\beta}$ and from the definition of S_{max}, we have:

$$\Phi^c\left(\frac{\alpha_{P,P^*}}{\sigma\sqrt{|S_{P,P^*}|}}\right) \leq \Phi^c\left(\frac{\beta}{\sigma\sqrt{S_{max}}}\right) \quad \forall\ P \in \mathcal{P}_{ij}^{\geq\beta}.$$

This helps us simplify the upper bound even further and obtain the final result:

$$q_\beta \leq \sum_{P \in \mathcal{P}_{ij}^{\geq\beta}} \Phi^c\left(\frac{\alpha_{P,P^*}}{\sigma\sqrt{|S_{P,P^*}|}}\right) \leq \left|\mathcal{P}_{ij}^{\geq\beta}\right| \cdot \Phi^c\left(\frac{\beta}{\sigma\sqrt{S_{max}}}\right).$$

□

A.4 Proof of Corollary 4.7

Proof. Note that showing $\mathbb{P}\left[B_{ij} < \sqrt{2}\left(\sigma z^* \sqrt{S}\right)\right] \geq 1 - \gamma$ is equivalent to showing that:

$$\mathbb{P}\left[B_{ij} \geq \sqrt{2}\left(\sigma z^* \sqrt{S}\right)\right] \leq \gamma,$$

which again, is equivalent to showing $q_\beta \leq \gamma$ where $\beta = \sqrt{2}\left(\sigma z^* \sqrt{S}\right)$. Now, recall that we have already shown in Theorem 4.5 that for any $\beta > 0$, we have:

$$q_\beta \leq \left|\mathcal{P}_{ij}^{\geq \beta}\right| \cdot \Phi^c\left(\frac{\beta}{\sigma\sqrt{S_{max}}}\right).$$

We can construct a slightly more conservative upper bound on q_β by noting that $|\mathcal{P}_{ij}^{\geq\beta}| \leq |\mathcal{P}_{ij}|$ and $S_{max} \leq 2S$ (in the worst case, all paths in $\mathcal{P}_{ij}$ have S edges and have no overlapping edges which leads to $S_{max} = 2S$). Therefore,

$$q_\beta \leq |\mathcal{P}_{ij}| \cdot \Phi^c\left(\frac{\beta}{\sigma\sqrt{2S}}\right). \tag{3}$$

Hence, it is sufficient to show that when $\beta = \sqrt{2}\left(\sigma z^* \sqrt{S}\right)$, the revised upper bound in Equation 3 is $\leq \gamma$. This can be verified easily by plugging in the value of β, as follows:

$$\begin{aligned}
|\mathcal{P}_{ij}| \cdot \Phi^c\left(\frac{\beta}{\sigma\sqrt{2S}}\right) &= |\mathcal{P}_{ij}| \cdot \Phi^c\left(\frac{\sigma z^* \sqrt{2S}}{\sigma\sqrt{2S}}\right) \\
&= |\mathcal{P}_{ij}| \cdot \Phi^c\left(z^*\right) \\
&= |\mathcal{P}_{ij}| \cdot (1 - \Phi(z^*)) \\
&= |\mathcal{P}_{ij}| \cdot \frac{\gamma}{|\mathcal{P}_{ij}|} \\
&= \gamma.
\end{aligned}$$

This concludes the proof of the corollary. □

A.5 Proof of Theorem 4.6

Proof. The proof is constructive. We will construct a setting where q_β matches the proposed lower bound.

Consider a pair of nodes (i, j) on graph G such that:

1. For any two paths $P_1, P_2 \in \mathcal{P}_{ij}$, P_1 and P_2 have no over-lapping edges, i.e., $|S_{P_1,P_2}| = n_{P_1} + n_{P_2}$.
2. For any two paths $P_1, P_2 \in \mathcal{P}_{ij}$ such that $P_1 \neq P_{ij}^*$ and $P_2 \neq P_{ij}^*$, P_1 and P_2 are identical, i.e., they have exactly k edges, each with the same ground truth weight (this implies, $w_G(P_1) = w_G(P_2)$). The true shortest path P_{ij}^* also has exactly k edges. Let the common path gap with the true shortest path P_{ij}^* be β.

Thus, we have constructed a scenario where there is a unique shortest path and all the remaining paths are identical and equally worse by amount β. Then, our aim is to compute q_β. Observe that:

$$\begin{aligned}
q_\beta &= \mathbb{P}\left[\text{shortest path on } \widetilde{G} \text{ is } \beta\text{-worse}\right] \\
&\overset{(i)}{=} 1 - \mathbb{P}\left[P_{ij}^* \text{ is the shortest path on } \widetilde{G}\right] \\
&= 1 - Q.
\end{aligned}$$

Step (i) follows from our construction of the set $\mathcal{P}_{ij}$. Now, note that with probability $1-\delta$, $w_{\widetilde{G}}(P_{ij}^*) \geq w_G(P_{ij}^*) - \gamma$ where $\gamma(\delta) \triangleq -\sigma\sqrt{k} \cdot \Phi^{-1}(\delta)$. This in particular implies that:

$$\begin{aligned}
Q &= \mathbb{P}\left[P_{ij}^* \text{ is the shortest path on } \widetilde{G}\right] \\
&= \mathbb{P}\left[w_{\widetilde{G}}(P_{ij}^*) < w_{\widetilde{G}}(P) \quad \forall\ P \in \mathcal{P}_{ij}^{\geq\beta}\right] \\
&\leq (1-\delta) \cdot \mathbb{P}\left[w_G(P_{ij}^*) - \gamma < w_{\widetilde{G}}(P) \quad \forall\ P \in \mathcal{P}_{ij}^{\geq\beta}\right] + \delta \\
&= (1-\delta) \cdot \prod_{P \in \mathcal{P}_{ij}^{\geq\beta}} \mathbb{P}\left[w_G(P_{ij}^*) - \gamma < w_{\widetilde{G}}(P)\right] + \delta.
\end{aligned}$$

where the second-to-last step follows from conditioning on $w_{\widetilde{G}}(P^*_{ij}) \geq w_G(P^*_{ij}) - \gamma$ for the first term and $w_{\widetilde{G}}(P^*_{ij}) \leq w_G(P^*_{ij}) - \gamma$ for the second term; and the last step follows from independence because the paths do not overlap.

Since $w_{\widetilde{G}}(P) \sim N\left(w_G(P^*_{ij}) + \beta, k\sigma^2\right)$, we then note that:

$$\mathbb{P}\left[w_G(P^*_{ij}) - \gamma < w_{\widetilde{G}}(P)\right] = \Phi^c\left(\frac{-\gamma - \beta}{\sigma\sqrt{k}}\right) = \Phi\left(\frac{\gamma + \beta}{\sigma\sqrt{k}}\right).$$

Finally, since all paths $P \in \mathcal{P}_{ij}^{\geq \beta}$ are identical, we have:

$$Q \leq (1-\delta) \cdot \left[\Phi\left(\frac{\gamma + \beta}{\sigma\sqrt{k}}\right)\right]^{|\mathcal{P}_{ij}^{\geq \beta}|} + \delta,$$

which implies that:

$$\begin{aligned} q_\beta &\geq 1 - \delta - (1-\delta) \cdot \left[\Phi\left(\frac{\gamma + \beta}{\sigma\sqrt{k}}\right)\right]^{|\mathcal{P}_{ij}^{\geq \beta}|} \\ &= (1-\delta)\left(1 - \left(1 - \Phi^c\left(\frac{\gamma + \beta}{\sigma\sqrt{k}}\right)\right)^{|\mathcal{P}_{ij}^{\geq \beta}|}\right). \end{aligned}$$

□

B Additional Experiments

We also examine wheel graphs which closely emulate road networks in cities like Paris and Rome. These graphs have two types of edges: **i)** circumference edges and **ii)** spoke edges. All circumference edges have their ground truth weights drawn independently from a *Uniform*$[0, 1]$ distribution. Since spoke edges are expected to accommodate larger flows, their ground truth weights are drawn independently from *Uniform*$[0, r]$ where $r \geq 1$. Thus, r represents the ratio of mean edge weights for the two groups of edges. For numerical experiments, our parameters of interest are the following: i) size of the graph N and ii) ratio r. However, wheel graphs have circular symmetry which means that N does not affect the outcomes independently. So, we fix $N = 101$ for all experiments and only vary r from the following set: $\{1, 20, 50, 100\}$. Additionally, like all previous experiments, we also consider different levels of privacy noise: 20 %, 50 % and 100 %. Refer to Figure 7 for a graphical representation of all results, based on which we make the following observations:

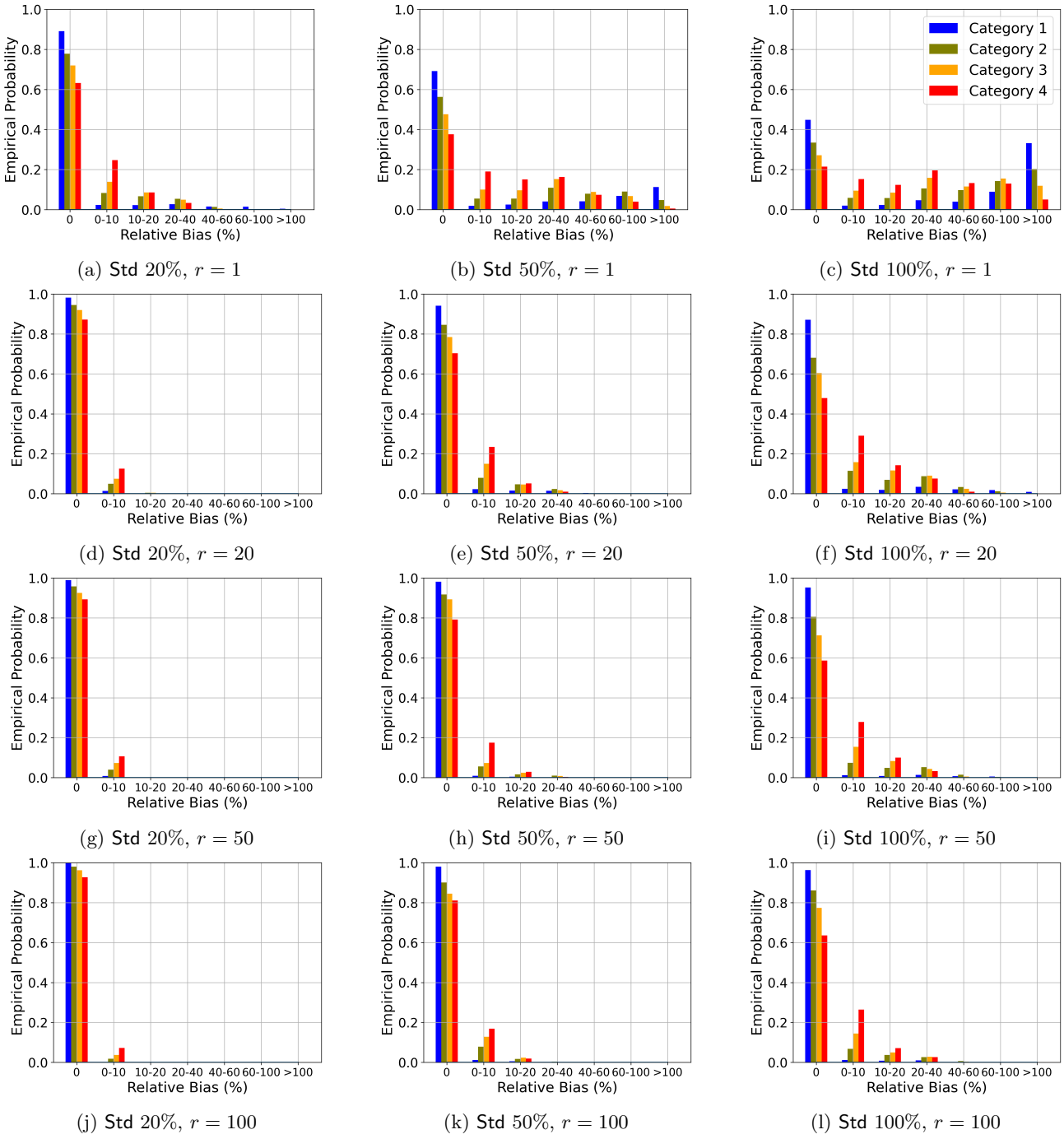
Similar to the observations for 2-D grid graphs, as the levels of noise increase, node pairs of all categories are more likely to be affected. Once again, Category 1 pairs are significantly more robust against privacy noise compared to Category 4 pairs, for the same reasons as highlighted earlier.

The most striking observation is that the ratio r greatly influences the degree to which bias is realized. As r increases, all node pairs become more and more robust to privacy noise. This is a direct consequence of the topology of a wheel graph. Note that there are only two kinds of source-destination pairs: **i)** between a central node and an outer node, and **ii)** between two outer nodes. In both cases, with high r, there is only one candidate path that is the most viable shortest path. For case **i)**, it involves identifying the spoke edge with the least weight, traversing it to reach the corresponding outer node, and then traveling along the circumference to reach the destination. For case **ii)**, the only feasible least-cost path is to travel along the low-weight paths on the circumference (any trip to the center involves traversing a high-weight spoke edge and is sub-optimal). This result follows from Theorem 4.5: in this case, the large gap β between the best path and all other paths drives the probability q_β to very low levels, leading to a high degree of robustness.

C Missing Plots

C.1 Graphical representation of the Effective Relative Noise Effect

We graphically demonstrate in Figure 8 the effects of the path gap α, sensitivity Δf and the size of the disjoint edge set $|S|$ on probability q, as predicted by Lemma 4.3.



(a) Std 20%, $r = 1$ (b) Std 50%, $r = 1$ (c) Std 100%, $r = 1$

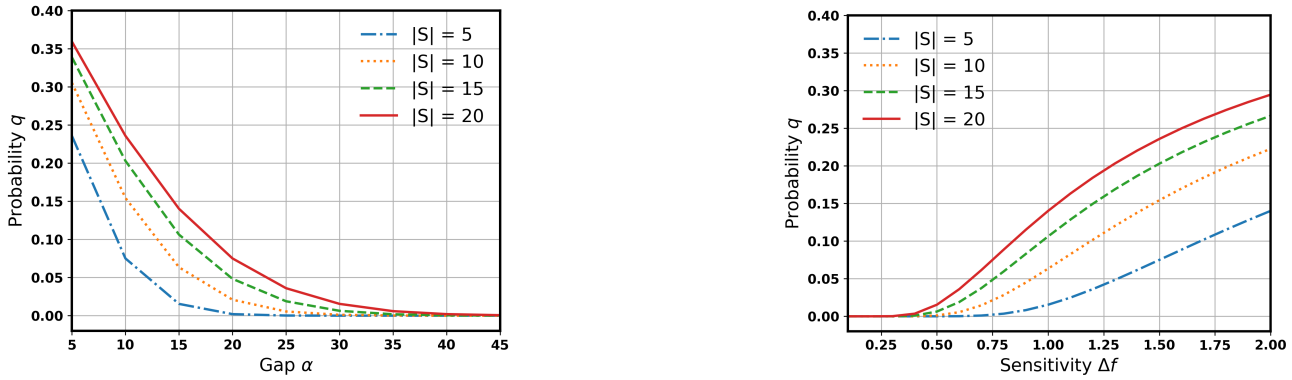
(d) Std 20%, $r = 20$ (e) Std 50%, $r = 20$ (f) Std 100%, $r = 20$

(g) Std 20%, $r = 50$ (h) Std 50%, $r = 50$ (i) Std 100%, $r = 50$

(j) Std 20%, $r = 100$ (k) Std 50%, $r = 100$ (l) Std 100%, $r = 100$

Figure 7: Statistics for wheel graphs with $N = 101$ nodes. In each row (from left to right), we generate results for 3 different levels of noise: i) 20%; ii) 50%; and iii) 100%. On the other hand, in each column (from top to bottom), we plot results for different values of r: i) $r = 1$; ii) $r = 20$; iii) $r = 50$; and iv) $r = 100$.

Note: q also depends on the local network topology of paths P' and P^* as we illustrate with the following example. Let there be two users traveling between two different node pairs, each of them has two path choices, one which is the true best and another which is strictly worse. For ease of comparison, we assume that for both node pairs, the worse path is off the respective true best by the same amount α. Now, suppose that user 1 faces a scenario where both of her paths have a large degree of overlap, leading to a smaller $|S|$, while for user 2, the paths are largely distinct. In this case, user 2 has a higher chance of deviating to the *worse* path, simply because noise on shared edges affects both paths equally. This example demonstrates that despite the number



(a) Variation with gap α

(b) Variation with sensitivity Δf

Figure 8: Variation of probability q as a function of gap α_{P',P^*} in (a) and sensitivity Δf in (b) for different values of $|S_{P',P^*}|$. We set $(\varepsilon, \delta) = (1, 0.01)$. Additionally, for (a), we fix $\Delta f = 1$ and for (b), we fix $\alpha_{P',P^*} = 15$.

of alternative paths and the path gap being identical, unfairness can also arise due to network topology wherein privacy has a much more adverse effect on some users compared to others.

C.2 Evolution of the Upper Bound in Theorem 4.5

We demonstrate in Figure 9 how the upper bound on q_β derived in Theorem 4.5 evolves as a function of β. We use a wheel graph with $N = 21$ nodes. All ground truth edge weights drawn independently from $U[0, 1]$. We plot results for two types of source-destination pairs: the blue legend is for a pair of nodes which lie on diametrically opposite sides of the wheel graph, the red legend is for a pair of nodes consisting of the central node and a circumference node. The noise is sampled from a mean zero Gaussian distribution with standard deviation $\sigma = 0.3$. For very small values of β, the bound is vacuous. However, once the bound becomes non-trivial, it decreases rapidly and can be expected to approximate q_β very accurately.

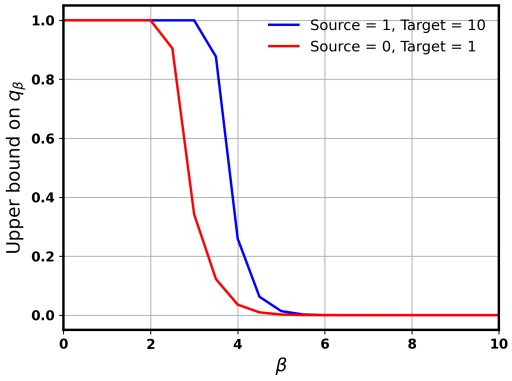
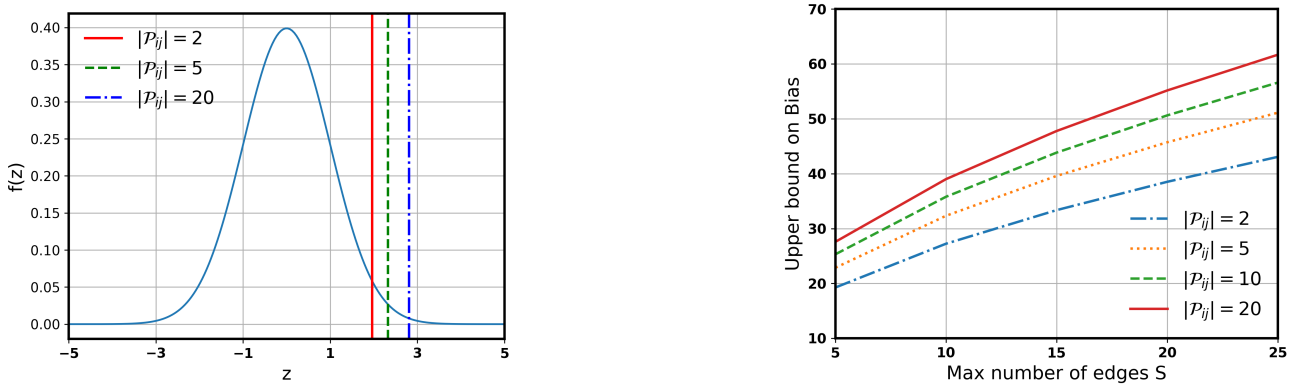


Figure 9: Evolution of the upper bound on q_β as a function of β for a wheel graph with $N = 21$.

C.3 How does the Upper Bound on Realized Bias grow in Corollary 4.7?

In Figure 10 (a), we show how the z-scores change with the cardinality of $\mathcal{P}_{ij}$. Higher values of $|\mathcal{P}_{ij}|$ leads to higher z-scores. For all cases, we use $\gamma = 0.05$, i.e., we desire 95% coverage. In (b), we illustrate how the bounds on bias B_{ij} calculated in Corollary 4.7 vary with S and $|\mathcal{P}_{ij}|$. The bound clearly grows as $O(\sqrt{S})$ in S.



(a) (b)

Figure 10: Graphical representation of the upper bound in Corollary 4.7