Controlling the Spread of Epidemics on Networks with Differential Privacy: Supplementary Material

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Designing effective strategies for controlling epidemic spread by vaccination is an important question in epidemiology, especially in the early stages when vaccines are limited. This is a challenging question when the contact network is very heterogeneous, and strategies based on controlling network properties, such as the degree and spectral radius, have been shown to be effective. Implementation of such strategies requires detailed information on the contact structure, which might be sensitive in many applications. Our focus here is on choosing effective vaccination strategies when the edges are sensitive and differential privacy guarantees are needed. Our main contributions are (ε, δ)-differentially private algorithms for designing vaccination strategies by reducing the maximum degree and spectral radius. Our key technique is a private algorithm for the multi-set multi-cover problem, which we use for controlling network properties. We evaluate privacy-utility tradeoffs of our algorithms on multiple synthetic and real-world networks, and show their effectiveness.

1 Introduction

A fundamental public health problem is to implement interventions such as vaccination to control the spread of an outbreak, e.g., [37, 8]. This is especially important in the early stages of an outbreak, when resources are limited. Here, we focus on network based models for epidemic spread, such as SI/SIS/SIR models, in which the disease spreads on a contact network $G = (V, E)$ from an infected node $u \in V$ to each susceptible neighbor v of u independently with some probability, e.g., [32, 1, 16, 36]; such models (which are simplifications of agent based models) have been used extensively in public health analyses in recent years. Interventions such as vaccination and isolation, can be modeled as node removal in such models [32]. The *Vaccination Problem* (VP), introduced in [15] for an SI type model (a similar version was considered in [21]), formalizes the design of an optimal vaccination strategy as choosing a subset $S \subset V$ so that the expected number of infections in the residual graph $G[V \setminus S]$ is minimized. This problem remains a challenging computational problem, and is NP-hard, in general [21, 15, 45].

Due to the computational hardness of the vaccination problem, a number of heuristics have been proposed for choosing a set S to vaccinate, which involve choosing nodes based on properties related to the underlying contact network, such as degree and different notions of centrality, e.g., betweenness, pagerank and eigenscore [8, 9, 14, 16]; such heuristics have been shown to be much more effective that picking nodes randomly. In particular, choosing nodes which lead to a reduction in certain network properties of the residual network (i.e., after the vaccinated nodes are removed), below a critical threshold are quite effective. Examples of such strategies are reducing the maximum degree (the MaxDeg problem) [4, 36, 9], and the spectral radius (the MinSR problem) [46, 38]; we note that heuristics for the MaxDeg and MinSR problems have been used in many network based epidemic models, such as SIS, SIR, SEIR, etc. [38], as well as other contagion models, such as spread of influence [27, 43]. Optimal choice of such nodes (whose removal leads to the maximum reduction in such metrics) is also a difficult computational problem. There is a lot of work on approximation algorithms, e.g., [44, 46, 39, 41, 40], and is our focus here.

In most settings, data privacy is a fundamental challenge, due to the risk of revealing sensitive private information of users. For instance, individuals might wish to keep certain kinds of contacts private, since these might reflect sensitive activities they participate in. Privacy concerns were a major factor limiting user adoption of digital contact tracing apps [5]. Differential Privacy (DP) [13] has emerged as a very popular notion for supporting queries on private and sensitive data. Here, we study the problems of choosing nodes with edge DP guarantees to minimize the maximum degree (PrivMaxDeg) and the spectral radius (PrivMinSR); we note that the edge DP model has been studied quite extensively (the other commonly used model of node DP, e.g., [26, 49, 25], is not suitable for the PrivMaxDeg and PrivMinSR problems, since the goal is to output selected nodes). There has been recent work on different kinds of epidemic analyses with privacy, e.g., [7, 31, 30], and for more general problems of network science and graph mining, e.g., [34, 12, 10, 26, 49, 25]. However, *the* PrivMaxDeg *and* PrivMinSR *problems have not been studied so far*.

The PrivMaxDeg and PrivMinSR problems are closely related to a fundamental problem in combinatorial optimization, namely multi-set multi-cover. While there has been some work on covering problems with privacy, e.g., [19, 30, 11, 18], the version we study has not been considered before. Further, most of the prior work on covering problems with privacy, except [30], considers an *implicit* or *blackboard* model, which does not make the solution explicit; instead, sets which are part of the solution know this implicitly. This kind of implicit solution is not suitable for problems of epidemic control we consider here, and we design techniques to make our private solutions explicit. Our main contributions are summarized below.

1. Minimizing the maximum degree with edge DP (PrivateMaxDeg). We design Algorithm 1 (Section 4.2) for this problem, and show that it gives an $O(\ln n \ln(e/\delta)/\epsilon)$-approximation, with high probability. PrivMaxDeg can be reduced to the private multi-set multi-cover problem (PrivateMulSet), a generalization of the set cover problem with privacy, which hasn't been considered before. We show that the iterative exponential mechanism can be used for PrivateMulSet (Section 4), and discuss how PrivMaxDeg can be solved by reduction to it (Algorithm 1). We also show how to construct explicit solutions for PrivMaxDeg (Algorithm 2) using the sparse vector technique [13].

2. Minimizing the spectral radius with edge DP (PrivMinSR). This turns out to be a much harder problem because non-private algorithms use metrics (e.g., number of walks through a node) which have high sensitivity [44]. While the spectral radius satisfies $\rho(G) \leq \Delta$, where $\rho(G)$ and Δ denote the spectral radius and maximum degree, respectively, this bound can be quite weak in many graphs. We present two algorithms which lead to stronger bounds on $\rho(G)$ under different regimes (Section 5); the first is based on reducing the number of walks of a certain length, as in [44], and the second is in terms of the average degree of neighbors [17].

3. Lower bounds. It is well-known that for the covering problems, no differentially private algorithms can both output a non-trivial explicit solution and satisfy the covering requirement at the same time. We derive the lower bounds for even outputting an explicit partial coverage requirement, stating that any (ϵ, δ)-differentially private algorithm using no more than $O(\log n) + |OPT|$ must incur an additive partial coverage requirement error of at least $\Omega(\log n)$. Similarly, for the PrivateMaxDeg, the explicit solution must have an additive error of at least $\Omega(\log n)$ for the target maximum degree.

4. Experimental results. We evaluate our methods on realistic and random networks. Our solutions lead to good bounds on both the maximum degree and the spectral radius. We find that implicit solutions have a higher cost relative to the non-private solutions, while the explicit solutions are quite sensitive to the privacy parameters, highlighting the need for carefully choosing the privacy parameters. We observe that our empirical results for the PrivateMaxDeg problem are consistent with the theoretical bounds we prove for our algorithms.

Some algorithms, proofs, and experimental results are provided in the supplementary material due to space constraints.

2 Related Work

As mentioned earlier, the PrivMaxDeg and PrivMinSR problems have not been studied earlier. We briefly summarize prior work on two areas directly related to our work: (1) network-based epidemic control and (2) differential privacy for network and graph problems; additional discussion is presented in Section A in the appendix. There has been a lot of work on non-private algorithms for controlling epidemic spread on networks, e.g., [48, 14, 9, 33]. As mentioned earlier, strategies based on degree or centrality, e.g., [9, 33], have been shown to be quite effective in many classes of networks (including random graphs). There has also been prior work on reducing the spectral radius of the contact network, e.g., [41, 39, 40, 44, 35], which is closely related to the concept of epidemic threshold–a quantity that determines if there will be a large outbreak or not.

While there is a lot of work on private computation of different kinds graph properties (e.g., degree distribution, subgraph counts and community detection), e.g., [26, 22, 3, 23, 49], there is no prior work on the problems of controlling metrics related to epidemic spread. The most relevant work involves private algorithms for other problems in computational epidemiology , e.g., computing the reproductive number [7], estimation of the number of infections [31], and determining facility locations for vaccine distribution [30]. However, none of these methods imply solutions for the problems we study here.

3 Preliminaries

Definition 3.1. *A mechanism $M : \mathcal{X} \to \mathcal{Y}$ is (ϵ, δ)-differentially private if for any two neighboring inputs $X_1 \sim X_2$, and any measurable subset of the output space $S \subseteq \mathcal{Y}$, the following holds:* $\Pr[M(X_1) \in S] \leq e^{\epsilon} \Pr[M(X_2) \in S] + \delta$ *[13].*

When $\delta = 0$, we say that M is ϵ-differentially private. We study graph datasets, i.e., $\mathcal{X}$ corresponds to the set graphs with n nodes. We consider the edge-DP model, where V, the set of nodes, is public and E, the set of edges, is kept private. More formally, two networks $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$, are considered neighbors if $V_1 = V_2$ and there exists an edge e such that $E_1 = E_2 \cup \{e\}$ or $E_2 = E_1 \cup \{e\}$ (i.e. they differ in the existence of a single edge). We note that there are other models of privacy in graphs, such as node DP, e.g., [26]; since our problems involve choosing subsets of nodes to be vaccinated, this model is not relevant here, and we only focus on edge DP.

We also utilize some standard privacy techniques and notations, such as the Exponential mechanism, Laplace mechanism, and AboveThreshold. See Appendix B for their definitions.

3.1 Problem Formulations

We study interventions for epidemic control, such as vaccination or isolation, which can be modeled as removing nodes from a contact network $G = (V, E)$ under the SIR model [32, 1, 16, 36]. Reducing structural properties of the contact network—such as the maximum degree $\Delta(G)$ or the spectral radius $\rho(G)$ – can help limit epidemic spread [32, 38].

Let $n = |V|$ and $m = |E|$. For a graph G, let $d(v, G)$ denote the degree of a node v, and let $\Delta(G) = \max_v d(v, G)$ be the maximum degree in G. Let $\rho(G)$ denote the largest eigenvalue of the adjacency matrix of G. We also consider weighted graphs where $w(v)$ is the weight of node v.

Definition 3.2. *(*PrivMaxDeg *problem) Given a graph $G = (V, E)$, a target max degree $D < \Delta(G)$, and privacy parameters ϵ, δ, the goal is to compute the smallest subset $S \subseteq V$ to remove (or vaccinate), such that the induced subgraph $G' = G[V \setminus S]$ satisfies $\Delta(G') \leq D$, while satisfying edge-DP.*

We refer to the non-private version of this problem as MaxDeg, and use $OPT_{\text{MaxDeg}}(G, D) = \min\{|S| : S \subseteq V, , \Delta(G[V \setminus S]) \leq D\}$ to denote the optimal solution of the non-private version.

Definition 3.3. *(*PrivMinSR *problem) Given a graph G, a target threshold τ, and privacy parameters ϵ, δ, the goal is to compute the smallest subset $S \subseteq V$ to remove, such that $\rho(G[V \setminus S]) \leq \tau$, while satisfying edge-DP.*

We refer to the non-private version of this problem by MinSR. Many bounds are known for the spectral radius, including: $\rho(G) \le \Delta(G)$ and $\rho(G) \le \max_v \sqrt{d(v,G)d_2(v,G)}$, where $d_2(v,G) = \sum_{u\sim v} d(u,G)/d(v,G)$ [24].

Explicit and implicit solutions. For the problems discussed above, the *explicit* version outputs an actual solution S that satisfies edge-DP. However, for covering type problems, this is often challenging under DP [19]. We therefore also consider *implicit* solutions – these output a differentially private quantity π such that each node v can determine whether it is part of the solution based on π and G.

3.2 Multi-set Multi-cover problem

To solve some of the above problems, we reduce them to the Multi-set Multi-cover problem, which we formally define as follows:

Definition 3.4. *(*MULSET *problem) Let $U = \{e_1, \dots, e_n\}$ be a universe set on n distinct elements. For each element $e \in U$, let the* covering requirement *r_e be the minimum number of times e must be covered, and let $R = \{r_e\}_{e\in U}$. Let $\mathcal{S} = \{S_1, \dots, S_m\}$ be a collection of multi-sets, where each set S_i contains $m(S_i, e)$ copies of element e. We refer to $m(S_i, e)$ as the* multiplicity *of e in S_i. The* MULSET$(U, \mathcal{S}, R)$ *asks to find the smallest sub-collection $\mathcal{S}' \subseteq \mathcal{S}$ such that each element e in U is covered at least r_e times by the sets in $\mathcal{S}'$.*

In the WEIGHTEDMULSET *problem $(U, \mathcal{S}, R, C)$, each set $S \in \mathcal{S}$ has a cost, given by the function $C : \mathcal{S} \to \mathbb{R}$. The objective is to find a cover $\mathcal{S}'$ that minimizes the total cost, i.e., $\sum_{S\in\mathcal{S}'} C(S)$.*

Now, we consider the differentially private version of this problem, denoted PRIVATEMULSET. To match the edge-DP model described earlier, we define neighboring instances of the Multi-set Multi-cover problem as follows. Two instances $(U, \mathcal{S}, R)$ and $(U, \mathcal{S}', R')$ are said to be *neighbors* if one of the following conditions holds:

- There exists an element $e \in U$ such that $|r_e - r'_e| = 1$, and all other coverage requirements and sets are identical. That is, $\mathcal{S} = \mathcal{S}'$ and $R \triangle R' = \{r_e, r'_e\}$ for some $e \in U$.
- There exists an element $e \in U$ and an index $i \in [m]$ such that the multi-sets S_i and S'_i differ only in the multiplicity of e: $|m(S_i, e) - m(S'_i, e)| = 1$. All other sets and coverage requirements remain unchanged, i.e., $\mathcal{S} \triangle \mathcal{S}' = \{S_i, S'_i\}$ and $R = R'$.

Reducing a graph's degree-based objective – such as $\max_v d(v,G)$ or $\max_v d(v,G) \cdot d_2(v,G)$ – below a target threshold D can be naturally formulated as an instance of the MULSET problem. Specifically, we define the universe as $U = V(G)$ and associate each vertex $u \in V(G)$ with a multi-set S_u containing u and its neighbors. The covering requirements R are then defined to reflect how much the degree-related quantity, such as $r_v = \max(d(v,G) - D, 0)$ or $r_v = \max(d(v,G) \cdot d_2(v,G) - D, 0)$, must be reduced at each vertex. These reductions are described formally in the corresponding sections.

4 PRIVATEMULSET and PRIVATEMAXDEG Problems

We now describe private algorithms for reducing degree-based graph properties under the edge-DP model. These problems are reduced to instances of the PRIVATEMULSET framework introduced earlier. The intuition is the following: for example, in the MAXDEGREE problem, the utility of removing a node v should naturally depend on how much its degree exceeds the threshold D, i.e., $\max(d(v,G) - D, 0)$. This translates naturally into the MULSET framework, where each element (e.g., an edge or neighborhood constraint) has a coverage requirement, and sets (vertices) contribute to meeting them. More generally, any problem where elements contribute toward satisfying some threshold-based constraints can be reduced to an instance of MULSET. We apply the same reduction principle to the SPECTRALRADIUS problem as well. We present some of the main ideas and results here, while deferring all formal details to the appendix.

4.1 Multi-set Multi-cover Problem: Algorithm and Analysis

In this section we discuss the **Unweighted** case. The algorithm and analysis of the **Weighted** case are similarly constructed, and are discussed in Appendix C.1.2. Our differentially private algorithm for the PRIVATEMULSET problem is inspired by [19]. The idea is that we assign a utility score to each set based on how much it contributes toward unmet coverage requirements, and let the algorithm repeatedly samples a set based on its current utility. Specifically, for a set $S_i \in \mathcal{S}$ and element $e \in U$, the marginal utility is $A(S_i, e) := \min(m(S_i, e), r_e)$, and total utility is $A(S_i) = \sum_{e \in S_i} A(S_i, e)$.

By iteratively sampling out a set until no sets are left, the algorithm outputs an *implicit* solution — a permutation $\pi \in \sigma(\mathcal{S})$ over the sets – rather than an explicit cover. The permutation defines a valid solution: for each element $e \in U$, we select the first sets in π that together satisfy r_e. Formally, let $\pi_e := \{\pi(i) \mid 1 \le i \le n : \min(\sum_{j=1}^{i} m(S_{\pi(j)}, e), r_e) - \min(\sum_{j=1}^{i-1} m(S_{\pi(j)}, e), r_e) > 0\}$ be the indices of sets that contribute to covering e according to π. Then $\{S_j : j \in \bigcup_{e \in U} \pi_e\}$ forms a valid multi-cover of U. The algorithm and its analysis are provided in Appendix C.1.1 (Algorithm 4).

Lemma 4.1. *Algorithm 4 is (ϵ, δ)-differentially private, runs in time $\tilde{O}(qf|\mathcal{S}|)$ where q is the maximum set size and f is the maximum frequency of any element (ignoring multiplicity), and outputs a solution of cost at most $O((\ln m)/\epsilon' + \ln q) \cdot |OPT|$ with probability at least $1 - 1/m$, where $|OPT|$ denotes the cost of an optimal non-private solution.*

4.2 The Private MaxDegree (PRIVATEMAXDEG) problem

We now reduce PRIVATEMAXDEG to an instance of PRIVATEMULSET, then apply the private multi-set algorithm as shown in Algorithm 1 in Step 8. The edge-privacy model of PRIVATEMAXDEG is equivalent to the privacy model ofPRIVATEMULSET under the transformation in Algorithm 1. Specifically, if $G \sim G'$ differ by a single edge (u, v), then the corresponding PRIVATEMULSET instances are at most 4-step neighbors: the covering requirements for u and v change by at most 1, i.e., $|r_u - r'_u| \le 1$ and $|r_v - r'_v| \le 1$, and the multiplicities $m(S_u, v)$ and $m(S_v, u)$ change by at most 1 as well.

Algorithm 1 Private algorithm for PRIVATEMAXDEG

1: **Input:** (ϵ, δ), graph G, target degree D
2: Initialize set system $\mathcal{S} \leftarrow \emptyset$, requirements $R \leftarrow \emptyset$
3: **for** each $v \in V$ **do**
4:     Define multiset S_v with $m(S_v, v) = \infty$ and $m(S_v, u) = 1$ for all $u \sim v$
5:     $\mathcal{S} \leftarrow \mathcal{S} \cup \{S_v\}, \quad R \leftarrow R \cup \{r_v = \max(\deg(v) - D, 0)\}$
6: **end for**
7: Set $\epsilon' \leftarrow \epsilon/4$, $\delta' \leftarrow \delta/4e^{3\epsilon'}$
8: **Return:** Algorithm 4$(\epsilon', \delta', \mathcal{S}, R)$ /*Applying the private multi-set algorithm*/

Utility analysis. Since we reduce the PRIVATEMAXDEG problem to an instance of the PRIVATEMULSET and apply Algorithm 4 to solve it, the utility of Algorithm 1 (stated by Theorem 4.2) follows the utility of Algorithm 4, by setting $m = |V|, q = 2GS_{\text{MAXDEG}} < 2|V|$ in Lemma 4.1, where GS_{MAXDEG} is the global sensitivity of MAXDEG. The analysis for the weighted PRIVATEMAXDEG problem follows identically to the unweighted case, using Algorithm 5 for the weighted version of PRIVATEMULSET discussed earlier. Consequently, all arguments and results discussed previously are applicable with minimal modifications required for the utility bounds.

Theorem 4.2. *Let $\hat{B}$ be the cost of the output of Algorithm 1. W.h.p., $\hat{B} < |OPT_{\text{MAXDEG}}| \cdot O((1 + 1/\epsilon') \ln |V|)$.*

Explicit Solution. We also provide an explicit solution of which nodes to remove, incurring an additional privacy cost of $4\epsilon_1$. Unlike the implicit solution, the explicit output may allow some remaining nodes whose degrees exceed the the target degree threshold D. This approach builds on the permutation π produced by the implicit algorithm: we apply the AboveThreshold mechanism to π to find an index k such that removing the nodes $\pi_1, \pi_2, \ldots, \pi_k$ reduces the maximum degree to at most $D + O(\log m/\epsilon)$. The resulting solution removes at most $O(|\text{OPT}| \cdot \log k)$ nodes, where |OPT| is defined as above.

Utility and Runtime Analysis. Theorem 4.3 states the utility of the explicit solution output by Algorithm 2. We first observe that if we stop the algorithm at some iteration $\hat{k}$ where the selected node (and its equivalent set) no longer improves the coverage requirement by an amount $T = 6 \log n/\epsilon'$, then the maximum degree of the remaining graph is off from the target D by an amount at most $O(\log n)$. Steps $2 - 6$ of the algorithm follows the AboveThreshold technique to select the first index

Algorithm 2 Explicit solution algorithm for PRIVATEMAXDEG

1: **Input:** Instance of PRIVATEMAXDEG and a permutation π obtained from the exponential mechanism
2: $T' \leftarrow 6 \ln n/\epsilon' - Lap(2/\epsilon_1)$
3: **for** $i = 1$ to n **do**
4: $\quad \gamma_i \leftarrow L_i - Lap(4/\epsilon_1)$
5: **end for**
6: Let k be the first index such that $\gamma_k \leq T'$
7: **Output:** $\{\pi_1, \ldots, \pi_k\}$

k that approximately satisfies the covering requirement of $\hat{k}$, i.e., the noisy utility γ_k of the set chosen at step k is relatively small enough (less than the noisy threshold T' of the true target threshold T). We then utilize the accuracy guarantee of the AboveThreshold routine to argue that the selected index k is in fact not too far away from the "true" stopping iteration $\hat{k}$ where its utility truly falls below the threshold T. Finally, as an immediate corollary of Lemma 4.1, the runtimes of Algorithm 1 and 2 are stated in Theorem 4.4.

Theorem 4.3. *The output k of from Algorithm 2 satisfies $\Delta(G - \cup_{i=1}^{k}\{\pi_i\}) \leq D + O(\log n/\epsilon')$ with high probability. In addition, $k = O(OPT \cdot \log n/\epsilon')$ with high probability.*

Theorem 4.4. *Algorithms 1 and 2 each run in time $O(n\Delta^2)$, where Δ is the maximum degree of the input graph.*

Lower bounds. The explicit solutions cannot guarantee the coverage for the PRIVATEMULSET under DP guarantee. In this section, we argue that any explicit solution of the PRIVATEMULSET containing no more than $|OPT| + O(\log n)$ sets can only guarantees some partial covering with the additive error at least $\Omega(\log n)$. For the PRIVATEMAXDEG, the following lemma states the additive error of the target maximum degree, similar to the lower bounds of the PRIVATEMULSET, which we present in Appendix C.3.

Lemma 4.5. Lower bound of PRIVATEMAXDEG. *Any explicit (ϵ, δ)-differentially private algorithm for the* PRIVATEMAXDEGREE *removing at most $O(\log n) + |OPT|$ nodes with probability at least $1 - C, C = n^{-\Omega(1)}$, must incur an additive error $\Delta(G - \cup_{i=1}^{k}\{\pi_i\}) = D + \tilde{\Omega}(\log n)$, where $\pi_1, \ldots, \pi_k$ are the removed nodes.*

5 Private SPECTRALRADIUS

In this section, we introduce two algorithms designed to reduce the spectral radius, $\rho(G)$, of a given graph $G = (V, E)$. In particular, these algorithms minimize specific graph metrics that upper bound $\rho(G)$. The proofs of the results in this section are deferred to Appendix D.

5.1 Bound via PARTIALSETCOVER

The idea for our first approach is based on reducing the number of walks of length four, denoted by $|W_4(G)|$, where $W_4(G)$ is the set of all such walks. Reducing $|W_4(G)|$ below nT^4 implies a bound on the spectral radius $\rho(G) \leq O(n^{1/4}T)$. Setting $T = \Delta^{1/2}$ thus achieves a bound of $O(n^{1/4}\Delta^{1/2})$, which is significantly improves over the bound $\rho(G) \leq \Delta$ when $\Delta = \Omega(\sqrt{n})$.

We employ the GREEDYWALK node selection algorithm from [44], which follows a greedy strategy to reduce the number of paths of a specified length. This algorithm, combined with the exponential mechanism, forms the first part of our approach. Further, we reduce our problem to an instance of the Partial Set Cover problem: each vertex in the graph corresponds to a set, and removing a vertex "hits" (or covers) a collection of walks that include it. Specifically, the utility of removing a vertex v is defined as the number of walks of length 4 that pass through v, formally given by: $A(v) = |\{w \in W_4(G) : v \in w\}|$. A differentially private algorithm for Partial Set Cover problem was introduced in [30], using an approach similar to Algorithm 3. However, it is important to note that in this context, the sensitivity of $|W_4(G)|$ is Δ^2.

Lemma 5.1. *If $T^4 \geq 6 \ln n/\epsilon'$, the output $V' = \{\pi_1, \ldots, \pi_k\}$ of Algorithm 3 satisfies $W_4(G[V \setminus V']) \leq nT^4 + O(\log n/\epsilon')$ and gives an $O(\log n)$ approximation with high probability; the algorithm*

Algorithm 3 Private Hitting Walks Algorithm for PRIVMINSR

Input: Graph $G = (V, E)$, privacy parameters (ϵ, δ)
Set $\epsilon' \leftarrow \epsilon/(2\ln(e/\delta))$, initialize permutation $\pi \leftarrow \emptyset$
for $i = 1$ to n **do**
  Sample $v \in V$ with prob. $\propto \exp(\epsilon' \cdot A(v))$, append v to π and remove v from V
end for
Set $T \leftarrow \Delta^{1/2}$, $\theta \leftarrow 4nT^4$, $\hat{\theta} \leftarrow \theta - Lap(2/\epsilon_1)$
for $i = 1$ to n **do**
  $\gamma_i \leftarrow W_4(G[V - \{\pi(1), \dots, \pi(i)\}]) - Lap(4/\epsilon_1)$
end for
Let k be the first iteration such that $\gamma_k \leq \hat{\theta}$
Output: $(\pi(1), \dots, \pi(k))$

is $(\Delta^2(\epsilon+\epsilon_1), \Delta^2\delta e^{(\Delta^2-1)\epsilon})$-differentially private and runs in time $\tilde{O}(n\Delta^4\omega^4)$, where ω is the matrix multiplication exponent for $n \times n$ matrices.

5.2 Bound via PRIVATEMULSET

We can also apply our private Algorithm 4 for PRIVATEMULSET to indirectly reduce the spectral radius, $\rho(G)$, of a graph $G = (V, E)$. According to [17], the spectral radius is bounded by $\rho(G) \leq \max_{u \in V(G)} \sqrt{\sum_{v \sim u} d(v, G)}$, which is always better than the trivial bound $\rho(G) \leq \Delta$. This improvement is especially significant in degree-disassortative graphs, where high-degree vertices are typically adjacent to many low-degree vertices.

We approach the problem of reducing $\max_{u \in V(G)} \sqrt{\sum_{v \sim u} d(v, G)}$ using a similar strategy as in the PRIVATEMAXDEG case – by reformulating it as an instance of PRIVATEMULSET. First, we define sets $\{S_u\}_{u \in V}$, so that $m(S_u, u) = \infty$ and $m(S_u, v) = d(u, G)$ for all vertices v adjacent to u. Additionally, for each vertex $u \in V$, set $r_u = \max(0, \sum_{v \sim u} d(v, G) - D)$, where $\sqrt{D}$ is a target upper bound. We can then apply the same analysis used in the PRIVATEMAXDEG case. However, we must adjust our edge-privacy model for this scenario. In the worst case, adding an edge (u, v) could cause neighboring graphs in the PRIVATEMULSET formulation to become 4Δ-neighbors. This happens because such an edge addition can increase both the covering requirements r_u and r_v as well as the multiplicities $m(S_u, v)$ and $m(S_v, u)$ by up to Δ. Thus, the algorithmic approach and results from PRIVATEMAXDEG largely carry over, but the sensitivity needs to be adjusted from 4 to 4Δ. The details can be found in Appendix D.2.

6 Experimental evaluation

We evaluate the performance of our algorithms on different realistic and random networks in terms of the following questions
- Effects of privacy budgets on the utility of our algorithm (both in terms of vaccination budget and epidemic metrics $\Delta(G)$ and $\rho(G)$).
- Tradeoff between vaccination cost, different epidemic metrics, and privacy parameters.
- Comparison between the implicit and explicit solutions.

Graph Name	**#nodes**	**#edges**
Subgraph of digital twin of contact network for Montgomery, VA [16]	10,000	83842, 84025, 84549
BTER [28] with Power Law Degree ($\gamma = 0.5, \rho = 0.95, \eta = 0.05$)	1000	31530, 31582, 31621

Table 1: Network datasets used in evaluation

Datasets and setup. We consider two classes of networks, as summarized in Table 1. The digital twin of a contact network [2, 16] is a model of real world activity based contact networks; we

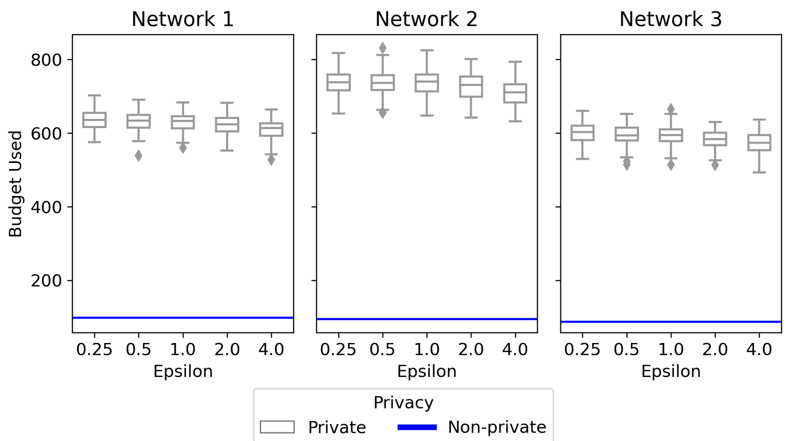

Figure 1: Effect of Privacy on Budget Requirements on Montgomery County Subnets

Table 2: Comparison of Average Performance of Implicit vs Explicit Solutions ($\epsilon = 4.0$)

γ	ρ	η	EXPLICIT?	BUDGET	MAX DEGREE	SPECTRAL RADIUS
0.3	0.95	0.05	YES	83.89	92.78	72.28
			NO	506.62	20	18.35
0.5	0.95	0.05	YES	66.19	92.80	77.99
			NO	430.36	20	18.55

consider three subgraphs with 10,000 nodes of the network for Montgomery county VA. The BTER model [28] is a random graph model, which preserves both degree sequence and clustering; we consider three randomly generated networks. Both classes of networks have been used in a number of epidemiological analyses, e.g., [32, 8, 1].

Effect of privacy on solution cost for the PRIVATEMAXDEG problem. Figure 1 shows the cost of the implicit solutions computed using Algorithm 1 for the three subgraphs of the Montgomery county networks (labeled as Network 1-3). We use a privacy budget of $\delta = 10^{-6}$ and $\epsilon \in \{0.25, 0.5, 1, 2, 4\}$, and set a target degree of $D = 45$. For each ϵ, we show a distribution over results computed by multiple runs of the algorithm. As described in the implicit Algorithm 4 for PRIVATEMULSET, the implicit solution is computed and plotted here. The cost of the solution to a non-private greedy algorithm for the multi-set multi-cover problem (which has a H_Δ-approximation[42], where H_n denotes the n-th harmonic number) is shown as the **baseline**. We note that the solution of Algorithm 4 is within a factor of about 10 of the non-private baseline, which could be viewed as being consistent with Theorem 4.2; further, the cost of the private solutions has a slight reduction with ϵ.

Figure 2 shows the impact of privacy cost on the cost of the explicit solution for PRIVATEMAXDEG for the three BTER networks (Table 1) computed by Algorithm 2 with a target $D = 20$. We pick $\delta = 1/n = 10^{-3}$ here, and have relaxed the privacy to the multi-set multi-cover definition rather than the edge private definition of neighboring datasets. The results show, somewhat counter-intuitively, that the solution cost actually increases with ϵ. Since in explicit solutions the solution cost is mainly determined by Above Threshold step (in Algorithm 2), which allows lower ϵ to halt set selection earlier (before certain vertices meet their cover requirements), the algorithm is closer to fulfilling the entire covering requirement as in the non-private version as ϵ increases, which explains this behavior. This is also consistent with Figure 3, which shows that the resulting violation in the maximum degree from the target decreases significantly with ϵ. This suggests that the choice of the privacy budget needs to be done carefully.

Implicit vs Explicit solutions. We investigate the performance of the implicit and explicit solutions (Table 2). The main difference between the two methods lies in when the permutations terminate, explicit would halt before the target degree is fully satisfied whereas implicit would not. This is demonstrated in that implicit solutions perform much better with metrics like max degree whereas explicit solutions have significantly lower vaccination costs.

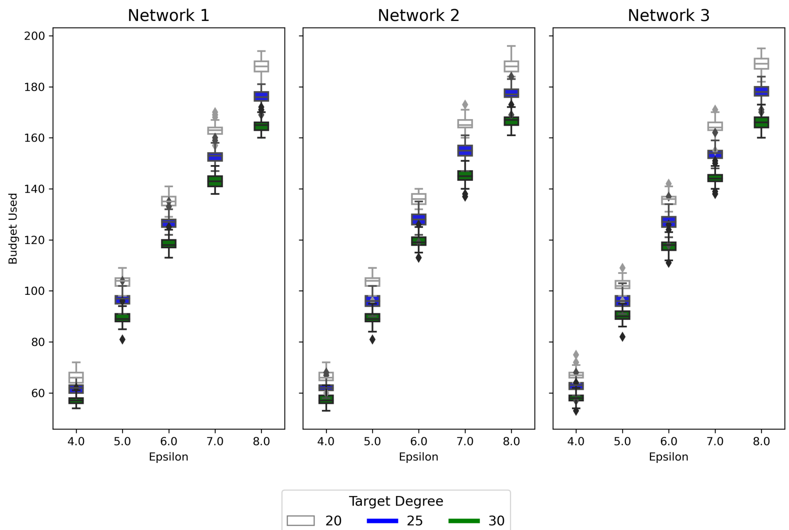

Figure 2: Effect of Privacy on Budget Requirements on BTER Graphs

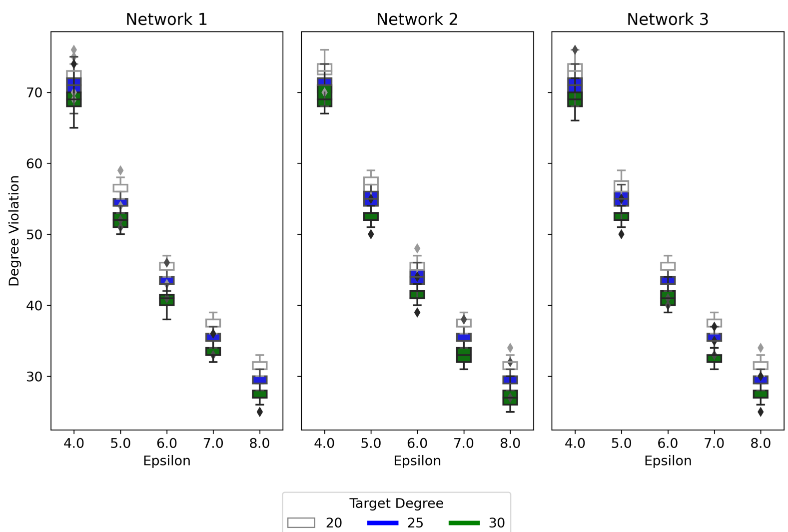

Figure 3: Effect of Privacy on Max Degree Violation on BTER Graphs

7 Conclusion

We initiate the study of the challenging and largely unexplored problems of epidemic control on networks under differential privacy. Our focus is on the approach of removing nodes from a graph to optimize certain properties, such as the maximum degree and spectral radius of the residual graph, which models the vaccination effect on a contact network. We design the first set of algorithms along with rigorous utility analyses for minimizing the maximum degree and spectral radius under the edge differential privacy model. One of our main techniques involves transforming these problems into a multi-set multi-cover problem and using its private solution to determine the sets of nodes to be removed (or vaccinated). While providing explicit solutions for covering-type problems is challenging, we employ the sparse vector technique to relax the covering requirement, allowing for approximate explicit solutions that can be used to design vaccination strategies. The experimental results of our algorithms, evaluated on multiple realistic and random networks, demonstrate good privacy-utility trade-offs.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our abstract claims a (ϵ, δ) differential private algorithm for minimize maximum degree and spectral radius, which is the result of this paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We recognize that the problem of minimizing the number of infections is NP-hard and hence used the heuristic of minimizing max degree and spectral radius, which is shown to be useful in decreasing infections. In addition, we present the setups of our experiments and the limitations that they are run with.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: We have clearly stated the requirements of theorems and provided proofs for each.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The steps taken to produce the experiments are detailed in the algorithms of the paper and the datasets are clearly described with the appropriate setups.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
    (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
    (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
    (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
    (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The instructions to reproduce the experiments are detailed in the algorithms. All datasets are publicly available with their sources cited.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All the parameters to set up the experiments are listed in the respective sections.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The experimental results and analyses include the error bars of the distribution of results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: We presented the computational complexity of our algorithms, which was quite reasonable for the networks we considered in our experiments. Since the experiments can be performed by typical workstations without specialized hardware and in relatively short amount of time, we have decided not to include such information.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: There is no violation of the Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: This paper is aimed to increase the effectiveness of vaccination strategies under the privacy-preserving of sensitive data. There may exist other societal impacts that are out of the scope of this work. Currently, we find no negative societal impacts of this work.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA] .

Justification: No additional data nor trained models are generated.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All assets are properly cited.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: No new assets produced.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA] .

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA] .

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA] .

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

Supplementary Material

A Related Work

We include additional discussion on related work here.

[7] studied the problem of estimating the reproductive number R_0 of an epidemic on its contact network in the SIS and SIR models. The reproductive number R_0 is closely related to the spectral radius, i.e., the reproductive number R_0 can be expressed as a function of the first eigenvalue of the adjacency matrix. Their privacy model protected the "weights" of the weighted contact network. Moreover, the work did not specify or imply any approach to modify the contact network to reduce such quantity, in order to reduce the spread of the pandemic.

There has been several work to calculate the spectral radius of an input graph–that is of independent interest from the perspective of epidemic control. [47] computed the eigenvalues and eigenvectors of an input graph under the edge-differential privacy. [6], also under the egde-differential privacy model, estimated the second smallest eigenvalues (λ_2), which is also commonly refered to as "algebraic connectivity". Similarly, [20] studied the same problem, but also considered the problem under the node-differential privacy model, in which two neighbor graphs differ by a node and its adjacent edges. None of the work suggested a method to reduce the spectral radius of the input network.

B Background

We briefly discuss the basic ideas of DP here; see [13] for more details.

Definition B.1 (Exponential mechanism). *Given a utility function $u : \mathcal{X}^n \times \mathcal{R} \to \mathbb{R}$, let $\mathrm{GS}_u = \max_{r\in\mathcal{R}} \max_{x\sim x'} |u(x,r) - u(x',r)|$ be the global sensitivity of u. The exponential mechanism $M(x,u,\mathcal{R})$ outputs an element $r \in \mathcal{R}$ with probability $\propto \exp(\frac{\epsilon u(x,r)}{2\,\mathrm{GS}_u})$.*

Lemma B.2. *The exponential mechanism is ϵ-differentially private. Furthermore, for a fixed dataset $x \in \mathcal{X}^n$, let $OPT = \max_{r\in\mathcal{R}} u(x,r)$, then the exponential mechanism satisfies*

$$\Pr[u(x, M(x,u,\mathcal{R}) \le OPT - \frac{2\,\mathrm{GS}_u}{\epsilon}(\ln|\mathcal{R}| + t)] \le e^{-t}.$$

Definition B.3 (Laplace mechanism). *Let $f : \mathcal{X}^n \to \mathbb{R}^d$ be a function with global ℓ_1-sensitivity $\Delta_f = \max_{x\sim x'} |f(x) - f(x')|_1$. The Laplace mechanism releases*

$$M(x) = f(x) + (Z_1, \ldots, Z_d),$$

where $Z_i \sim \mathrm{Lap}(\Delta_f/\epsilon)$ are independent random variables drawn from the Laplace distribution with scale parameter Δ_f/ϵ.

Lemma B.4. *The Laplace mechanism is ϵ-differentially private. Moreover, each coordinate of the output is concentrated around the true value of $f(x)$, with noise magnitude proportional to Δ/ϵ.*

Definition B.5 (AboveThreshold). *Let $f_1, \ldots : \mathcal{X}^n \to \mathbb{R}$ be a sequence of queries with sensitivity 1. Given a threshold τ and a privacy parameter ϵ, the AboveThreshold mechanism adds Laplace noise $\mathrm{Lap}(2/\epsilon)$ and $\mathrm{Lap}(4/\epsilon)$ to the threshold and to each query, and returns the first index i such that the noisy query exceeds the noisy threshold.*

Lemma B.6. *The AboveThreshold mechanism is $(\epsilon, 0)$-differentially private.*

C PrivateMulSet and PrivateMaxDeg Problems

C.1 PrivateMulSet

C.1.1 Unweighted case.

We present an algorithm for the PrivateMulSet problem, building upon the framework and analysis from [19]. First, we define a utility function $A : \mathcal{S} \to \mathbb{R}_{\ge 0}$. For a set $S_i \in \mathcal{S}$ and an element $e \in U$, the marginal utility is defined as $A(S_i, e) := \min(m(S_i, e), r_e)$. The total utility of a set S_i is then given by $A(S_i) = \sum_{e\in S_i} A(S_i, e)$.

It is important to note that directly outputting an explicit solution – i.e., listing only the sets that form a valid cover—would violate differential privacy. In particular, this is because the solution to the vertex cover problem, which is a special case of the set cover problem, is known to retain privacy iff the output contains at least $|V| - 1$ vertices[19]. Therefore, in order to preserve privacy, the algorithm must produce an *implicit* solution, typically in the form of a permutation $\pi \in \sigma(\mathcal{S})$ over the sets in $\mathcal{S}$. Intuitively, π should have the sets arranged in the order of decreasing utility. This ordering implicitly defines a cover: for each element $e \in U$, we select the first few sets in π that would fully cover e. Formally, let $\pi_e := \{\pi(i) \mid 1 \le i \le n : \min(\sum_{j=1}^{i} m(S_{\pi(j)}, e), r_e) - \min(\sum_{j=1}^{i-1} m(S_{\pi(j)}, e), r_e) > 0\}$ be the indices of sets that contribute to covering e according to π. Then $\{S_j : j \in \bigcup_{e \in U} \pi_e\}$ forms a valid multi-cover of U.

Algorithm 4 Private algorithm for PRIVATEMULSET

```
Input: privacy parameters (ε, δ), set system S, covering requirement R
Set ε' ← ε / (2 ln(e/δ))
Initialize empty permutation π ← ∅
Initialize r_e^(0) ← r_e for all e ∈ U, S^(1) ← S
for i = 1 to |S| do
    Define A^(i)(S_j) := Σ_{e∈S_j} min(m(S_j, e), r_e^(i−1))
    Sample S_j ∈ S^(i) with probability ∝ exp(ε' A^(i)(S_j))
    Append j to π: π(i) ← j
    Update available set system: S^(i+1) ← S^(i) \ {S_j}
    for e ∈ S_j do
        Update covering requirement: r_e^(i) ← max(0, r_e^(i−1) − m(S_j, e))
    end for
end for
Output permutation π
```

Lemma C.1. *The output of the Algorithm 4 is at most $O(\ln m/\epsilon' + \ln q)OPT$ with probability at least $1 - 1/m$, where $q = \max_S \sum_e A(S, e)$ is the size of the largest set, and OPT denotes the cost of an optimal non-private solution.*

Proof. Without loss of generality, we may assume that the permutation π output by the Algorithm 4 $\pi = (1, 2, \ldots, m)$. In other words, the sets $S_1, S_2, \ldots, S_m$ are sequentially added to the cover in that exact order.

Let $L_i = \max_{j \ge i} A_i(S_j)$ be the maximum utility possible at step i (this implies that there is a multi-set of that utility). Then the probability of selecting a set of utility $< L_i - 3\frac{\ln m}{\epsilon'}$ (of which there are at most m) is less than $\frac{m \cdot exp(\epsilon' L_i - 3\ln m)}{exp(\epsilon' L_i) + m \cdot exp(\epsilon' L_i - 3\ln m)} = \frac{1/m^2}{1 + 1/m^2} \le 1/m^2$. Next, consider two cases:

$\mathbf{L_i > 6\frac{\ln m}{\epsilon'}}$. The probability that every multi-set selected has utility at least $L_i - 3\frac{\ln m}{\epsilon'} > L_i/2$ is $\ge (1 - 1/m^2)^m \ge (1 - 1/m)$. Because the greedy approximation is a $O(\ln q)$ approximation, Algorithm 4 can cover this region in at most $O(OPT \ln q)$ multi-sets with high probability.

$\mathbf{L_i \le 6\frac{\ln m}{\epsilon'}}$. At this point there are at most $OPT \cdot L_i$ elements that require covering, and $OPT \cdot L_i \le OPT \cdot O(\frac{\ln m}{\epsilon'})$. Since the post-processing of the implicit solution selects only sets that cover at least one element, covering the remaining $O(OPT\frac{\ln m}{\epsilon'})$ elements takes an additional $O(OPT \ln m/\epsilon')$ sets.

Therefore, the algorithm uses at most $O(OPT(\ln m/\epsilon' + \ln q)$ sets. □

Lemma C.2. *Algorithm 4 is (ϵ, δ)-DP.*

Proof. First, we consider neighboring problems $\mathcal{A} = (U, \mathcal{S}, R)$, $\mathcal{A}' = (U, \mathcal{S}', R)$ that share the same coverage requirements $R = R' = \{r_e : e \in U\}$ but differ in the multiplicity of a particular element

e_0 in one set, such that $|m(S_k, e_0) - m(S'_k, e_0)| = 1$ for some $k \in [m]$. Define t as the epoch when element e_0 is fully covered in both instances $\mathcal{A}$, $\mathcal{A}'$. Let $A_i(S)$ and $A_i(S')$ (which were written as $A^{(i)}(\cdot)$ in Algorithm 4) denote the remaining aggregate coverage requirement of set $S \in \mathcal{S}$ and $S' \in \mathcal{S}'$ after the first $i-1$ sets in π have been added to the cover in instances $\mathcal{A}$ and $\mathcal{A}'$, respectively.

We wish to establish a bound for $\frac{Pr[M(\mathcal{A})=\pi]}{Pr[M(\mathcal{A}')=\pi]}$. Expressing it explicitly, we derive the following:

$$\begin{aligned}\frac{\Pr[M(\mathcal{A})=\pi]}{\Pr[M(\mathcal{A}')=\pi]} &= \prod_{i=1}^{n}\left(\frac{e^{\epsilon' A_i(S_i)}}{\sum_{j=i}^{n} e^{\epsilon' A_i(S_j)}}\right) / \left(\frac{e^{\epsilon' A_i(S'_i)}}{\sum_{j=i}^{n} e^{\epsilon' A_i(S'_j)}}\right)\\ &= \frac{e^{\epsilon'(\sum_{i=1}^{n} A_i(S_i))}}{e^{\epsilon'(\sum_{i=1}^{n} A_i(S'_i))}} \cdot \prod_{i=1}^{n} \frac{\sum_{j\geq i} e^{\epsilon' A_i(S'_j)}}{\sum_{j\geq i} e^{\epsilon' A_i(S_j)}}\\ &= \prod_{i=1}^{t} \frac{\sum_{j\geq i} e^{\epsilon' A_i(S'_j)}}{\sum_{j\geq i} e^{\epsilon' A_i(S_j)}},\end{aligned}$$

where the last equality holds because if $i > t$, then the element e_0 was fully covered by iteration i, implying that $A_i(S_j) = A_i(S'_j)$ for all $j \geq i$. Also, given that all elements are eventually covered, we have $\sum_{i=1}^{n} A_i(S_i) = \sum_{i=1}^{n} A_i(S'_i)$.

Assuming $k \leq t$, we break up the product $\prod_{i=1}^{t} \frac{\sum_{j\geq i} e^{\epsilon' A_i(S'_j)}}{\sum_{j\geq i} e^{\epsilon' A_i(S_j)}}$ into three terms I_1, I_2, I_3 as follows:

$$\begin{aligned}I_1 \cdot I_2 \cdot I_3 := \left(\prod_{i=1}^{k-1} \frac{\sum_{j\geq i} e^{\epsilon' A_i(S'_j)}}{\sum_{j\geq i} e^{\epsilon' A_i(S_j)}}\right) \cdot \left(\frac{\sum_{j\geq k} e^{\epsilon' A_t(S'_j)}}{\sum_{j\geq k} e^{\epsilon' A_t(S_j)}}\right) \cdot\\ \left(\prod_{i=k+1}^{t} \frac{\sum_{j\geq i} e^{\epsilon' A_i(S'_j)}}{\sum_{j\geq i} e^{\epsilon' A_i(S_j)}}\right).\end{aligned}$$

In the case when $t < k$ (or $t \leq k$) the terms I_2 and I_3 (or just I_3) vanish, and k is replaced by t. However, the argument by enlarge would remain unaffected by this adjustment.

We proceed by considering two possible cases: $m(S_k, e_0) > m(S'_k, e_0)$ and $m(S_k, e_0) < m(S'_k, e_0)$.

$\mathbf{m(S_k, e_0) > m(S'_k, e_0)}$. In this scenario, both I_1 and I_2 are less than or equal to 1. Therefore, it is sufficient to focus on upper bounding I_3. Define an index-set $S^{I,i} := \{j : A_i(S_j) \neq A_i(S'_j)\}$. We then find that

$$\begin{aligned}I_3 &= \prod_{i=k+1}^{t} \frac{\sum_{j\geq i} e^{\epsilon' A_i(S'_j)}}{\sum_{j\geq i} e^{\epsilon' A_i(S_j)}}\\ &= \prod_{i=k+1}^{t} \frac{(e^{\epsilon'}-1)\sum_{j\in S^{I,i}} e^{\epsilon' A_i(S_j)} + \sum_{j\geq i} e^{\epsilon' A_i(S_j)}}{\sum_{j\geq i} e^{\epsilon' A_i(S_j)}}\\ &= \prod_{i=k+1}^{t} \left(1 + (e^{\epsilon'}-1)\cdot \Pr[\pi_i \in S^{I,i}]\right)\\ &= \prod_{i=k+1}^{t} (1 + (e^{\epsilon'}-1)\cdot \Pr[S^{I,i}]).\end{aligned}$$

Observe that to sample a set from $S^{I,i}$ means to fully cover e_0, which can only occur at step t. Recall the following lemma from [19]

Lemma C.3. *The probabilistic process is modeled by flipping a coin over t rounds. p_i is the probability that it would come up heads in round i, p_i can be chosen adversarially based on* *the previous $i-1$ rounds. Let Zi be the indicator for the event that no coin comes up heads in the first i steps. Let $Y = \sum_{i=1}^{t} p_i Z_i$. Then for any q, $\Pr[Y > q] \le \exp(-q)$.*

In our setup, Z_i corresponds to the indicator of the event "e_0 is fully covered at round i". If $\sum_{i=k+1}^{t-1} \Pr[S^{I,i}] Z_i \le \ln \delta^{-1}$, then we obtain

$$\begin{aligned}
\frac{\Pr[M(\mathcal{A}) = \pi]}{\Pr[M(\mathcal{A}') = \pi]} &\le I_3 \\
&\le \prod_{i=k+1}^{t} \exp((e^{\epsilon'} - 1) \Pr[S^{I,i}]) \\
&\le \exp(2\epsilon' \sum_{i=k+1}^{t} \Pr[S^{I,i}]) \\
&\le \exp(2\epsilon'(\ln(1/\delta) + \Pr[S^{I,t}])) \\
&\le \exp(2\epsilon'(\ln(1/\delta) + 1)).
\end{aligned}$$

Now, by Lemma C.3, the probability of the event $\sum_{i=k+1}^{t-1} \Pr[S^{I,i} Z_i] > \ln \delta^{-1}$ is upper bounded by δ. Consequently, if $\mathcal{P}$ denotes the set of outcomes, we conclude that

$$\Pr[M(\mathcal{A}) \in \mathcal{P}] \le \exp(\epsilon) \Pr[M(\mathcal{A}') \in \mathcal{P}] + \delta.$$

We refer to [19] for a detailed proof.

$\mathbf{m(S_k, e_0) < m(S'_k, e_0)}$. In this scenario, $I_3 \le 1$. Our focus then shifts to $I_1 \cdot I_2$, following an analogous argument to the one discussed above, we obtain

$$\begin{aligned}
I_1 \cdot I_2 &= \prod_{i=1}^{k} \frac{\sum_{j \ge i} e^{\epsilon' A_i(S'_j)}}{\sum_{j \ge i} e^{\epsilon' A_i(S_j)}} \\
&= \prod_{i=1}^{k} (1 + (e^{\epsilon'} - 1) \cdot \Pr[\pi_i \in S^{I,i}) \\
&= \prod_{i=1}^{k} 1 + (e^{\epsilon'} - 1) \cdot \Pr[\pi_i = S_k] \\
&\le \prod_{i=1}^{k} \exp((e^{\epsilon'} - 1) \Pr[\pi_i = S_k]) \\
&\le \exp(2\epsilon' \sum_{i=1}^{k} \Pr[\pi_i = S_k]),
\end{aligned}$$

where to justify the third equality, we observe that $S^{I,i} = \{k\}$. To obtain the inequalities, we apply the approximation $1 + x \le e^x \le 1 + 2x$ for sufficiently small positive values of x.

We then apply Lemma C.3 analogous to the above discussion, which completes the proof.

We now turn to the instance where the neighboring problems have differing covering constraints $R \ne R'$. As before, let t denote the epoch at which the covering constraint for e_0 is satisfied by both $M(\mathcal{A})$ and $M(\mathcal{A}')$. Although $\mathcal{S} = \mathcal{S}'$, we refer to the sets in $\mathcal{S}'$ as S' for for clarity.

$\mathbf{r_{e_0} > r'_{e_0}}$. This case is straightforward, since $\sum_{i=1}^{n} A_i(S_i) - \sum_{i=1}^{n} A_i(S'_i) = r_{e_0} - r'_{e_0} = 1$, and we obtain:

$$\frac{\Pr[M(\mathcal{A}) = \pi]}{\Pr[M(\mathcal{A}') = \pi]} = e^{\epsilon'} \prod_{i=1}^{t} \frac{\sum_{j \ge i} e^{\epsilon' A(S'_j)}}{\sum_{j \ge i} e^{\epsilon' A(S_j)}} \le \exp(\epsilon').$$

$\mathbf{r_{e_0}} < \mathbf{r'_{e_0}}$. In this case we have

$$\frac{\Pr[M(\mathcal{A})=\pi]}{\Pr[M(\mathcal{A}')=\pi]} = e^{-\epsilon'} \prod_{i=1}^{t} \frac{\sum_{j\geq i} e^{\epsilon' A(S_j')}}{\sum_{j\geq i} e^{\epsilon' A(S_j)}}$$
$$\leq e^{-\epsilon'} \prod_{i=1}^{t} \Big(1 + (e^{\epsilon'}-1)\cdot \Pr[\pi_i \in S^{I,i}]\Big)$$
$$= e^{-\epsilon'} \prod_{i=k+1}^{t} (1 + (e^{\epsilon'}-1)\cdot \Pr[S^{I,i}]).$$

The remainder of the proof utilizes the same arguments as previously discussed. □

Lemma C.4. *Algorithm 4 runs in $\tilde{O}(qf|\mathcal{S}|)$, where q is the maximum set size and f is the maximum frequency of any element (ignoring multiplicity).*

Proof. Initially, the algorithm computes $A^{(1)}(\cdot)$ for all sets in $\mathcal{S}$, which can be done in $O(|\mathcal{S}|q)$. Then the algorithm runs for $|\mathcal{S}|$ iterations, once per set, contributing the $|\mathcal{S}|$ factor. In each iteration, a set is sampled according to the exponential mechanism, where probabilities are proportional to $\exp(\varepsilon' A(S))$. Sampling can be done in $\tilde{O}(1)$.

After a set S_j is selected, the algorithm updates the covering requirements for each element $e \in S_j$, which affects the utilities $A(S')$ for all other sets S' containing e. Since each element appears in at most f sets, and each set contains at most q elements, the number of affected utilities per iteration is at most qf. □

C.1.2 Weighted case.

Here we briefly discuss the weighted version of PRIVATEMULSET, and adapt the methodology of [19] with some minor modifications. First, we may assume without loss of generality that $\min_{S\in\mathcal{S}} C(S) = 1$, and $W = \max_{S\in\mathcal{S}} C(S)$ with $n = |\mathcal{S}|$. Let $M = \sum_{e\in U} r_e$. Similar to the unweighted version, we define $A(S) = \sum_{e\in S} \min(r_e, m(S,e))$ for a set $S \in \mathcal{S}$, and we say that the *utility* $u(S)$ is defined to be equal to $A(S) - C(S)$. Additionally, we add a dummy set `halve` to $\mathcal{S}$ with utility $u(\texttt{halve}) = -T$ for $T = \Theta(\frac{\log n + \log\log(MW)}{\epsilon'})$. When `halve` is selected by Algorithm 5, it indicates that no set was actually chosen. Additionally, unlike other selections, `halve` is never removed from $\mathcal{S}$.

Lemma C.5. *The cost of the output of 5 is at most $O(T \log n \cdot \mathit{OPT})$ with probability at least $1 - 1/\mathit{poly}(n)$.*

Proof. This follows from a verbatim argument in [19] with n replaced by M and m replaced by n in our notation. □

Lemma C.6. *5 is (ϵ, δ)-differentially private.*

Proof. The proof is identical to the privacy proof of the algorithm in the unweighted case, with $A(S)$ replaced by $u(S)$. □

Identically to [19], we can remove the dependency on W to obtain an $O(\log M(\log n + \log\log M/\epsilon))$-approximation.

C.2 MAXDEGREE

Theorem 4.2. *Let $\hat{B}$ be the cost of the output of Algorithm 1. W.h.p., $\hat{B} < |OPT_{\textsc{MaxDeg}}| \cdot O((1 + 1/\epsilon')\ln|V|)$.*

Proof. Since PRIVATEMAXDEG reduces to PRIVATEMULSET, the optimal solutions for both problems are equivalent. In addition, since Algorithm 4 outputs a $O(\ln m/\epsilon' + \ln q)$-approximation, and $m = |V|$, $q = 2\,\mathrm{GS}(G) \leq 2|V|$, we have $O(\ln m/\epsilon' + \ln q) \leq O((1 + 1/\epsilon')\ln|V|)$. □

Algorithm 5 Private algorithm for WEIGHTEDPRIVATEMULSET

Input: (ϵ, δ), set system $\mathcal{S}$, covering requirement $R = \{r_e\}_{e \in U}$
$\epsilon' \leftarrow \frac{\epsilon}{2\ln(e/\delta)}$, initialize permutation $\pi \leftarrow \emptyset$
$\theta \leftarrow M, T = \Theta\left(\frac{\log n + \log\log(MW)}{\epsilon'}\right)$
$i \leftarrow 1, r_e^{(0)} \leftarrow r_e$ for all $e \in U$, $\mathcal{S}^{(1)} \leftarrow \mathcal{S}$
while $\theta \geq 1/W$ **do**
  Define $u^{(i)}(S) := \sum_{e \in S} \min(m(S,e), r_e^{(i-1)}) - C(S)/\theta$
  Sample $S \in \mathcal{S}^{(i)}$ with probability $\propto \exp(\epsilon' u^{(i)}(S))$
  if $S = \texttt{hal}$ **then**
    $\theta \leftarrow \theta/2$
    $\mathcal{S}^{(i+1)} \leftarrow \mathcal{S}^{(i)}$
    $r_e^{(i)} \leftarrow r_e^{(i-1)}$ for all $e \in U$
  else
    Append S to π
    $\mathcal{S}^{(i+1)} \leftarrow \mathcal{S}^{(i)} \setminus \{S\}$
    for $e \in S$ **do**
      $r_e^{(i)} \leftarrow \max(0, r_e^{(i-1)} - m(S,e))$
    end for
  end if
end while
Output π concatenated with a random permutation of $\mathcal{S}^{(i)} \setminus \{\texttt{hal}\}$

Algorithm 1 (restated). Private algorithm for PRIVATEMAXDEG

Input: (ϵ, δ), graph G, target degree D
Initialize set system $\mathcal{S} \leftarrow \emptyset$, requirements $R \leftarrow \emptyset$
for each $v \in V$ **do**
  Define multiset S_v with $m(S_v, v) = \infty$ and $m(S_v, u) = 1$ for all $u \sim v$
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{S_v\}, \quad R \leftarrow R \cup \{r_v = \max(\deg(v) - D, 0)\}$
end for
Set $\epsilon' \leftarrow \epsilon/4$, $\delta' \leftarrow \delta/4e^{3\epsilon'}$
Return: Algorithm $4(\epsilon', \delta', \mathcal{S}, R)$ /*Applying the private multi-set algorithm*/

C.2.1 Explicit solution for MAXDEGREE

Algorithm 2 (restated). Explicit solution algorithm for PRIVATEMAXDEG

Input: Instance of PRIVATEMAXDEG and a permutation π obtained from the exponential mechanism
$T' \leftarrow 6\ln n/\epsilon' - Lap(2/\epsilon_1)$
for $i = 1$ to n **do**
  $\gamma_i \leftarrow L_i - Lap(4/\epsilon_1)$
end for
Let k be the first index such that $\gamma_k \leq T'$
Output: $\{\pi_1, \ldots, \pi_k\}$

Theorem 4.3. *The output k of from Algorithm 2 satisfies $\Delta(G - \cup_{i=1}^k \{\pi_i\}) \leq D + O(\log n/\epsilon')$ with high probability. In addition, $k = O(OPT \cdot \log n/\epsilon')$ with high probability.*

Proof. It is well established that the AboveThreshold algorithm is (α, β) accurate, i.e., $\Pr[|L_k - 6\ln n/\epsilon'| > \alpha] \leq \beta$, with

$$\alpha = \frac{8(\log n + \log(2/\beta))}{\epsilon_1}.$$

Then, for $\beta = 1/n$, we obtain $\alpha = \frac{16\ln n + 8\ln 2}{\epsilon} = O(\log n/\epsilon')$. Thus, $L_i \leq O(\log n/\epsilon')$ with high probability.

On the other hand, observe that if $\Delta(G - \cup_{i=1}^{k}\{\pi_i\}) > D + x$, then there is a node j that has degree at least $D + x$. The multi-set corresponding to this node would have size at least x, since removing this node would satisfy its covering requirement completely. Therefore, $x \leq L_k$, and hence, $\Delta(G - \cup_{i=1}^{k}\{\pi_i\}) \leq D + x$ with probability at least $1 - 1/n$.

Let $\hat{k}$ be the "true" stopping point $L_{\hat{k}} \leq 6 \ln n/\epsilon'$. Using the proof for Lemma C.1, the exponential mechanism satisfies $\hat{k} \leq O(OPT \cdot \log n/\epsilon')$ when $L_i \geq 6 \log n/\epsilon'$. It is sufficient to show that $k - \hat{k} \leq O(\log n/\epsilon)$. Observe that for $i \geq \hat{k}$, $L_i \leq 6 \ln n/\epsilon'$ but $\gamma_i \geq T'$, the Laplace noise added to L_i is greater than that added to T, this occurs with probability at most $1/2$ (since the Laplace distribution is symmetric about 0). Then the probability $\Pr[k - \hat{k} \geq \log_2 n] \leq 1/n$, so $k \leq O(OPT \cdot \log n/\epsilon')$ with high probability. □

C.3 Lower Bounds

In this section, we state the lower bounds of even outputting an explicit partial coverage requirements, that (1) any (ϵ, δ)-differentially private algorithm outputting (explicitly) a multiplicative coverage requirements (covers at least αr_e for all $e, \alpha < 1$) must output at least $m - 1$ sets, and (2) any (ϵ, δ)-differentially private algorithm outputting (explicitly) an additive coverage requirements (covers more than $r_e - \beta$ for all e) using no more than $O(\log n) + |OPT|$, where OPT indicates the optimal solution without privacy, must do so with $\beta = \tilde{\Omega}(\log n)$. The multiplicative case is straightforward to verify, as setting $r_e = 1$ for some element e. Any multiplicative partial cover must cover at least a total copy of e. This impossibility of this instance is reduced to the impossibility of the set cover problem as stated by [19].

Theorem C.7. *Any (ϵ, δ)-differentially private algorithm outputting an additive coverage requirements explicitly (covers at least $r_e - \beta$ for all e) using less than $O(\log n) + |OPT|$ with probability at least $1 - C, C = n^{-\Omega(1)}$ must do so with $\beta = \tilde{\Omega}(\log n)$.*

Proof. Assume an algorithm M that is (ϵ, δ)-DP with $\delta = O(1/poly(n))$ that outputs an explicit cover that can partially cover at least $r_e - \beta$, using less than $O(\log n) + OPT$ sets for all e with probability $\Theta(1)$.

Let $\mathcal{U} = \{e\}$.

Let α be a positive constant, such that the number of sets that M outputs no more $|OPT| + \alpha \log n$ with probability at least $1 - C$. Let $r_e^0 > \beta + 3\alpha \log n)$.

Let $S_1 = \{e \times (r_e^0 - \beta - \alpha \log n)\}$, i.e., a set with $(r_e^0 - \beta - \alpha \log n)$ copies of e. Let the set system be $\mathcal{S} = \{S_1, \{e\} \times (\beta + \alpha \log n)\}$.

Consider four instances of the the input with coverage requirements $I_1 = (r_e = r_e^0, \mathcal{S}), I_2 = (r_e = r_e^0 - \beta, \mathcal{S}), I_3 = (r_e = r_e^0 - \beta - 1, \mathcal{S}), I_4 = (r_e = r_e^0 - 1, \mathcal{S})$ respectively, with the set system $\mathcal{S}$. It is clear that $\mathcal{S}$ is enough to fully cover all the instances.

Let $S^* = \{S \subset \mathcal{S} : S_1 \in S, |S| \leq \alpha \log n\}$. In other words, each S contains S_1 and up to $\alpha - 1$ copies of $\{e\}$. Then every $S \in S^*$ covers at most $r_e^0 - \beta - 1$ copies of e.

Consider instance I_1. M guarantees to cover at least $r_e^0 - \beta$ copies of e with probability at least $1 - C$. Therefore, $\Pr[M(I_1) \in S^*] \leq C$.

Consider instance I_2. Without privacy, S_1 is the optimal solution, hence $|OPT = 1|$ for I_2. If $S_1 \notin M(I_2)$, $M(I_2)$ must use at least $r_1^0 - \beta - \alpha \log n > \alpha \log n$ sets. With probability at least $1 - C$, the output contains S_1 and using no more than $\alpha \log n$ sets, hence $\Pr[M(I_2) \in S^*] \geq 1 - C$.

Because I_1, I_2 are β-step neighbors, using group privacy we have:

$$\begin{aligned}
1 - C &\leq \Pr[M(I_2) \in S^*] \\
&\leq e^{\beta\epsilon} \Pr[M(I_1) \in S^*] + \beta e^{\beta\epsilon}\delta \\
&\leq e^{\beta\epsilon} C + \beta e^{\beta\epsilon}\delta.
\end{aligned}$$

Therefore $\frac{1 - e^{\beta\epsilon} C}{\beta e^{\beta\epsilon}} \leq 1/poly(n)$. It is clear that $\beta = \tilde{\Omega}(\log n)$.

$\square$

Lemma 4.5. Lower bound of PRIVATEMAXDEG. *Any explicit (ϵ, δ)-differentially private algorithm for the* PRIVATEMAXDEGREE *removing at most $O(\log n) + |OPT|$ nodes with probability at least $1 - C, C = n^{-\Omega(1)}$, must incur an additive error $\Delta(G - \cup_{i=1}^{k}\{\pi_i\}) = D + \tilde{\Omega}(\log n)$, where $\pi_1, \ldots, \pi_k$ are the removed nodes.*

Using the same setup as in Theorem C.7, setting $r_v = \max(d(v, G) - D, 0)$ for all nodes v. Similar to Theorem C.7, any explicit (ϵ, δ)-DP algorithm removing fewer than $O(\log n) + |OPT|$ nodes will guarantee to cover each node v no more than $d(v, G) - D - \tilde{\Omega}(\log n)$ times, i.e., the maximum degree of the remaining graph is $\Delta - (\Delta - D - \tilde{\Omega}(\log n)) = D + \tilde{\Omega}(\log n)$.

D SPECTRALRADIUS

D.1 Bound via PARTIALSETCOVER

Algorithm 3 (restated). Private Hitting Walks Algorithm for PRIVMINSR

1: **Input:** Graph $G = (V, E)$, privacy parameters (ϵ, δ)
2: Set $\epsilon' \leftarrow \epsilon/(2\ln(e/\delta))$, initialize permutation $\pi \leftarrow \emptyset$
3: **for** $i = 1$ to n **do**
4:     Sample $v \in V$ with prob. $\propto \exp(\epsilon' \cdot A(v))$,   append v to π and remove v from V
5: **end for**
6: Set $T \leftarrow \Delta^{1/2}$, $\theta \leftarrow 4nT^4$, $\hat{\theta} \leftarrow \theta - Lap(2/\epsilon_1)$
7: **for** $i = 1$ to n **do**
8:     $\gamma_i \leftarrow W_4(G[V - \{\pi(1), \ldots, \pi(i)\}]) - Lap(4/\epsilon_1)$
9: **end for**
10: Let k be the first iteration such that $\gamma_k \leq \hat{\theta}$
11: **Output:** $(\pi(1), \ldots, \pi(k))$

Lemma D.1. *If $T^4 \geq 6\ln n/\epsilon'$, the output $V' = \{\pi_1, \ldots, \pi_k\}$ of Algorithm 3 satisfies $W_4(G[V \setminus V']) \leq nT^4 + O(\log n/\epsilon')$ and is a $O(\log n)$ approximation with high probability.*

Proof. Since AboveThreshold is (α, β)-accurate, for $\beta = 1/n$, we obtain $W_4[V \setminus V'] \leq nT^4 + O(\log n/\epsilon')$ whp, similar to the proof for Theorem 4.3.

Let L_i denote the utility of the largest set after the $V_i = \{\pi_1, \ldots, \pi_i\}$ have been removed (i.e. $L_i = \max_v A(v)$). For $i < k$, $W_4(V \setminus V_i) \geq nT^4 \geq n \cdot 6\ln n/\epsilon'$, and $W_4(V \setminus V_i) \leq \sum_{v \in V} A(v) \leq nL_i$. Hence, $L_i \geq 6\ln n/\epsilon'$. By the same argument as in Proof 4.3, $A(\pi_i) \geq L_i/2$ whp. In other words, the utility of the chosen set is at least half of that chosen by a non-private greedy algorithm. Since the greedy algorithm is a $O(\ln n)$ approximation, Algorithm 3 would be a $O(2\ln n) = O(\ln n)$-approximation. $\square$

Lemma D.2. *Algorithm 3 is $(\Delta^2(\epsilon + \epsilon_1), \Delta^2\delta e^{(\Delta^2-1)\epsilon})$-private.*

Proof. Since $A(v)$ has a sensitivity of Δ^2, neighboring datasets in Private Hitting Walks would be Δ^2-step neighbors in Partial Set Cover and Above Threshold instead.

Algorithm 3 is the composition of a $(\Delta^2\epsilon, \Delta^2\delta e^{(\Delta^2-1)\epsilon})$-private set cover algorithm and $\Delta^2\epsilon_1$-private AboveThreshold process, hence the overall privacy budget would be $(\Delta^2(\epsilon+\epsilon_1), \Delta^2\delta e^{(\Delta^2-1)\epsilon})$. $\square$

D.2 PRIVATESPECTRALRADIUS via PRIVATEMULSET

Theorem D.3. *Let $\hat{B}$ be the cost of the output of Algorithm 6. W.h.p., $\hat{B} < |OPT| \cdot O((1 + 1/\epsilon')\ln|V|)$.*

Algorithm 6 Private algorithm for PRIVATEMAXDEG

Input: (ϵ, δ), input graph G, target max degree D

$\mathcal{S} = \{S_v : v \in V\}$, such that S_v contains ∞ copies of v and $d(u, G)$ copies of each u that is adjacent to v

$R = \{r_V : v \in V\}, \forall v \in V : r_v \leftarrow \max(\sum_{u \sim v} d(u, G) - D, 0)$

$\epsilon' \leftarrow \epsilon/4\Delta$

$\delta' \leftarrow \delta/4\Delta e^{(4\Delta-1)\epsilon'}$

Return Algorithm 4$(\epsilon', \delta', \mathcal{S}, R)$

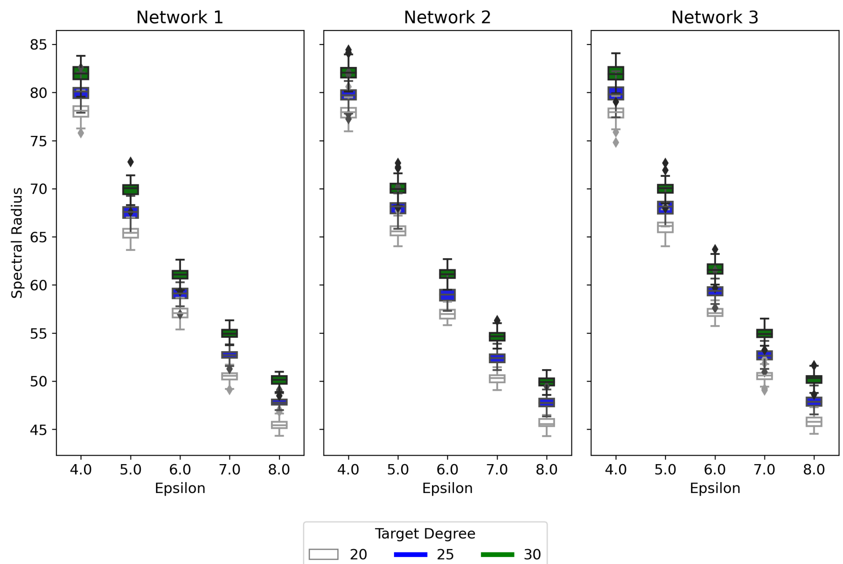

Figure 4: Effect of Privacy on Spectral Radius on BTER Graphs

E Additional Experiments

Effect of the ϵ on the spectral radius. Figure 4 shows the $\rho(G[V - S])$ for the explicit solutions S computed using Algorithm 2 for BTER networks, for the same parameters and privacy budgets mentioned earlier. The results here show that the resulting spectral radius is quite a bit smaller than the maximum degree. As expected, the resulting spectral radius of the residual graphs follow a similar trend as the max degree, with higher ϵ budgets obtaining better metrics due to less privacy constraints.

Cost of achieving different epidemic metrics. Figures 5 and 6 show the violation in the target degree (for $D = 20$) and the spectral radius vs the explicit solution cost (computed using Algorithm 2), for different ϵ in the BTER networks. As noted earlier, the violation and spectral radius decrease significantly as the solution cost increases, which is achieved for higher ϵ.

Privacy vs Vaccination Cost in BTER. We also investigated the tradeoff of privacy and vaccination cost in the 3 BTER graphs ($\gamma = 0.5, \rho = 0.95, \eta = 0.05$) for implicit PRIVATEMAXDEG, with target degree $D = 20$, as shown in Figure 7. The non-private greedy algorithm is used as a baseline comparison. Due to the relaxed privacy budget of $\delta = 0.01$, the variation of ϵ has a much more pronounced effect on vaccination budget, and the algorithm's performance is much closer to that of the non-private greedy as compared to Figure 1, and are within the bounds expected from Lemma 4.2.

Effect on Infection Simulation. Finally, we computed the 300 explicit solutions using various privacy budgets ϵ (and $\delta = 0.01$) and target max degree 10 for 3 "social circles" in the SNAP Facebook datasets [29], we then performed 200 simulations of SIR with transmission probability 0.2 and 20 initial infections to determine the average vaccination budget and infection size and demonstrate the effectiveness of the solutions to minimize infection spread. Note that we used the more relaxed mutltiset multicover version of differential privacy for these experiments.

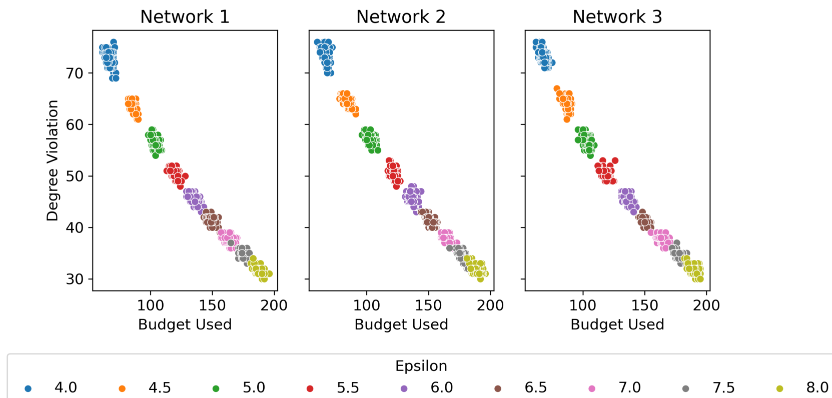

Figure 5: Tradeoff of Degree Violation vs Budget on BTER Graphs

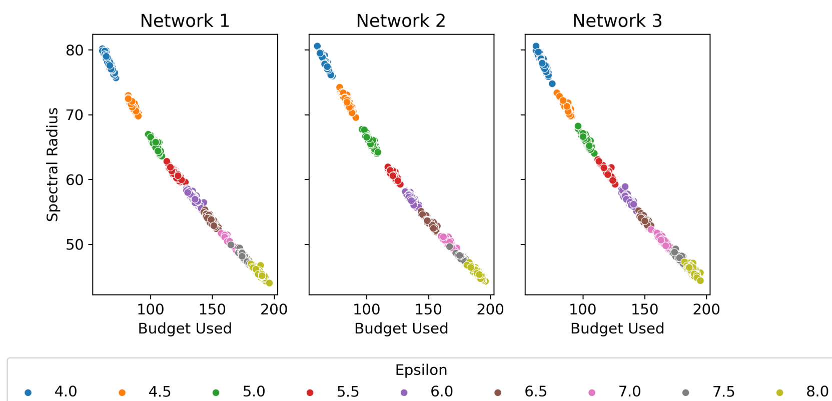

Figure 6: Tradeoff of Spectral Radius vs Budget on BTER Graphs

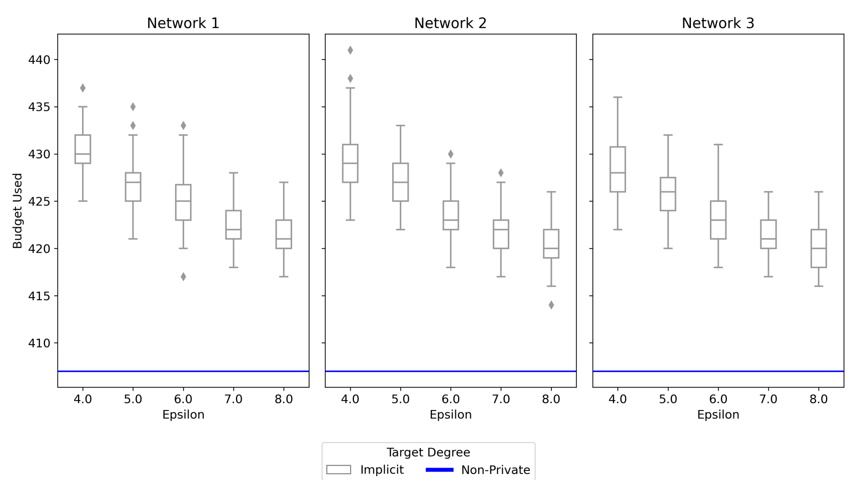

Figure 7: Tradeoff of Spectral Radius vs Budget on BTER Graphs for Implicit Solution ($D = 20$)

Table 3: Infection Spread on Facebook Social Circles

Network	$\epsilon = 4$		$\epsilon = 6$		$\epsilon = 8$	
	Budget	Spread	Budget	Spread	Budget	Spread
Circle 0	14.52	205.18	30.48	171.55	42.28	138.02
Circle 1	311.70	586.99	411.53	413.50	546.56	251.49
Circle 2	45.52	138.29	73.45	90.07	94.57	60.38

References

[1] Aniruddha Adiga, Devdatt Dubhashi, Bryan Lewis, Madhav Marathe, Srinivasan Venkatramanan, and Anil Vullikanti. Mathematical models for covid-19 pandemic: a comparative analysis. *Journal of the Indian Institute of Science*, pages 1–15, 2020.

[2] C. L. Barrett, R. J. Beckman, M. Khan, et al. Generation and analysis of large synthetic social contact networks. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pages 1003–1014, 2009.

[3] Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. Differentially private data analysis of social networks via restricted sensitivity. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, ITCS '13, page 87–96, New York, NY, USA, 2013. Association for Computing Machinery.

[4] B. Bollobás and O. Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, 2004.

[5] Eugene Chan and Najam Saqib. Privacy concerns can explain unwillingness to download and use contact tracing apps when covid-19 concerns are high. *Computers in Human Behavior*, 119:106718, 01 2021.

[6] Bo Chen, Calvin Hawkins, Kasra Yazdani, and Matthew Hale. Edge differential privacy for algebraic connectivity of graphs. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 2764–2769. IEEE, 2021.

[7] Bo Chen, Baike She, Calvin Hawkins, Alex Benvenuti, Brandon Fallin, Philip E Paré, and Matthew Hale. Differentially private computation of basic reproduction numbers in networked epidemic models. In *2024 American Control Conference (ACC)*, pages 4422–4427. IEEE, 2024.

[8] Jiangzhuo Chen, Stefan Hoops, Achla Marathe, Henning Mortveit, Bryan Lewis, Srinivasan Venkatramanan, Arash Haddadan, Parantapa Bhattacharya, Abhijin Adiga, Anil Vullikanti, et al. Effective social network-based allocation of covid-19 vaccines. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4675–4683, 2022.

[9] Reuven Cohen, Shlomo Havlin, and Daniel ben Avraham. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.*, 91:247901, Dec 2003.

[10] Vincent Cohen-Addad, Chenglin Fan, Silvio Lattanzi, Slobodan Mitrovic, Ashkan Norouzi-Fard, Nikos Parotsidis, and Jakub M Tarnawski. Near-optimal correlation clustering with privacy. *Advances in Neural Information Processing Systems*, 35:33702–33715, 2022.

[11] Laxman Dhulipala and George Z Li. Fine-grained privacy guarantees for coverage problems. *arXiv preprint arXiv:2403.03337*, 2024.

[12] Laxman Dhulipala, Quanquan C Liu, Sofya Raskhodnikova, Jessica Shi, Julian Shun, and Shangdi Yu. Differential privacy from locally adjustable graph algorithms: k-core decomposition, low out-degree ordering, and densest subgraphs. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 754–765. IEEE, 2022.

[13] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[14] Ken T.D. Eames, Jonathan M. Read, and W. John Edmunds. Epidemic prediction and control in weighted networks. *Epidemics*, 1(1):70 – 76, 2009.

[15] S. Eubank, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, and N. Wang. Structure of Social Contact Networks and Their Impact on Epidemics. In *Discrete Methods in Epidemiology*, volume 70, pages 179–200. American Math. Soc., Providence, RI, 2006.

[16] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.

[17] Odile Favaron, Maryvonne Mahéo, and Jean-François Saclé. Some eigenvalue properties in graphs (conjectures of graffiti - ii). *Discret. Math.*, 111:197–220, 1993.

[18] Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, and Adam Sealfon. Individualized privacy accounting via subsampling with applications in combinatorial optimization. *arXiv preprint arXiv:2405.18534*, 2024.

[19] Anupam Gupta, Katrina Ligett, Frank McSherry, Aaron Roth, and Kunal Talwar. Differentially private combinatorial optimization, 2009.

[20] Calvin Hawkins, Bo Chen, Kasra Yazdani, and Matthew Hale. Node and edge differential privacy for graph laplacian spectra: Mechanisms and scaling laws. *IEEE Transactions on Network Science and Engineering*, 11(2):1690–1701, 2023.

[21] Ara Hayrapetyan, David Kempe, Martin Pál, and Zoya Svitkina. Unbalanced graph cuts. In *Proceedings of the 13th Annual European Conference on Algorithms*, ESA'05, page 191–202, Berlin, Heidelberg, 2005. Springer-Verlag.

[22] Jacob Imola, Takao Murakami, and Kamalika Chaudhuri. Locally differentially private analysis of graph statistics. In *30th USENIX Symposium on Security*, 2021.

[23] Tianxi Ji, Changqing Luo, Yifan Guo, Jinlong Ji, Weixian Liao, and Pan Li. Differentially private community detection in attributed social networks. In *Asian Conference on Machine Learning*, pages 16–31. PMLR, 2019.

[24] M Kargar and T Sistani. New upper bounds on the spectral radius of graphs. *Journal of Mathematical Extension*, 14(4), 2020.

[25] Vishesh Karwa, Sofya Raskhodnikova, Adam Smith, and Grigory Yaroslavtsev. Private analysis of graph structure. *ACM Transactions on Database Systems (TODS)*, 39(3):1–33, 2014.

[26] Shiva Prasad Kasiviswanathan, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Analyzing graphs with node differential privacy. In *Proceedings of the 10th Theory of Cryptography Conference on Theory of Cryptography*, TCC'13, pages 457–476, Berlin, Heidelberg, 2013. Springer-Verlag.

[27] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.

[28] Tamara Gibson Kolda, Ali Pinar, and Seshadhri Comandur. The bter graph model: Blocked two-level erdos-renyi. Technical report, Sandia National Lab.(SNL-CA), Livermore, CA (United States), 2011.

[29] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014.

[30] George Z Li, Dung Nguyen, and Anil Vullikanti. Differentially private partial set cover with applications to facility location. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 4803–4811, 2023.

[31] George Z Li, Dung Nguyen, and Anil Vullikanti. Computing epidemic metrics with edge differential privacy. In *International Conference on Artificial Intelligence and Statistics*, pages 4303–4311. PMLR, 2024.

[32] Madhav Marathe and Anil Vullikanti. Computational epidemiology. *Communications of the ACM*, 56(7):88–96, 2013.

[33] Joel C. Miller and James Mac Hyman. Effective vaccination strategies for realistic social networks. pages 780–785, 2007.

[34] Dung Nguyen and Anil Vullikanti. Differentially private densest subgraph detection. In *38th International Conference on Machine Learning (ICML)*. PMLR, July 2021.

[35] Masaki Ogura and Victor M. Preciado. *Optimal Containment of Epidemics in Temporal and Adaptive Networks*, pages 241–266. Springer Singapore, Singapore, 2017.

[36] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.

[37] Govind Persad, Ezekiel J Emanuel, Samantha Sangenito, Aaron Glickman, Steven Phillips, and Emily A Largent. Public perspectives on covid-19 vaccine prioritization. *JAMA network open*, 4(4):e217943–e217943, 2021.

[38] B Aditya Prakash, Deepayan Chakrabarti, Nicholas C Valler, Michalis Faloutsos, and Christos Faloutsos. Threshold conditions for arbitrary cascade models on arbitrary networks. *Knowledge and information systems*, 33:549–575, 2012.

[39] Victor M. Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George J. Pappas. Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks. In *IEEE Conference on Decision and Control*. IEEE, 2013.

[40] Victor M. Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George J. Pappas. Optimal resource allocation for network protection against spreading processes. In *IEEE Transactions on Control of Network Systems*, pages 99 – 108. IEEE, 2014.

[41] Victor M. Preciado, Michael Zargham, and David Sun. A convex framework to control spreading processes in directed networks. In *Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2014.

[42] S. Rajagopalan and V.V. Vazirani. Primal-dual rnc approximation algorithms for (multi)-set (multi)-cover and covering integer programs. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 322–331, 1993.

[43] Daniel M Romero, Brendan Meeder, and Jon Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 695–704, 2011.

[44] Sudip Saha, Abhijin Adiga, B Aditya Prakash, and Anil Kumar S Vullikanti. Approximation algorithms for reducing the spectral radius to control epidemic spread. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 568–576. SIAM, 2015.

[45] Zoya Svitkina and Éva Tardos. Min-max multiway cut. In Klaus Jansen, Sanjeev Khanna, José D. P. Rolim, and Dana Ron, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 207–218, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[46] Piet Van Mieghem, Dragan Stevanović, Fernando Kuipers, Cong Li, Ruud Van De Bovenkamp, Daijie Liu, and Huijuan Wang. Decreasing the spectral radius of a graph by link removals. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 84(1):016101, 2011.

[47] Yue Wang, Xintao Wu, and Leting Wu. Differential privacy preserving spectral graph analysis. In *Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II 17*, pages 329–340. Springer, 2013.

[48] Yingrui Yang, Ashley McKhann, Sixing Chen, Guy Harling, and Jukka-Pekka Onnela. Efficient vaccination strategies for epidemic control using network information. *Epidemics*, 27:115 – 122, 2019.

[49] Jun Zhang, Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava, and Xiaokui Xiao. Private release of graph statistics using ladder functions. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, SIGMOD '15, page 731–745, New York, NY, USA, 2015. Association for Computing Machinery.