7. Appendix

Theorem 3.3. LocalAlg is (ϵ, δ)-differentially private.

Proof. We show this by proving that each iteration of the `for` loop in line 7 of LocalAlg is (ϵ', δ')-differentially private, where $\epsilon' = \epsilon/m_C$ and $\delta' = \delta/m_C$. Since there are at most m_C iterations of this loop for each client, composition of differentially private algorithms [17] guarantees that LocalAlg ensures (ϵ, δ)-differential privacy for each client.

Denote each iteration of the `for` loop in line 7 of LocalAlg by L; it takes as input a record $\langle q, u\rangle \in D$, and returns a record, which we denote $L(\langle q, u\rangle)$. If q is not in HL or u is not in $HL[q]$, then they immediately get transformed into a default value ($\star$) that is in the head list. Since L outputs only values that exist in the head list, to confirm differential privacy we need to prove that for any arbitrary neighboring datasets $\langle q, u\rangle$ and $\langle q', u'\rangle$, $\Pr\big[L(\langle q, u\rangle) \in Y\big] \leq e^{\epsilon'} \Pr\big[L(\langle q', u'\rangle) \in Y\big] + \delta'$ holds for all sets of head list records Y.

Whenever $k = 1$ or $k_q = 1$, the only query (or URL for a specific query) is $\star$, which will be output with probability 1. Thus, differential privacy trivially holds, since the reported values then do not rely on the client's data. Thus, we'll assume $k \geq 2$ and $k_q \geq 2$. Note that there is a single decision point where it is determined whether q will be reported truthfully or not. Thus, we can split the privacy analysis into two parts: 1) Usage of the f_C fraction of the privacy budget to report a query, and 2) Usage of the remainder of the privacy budget to report a URL (given the reported query). This decomposes a simultaneous two-item (ϵ', δ') reporting problem into two single-item reporting problems with (ϵ'_Q, δ'_Q) and (ϵ'_U, δ'_U) respectively, where $\epsilon'_Q = f\epsilon'$, $\delta'_Q = f\delta'$, $\epsilon'_U = (1 - f_C)\epsilon'$, and $\delta'_U = (1 - f_C)\delta'$.

1. Privacy of query reporting: Consider the query-reporting case first. Overloading our use of L, let $L(q)$ be the portion of L that makes use of q. We first ensure that

$$\Pr[L(q) = q_{HL}] \leq \exp(\epsilon'_Q) \Pr[L(q') = q_{HL}] + \frac{\delta'_Q}{2} \tag{7.1}$$

holds for all q, q', and $q_{HL} \in HL$. This trivially holds when $q_{HL} = q = q'$ or $q_{HL} \notin \{q, q'\}$. The remaining scenarios to consider are: 1) $q \neq q_{HL}, q' = q_{HL}$ and 2) $q = q_{HL}, q' \neq q_{HL}$. By the design of the algorithm, $\Pr[L(q_{HL}) = q_{HL}] = t$ and $\Pr[L(\bar{q}_{HL}) = q_{HL}] = (1 - t)(\frac{1}{k-1})$, where $\bar{q}_{HL}$ represents any query not equal to q_{HL}. With $t = \frac{\exp(\epsilon'_Q) + (\delta'_Q/2)(k-1)}{\exp(\epsilon'_Q) + k - 1}$, it is simple to verify that inequality (7.1) holds.

Consider an arbitrary set of head list queries Y.

$$\begin{aligned}
\Pr[L(q) \in Y] &= \sum_{q_{HL} \in Y} \Pr[L(q) = q_{HL}] \\
&= \sum_{q_{HL} \in Y \setminus \{q, q'\}} \Pr[L(q) = q_{HL}] + \sum_{q_{HL} \in Y \cap \{q, q'\}} Pr[L(q) = q_{HL}] \\
&= \sum_{q_{HL} \in Y \setminus \{q, q'\}} \Pr[L(q') = q_{HL}] + \sum_{q_{HL} \in Y \cap q, q'} \Pr[L(q) = q_{HL}] \qquad (7.2) \\
&\leq \sum_{q_{HL} \in Y \setminus \{q, q'\}} \Pr[L(q') = q_{HL}] + \sum_{q_{HL} \in Y \cap \{q, q'\}} \big(e^{\epsilon'_Q} \Pr[L(q') = q_{HL}] + \frac{\delta'_Q}{2}\big) \qquad (7.3) \\
&\leq e^{\epsilon'_Q} \sum_{q_{HL} \in Y} \Pr[L(q') = q_{HL}] + 2 \cdot \frac{\delta'_Q}{2} \\
&= e^{\epsilon'_Q} \Pr[L(q') \in Y] + \delta'_Q,
\end{aligned}$$

Equality (7.2) stems from the fact that the probability of reporting a false query is independent of the user's true query. The inequality (7.3) is a direct application of inequality (7.1). Thus, L is (ϵ'_Q, δ'_Q)-differentially private for query-reporting.

2. Privacy of URL reporting: With t_q defined as $t_q = \frac{\exp(\epsilon'_U)+0.5\delta'_U(k_q-1)}{\exp(\epsilon'_U)+k_q-1}$, an analogous argument shows that the (ϵ'_U, δ'_U)-differential privacy constraints hold if the original q is kept. On the other hand, if it is replaced with a random query, then they trivially hold as the algorithm reports a random element in the URL list of the reported query, without taking into consideration the client's true URL u.

By composition [17], each of the at most m_C iterations of L is $(\epsilon'_Q+\epsilon'_U, \delta'_Q+\delta'_U) = (\epsilon', \delta')$-differentially private. □

Observation 3.4. $\hat{p}_C$ gives the unbiased estimate of record and query probabilities under EstimateClientProbabilities.

Proof. Reporting records is a two-stage process (first, decide which query to report, then report a record); similarly, denoising is also done in two stages.

Denoising of query probability estimates: Let $r_{C,q}$ denote the probability that the algorithm has received query q as a report, and let p_q be the true probability of a user having query q. We want to learn p_q based on $r_{C,q}$. By the design of our algorithm,

$$\begin{aligned} r_{C,q} &= t \cdot p_q + \sum_{q' \neq q} p_{q'}(1-t)\frac{1}{k-1} \\ &= t \cdot p_q + \frac{1-t}{k-1}\sum_{q' \neq q} p_{q'} \\ &= t \cdot p_q + \frac{1-t}{k-1}(1-p_q). \end{aligned}$$

Solving for p_q in terms of $r_{C,q}$ yields $p_q = \frac{r_{C,q}-\frac{1-t}{k-1}}{t-\frac{1-t}{k-1}}$. Using the obtained data for the query $\hat{r}_{C,q}$, we estimate $p_{C,q}$ as $\hat{p}_{C,q} = \frac{\hat{r}_{C,q}-\frac{1-t}{k-1}}{t-\frac{1-t}{k-1}}$.

Denoising of record probability estimates: Analogously, denote by $r_{C,\langle q,u\rangle}$ the probability that the algorithm has received a record $\langle q, u\rangle$ as a report, and recall $p_{\langle q,u\rangle}$ is the record's true probability in the dataset. Then $r_{C,\langle q,u\rangle} = t \cdot t_q \cdot p_{\langle q,u\rangle} + \left(t\frac{1-t_q}{k_q-1}\right)(p_q - p_{\langle q,u\rangle}) + \left(\frac{1-t}{k-1}\cdot\frac{1}{k_q}\right)(1-p_q)$, recalling from the algorithm that k_q is the number of URLs associated with query q and t_q is the probability of truthfully reporting u given that query q was reported. Solving for $p_{\langle q,u\rangle}$ yields $p_{\langle q,u\rangle} = \frac{r_{C,\langle q,u\rangle} - \left(t\frac{1-t_q}{k_q-1}p_q + \frac{(1-t)(1-p_q)}{(k-1)k_q}\right)}{t(t_q-\frac{1-t_q}{k_q-1})}$.

Using the obtained data for the empirical report estimate $\hat{r}_{C,\langle q,u\rangle}$ together with the query estimate $\hat{p}_{C,q}$, we estimate $p_{\langle q,u\rangle}$ as $\hat{p}_{C,\langle q,u\rangle} = \frac{\hat{r}_{C,\langle q,u\rangle} - \left(t\frac{1-t_q}{k_q-1}\hat{p}_{C,q} + \frac{(1-t)(1-\hat{p}_{C,q})}{(k-1)k_q}\right)}{t(t_q-\frac{1-t_q}{k_q-1})}$. □

Theorem 3.5. If $m_O = 1$ then the unbiased variance estimate for the opt-in group's record probabilities can be computed as: $\hat{\sigma}^2_{O,\langle q,u\rangle} = \frac{|D_T|}{|D_T|-1}\left(\frac{\hat{p}_{O,\langle q,u\rangle}(1-\hat{p}_{O,\langle q,u\rangle})}{|D_T|} + 2\left(\frac{b_T}{|D_T|}\right)^2\right)$.

Proof. Given the head list, the distribution of EstimateOptinProbabilities' estimate for a record $\langle q,u\rangle$ is given by $r_{O,\langle q,u\rangle} = p_{\langle q,u\rangle} + \frac{Y}{|D_T|}$, where $Y \sim \text{Laplace}(b_T)$ with b_T being the scale parameter and recalling that $|D_T|$ is the total number of records from the opt-in users used to estimate probabilities. The empirical estimator for $r_{O,\langle q,u\rangle}$ is $\hat{r}_{O,\langle q,u\rangle} = \frac{1}{|D_T|}\sum_{j=1}^{|D_T|} X_j + Y$, where $X_j \sim \text{Bernoulli}(p_{\langle q,u\rangle})$ is the random variable indicating whether report j was record $\langle q,u\rangle$.

The expectation of this estimator is given by $\mathrm{E}[\hat{r}_{O,\langle q,u\rangle}] = p_{\langle q,u\rangle}$. Thus, $\hat{r}_{O,\langle q,u\rangle}$ is an unbiased estimator for $p_{\langle q,u\rangle}$. We denote $\hat{p}_{O,\langle q,u\rangle} = \hat{r}_{O,\langle q,u\rangle}$ to explicitly reference it as the estimator of $p_{\langle q,u\rangle}$. The variance for this estimator is

$$\begin{aligned}
\sigma^2_{O,\langle q,u\rangle} &= \mathrm{Var}[\hat{p}_{O,\langle q,u\rangle}] && (7.4)\\
&= \mathrm{Var}\Big[\frac{1}{|D_T|}\Big(\sum_{j=1}^{|D_T|} X_j + Y\Big)\Big] \\
&= \frac{1}{|D_T|^2}\Big(\mathrm{Var}\Big[\sum_{j=1}^{|D_T|} X_j\Big] + \mathrm{Var}\,[Y]\Big) && (7.5)\\
&= \frac{1}{|D_T|^2}\Big(\sum_{j=1}^{|D_T|} \mathrm{Var}\,[X_j] + \mathrm{Var}\,[Y]\Big) && (7.6)\\
&= \frac{1}{|D_T|^2}\big(|D_T|\cdot p_{\langle q,u\rangle}(1-p_{\langle q,u\rangle})\big) + 2\Big(\frac{b_T}{|D_T|}\Big)^2 \\
&= \frac{p_{\langle q,u\rangle}(1-p_{\langle q,u\rangle})}{|D_T|} + 2\Big(\frac{b_T}{|D_T|}\Big)^2.
\end{aligned}$$

Equality 7.5 comes from the independence between Y and all X_j. Equality 7.6 relies on an assumption of independence between X_j, X_k for all $j \neq k$ (i.e., the iid assumption discussed prior to the theorem statements).

To compute this variance, we need to use the data in place of the unknown $p_{\langle q,u\rangle}$. Using $\hat{p}_{O,\langle q,u\rangle}$ directly in place of $p_{\langle q,u\rangle}$ requires a $\frac{|D_T|}{|D_T|-1}$ factor correction (known as "Bessel's correction[11]") to generate an unbiased estimate. Thus, the variance of each opt-in record probability estimate is: $\hat{\sigma}^2_{O,\langle q,u\rangle} = \frac{|D_T|}{|D_T|-1}\left(\frac{\hat{p}_{O,\langle q,u\rangle}(1-\hat{p}_{O,\langle q,u\rangle})}{|D_T|} + 2\left(\frac{b_T}{|D_T|}\right)^2\right)$. □

Theorem 3.6. If $m_C = 1$ then the unbiased variance estimate for the client group's record probabilities can be computed as:

$$\hat{\sigma}^2_{C,\langle q,u\rangle} = \frac{|D_C|}{t^2\big(t_q - \frac{1-t_q}{k_q-1}\big)^2(|D_C|-1)}\cdot$$
$$\Big(\frac{\hat{r}_{C,\langle q,u\rangle}(1-\hat{r}_{C,\langle q,u\rangle})}{|D_C|} + \big(\frac{1-t}{(k-1)k_q} - t\frac{1-t_q}{k_q-1}\big)^2\hat{\sigma}^2_{C,q} + 2\big(\frac{1-t}{(k-1)k_q} - t\frac{1-t_q}{k_q-1}\big)\frac{\hat{r}_{C,\langle q,u\rangle}(1-\hat{r}_{C,q})}{|D_C|(t-\frac{1-t}{k-1})}\Big).$$

[11] https://en.wikipedia.org/wiki/Bessel's_correction

Proof. We'll first derive the variance estimate for the client group's query probabilities, then move on to the variance estimate for their record probabilities.

From the proof of Observation 3.4, the distribution of the reported query q from the client algorithm is given by $r_{C,q} = t \cdot p_q + \frac{1-t}{k-1}(1-p_q)$, and so the true probability of query q is distributed as $p_q = \frac{r_{C,q} - \frac{1-t}{k-1}}{t - \frac{1-t}{k-1}}$. The empirical estimator for p_q is $\hat{p}_{C,q} = \frac{\hat{r}_{C,q} - \frac{1-t}{k-1}}{t - \frac{1-t}{k-1}}$, where $\hat{r}_{C,q}$ is the empirical estimator of $r_{C,q}$ defined explicitly as $\hat{r}_{C,q} = \frac{1}{|D_C|}\sum_{j=1}^{|D_C|} X_j$, where $X_j \sim \text{Bernoulli}(r_{C,q})$ is the random variable indicating whether report j was query q and recalling that $|D_C|$ is the total number of records from the client users.

The variance of $\hat{r}_{C,q}$ is

$$\begin{aligned}
\text{Var}[\hat{r}_{C,q}] &= \text{Var}\Big[\frac{1}{|D_C|}\sum_{j=1}^{|D_C|} X_j\Big] \\
&= \Big(\frac{1}{|D_C|}\Big)^2 \sum_{j=1}^{|D_C|} \text{Var}\left[X_j\right] && (7.7) \\
&= \big(\frac{1}{|D_C|}\big)^2 \big(|D_C| \cdot r_{C,q}(1-r_{C,q})\big) && (7.8) \\
&= \frac{r_{C,q}(1-r_{C,q})}{|D_C|},
\end{aligned}$$

where equality 7.7 relies on an assumption of independence between X_j, X_k for all $j \neq k$ (i.e., the iid assumption discussed prior to the theorem statements).

Then, the variance of $\hat{p}_{C,q}$ is

$$\sigma^2_{C,q} = \text{Var}[\hat{p}_{C,q}] = \text{Var}\Big[\frac{\hat{r}_{C,q} - \frac{1-t}{k-1}}{t - \frac{1-t}{k-1}}\Big] = \frac{r_{C,q}(1-r_{C,q})}{|D_C|\big(t - \frac{1-t}{k-1}\big)^2}.$$

To compute this variance, we need to use the data in place of the unknown $r_{C,q}$. Using $\hat{r}_{C,q}$ directly in place of $r_{C,q}$ requires including Bessel's $\frac{|D_C|}{|D_C|-1}$ factor correction to yield an unbiased estimate. Thus, the variance of the query probability estimates by the client algorithm is: $\hat{\sigma}^2_{C,q} = \left(\frac{1}{t - \frac{1-t}{k-1}}\right)^2 \frac{\hat{r}_{C,q}(1-\hat{r}_{C,q})}{|D_C|-1}$.

Now, we'll derive the variance estimate for the record probabilities. For a given query q and corresponding URL u in head list, denote X_i^q as the indicator random variable that is 1 if user i reported query q and 0 otherwise, and similarly denote $X_i^{\langle q,u\rangle}$ as the indicator random variable that is 1 if user i reported query q and URL u and 0 otherwise. Note that $X_i^q \sim \text{Bern}(r_{C,q})$ and $X_i^{\langle q,u\rangle} \sim \text{Bern}(r_{C,\langle q,u\rangle})$. The covariance between these two random variables is given by

$$\text{Cov}[X_i^q, X_i^{\langle q,u\rangle}] = \text{E}[X_i^q X_i^{\langle q,u\rangle}] - \text{E}[X_i^q]\text{E}[X_i^{\langle q,u\rangle}] = r_{C,\langle q,u\rangle} - r_{C,\langle q,u\rangle} r_{C,q} = r_{C,\langle q,u\rangle}(1 - r_{C,q}).$$

Also note that due to the iid assumption, for any other user j, we have $\mathrm{Cov}(X_i^q, X_j^{\langle q,u\rangle}) = 0$. Thus, we have the covariance between our empirical query and record estimates as

$$\begin{aligned}
\mathrm{Cov}[\hat{r}_q, \hat{r}_{\langle q,u\rangle}] &= \mathrm{Cov}\left[\frac{1}{|D_C|}\sum_{i\in D_C} X_i^q, \frac{1}{|D_C|}\sum_{i\in D_C} X_i^{\langle q,u\rangle}\right] \\
&= \frac{1}{|D_C|^2}\mathrm{Cov}\left[\sum_{i\in D_C} X_i^q, \sum_{i\in D_C} X_i^{\langle q,u\rangle}\right] \\
&= \frac{1}{|D_C|^2}\sum_{i,j\in D_C} \mathrm{Cov}[X_i^q, X_j^{\langle q,u\rangle}] \\
&= \frac{1}{|D_C|^2}\sum_{i\in D_C} \mathrm{Cov}[X_i^q, X_i^{\langle q,u\rangle}] \\
&= \frac{r_{C,\langle q,u\rangle}(1-r_{C,q})}{|D_C|}.
\end{aligned}$$

Utilizing this covariance expression, we can now compute the desired variance estimate as:

$$\begin{aligned}
&\sigma^2_{C,\langle q,u\rangle} = \mathrm{Var}[\hat{p}_{C,\langle q,u\rangle}] \\
&= \mathrm{Var}\left[\frac{\hat{r}_{C,\langle q,u\rangle} - \big(t\frac{1-t_q}{k_q-1}\hat{p}_{C,q} + \frac{(1-t)(1-\hat{p}_{C,q})}{(k-1)k_q}\big)}{t(t_q - \frac{1-t_q}{k_q-1})}\right] \\
&= \frac{1}{t^2(t_q - \frac{1-t_q}{k_q-1})^2}\mathrm{Var}\left[\hat{r}_{C,\langle q,u\rangle} - \big(t\frac{1-t_q}{k_q-1}\hat{p}_{C,q} + \frac{(1-t)(1-\hat{p}_{C,q})}{(k-1)k_q}\big)\right] \\
&= \frac{1}{t^2(t_q - \frac{1-t_q}{k_q-1})^2}\mathrm{Var}\left[\hat{r}_{C,\langle q,u\rangle} - \hat{p}_{C,q}\big(\frac{1-t}{(k-1)k_q} - t\frac{1-t_q}{k_q-1}\big)\right] \\
&= \frac{1}{t^2(t_q - \frac{1-t_q}{k_q-1})^2}\cdot \\
&\quad\Big(\mathrm{Var}\left[\hat{r}_{C,\langle q,u\rangle}\right] + \big(\frac{1-t}{(k-1)k_q} - t\frac{1-t_q}{k_q-1}\big)^2\mathrm{Var}\left[\hat{p}_{C,q}\right] + 2\big(\frac{1-t}{(k-1)k_q} - t\frac{1-t_q}{k_q-1}\big)\mathrm{Cov}[\hat{p}_{C,q}, \hat{r}_{C,\langle q,u\rangle}]\Big) \\
&= \frac{1}{t^2(t_q - \frac{1-t_q}{k_q-1})^2}\cdot \\
&\quad\left(\frac{r_{C,\langle q,u\rangle}(1-r_{C,\langle q,u\rangle})}{|D_C|} + \big(\frac{1-t}{(k-1)k_q} - t\frac{1-t_q}{k_q-1}\big)^2\sigma^2_{C,q} + 2\big(\frac{1-t}{(k-1)k_q} - t\frac{1-t_q}{k_q-1}\big)\frac{1}{t-\frac{1-t}{k-1}}\mathrm{Cov}[\hat{r}_{C,q}, \hat{r}_{C,\langle q,u\rangle}]\right) \\
&= \frac{1}{t^2(t_q - \frac{1-t_q}{k_q-1})^2}\cdot \\
&\quad\left(\frac{r_{C,\langle q,u\rangle}(1-r_{C,\langle q,u\rangle})}{|D_C|} + \big(\frac{1-t}{(k-1)k_q} - t\frac{1-t_q}{k_q-1}\big)^2\sigma^2_{C,q} + 2\big(\frac{1-t}{(k-1)k_q} - t\frac{1-t_q}{k_q-1}\big)\frac{1}{t-\frac{1-t}{k-1}}\frac{r_{C,\langle q,u\rangle}(1-r_{C,q})}{|D_C|}\right).
\end{aligned}$$

Using our already-computed estimates $\hat{r}_{C,q}, \hat{r}_{C,\langle q,u\rangle}$, and $\hat{\sigma}^2_{C,\langle q,u\rangle}$ (in place of $r_{C,q}, r_{C,\langle q,u\rangle}$, and $\sigma^2_{C,\langle q,u\rangle}$ respectively) and applying Bessel's correction, we obtain the stated result. □

Theorem 3.7. If $\hat{\sigma}^2_{O,\langle q,u\rangle}$ and $\hat{\sigma}^2_{C,\langle q,u\rangle}$ are sample variances of $\hat{p}_{O,\langle q,u\rangle}$ and $\hat{p}_{C,\langle q,u\rangle}$ respectively, and the blended estimate is the convex combination $\hat{p}_{\langle q,u\rangle} = w_{\langle q,u\rangle} \cdot \hat{p}_{O,\langle q,u\rangle} + (1 - w_{\langle q,u\rangle}) \cdot \hat{p}_{C,\langle q,u\rangle}$, then the sample variance optimal weighting is given by $w_{\langle q,u\rangle} = \frac{\hat{\sigma}^2_{C,\langle q,u\rangle}}{\hat{\sigma}^2_{O,\langle q,u\rangle} + \hat{\sigma}^2_{C,\langle q,u\rangle}}$.

Proof. With the record probability and variance estimates for each group fully computed, the blended estimate of $p_{\langle q,u\rangle}$ is given by $\hat{p}_{\langle q,u\rangle} = w_{\langle q,u\rangle} \cdot \hat{p}_{O,\langle q,u\rangle} + (1 - w_{\langle q,u\rangle}) \cdot \hat{p}_{C,\langle q,u\rangle}$. The sample variance of $\hat{p}_{\langle q,u\rangle}$ is given by $\hat{\sigma}^2_{\langle q,u\rangle} = w^2_{\langle q,u\rangle} \cdot \hat{\sigma}^2_{O,\langle q,u\rangle} + (1 - w_{\langle q,u\rangle})^2 \cdot \hat{\sigma}^2_{C,\langle q,u\rangle}$. Minimizing $\hat{\sigma}^2_{\langle q,u\rangle}$ with respect to $w_{\langle q,u\rangle}$ yields the stated result. □