Publicação Diferencialmente Privada de Dados de Pacientes de COVID-19

Manuel E. B. Filho[1], **Eduardo R. Duarte Neto** [1], **Javam C. Machado** [1]

[1]Laboratório de Sistemas e Banco de Dados (LSBD)
DC/UFC – UFC – CEP 60440-900 – Fortaleza – CE – Brazil

{edvar.filho,eduardo.rodrigues,javam.machado}@lsbd.ufc.br

Abstract. *The pandemic of the new corona virus (COVID-19) has brought new challenges to health systems in almost every corner of the world, many of them overloaded. Data analysis has played a key role in combating the corona virus, guiding both health professionals and government officials in the strategies adopted. However, private information of individuals must be preserved, and a balance between privacy and utility must be achieved. This work will demonstrate that it is possible to guarantee the privacy of infected patients and maintain the usefulness of the data, allowing an analysis on them with quality.*

Resumo. *A pandemia do novo corona vírus (COVID-19) trouxe novos desafios aos sistemas de saúde por quase todos os cantos do mundo, muitos deles sobrecarregados. A análise de dados tem desempenhado um papel fundamental no combate ao corona vírus, guiando tanto profissionais de saúde, como autoridades governamentais nas estratégias adotadas. Entretanto, informações privadas dos indivíduos devem ser preservadas, e um equilíbrio entre privacidade e utilidade deve ser alcançado. Este trabalho irá demonstrar que é possível garantir a privacidade dos pacientes infectados e manter a utilidade dos dados, permitindo uma análise de qualidade sobre os mesmos.*

1. Introdução

A pandemia do novo Coronavírus (COVID-19) trouxe novos desafios aos sistemas de saúde pelo mundo. Organizações de saúde tem publicado dados para facilitar a realização de pesquisa e informar a população sobre a evolução da doença. A análise dos dados de pacientes identifica padrões que permitem dar suporte aos profissionais de saúde no tratamento dos seus pacientes, e também apoia as autoridades governamentais na implementação de medidas de controle dos efeitos da pandemia. Estas informações sobre os pacientes podem estar sob a custódia das próprias unidades de saúde em que foram atendidos ou de forma integrada entre as instituições de saúde [Haas et al. 2011]. Em virtude do combate a pandemia no Brasil, inúmeras instituições publicam dados sobre os pacientes com COVID-19 [SUS 2020], por exemplo, o Governo do Estado do Ceará.

Embora a análise dos dados seja uma técnica fundamental, ela deve respeitar a *privacidade* dos pacientes. Registros de saúde contém, com frequência, informações sensíveis dos pacientes, *e.g.,* nomes, CPFs, endereço e particularmente o acometimento de doenças. O acesso indevido a essas informações dos pacientes tem gerado discussão sobre como garantir a análise desses dados de forma privada [Machado et al. 2019]. Diversas técnicas de privacidade e anonimização de dados já foram propostas [SWEENEY 2002,

Erlingsson et al. 2014, Dwork et al. 2006], dentre elas a Privacidade Diferencial (PD), que fornece sólidas garantias de privacidade, e sua implementação se dá por meio de um algoritmo aleatório, denominado mecanismo. A escolha do mecanismo e os ajustes dos parâmetros de privacidade são fundamentais para se alcançar o maior equilíbrio entre privacidade e utilidade dos dados e dependem fortemente da consulta a ser realizada.

Nesse trabalho investigamos o impacto da privacidade diferencial na publicação de dados da COVID-19 sobre a utilidade dos dados. Procuramos entender a utilidade da resposta de consultas numéricas sobre esses dados quando são dadas respostas aproximadas decorrentes da aleatoriedade introduzida pelo mecanismo de Laplace e verificar qual o nível de privacidade que tem uma alta utilidade quando comparamos com a resposta original das consultas. Experimentos exaustivos são descritos na Seção 3, que indicam quanto alcançamos de privacidade a medida que calibramos seu *budget* na execução da consulta. A Seção 4 apresenta nossas considerações finais.

2. Privacidade de Dados

A Privacidade Diferencial é uma técnica de preservação de privacidade bastante popular na análise de dados [Dwork et al. 2006, McSherry and Talwar 2007, Dwork 2008, Lecuyer et al. 2019]. Sua definição de privacidade é independente do conhecimento prévio dos adversários, indivíduos que procuram obter as informações sensíveis dos donos dos dados. A PD assegura que a distribuição da resposta de um mecanismo em uma consulta ao conjunto de dados não é afetada pela ausência ou presença de um indivíduo no mesmo, sendo assim, o conhecimento a priori sobre os dados não é alterado pela consulta.

Definição 2.1 (Privacidade Diferencial) *Um mecanismo M é ϵ-diferencialmente privado (DP) se para quaisquer conjuntos de dados D_1 e D_2 que se diferenciam em no máximo um elemento, e para todo conjunto S de todas as possíveis saídas de M,*

$$Pr[M(D_1) \in S] \leq \exp(\epsilon) \times Pr[M(D_2) \in S] \tag{1}$$

Um mecanismo DP garante que a remoção ou adição de um indivíduo no *dataset*, não vai ter um impacto significativo na probabilidade da saída, estando limitada ao ϵ, normalmente chamado de *budget*. Ele é responsável pela adição do ruído aleatório controlado à resposta da consulta a fim de garantir ϵ-*Privacidade Diferencial*. A quantidade de ruído necessária é calculada em função da sensibilidade da consulta f aplicada sobre o conjunto de dados D. Podemos definir a sensibilidade de uma consulta f como sendo:

$$\Delta f = \max_{D_1, D_2 \in D} \parallel f(D_1) - f(D_2) \parallel_1,$$

para todo D_1, D_2 diferindo em no máximo um elemento [Dwork et al. 2006].

O mecanismo de Laplace tem sido amplamente utilizado em abordagens de privacidade diferencial, particularmente em consulta sobre dados numéricos que retornam valores agregados de contagem. Dada uma função $f : D \rightarrow \Re$, o mecanismo de Laplace $M_f(D) = f(D) + Laplace(0, \Delta f/\epsilon)$ fornece ϵ-Privacidade Diferencial, onde $Laplace(0, \Delta f/\epsilon)$ retorna uma variável aleatória da distribuição de Laplace com média zero e escala $\Delta f/\epsilon$. A ideia por trás deste mecanismo é adicionar à resposta da consulta um ruído aleatório que irá garantir que um adversário não seja capaz de extrair conhecimento que lhe permita identificar indivíduos dentro do conjunto de dados. Por exemplo,

considere uma consulta sobre a quantidade de indivíduos com mais de 50 anos em um conjunto de dados, onde a resposta exata seria 20. Ao aplicar o mecanismo de Laplace, é calculado um ruído r a ser adicionado à resposta da consulta. Assim, a resposta retornada pelo mecanismo é dada por $20 \pm r$.

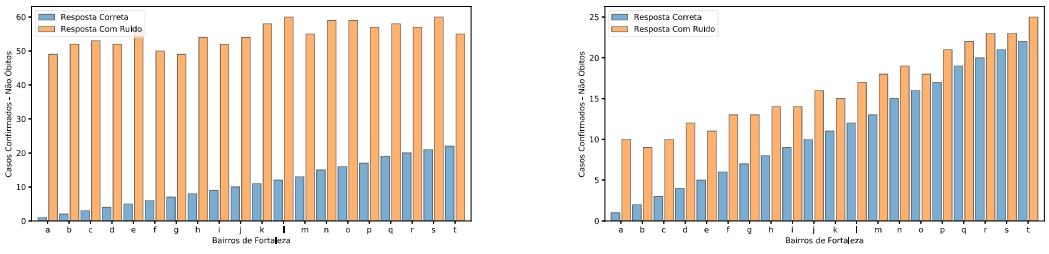
3. Anonimização de Dados de Saúde via Privacidade Diferencial

A plataforma Integra SUS [SUS 2020] contém dados coletados de pacientes de diversas instituições da rede de saúde do Estado do Ceará. Embora estes dados tenham passado por um processo de anonimização prévia com a supressão de alguns atributos identificadores [Fung et al. 2010], informações como a localização de residência dos pacientes, acometimento de comorbidades, e diferentes datas continuam disponíveis, e estão sujeitas a ataques que levam a exposição de informações sensíveis dos indivíduos [Narayanan and Shmatikov 2006]. Suprimir tais atributos impediria uma análise necessária sobre a população de pacientes. Muitas análises são realizadas através de consultas de agregação de dados numéricos, como por exemplo, o total de pessoas infectadas na cidade de Fortaleza. Dessa forma a adição do ruído de Laplace parece ser a estratégia mais indicada para garantir uma análise útil sem perda privacidade.

Para demonstrar a eficácia do mecanismo adicionamos o ruído de Laplace sobre a resposta da consulta: "Total de pacientes vivos infectados pelo COVID-19 por bairro de Fortaleza-CE" aplicada ao dataset sobre COVID-19 do Estado do Ceará, coletado até 12 de junho de 2020. Nesta data o dataset tinha 237.853 registros e, após limpeza dos dados ficou com 199.225 registros. Avaliamos o ajuste do *budget* ϵ no equilíbrio entre privacidade e utilidade, quanto menor o *budget*, maior é a garantia de privacidade, e menor a utilidade da resposta. Medimos isso através do erro quadrático médio entre a resposta original e a reposta com ruído. Utilizamos valores de *budget*, $\epsilon = [0.01, 0.05, 0.1, ln(2)]$. Os gráficos que mostram os resultados experimentais contêm histogramas com o número de casos positivos de COVID-19 reportados por bairro. Em azul temos a resposta correta e em amarelo a resposta adicionada do ruído. Para facilitar a apresentação, mostramos os resultados em três grupos de bairros que representam os números de casos positivos.

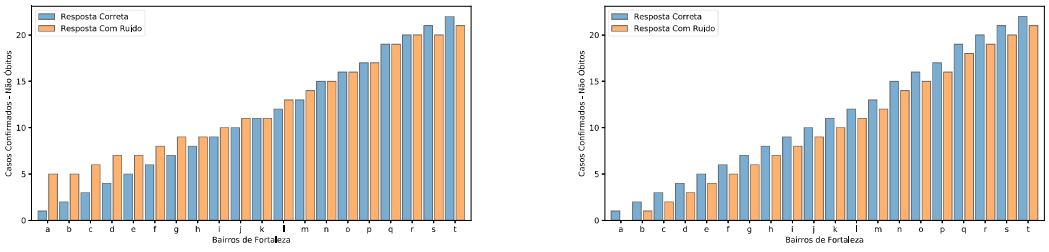
A Figura 1 apresenta os resultados da consulta sobre os 20 bairros com menor número de ocorrências, com um intervalo entre 1 e 22 casos. Com um número de casos tão baixos, estes são os bairros mais sujeitos a expor seus moradores. Podemos observar pela Figura 1(a) que o ruído necessário para garantir a privacidade é bem elevado quando o ϵ é muito baixo. Isso é um resultado direto do mecanismo de privacidade, que para proteger os indivíduos, modifica muito o resultado da consulta com a finalidade de confundir um adversário. Nestes casos, o ajuste do ϵ é essencial. Na Figura 1(c), cujo valor aplicado do ϵ é de 0.1, é possível obter um bom equilíbrio entre privacidade e utilidade. A Figura 4(a) apresenta o erro quadrático médio nos bairros com menos ocorrência. Como podemos observar o erro já tem uma redução muito grande para um $\epsilon = 0.05$.

A Figura 2 apresenta os resultados da consulta sobre os 20 bairros com número de ocorrências mediano, com um intervalo reportado entre aproximadamente 25 e 59 casos. Nas consultas sobre estes bairros obtemos um bom nível de utilidade dos dados com um valor de ϵ próximo e acima de 0.5, ou seja, a quantidade de ruído necessário para garantir a privacidade dos pacientes nestes casos é baixa. Observe a Figura 2(a). O valor de $\epsilon = 0.01$ ainda exige uma quantidade de ruído alta para os bairros com menos



(a) $\epsilon = 0.01$

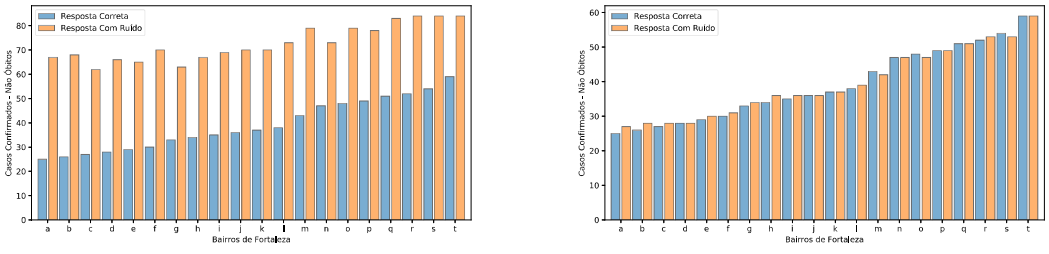
(b) $\epsilon = 0.05$



(c) $\epsilon = 0.1$

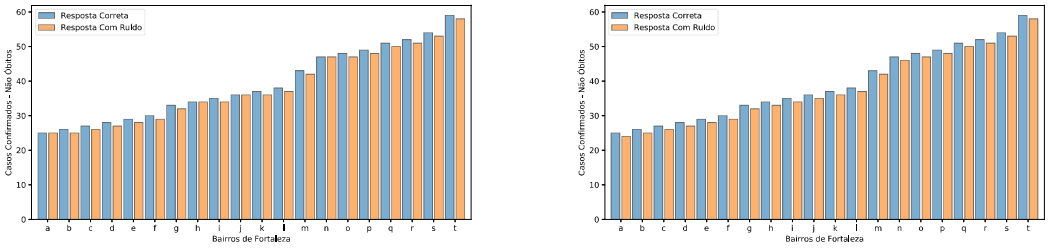
(d) $\epsilon = ln(2)$

Figura 1. Número de casos positivos nos 20 bairros com menos ocorrências de Fortaleza - CE



(a) $\epsilon = 0.01$

(b) $\epsilon = 0.05$



(c) $\epsilon = 0.1$

(d) $\epsilon = ln(2)$

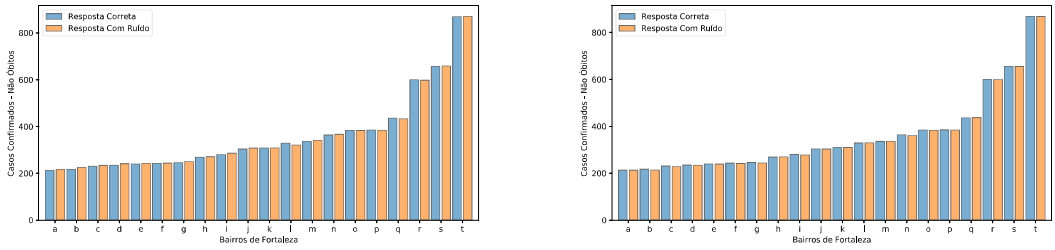
Figura 2. Número de casos positivos nos 20 bairros com ocorrências medianas de Fortaleza - CE

ocorrências da série, entretanto, quando comparados aos bairros da Figura 1(a), o ganho de utilidade já é significativo.

Nos 20 bairros com maior número de ocorrências, com valores variando entre 250 e 850 casos e cujos resultados são mostrados da Figura 3, o ruído necessário para garantir a privacidade dos pacientes é extremamente baixo, como podemos observar na

Figura 3(a), onde mesmo para um *budget* de 0.01 a resposta da consulta original é muito próximo da consulta com ruído. A Figura 4(c) demonstra esse comportamento, onde mesmo para $\epsilon = 0.01$ o erro já é muito baixo, e se aproxima muito de zero a medida que aumentamos esse *budget*.

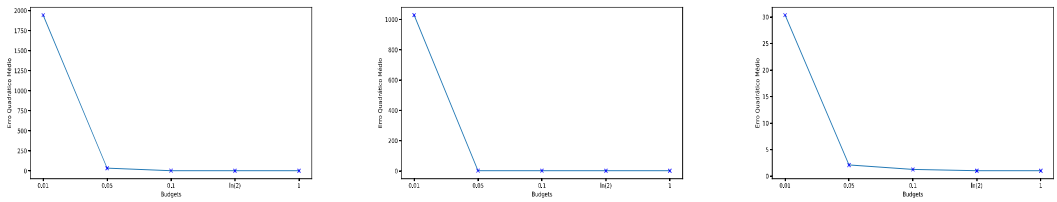
Analisando apenas a Figura 4 fica evidente que a escolha do *budget* é essencial para se obter uma boa utilidade dos dados. Quanto maior o ϵ, mais relaxadas estão as restrições para a garantia da privacidade diferencial, diminuindo a necessidade de se adicionar o ruído de Laplace na resposta das consultas. Nos três cenários observados, um $\epsilon = 0.05$ apresentou um erro quadrático médio baixo, sendo suficiente para garantir a privacidade e a utilidade dos dados. Este é o valor de budget indicado para a consulta aqui descrita, quando aplicada ao dataset sobre COVID-19 publicado pelo Estado do Ceará.



(a) $\epsilon = 0.01$ **(b)** $\epsilon = 0.05$

(c) $\epsilon = 0.1$ **(d)** $\epsilon = ln(2)$

Figura 3. Número de casos positivos nos 20 bairros com mais ocorrências de Fortaleza - CE



(a) Menores Ocorrências. **(b) Ocorrências medianas.** **(c) Maiores Ocorrências.**

Figura 4. Erro quadrático médio

4. Conclusão

A análise dos dados dos pacientes é fundamental no combate à pandemia de COVID-19, tanto para ajudar no tratamento, como para conter o avanço da doença. A privacidade é um

direito de cada cidadão e, portanto, deve ser observada. Através da aplicação de técnicas de privacidade, como a Privacidade Diferencial é possível garantir uma análise privada, sem que a qualidade da análise seja questionada, e nem que os dados dos pacientes sejam expostos. O mecanismo de Laplace quando aplicado em consultas de agregação sobre dados numéricos é capaz de atender com eficiência, sem perda de utilidade significativa, as restrições impostas pelas definições de privacidade diferencial, mesmo quando o *budget* ϵ é muito baixo. A consulta realizada neste trabalho é um exemplo dessa garantia, pois obtemos uma boa utilidade da resposta à consulta adicionada de ruído, com uma variação do *budget* entre $[0.05, 0.1]$.

Futuramente, pretendemos adicionar às nossas análises atributos categóricos, *e.g.,* tipo de comorbidade, faixa etária, gênero com a finalidade de realizarmos consultas mais complexas. Assim outras consultas e outros mecanismos de adição de ruído, como o mecanismo exponencial [McSherry and Talwar 2007], podem ser avaliados na busca de um bom equilíbrio entre privacidade e utilidade quando da publicação de dados de saúde.

Referências

Dwork, C. (2008). Differential privacy: A survey of results. In Agrawal, M., Du, D., Duan, Z., and Li, A., editors, *Theory and Applications of Models of Computation*, pages 1–19, Berlin, Heidelberg. Springer Berlin Heidelberg.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In Halevi, S. and Rabin, T., editors, *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg. Springer Berlin Heidelberg.

Erlingsson, Ú., Korolova, A., and Pihur, V. (2014). RAPPOR: randomized aggregatable privacy-preserving ordinal response. *CoRR*, abs/1407.6981.

Fung, B. C. M., Wang, K., Chen, R., and Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, 42(4):1–53.

Haas, S., Wohlgemuth, S., Echizen, I., Sonehara, N., and Müller, G. (2011). Aspects of privacy for electronic health records. *International journal of medical informatics*, 80(2):e26–e31.

Lecuyer, M., Atlidakis, V., Geambasu, R., Hsu, D., and Jana, S. (2019). Certified robustness to adversarial examples with differential privacy. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 656–672.

Machado, J. C., Neto, E. R. D., and Filho, M. E. B. (2019). Técnicas de privacidade de dados de localização. In *XXXIV SBBD, Fortaleza, CE, Brazil, October 7-10, 2019.*

McSherry, F. and Talwar, K. (2007). Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science*, pages 94–103.

Narayanan, A. and Shmatikov, V. (2006). How to break anonymity of the netflix prize dataset. *arXiv preprint cs/0610105*.

SUS (2020). Boletim epidemiológico novo coronavírus (covid-19). Acessado: 19-06-20.

SWEENEY, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570.