*Artificial Intelligence Research and Development*
*I. Sanz et al. (Eds.)*

*doi:10.3233/FAIA230693*

# Adding Edge Local Differential Privacy to the Dynamic Stochastic Block Model

Sudipta PAUL [a,1], Julián SALAS [b]and Vicenç TORRA [a]
[a] *Department of Computing Science, Umeå Universitet, Umea, Sweden*
[b] *Internet Interdisciplinary Institute,Universitat Oberta de Catalunya, Barcelona, Spain*

**Abstract.** In today's networked systems a massive amount of data is produced every day. These data can be modelled using graphs, where the nodes typically correspond to users or devices and the edges to the connections between them. Almost all networks change over time, with new nodes and edges appearing or disappearing as the system matures. Therefore dynamic graph models are more adequate to analyse such networks than static graphs, and appropriate tools need to be implemented to protect them. In this paper we obtain an edge differentially private version of dynamic stochastic block model. We show experimentally that the trends in the dynamic stochastic block model obtained from the original data are well preserved with the additional privacy guarantees.

**Keywords.** Local Differential Privacy, Dynamic Stochastic Block Model, Dynamic Graph, Social Network, Edge privacy

## 1. Introduction

A massive amount of data is produced every day in today's networked systems, such as social networks, biological networks, internet peer-to-peer networks, and other technical networks [1,2]. These data can be modelled using graphs, where the nodes typically correspond to users or systems and the edges to the connections between them. However, data changes with time, as well as the corresponding network models. Therefore static graphs are inadequate to represent such complex network architectures.

It is well recognized that naive anonymization of a graph can result in disclosure. Attackers can utilize their side-knowledge to infer private information of the graph. For example, through de-anonymization [3], degree [4], 1-neighbourhood [5], or sub-graph [6] attacks. For static graphs, appropriate privacy models have been devised and implemented, which can be broadly divided into two categories: those that adhere to $k$-anonymity [7] and those that adhere to differential privacy [8]. While using $k$-anonymity the existing solutions explore the neighbourhood of the nodes where the anonymised graph has at least $k-1$ nodes with the same degree for each node, or, the same neighbourhood, or, $k-1$ number of automorphic subgraphs etc. As for, differential privacy the approaches are mainly centered on triangle counting, smoothing the sensitivity, subgraph counting, clique counting or generating synthetic graphs [9,10,11]. Also, edge privacy is more extensively used than node privacy in real-world applications as it provides effi-

[1] Corresponding Author: Sudipta Paul, spaul@cs.umu.se

cient privacy protection, while keeping good data utility. Nevertheless, for the dynamic graphs, there is not so much literature on privacy models or solutions.

In this work, we focus on edge differential privacy for dynamic graphs. We apply the parallel mechanism [12] to guarantee edge differential privacy in the dynamic stochastic block model [13]. We carry out an empirical evaluation on the Enron dataset [14] to demonstrate how such algorithm preserves the edges' privacy while maintaining the utility of the dynamic stochastic block model.

## 2. Noise-graph Mechanism

In this section, we define the noise-graph mechanism [15] that we use to randomize the edges in the snapshot graphs an thus protect their privacy.

We denote by $G(V,E)$ the graph with the set of nodes $V$ and set of edges $E = E(G)$.

**Definition 2.1.** Let $G_1(V,E_1)$ and $G_2(V,E_2)$ be two graphs with the same set of nodes. Then the addition $G = G_1 \oplus G_2$ is the graph $G = (V,E)$ where $E = (E_1 \setminus E_2) \cup (E_2 \setminus E_1)$.

We denote by $G' \in \mathscr{G}(n,p)$, a random graph drawn from the *Gilbert model* (or the Erdös-Renyi model), in which there are $n$ nodes and each edge in $G'$ is chosen with probability $p$.

**Definition 2.2** (Noise-graph mechanism [15])**.** For any graph $G$ with $n$ nodes, and two probabilities $p_0$ and $p_1$, we define the following noise-graph mechanism:

$$\mathscr{A}_{p_0,p_1}(G) = G \oplus G_0 \oplus G_1,$$

where $E(G_0) = E(G') \setminus E(G)$ for $G' \in \mathscr{G}(n, 1-p_0)$ and $E(G_1) = E(G'') \cap E(G)$ for $G'' \in \mathscr{G}(n, 1-p_1)$.

## 3. Differential Privacy for Dynamic Graphs

In this section we provide the definitions of dynamic graph and local differential privacy applied specifically to edges in a dynamic graph. Then, we present the definition of the parallel protection mechanism and a condition so that these mechanism is $\varepsilon$-edge locally differentially private [12].

We follow the definition of a dynamic graph [16] which is a network observed at an initial state $G_0$ which is a graph, and at a set of $T$ further snapshots, evenly spaced at integer times $t = 1,\ldots,T$. These other graphs are denoted as $G_1, G_2, \ldots, G_T$. Note that it may be assumed that all the graphs have the same set of nodes, considering that $V(G_i) = \bigcup_{i=0,\ldots,T} V(G_i)$.

We denote by $\mathbb{1}_{uv(t)}$ the indicator function of edge $uv$ in $G_t$, that is $\mathbb{1}_{uv(t)} = 1$ if $uv \in E_t$, and $\mathbb{1}_{uv(t)} = 0$ otherwise. Similarly, $\mathbb{1}_{\mathscr{A}(uv(t))}$ is the indicator function of edge $uv$ in $\mathscr{A}(G_t)$, the randomized graph.

**Definition 3.1** (Edge-local differential privacy for dynamic graphs)**.** An edge randomization algorithm $\mathscr{A} : \mathscr{G} \to \mathscr{G}$, satisfies $\varepsilon$-edge local differential privacy (is $\varepsilon$-eLDP) if for every pair of nodes $u, v \in V$, any timestamp $t \in \{1, \dots T\}$ and $x, x', y \in \{0, 1\}$:

$$P(\mathbb{1}_{\mathscr{A}(uv(t))} = y \mid \mathbb{1}_{uv(t)} = x) \leq e^{\varepsilon} P(\mathbb{1}_{\mathscr{A}(uv(t))} = y \mid \mathbb{1}_{uv(t)} = x')$$

**Definition 3.2** (Parallel protection mechanism)**.** Let $G = G_0, G_1, \dots, G_T$ be a dynamic graph. Then, we define the parallel protection of the dynamic graph with parameters $p_0$ and $p_1$ as the protection process that provides $\tilde{G} = \tilde{G}_0, \tilde{G}_1, \dots, \tilde{G}_T$ with $\tilde{G}_i = \mathscr{A}_{p_0, p_1}(G_i)$ for $i = 0, \dots, T$. We denote the parallel protection of a dynamic graph $G$ with parameters $p_0$ and $p_1$ as $\mathscr{A}^{||}_{p_0, p_1}(G)$.

**Remark 3.1.** The mechanism $\mathscr{A}^{||}_{p_0, p_1}$ is $\varepsilon$-eLDP if $e^{\varepsilon} \geq \max\left\{\frac{1-p_1}{p_0}, \frac{p_1}{1-p_0}, \frac{p_0}{1-p_1}, \frac{1-p_0}{p_1}\right\}$.

## 4. Experimental Results

We perform an experiment on a dynamic social network constructed from the Enron corpus [14], which consists of about 0,5 million email messages between 184 Enron employees from 1998 to 2002, with the same pre-processing as Xu and Hero [13]. In addition to the email data, the roles of most of the employees within the company (directors, CEOs, presidents, vice-presidents, managers, traders, and others) are available.

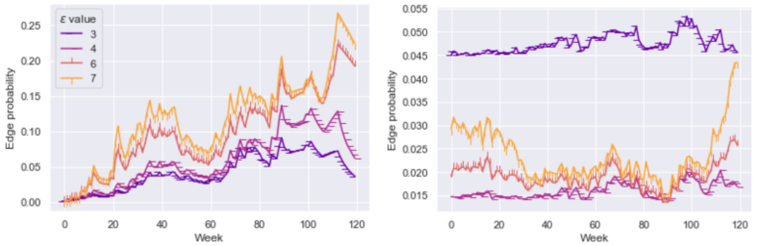
We fit the Dynamic-Stochastic-Block-Model [13] to the data protected with differential privacy for $\varepsilon$ values from 1 to 10. We note that for $\varepsilon \geq 7$ the estimated probabilities are the same as without any protection, and for $\varepsilon < 3$ the probabilities are almost constant, for clarity we included only few values in Fig. 1.

Note that in Fig. 1 the trends of the original data are preserved for large enough $\varepsilon$ and flattened for smaller values. Still, it can be observed in Fig. 1a the increase in edge probabilities from Enron CEOs to presidents as Enron's financial situation worsened, while in Fig. 1b other employees edge probabilities remain at their baseline levels until Enron fell under federal investigation. More specifically, Fig. 1a shows that the effects of CEO Skilling resignation in week 89 and the resignation of CEO Lay in week 111 are preserved in the dynamic stochastic block model with differential privacy, since the two most prominent peaks in the edge probabilities are present in the protected data.

## 5. Conclusions

In this work, we applied the parallel mechanism to guarantee edge differential privacy in the dynamic stochastic block model. We carried out an experimental evaluation on the Enron dataset, and observed that the trends in the edge probabilities obtained through the dynamic stochastic block model are well preserved for $\varepsilon$ values larger than 3. This shows that it is possible to analyse dynamic network data with the added privacy guarantees.

It remains as future work to carry out a more detailed evaluation of the differentially private mechanisms for dynamic networks proposed, applying them to different models and inference procedures to study the dynamics of other time-evolving networks.



(a) CEOs to presidents

(b) Others to others

**Figure 1.** Estimated edge probabilities obtained by fitting the Dynamic Stochastic Block Model to the protected data. Note that for $\varepsilon = 7$ (yellow line) the estimated probabilities are the same as those of the Dynamic Stochastic Block Model without any protection.

## References

[1] Backstrom L, Huttenlocher D, Kleinberg J, Lan X. Group formation in large social networks: membership, growth, and evolution. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining; 2006. p. 44-54.

[2] Asharov G, Demmler D, Schapira M, Schneider T, Segev G, Shenker S, et al. Privacy-preserving interdomain routing at internet scale. Cryptology ePrint Archive. 2017.

[3] Takbiri N, Shao X, Gao L, Pishro-Nik H. Improving privacy in graphs through node addition. In: 2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton). IEEE; 2019. p. 487-94.

[4] Pedarsani P, Grossglauser M. On the privacy of anonymized networks. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining; 2011. p. 1235-43.

[5] Zhou B, Pei J. Preserving privacy in social networks against neighborhood attacks. In: 2008 IEEE 24th International Conference on Data Engineering. IEEE; 2008. p. 506-15.

[6] Hay M, Miklau G, Jensen D, Towsley D, Weis P. Resisting structural re-identification in anonymized social networks. Proceedings of the VLDB Endowment. 2008;1(1):102-14.

[7] Samarati P. Protecting respondents identities in microdata release. IEEE transactions on Knowledge and Data Engineering. 2001;13(6):1010-27.

[8] Dwork C. Differential privacy: A survey of results. In: Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 25-29, 2008. Proceedings 5. Springer Berlin Heidelberg; 2008. p. 1-19.

[9] Jiang H, Pei J, Yu D, Yu J, Gong B, Cheng X. Applications of differential privacy in social network analysis: A survey. IEEE Transactions on Knowledge and Data Engineering. 2021;35(1):108-27.

[10] Hay M, Li C, Miklau G, Jensen D. Accurate estimation of the degree distribution of private networks. In: 2009 Ninth IEEE International Conference on Data Mining. IEEE; 2009. p. 169-78.

[11] Task C, Clifton C. What should we protect? Defining differential privacy for social network analysis. State of the Art Applications of Social Network Analysis. 2014:139-61.

[12] Paul S, Salas J, Torra V. Edge local differential privacy for dynamic graphs. In: Security and Privacy in Social Networks and Big Data. Springer; 2023. *to Appear*.

[13] Xu KS, Hero AO. Dynamic stochastic blockmodels for time-evolving social networks. IEEE Journal of Selected Topics in Signal Processing. 2014;8(4):552-62.

[14] Priebe CE, Conroy JM, Marchette DJ, Park Y. Scan statistics on Enron graphs. Computational & Mathematical Organization Theory. 2005;11:229-47.

[15] Salas J, González-Zelaya V, Torra V, Megías D. Differentially private graph publishing through noise-graph addition. In: International Conference on Modeling Decisions for Artificial Intelligence. Springer; 2023. p. 253-64.

[16] Zhang X, Moore C, Newman ME. Random graph models for dynamic networks. The European Physical Journal B. 2017;90(10):1-14.