Offre de thèse/Thesis Offer

Title: Robustness and privacy of graph neural networks: homomorphic encryption and randomization

Keywords: Graph Neural Networks, Adversarial attacks, Graph classification, Robustness, Randomized algorithms, Functional encryption

Publication Date: 15 March 2019
Beginning Date: October-November 2019
Duration: 36 months
Contact: cedric.gouy-pailler@cea.fr and renaud.sirdey@cea.fr

Context

In various domains, graphs represent a useful representation for many types of data. Prominent examples entail behavioural analyses performed in cybersecurity or social network analysis. In the former, user internet behaviour can be observed by monitoring DNS requests, interpreted as successive steps of a random walker on a graph in which nodes represent domain names and edges represent population-level average behaviour. Therefore studying user behaviour can be done by analyzing the subgraph induced by specific user movements. In the latter, graph representations naturally emerge from user interactions. For example nodes can represent users, and any relation between two users (messages or common interests) can be interpreted as edges. Understanding and analyzing graph structures appear to be a key tool in many real-world applications. It is thus essential to find efficient and robust methods for tasks such as node or graph classification.

In the last decade, deep neural networks have reached an outstanding level of accuracy in numerous areas, such as image classification [KSH12] or object detection [RHCS15]. These models have also recently been formulated in the context of graph-structured data, and now play an important role in node classification and graph classification problems [SGT+09, DDS16, KW16, HYL17, WPC+19]. Such formulations are currently explored in many domains, jointly exploiting classical features, as well as graph-structured information. Successful applications have been developed in physics, in which graph neural networks are able to predict physical properties of molecules based on their molecular graphs [CBG+17]. Graph neural networks also have convincing applications in material sciences [XG18], structural fore-
casting [LYSL17], natural language processing [MT17], or communication optimization in multi-agent systems [SSF16].

Despite these powerful representation properties, recent issues have been demonstrated in deep learning-based approaches regarding their robustness to adversarial attacks (small perturbations of the input) and the resulting training data privacy (due to overfitting and over-parameterization posing threats on data privacy). Adversarial attacks [SZS13] are small perturbations of an input that fools the results of classification for a network. Adversarial attacks raise questions of security and safety, and also responsibility in terms of law. Adversarial examples attacks against machine learning models have become a burning issue due to their efficiency, and the number of sensitive domains they could affect. Accordingly, both attacks and defenses are developed in a tight back-and-forth [GSS14, PMJ16, PMG17, DLT18, SKC18]. Recently, the idea of using randomization in the learning process to ensure robustness against adversarial examples attacks have been successfully used [XWZ17, MDST18, LCZH17, PMA19]. In the context of graph neural networks, these issues are of primary importance for various reasons: first graph-structured data often bear much more information than classical tabular data. For example in social networks, tabular based approaches classically summarize neighbourhood information in a few variables (number of friends/degree, node/edge betweenness, ...), while the topological information structurally bear much more sensitive information about individuals. Second the complexity of information present in graphs makes the graph neural network approaches much more sensitive to attacks. Basically this comes from the fact that small perturbations in graph data consist in adding/deleting nodes, or modifying edges weight. While adversarial modifications in images and sounds might be globally noticeable, graph-based perturbations will be hard to detect. Coincidentally with the fact that graph neural network approaches have shown superior results in various domains, their robustness have been investigated and attacks have been developed in the context of node classification and graph classification [ZAG18, DLT18, SWYL18].

In this context it is of primary importance to develop innovative approaches to ensure privacy and robustness of graph neural networks. Among possible approaches, lightweight approaches such as randomization will have to be adapted to these techniques. Depending on the criticity of the stage and needed privacy of data and models (learning or inference phases), randomization techniques should be complemented by homomorphic encryption approaches. With respect to privacy, a number of works have started to investigate how techniques for computing over encrypted data such as homomorphic cryptography (FHE) can be applied to the inference phase of deep neural network models with encouraging results when a clear-domain network is evaluated over an encrypted-domain input [BMMP18, CLM19, CdWM17, DGBL16]. Yet there are a number of practically interesting extensions most notably with respect to GNN regarding specific optimizations that may render them more amenable to better FHE-execution performances. Also investigating the relevance and practicality of using these techniques during the learning phase of such models is of high practical interest. On top of privacy, the connection between cryptographic theory and techniques and counter-measures against the aforementioned adversarial attacks is an another important research topic which can be considered as part of this PhD subject. The goal of this thesis is to explore robustness and privacy of graph-neural network-based approaches, by considering solutions combining randomization and homomor-
phic encryption to ensure a satisfying compromise between performance, robustness and data privacy.

CEA background in these fields

CEA LIST has been a key leader in fully homomorphic encryption techniques. In the context of FHE, machine learning applications appear as a killer application. Many key advances have yet to be considered to fully address machine learning applications using FHE technologies. Next technological barriers depend on the computational cost of the considered stage (training or inference) but the main approaches are: first to limit operators used in graph neural networks such that FHE associated computational cost is kept reasonable. Second FHE can be viewed as a building block, which could be activated in specific parts of the pipeline to ensure model or data privacy.

CEA LIST is also very active in the field of randomization algorithms to ensure data privacy and robustness to adversarial attacks. Past works include PhD thesis of Anne Morvan and Rafael Pinot.

Expected work

- Experimental study of state-of-the-art attacks.
- Theoretical approach of defenses.
- Corresponding implementation, and experimentation of defenses.

Required profile

- You are currently in the final year of engineering school or in M2 at the university with specialization in computer science and/or statistics.
- You have a strong background in applied mathematics/computer science (probability, statistics, graph theory).
- You have good programming skills (Python/R, torch/tensorflow, C++).
- Academic research interests you, but also applications to concrete problems.

Conditions

The doctoral will take place at CEA LIST in Saclay, where you will work with Renaud Sirdey and Cédric Gouy-Pailler. Access to the CEA is based on a daily bus network service covering the entire Paris region (several buses from Paris in particular).

References


