**Title:** DGCNN: A Convolutional Neural Network over Large-scale Labeled Graphs

**Abstract**

Exploiting graph-structured data has many real applications in domains including natural language semantics, programming language processing, and malware analysis. A variety of methods has been developed to deal with such data. However, learning graphs of large-scale, varying shapes and sizes is big challenges for any method. In this paper, we propose a multi-view multi-layer convolutional neural network on labeled directed graphs (DGCNN), in which convolutional filters are designed flexibly to adapt to dynamic structures of local regions inside graphs. The advantages of DGCNN are that we do not need to align vertices between graphs, and that DGCNN can process large-scale dynamic graphs which have up to greater than 157,000 nodes. To verify the effectiveness of DGCNN, we conducted experiments on two tasks: malware analysis and software defect prediction. The results show that DGCNN outperforms the baselines, including several deep neural networks.

**Key words**:

Labeled Directed Graphs; Convolutional Neural Networks (CNNs); Control Flow Graphs (CFGs); Abstract Syntax Trees (ASTs)

**Comments to editors:**

Dear Editors,

I am writing to submit our manuscript entitled: DGCNN: A Convolutional Neural Network over Large-scale Labeled Graphs, which is an improved and extended version of the paper: Convolutional Neural Networks over Control Flow Graphs for Software Defect Prediction, presented at International Conference on Tools with Artificial Intelligence (ICTAI) 2017, for the consideration of publication in Engineering Applications of Artificial Intelligence. All the necessary documents can be accessed at https://github.com/nguyenlab/DGCNN,

Thank you very much for receiving our manuscript and considering it for review. We appreciate your time and look forward to your response.

Sincerely,

Associate Professor Minh-Le Nguyen