# Causal QA with GNNs for text answer QA tasks

**Anonymous ACL submission**

## Abstract

In recent years, there has been a lot of progress in NLP with the development of language models (LMs) such as BERT. On the one hand, these LMs have been applied in the question-answering tasks, both multiple-choice questions answering tasks (CommonsenseQA, OpenBookQA) and text answer tasks (using the SQuAD dataset). On the other hand, knowledge graphs have been applied to question answering tasks too by obtaining embeddings that could be queried for generating answers for the multiple-choice QA tasks. Recently, Yasunaga et. al. proposed QA-GNN which combines GNN's networks on Knowledge graphs to extract knowledge for multiple-choice QA downstream tasks and this method achieves SOTA. We examine the performance of the vanilla QA-GNN on text answer tasks using the SQuAD dataset and propose a solution to use QA-GNN to achieve relevant results on these tasks.

## 1 Introduction

Last decade has witnessed a lot of progress on the language models (LMs). Pre-trained language models such as BERT have changed the approach to most NLP tasks. However, these models work based on the associations found in the training data, which is prone to spurious correlations. Recently, some researchers have tried to incorporate the causality point of view to improve the overall performance of the NLP tasks [1]. In the question-answering (QA) tasks, this goal has been persuaded mainly by incorporating the knowledge graph to reduce the chance of spurious correlations [2, 3]. Although this approach has been successful, they have been applied to the format of multiple-choice answering datasets. Applying these methods to other QA tasks requires modification of the existing datasets. We propose a method to apply the knowledge graph for text answer QA tasks using the SQuAD dataset. The datasets we will use for the project are the following:

- *Knowledge bases:*
  - CONCEPTNet
  - Freebase
  - Comet
- *QA dataset*
  - SQuAD
- Possible research directions:
  - Our research project will aim to accomplish two possible goals:
  - Apply QA GNN to other QA formats such as SQuAD If necessary (topics of SQuAD are distant from knowledge in KBs), we can build a graph using COMET and the context of SQuAD to generate KG for SQuAD task.

## 2 Relevant literature

The language models provide high coverage of the language. With the development of the transformer architecture and higher computation power, new formulations have been researched to provide question answering systems with relevant knowledge. In particular, Bosselut et. al. [4] developed COMET (COMmonsEnse Transformers) which is a framework to generate rich and diverse commonsense descriptions when receiving tuples of phrases with relations, ultimately developing a framework able to generate commonsense knowledge tuples. Furthermore, Petroni et. al. [5] analyzed the relational knowledge present in large transformer models, like BERT, before fine-tuning and after-tuning. The results suggest that large language models have the ability to retain a significant amount of relational knowledge and could be a good source for unsupervised QA systems. However, extraction information beyond the training corps has proven to be a challenging task. In particular, Kassner and Schütze [6] showed that the same models analyzed by Petroni et. al. showed severe difficulties in distinguishing between negated sentences and 'misprimed' sentences. In particular, the authors were able to reformulate Question Answering tasks as sentences to predict tokens, thus having the BERT models analyzed by Petroni et. al. readily available for testing the results on negated and misprimed sentences.

On the other hand, knowledge graphs can be used to provide inference and provide a variety of pieces of knowledge that could be used for a wide gamut of question-answering tasks. In particular, knowledge graphs such as Freebase, by Bollacker et. al. [7] and ConceptNet by Speer et. al. [8] are graphs condensing information from various different sources in a way where entities are represented as nodes and relations as edges. In particular, Ren et. al.[9] and Ren and Leskovec [10] were able to show that using embedding-based frameworks it is possible to obtain reasoning over arbitrary queries with AND, OR and EXISTS operators in knowledge graphs. However, by nature, the Knowledge Graphs from which these embeddings are learned are not complete, and as shown by Guu et. al. [11], knowledge graphs can have the issue that an error in the graph can cause cascading errors in the knowledge extracted from a graph.

Combining language models and knowledge graphs (KGs) has been a topic of various papers in recent years. Bao et al (2016)[12], introduced the idea of constraint-based QA with a knowledge graph. In their method, each query was transformed into a multi-constrained query graph. The query graph uses a bag of words as the feature of each graph node. A graph convolutional network is used to generate the response to the query. Sun et al (2018)[13] propose Graphs of Relations Among Facts and Text Networks (GRAFT-Net) to operate over KG and text sentences. Each node representation is initialized using an LSTM over the document in which the specific text is extracted (Sun et al. 2018). Similarly, Wang et al (2019a)[14] suggest an architecture with two subgraphs one graph capturing the premise, and another one modelling the hypothesis. An LSTM will provide the concept embedding which is used in a graph attention network. In the end, the outcome of the graph model and text model are concatenated and passed through a neural network. In an influential work, Lin et al. (2019)[15] suggested the idea of Knowledge-Aware Graph Network (KagNet) as the core of QA reasoning. A nondirectional graph is extracted from the KG, by ignoring the labels and direction on the edge, and becomes an input to graph attention The input to the KagNet is the path between concepts extracted from the KG. An LSTM-based path encoder is used on top of the graph neural net to capture the relational information. At the time, they were able to achieve state-of-the-art performance on the CommonsenseQA by using ConceptNet as the only external resource for BERT-based models. Feng et al (2020)[3] tried to convert the KagNet solution to a scalable scheme. Their suggestion, multi-hop graph relation network (MHGRN) performed multi-rational multi-hop reasoning over the extracted subgraphs from the external knowledge graphs. Just recently, Yasunaga et al (2021)[2] presented the latest effort on combining large KGs with a pre-trained language model in a method called: QA-GNN. In the QA-GNN, the LM output is added to the extracted graph as the context node. A graph attention network is used to model the representation of the joint graph. In the end, the final prediction is made by using LM representation, the context node, and an extracted graph representation. Currently, QA-GNN has the highest score on common reasoning tasks such as the CommonsenseQA.