Project

Rehan Ahmed

# Project Domain & Goals

I plan to create a knowledge graph for cryptocurrencies that will include information about their creation history, their prices over the years as structured data and unstructured data including news articles, social media (reddit & twitter).

The goal of the project is to analyze the impact of certain news or events on the price of a cryptocurrency. The main use of the knowledge graph will be to give a better insight on what impacts certain events have on a particular cryptocurrency and, more importantly, what level of impact. For example, Tesla’s recent decision to buy $1.5 billion of bitcoin pushed the price significantly up. Similarly, a subreddit like wallstreetbets was instrumental in the GameStop short squeeze and I want to extract similar information but for cryptocurrency prices. The goal is not prediction, but to develop a relation to better understand crypto performance.

I plan to create visualizations that will allow users to easily see the change in price over time, overlayed with news and social media posts (including stats like sentiment, interest, etc.)

# Datasets

The knowledge graph will include structured data of Cryptocurrency prices from services like [CoinBase](https://developers.coinbase.com/api/v2#api-key) and [CoinMarketCap](https://coinmarketcap.com/api/) which provide free APIs.

It will also include a lot of unstructured data about cryptocurrencies from microblogging sites like Twitter and [StockTwits](https://stocktwits.com/symbol/BTC.X). I also plan to use Reddit subreddits like [r/CryptoCurrency](https://www.reddit.com/r/CryptoCurrency/) and several others. Furthermore, I plan to scrape news articles about cryptocurrencies from a host of different sources (for e.g. [Investing.com](https://www.investing.com/news/cryptocurrency-news) ). I will perform information extraction on all this unstructured data to extract information like Sentiment, interest and use it to evaluate the performance of a cryptocurrency.

# Technical Challenge

There are a number of technical challenges, I have listed them down below,

* Information extraction from unstructured social media posts – there is a lot of noisy data i.e. (memes, sarcastic posts), which need to be ignored for the algorithm to extract only meaningful analysis/sentiment about a certain cryptocurrency
* Designing scrapers for a wide range of news sources
* Storing all the data, structured as well as unstructured, in an RDF model so it can be easily queried using SPARQL
* Visualizing an interactive linear timeline of a crypto’s prices along with news data and other important stats extracted from unstructured data using VisJs to allow a user to easily extract insights