HW2)

Q1)

Skip-gram

We see that the results are more consistent across different configurations, the closest word is always dogma. This makes sense and seems correct.

At larger window sizes, irrelevant words are included in the context and hence are considered similar to the work of interest, eg. window = 50, word 'apredtneatcstorontoedu' is considered similar to 'faith'. However is purely an artefact of our corpus.

The larger the embedding size, the more complex the model and the longer the time to generate the embeddings, however assuming we have a sufficiently large corpus, larger embedding size can mean better overall performance of the final model.

CBOW:

Cbow does a better job in that no irrelevent words appear even on larger window sizes, CBOW also performs significantly faster than skip gram.

however overall we see that skip gram gives us more consistent results across various parameter sizes. This is because skip-grams gives more weightage to words that are closer to the word of interest.

Link to Code and results:

https://github.com/sabnisshreyas91/sss_msia414_2019/blob/hw2/HW2.ipynb

Q2)

	Word2Vec	**Bert**
# Google scholar citations	15135	1731
Year of publication	2013	2018
Training Corpus	Google dataset ~ 1 billion words	English Wikipedia (2.5 Billion words) & TheBooks corpus (11000 books consisting of 800 million words)
Objective	Predict neighboring words based on winFdow size. Word order does not matter	Based on context, predict masked words. Word order matters