

Answer 7

```
Harshith Guru Prasad@hp MINGW64 ~/Desktop/NLP_HW5
$ python q7.py
Group               Best-1 Acc      Best-5 Acc      Best-10 Acc
capital             0.083          0.583           0.750
currency            0.100          0.100           0.100
city-in-state       0.333          0.611           0.833
family              0.705          0.910           0.968
adjective-to-adverb 0.011          0.122           0.222
comparative         0.533          0.762           0.800
superlative         0.429          0.762           0.810
nationality-adjective 0.453          0.744           0.872
```

From the table of best-n accuracies, it is evident that some group of relations are predicted more accurately, while some others are predicted less accurately using the dense vectors for the cosine similarity comparison metric.

Best 1 Accuracies:

family > comparative > nationality-adjective > superlative > city-in-state > currency > capital > adj-to-adverb

Best 5 Accuracies:

family > nationality-adjective > comparative = superlative > city-in-state > capital > adj-to-adverb > currency

Best 10 Accuracies:

family > nationality-adjective > city-in-state > superlative > comparative > capital > adj-to-adverb > currency

Conclusions

currency has a poor accuracy of 10%. Only 1 of the analogies were correctly predicted while the other analogies failed to produce semantically meaningful words that satisfied the semantics of the analogy for best-1,5,10 results. Currencies like 'won', 'real' tend to have different semantic meanings while dollar is used only in a financial context. The words 'won' is mostly used as a verb while 'real' is used as adjectives or adverbs. Hence currencies has poor prediction accuracies based on the given vectors.

adjective to adverb performs poorly as well with a top 10 vector accuracy of 22.22%. Adjectives and adverbs can be used in diverse contexts and hence don't provide good prediction accuracies. Need larger datasets from which better context scores or vector embeddings can be derived and used to compute similarity metrics.

family performs extremely well and predicts the missing word in the analogy with a best-10 accuracy of 96.795. Family tops the accuracy charts for best-1,5,10 (Family comes first always). Excellent predictions are on account of frequently used pronouns and common nouns that are often used in the same context as each other. Family relations are close in the semantic graph and hence have a high prediction accuracy.

city-in-state, comparative, superlative, nationality-adjective perform well as well and produce a best-10 prediction accuracy of 80-87%. Comparative and superlative forms of English words have essentially the same stem(lemma). They are often used in the same context with a varying degree of intensity. Hence, they have good prediction accuracies. City and state are all locations and hence have good accuracies on account of their limited context. Similarly, nationality and adjective have a common stem(lemma) and used in a similar context. Hence their relevance to each other in most contexts accounts for good accuracies.

Word predictions ultimately depends on the context scores / word embeddings of each word relative to the other words in a huge corpus. The prediction accuracy depends on the representation of the various contexts in which a word has a semantically related meaning.

Examples of Incorrect (Best-1) and Correct Predictions for Each Relation Group

1. capital

```
madrid : spain :: york : england
madrid : spain :: orleans : england
madrid : spain :: jersey : england
madrid : spain :: london : england
madrid : spain :: bronx : england
madrid : spain :: mexico : england
madrid : spain :: seattle : england
madrid : spain :: baghdad : england
madrid : spain :: jail : england
madrid : spain :: queens : england
```

```
paris : france :: london : england
paris : france :: york : england
paris : france :: orleans : england
paris : france :: seattle : england
paris : france :: chicago : england
paris : france :: jersey : england
paris : france :: n.y. : england
paris : france :: ma : england
paris : france :: bronx : england
paris : france :: boston : england
```

2. currency

```
korea : won :: china : dollar
korea : won :: fuel : dollar
korea : won :: iran : dollar
korea : won :: russia : dollar
korea : won :: it : dollar
korea : won :: oil : dollar
korea : won :: economy : dollar
korea : won :: nuclear : dollar
korea : won :: chinese : dollar
korea : won :: wal-mart : dollar
```

```
brazil : real :: canada : dollar
brazil : real :: japan : dollar
brazil : real :: china : dollar
brazil : real :: france : dollar
brazil : real :: europe : dollar
brazil : real :: russia : dollar
brazil : real :: quarter : dollar
brazil : real :: nation : dollar
brazil : real :: increase : dollar
brazil : real :: oil : dollar
```

3. city-in-state

```
houston : texas :: philadelphia : florida
houston : texas :: seattle : florida
houston : texas :: atlanta : florida
houston : texas :: miami : florida
houston : texas :: boston : florida
houston : texas :: dallas : florida
houston : texas :: chicago : florida
houston : texas :: detroit : florida
houston : texas :: ninth : florida
houston : texas :: calif. : florida
```

```
seattle : washington :: miami : florida
seattle : washington :: atlanta : florida
seattle : washington :: dallas : florida
seattle : washington :: kentucky : florida
seattle : washington :: philadelphia : florida
seattle : washington :: detroit : florida
seattle : washington :: liverpool : florida
seattle : washington :: houston : florida
seattle : washington :: woods : florida
seattle : washington :: las : florida
```

4. family

```
father : mother :: brother : wife
father : mother :: son : wife
father : mother :: husband : wife
father : mother :: girlfriend : wife
father : mother :: uncle : wife
father : mother :: grandfather : wife
father : mother :: sister : wife
father : mother :: cousin : wife
father : mother :: daughter : wife
father : mother :: boyfriend : wife
```

```
grandfather : grandmother :: husband : wife
grandfather : grandmother :: son : wife
grandfather : grandmother :: father : wife
grandfather : grandmother :: brother : wife
grandfather : grandmother :: sister : wife
grandfather : grandmother :: daughter : wife
grandfather : grandmother :: girlfriend : wife
grandfather : grandmother :: uncle : wife
grandfather : grandmother :: cousin : wife
grandfather : grandmother :: mother : wife
```


5. adjective – adverb

```
quick : quickly :: usually : typically
quick : quickly :: plus : typically
quick : quickly :: rare : typically
quick : quickly :: regular : typically
quick : quickly :: typical : typically
quick : quickly :: excellent : typically
quick : quickly :: required : typically
quick : quickly :: fancy : typically
quick : quickly :: minimum : typically
quick : quickly :: relatively : typically
```

```
rare : rarely :: typical : typically
rare : rarely :: limited : typically
rare : rarely :: large : typically
rare : rarely :: unusual : typically
rare : rarely :: small : typically
rare : rarely :: common : typically
rare : rarely :: expensive : typically
rare : rarely :: available : typically
rare : rarely :: significant : typically
rare : rarely :: included : typically
```

6.comparative

```
old : older :: itself : stronger
old : older :: within : stronger
old : older :: natural : stronger
old : older :: historic : stronger
old : older :: structure : stronger
old : older :: happening : stronger
old : older :: deep : stronger
old : older :: ocean : stronger
old : older :: tower : stronger
old : older :: whole : stronger
```

```
small : smaller :: strong : stronger
small : smaller :: weak : stronger
small : smaller :: huge : stronger
small : smaller :: large : stronger
small : smaller :: deep : stronger
small : smaller :: little : stronger
small : smaller :: simple : stronger
small : smaller :: healthy : stronger
small : smaller :: dangerous : stronger
small : smaller :: big : stronger
```

7. superlative

```
great : greatest :: huge : largest
great : greatest :: biggest : largest
great : greatest :: large : largest
great : greatest :: vast : largest
great : greatest :: big : largest
great : greatest :: capital : largest
great : greatest :: giant : largest
great : greatest :: nonprofit : largest
great : greatest :: corporation : largest
great : greatest :: small : largest
```

```
high : highest :: large : largest
high : highest :: small : largest
high : highest :: biggest : largest
high : highest :: huge : largest
high : highest :: giant : largest
high : highest :: industrial : largest
high : highest :: vast : largest
high : highest :: main : largest
high : highest :: big : largest
high : highest :: asian : largest
```

8. nationality-adjective

```
russia : russian :: china : indian
russia : russian :: europe : indian
russia : russian :: india : indian
russia : russian :: brazil : indian
russia : russian :: japan : indian
russia : russian :: germany : indian
russia : russian :: region : indian
russia : russian :: iran : indian
russia : russian :: britain : indian
russia : russian :: france : indian
```

```
france : french :: india : indian
france : french :: brazil : indian
france : french :: britain : indian
france : french :: ukraine : indian
france : french :: japan : indian
france : french :: germany : indian
france : french :: africa : indian
france : french :: europe : indian
france : french :: italy : indian
france : french :: canada : indian
```