Sangmin Woo
Dr. Anjalie Field
EN.601.472 NLP4CSS

HW1 Questions

1.

   d)

     i) The results from Part B are as follows:

```
More Republican
Word            Odds    R count    D count
--------------  --------  ---------  ---------
spending     -34.8438    40463      56799
obamacare    -30.988     13703      15294
government   -25.5483    67563     113791
trillion      -23.8218    18131      25215
taxes         -23.4738    20084      28717
going         -23.2429    93078     165287
administration -22.475     30833      48329
debt          -21.6624    29092      45709
that          -21.0222   836428    1685992
you           -20.996    156369     293367
```

```
More Democrat
Word            Odds    R count    D count
-----------  -------  ---------  ---------
republican   31.8057    12239      49158
women        22.6697    14424      46898
republicans  21.911     14028      45238
families     21.1251    19020      57190
communities  18.7554     8052      27362
gun          18.262      2080      10532
education    18.237     10970      34546
violence     17.9448     3707      15121
for          17.6597   403216     912669
proceed      17.1741     2978      12664
```

Evidently, Republicans tend to focus on words like spending, obamacare, government, and taxes along with other words shown in the table. On the other hand, Democrats use words like education, communities, families, women, and republican along with the other words shown in the table.

     ii) The results from Part C are as follows:

```
Changes over time in log odds with prior
Word            112       113       114
-----------  ---------  ---------  ----------
freedom      -6.97069  -7.00203  -7.07094
```

```
justice      8.67755   7.57113   6.98866
equality     5.35283   4.8281    5.1813
democracy    2.05073   2.90348   2.91989
abortion    -8.59186  -7.58392  -7.21186
immigration -2.68343  -2.31368  -0.971269
welfare     -1.82275  -1.8268   -2.1418
taxes       -28.549   -25.6366  -21.9471
democrat    -12.2592  -10.7022  -9.57163
republican   5.05622   9.61488  10.9884
```
The 2 words that have changed the most in usage are taxes and republican.

2.

a)

ii) The results of Part A are as follows:

Changes over time in log odds with prior

```
Word         112        113        114
--------  ----------  ----------  ---------
abortion  -4.9794     -1.26054    -0.24167
justice    0.0276949   0.0306654   0.483116
freedom   -1.84663    -0.899684   -0.975187
```

Abortion, was a mostly democratic topic, but it has quickly become a point of discussion important to both parties. On the other hand, justice is mostly talked about similarly across both parties in terms of healthcare. Freedom is somewhat of a term used mainly by Democrats. By focusing specifically on the topic of healthcare, we get to see which words are used by which parties in specific contexts, for example, justice in healthcare is very different compared to justice in law. Topics allow us to zoom in on word usage in specific debates.

3. a)

ii) We examine the nearest neighbors for democrats and republicans for the query 'taxes':

Republican near neighbors
tax 0.72828865
taxing 0.6082767
taxation 0.5939475
surtax 0.5514295
taxed 0.5315596
whammy 0.51430357
highincome 0.51288736
income 0.5055604
revenue 0.50201863
millionaires 0.49545932

Democrat near neighbors
tax 0.71427786
revenue 0.640787

taxing 0.61075485
taxation 0.60624295
taxed 0.60393935
revenues 0.6003616
pay 0.5764996
excise 0.57559955
paying 0.5415774
subchapter 0.53279096

While there exist some similarities, neighbors such as highincome and millionaire suggest that Republicans focus on how taxation affects wealthier constituents. On the other hand, neighbors such as pay, excise, and subchapter focus on taxation more in the way of policy.

c)

i) We examine the word similarities:

Political Words Similarities
freedom 0.81276953
justice 0.8045187
democracy 0.7845567
immigration 0.76778907
equality 0.73993695
abortion 0.7213996
taxes 0.690596
welfare 0.65670896
democrat 0.6179996
republican 0.5917814

To highlight some, we see that freedom, justice, and democracy are used most similarly among the keywords we studied.

ii) The change in polarization is shown below:

Polarization over time
Session 111=   0.6929959058761597
Session 112=   0.6881260275840759
Session 113=   0.6997675895690918
Session 114=   0.6725711822509766

The polarization is mostly consistent across all sessions, however we do see that there is somewhat of a downward trend, suggesting that the speeches are actually becoming less polarized.

iii) There are many limitations of using solely cosine similarity as our metric. For one, natural language is often very context dependent, which is not accounted for. Also, political language in particular can change and evolve greatly over time. This can be a result of various new phenomena (i.e. technology), social shifts, etc. This is also very

difficult to capture using just cosine similarity.