

# NYT News Text NLP Visualization

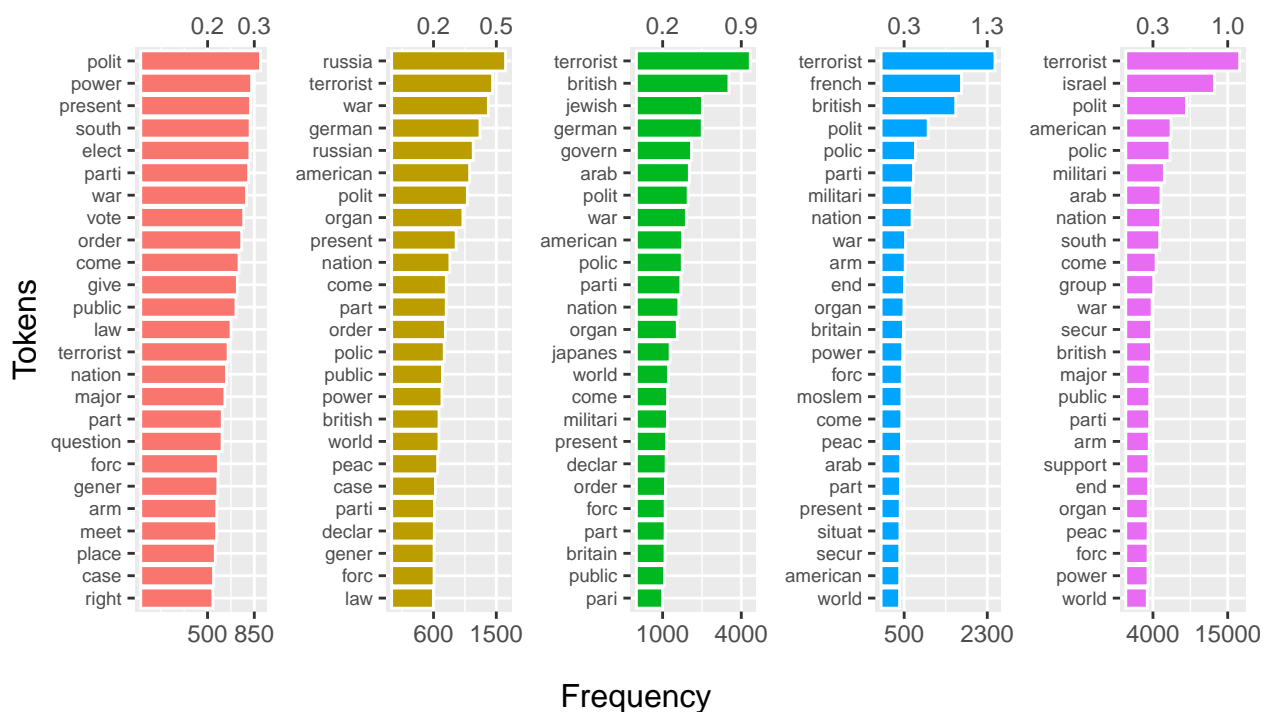
Yu-Hsuan Liu

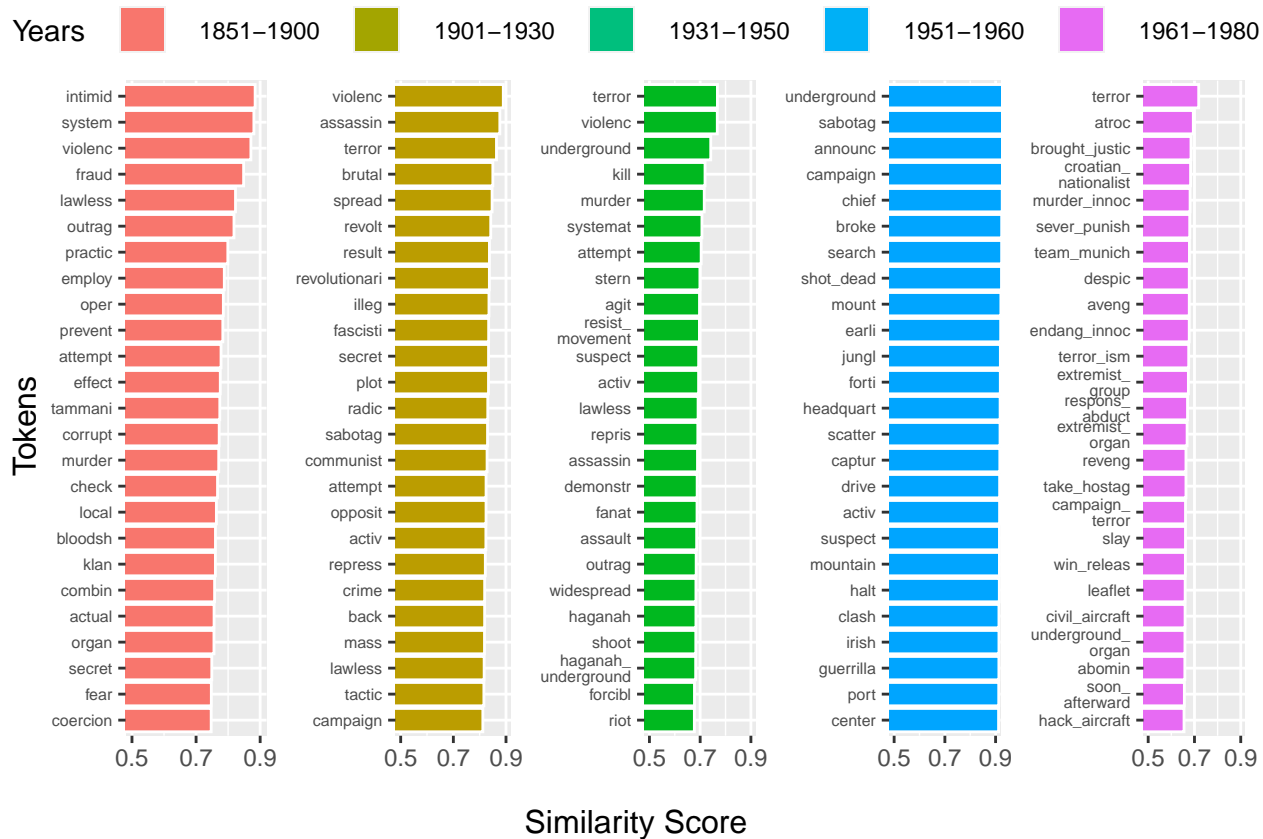
April 1st, 2022

Years	Number of Articles	Number of Tokens
1851-1900	594	272,458
1901-1930	1,127	295,846
1931-1950	2,816	468,335
1951-1960	1,501	190,501
1961-1980	8,878	1,583,450

Years ■ 1851-1900 ■ 1901-1930 ■ 1931-1950 ■ 1951-1960 ■ 1961-1980

Probability (%)





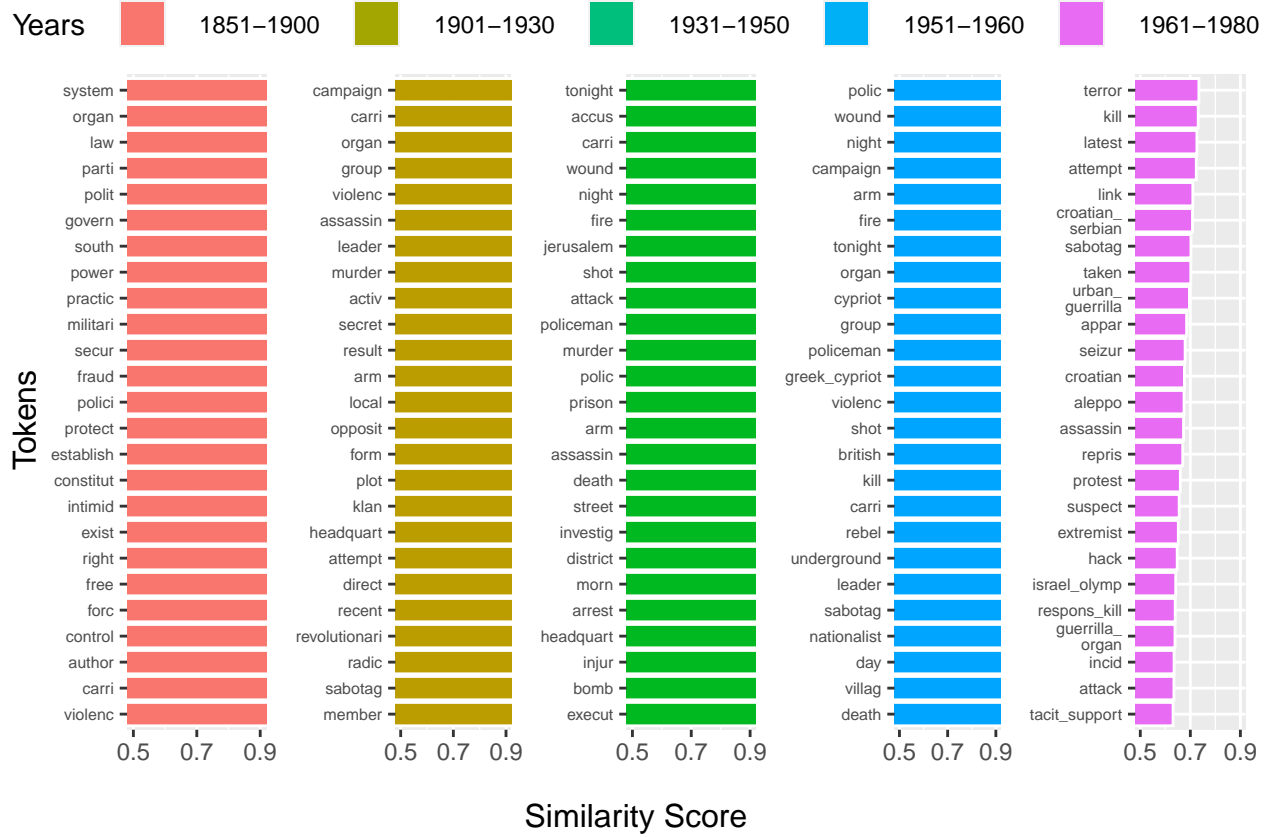


Table 2: Jaccard Similarity and TF-IDF Similarity of Top 100 Similar Words to "terrorist" Across the Five Historical Periods

Years	1851-1900	1901-1930	1931-1950	1951-1960	1961-1980
<b>Jaccard Similarity, SKNG Model</b>					
1851-1900	1.000				
1901-1930	0.099	1.000			
1931-1950	0.058	0.130	1.000		
1951-1960	0.026	0.058	0.099	1.000	
1961-1980	0.005	0.010	0.031	0.010	1
<b>TF-IDF Similarity, SKNG Model</b>					
1851-1900	1.000				
1901-1930	0.115	1.000			
1931-1950	0.063	0.144	1.000		
1951-1960	0.028	0.062	0.116	1.000	
1961-1980	0.002	0.007	0.037	0.009	1
<b>Jaccard Similarity, CBOW Model</b>					
1851-1900	1.000				
1901-1930	0.064	1.000			
1931-1950	0.020	0.136	1.000		
1951-1960	0.036	0.149	0.342	1.000	
1961-1980	0.000	0.058	0.053	0.070	1
<b>TF-IDF Similarity, CBOW Model</b>					
1851-1900	1.000				
1901-1930	0.081	1.000			

1931-1950	0.018	0.137	1.000		
1951-1960	0.040	0.151	0.371	1.000	
1961-1980		0.058	0.047	0.070	1

---

<sup>1</sup> TESTTESTTEST

```
## # A tibble: 270 x 5
## # Groups:   terrorist_stm [71]
##   terrorist_stm terrorist_Values_stm Years      count simi_score
##   <chr>          <dbl> <chr>      <int>      <dbl>
## 1 intimid      0.884 1851-1900 SKNG      4      0.922
## 2 violenc      0.871 1851-1900 SKNG      7      0.918
## 3 lawless      0.823 1851-1900 SKNG      4      0.832
## 4 oper         0.785 1851-1900 SKNG      3      0.893
## 5 attempt      0.778 1851-1900 SKNG      5      0.806
## 6 murder       0.771 1851-1900 SKNG      7      0.836
## 7 local        0.764 1851-1900 SKNG      3      0.919
## 8 klan         0.760 1851-1900 SKNG      4      0.887
## 9 organ        0.756 1851-1900 SKNG      5      0.921
## 10 secret      0.749 1851-1900 SKNG      5      0.844
## # ... with 260 more rows
```

Top 100 Similar Words to "terrorist" Counts Over Three Times  
within the Five Historical Periods(Summed from both SKNG & CBOW Models)

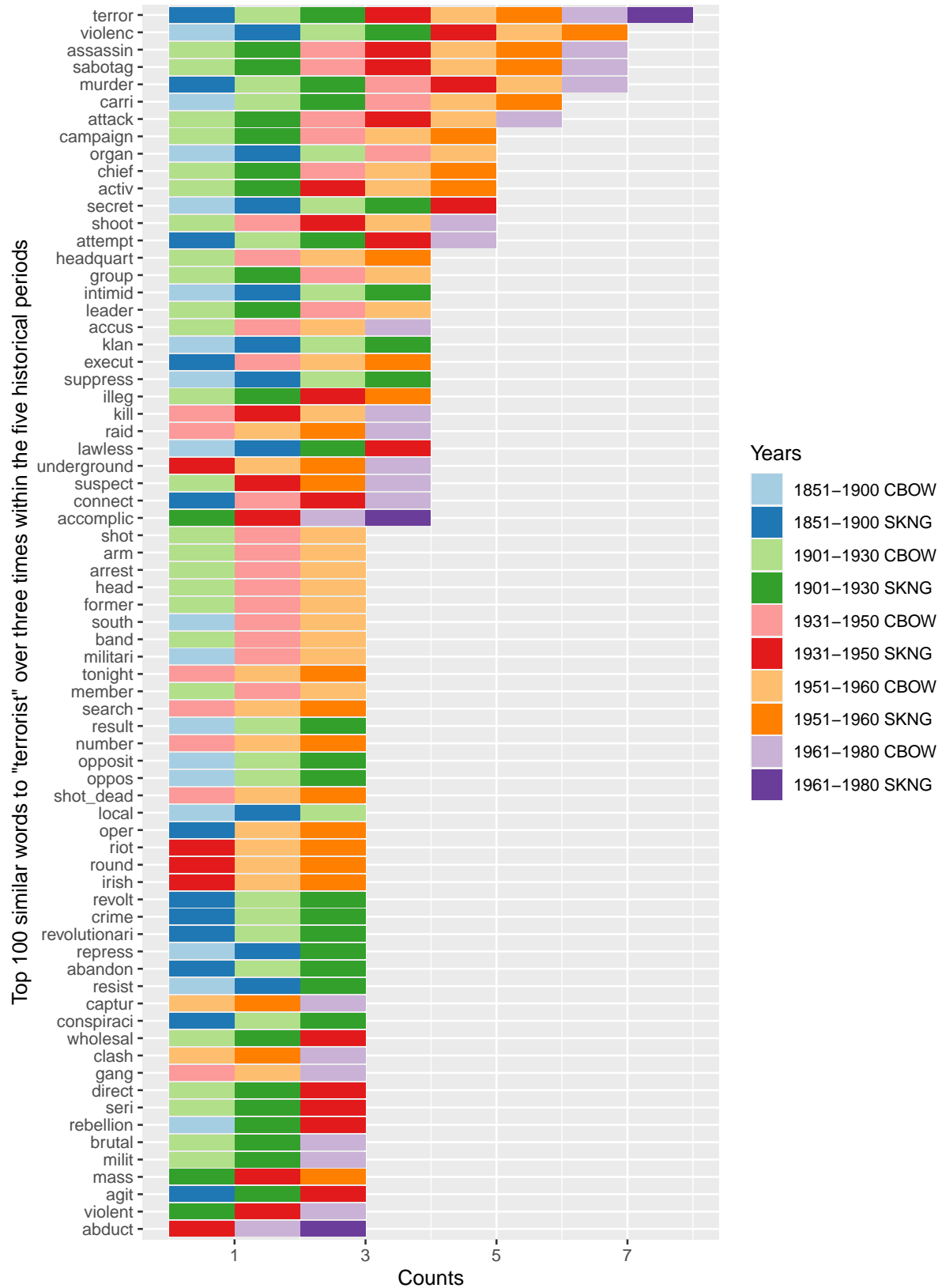
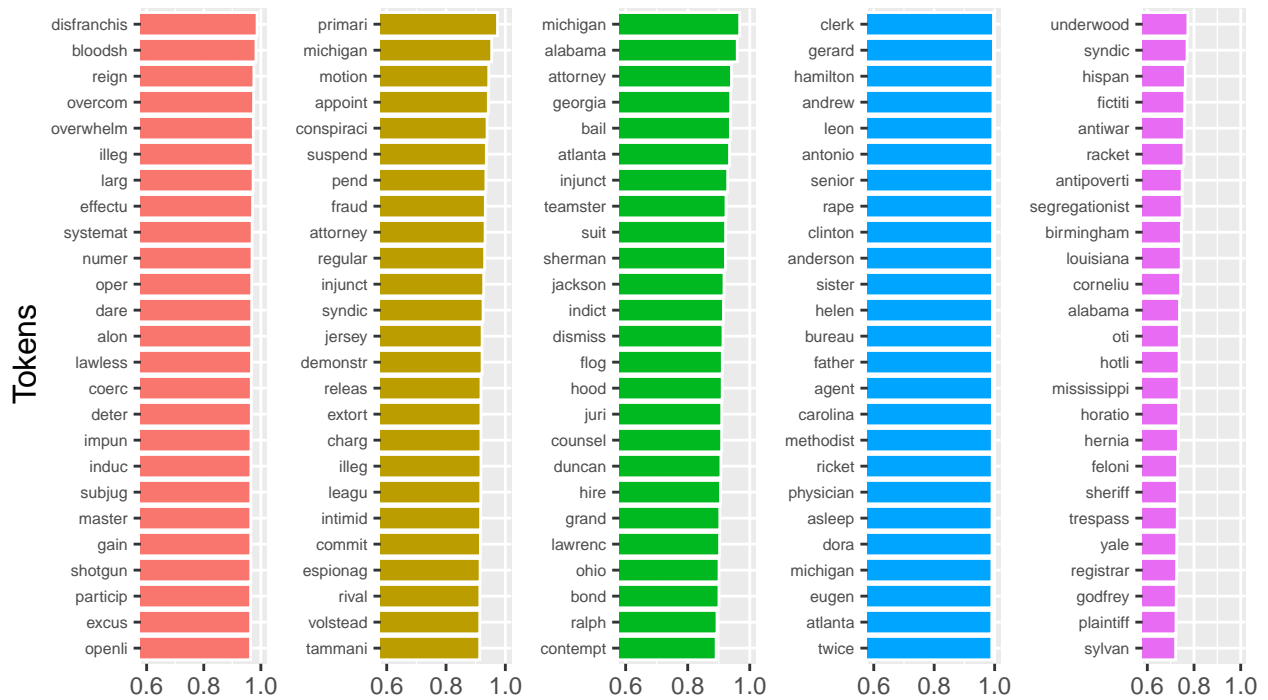


Table 3: Result of LDA Topic Modeling

Historical Periods	Rapoport (2001; 2004)	Hoffman (2006)	k	LDA Topic Words	Topic Distribution (%)
1851-1900	Anarchist Wave		1	polit, power, present, south, elect, parti, war, vote, order, come	100
1901-1930	Anarchist Wave	Government abusing power	3	terrorist, organ, polic, strike, american, polit, public, law, union, case	46.37
	Anti-Colonial Wave			terrorist, german, russian, polic, russia, organ, parti, polit, order, british	26.83
				russia, war, german, russian, american, nation, world, polit, peac, present	26.8
1931-1950	Anti-Colonial Wave	National liberation	3	german, terrorist, japanes, french, polic, govern, organ, pari, chines, union	99.02
		Self-determination waves		british, terrorist, jewish, arab, polic, jerusalem, militari, arm, tonight, night	0.49
		Revolutionary in WWII		polit, war, nation, american, world, parti, jewish, peac, british, terrorist	0.49
1951-1960	Anti-Colonial Wave	Liberation battles Nationalism and separatism	1	terrorist, french, british, polit, polic, parti, militari, nation, war, arm	100
1961-1980	New Left Wave	Destabalize the West	2	israel, polit, american, terrorist, nation, militari, world, support, arab, power	91.55
				terrorist, polic, israel, group, british, polit, organ, attack, secur, night	8.45

<sup>1</sup> TESTTESTTEST



Years      ■ 1851-1900      ■ 1901-1930      ■ 1931-1950      ■ 1951-1960      ■ 1961-1980

```
## null device
##           1
```