Using NLP to Explore ‘Love’ In TED Talks

By Sean Davern

Source Data

TED

kaggle

Filter, Clean

Tokenize

Corpus

Transcripts

Contain love+

Sentences with love+

2550

1132

1132 ‘Documents’
3160 Sentences

NLP Methods Used

Corpus

CV, Binary

Normalize

Evaluate Clusters

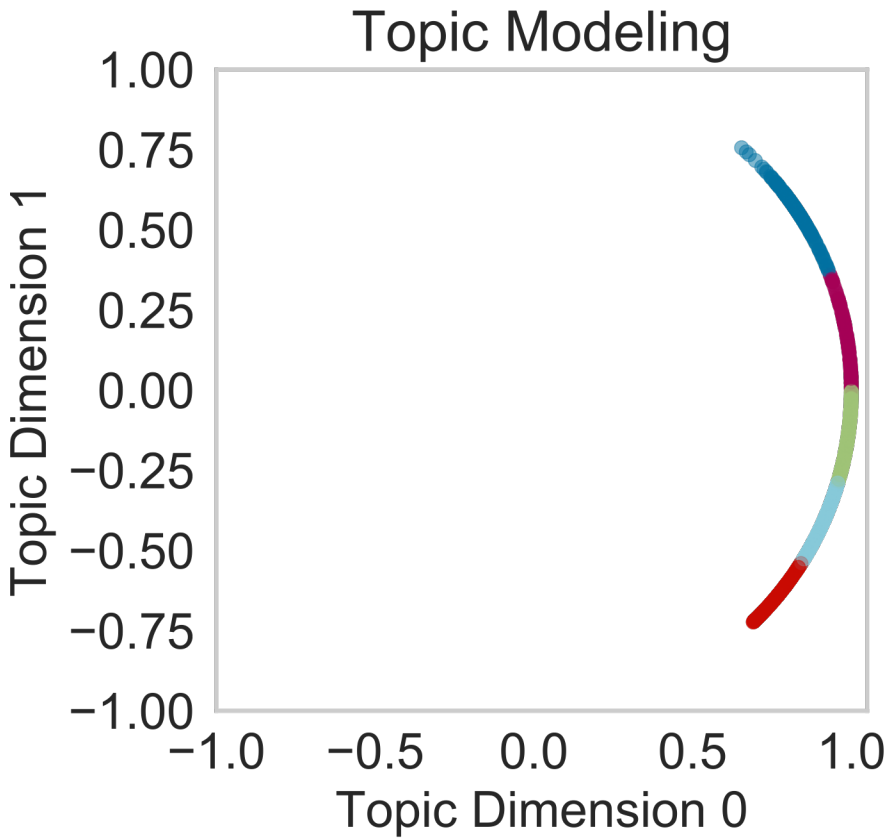
Vectorize → Reduce Dimensions → Cluster

CountVectorizer
TfidfVectorizer

LSA: 2-6 Topics

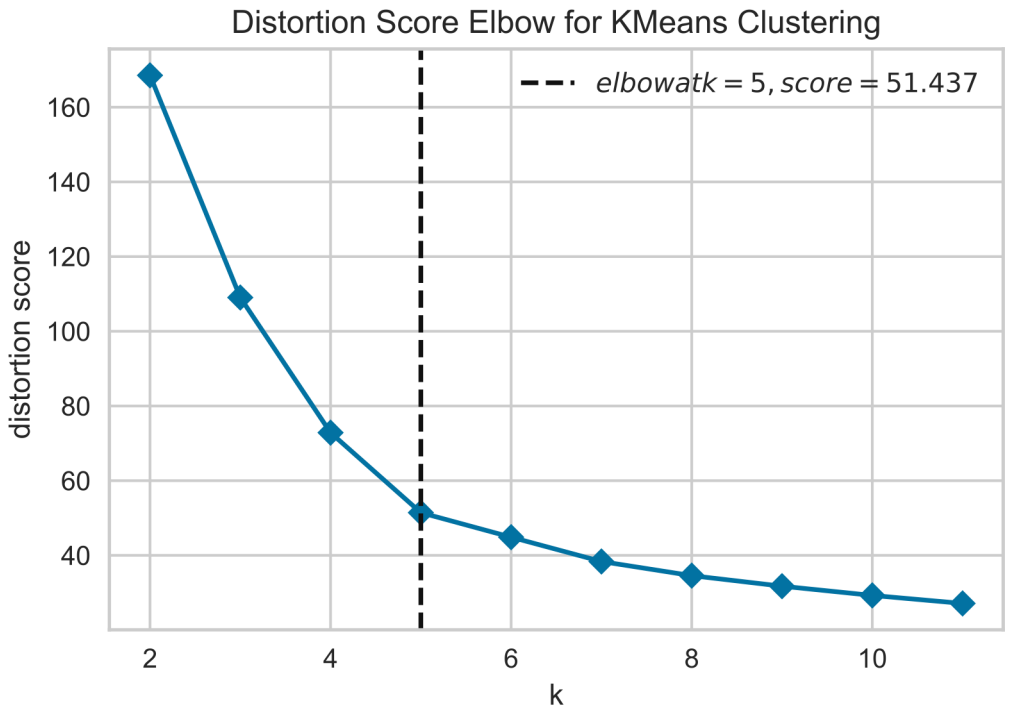
K-Means, DBScan, MeanShift
Hierarchical, Spectral

Normalize & Cluster

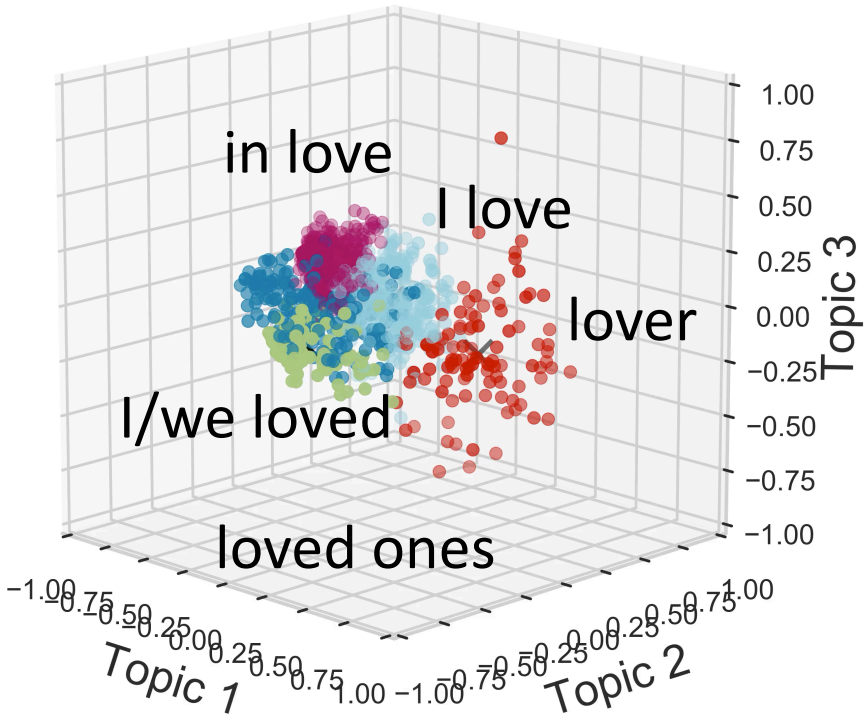


- Normalizing (projecting onto the unit circle) covers a relatively small arc
- Dimension 0 can effectively be neglected (visually)

4-Topic Model



Topic Modeling Clusters



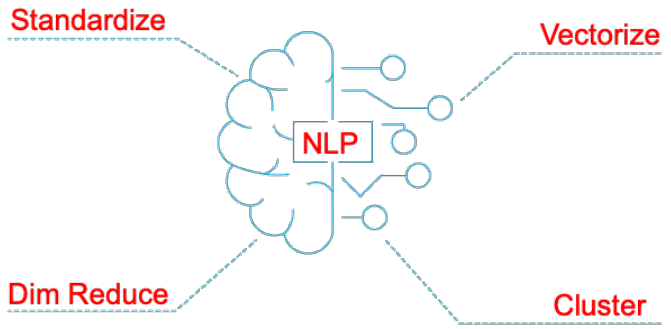
Conclusions & Future Work

- NLP was able to identify distinguishable clusters
- These clusters weren’t particularly insightful regarding exploring how authors talked about love.
- Tools to mathematically validate the distinguishability of clusters in higher dimensions would be helpful!

Questions?

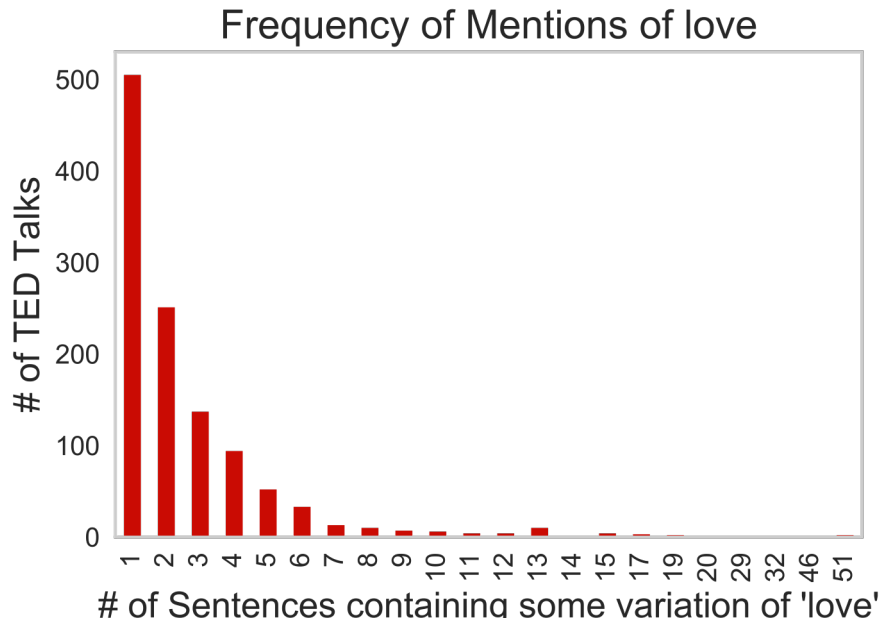
TED
Ideas worth spreading

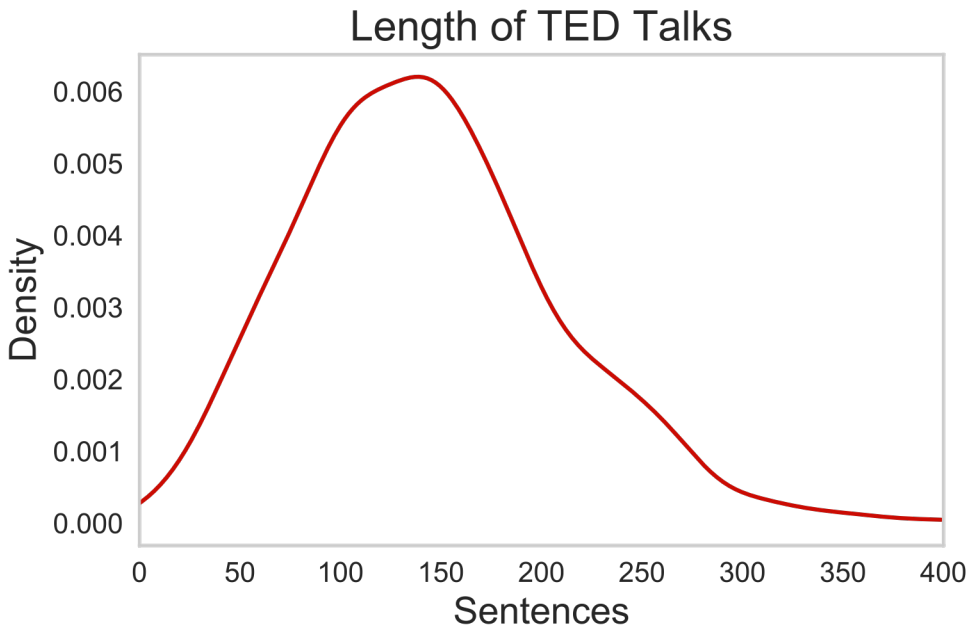
+



APPENDIX

Love Content





Cluster Sense Making

- (green): (obj) “in love”, “an expression of love”, I love (1/2)
- (mag): (present tense feeling) I love, his real love, - I love you, a lot of love in these passwords – expect to find love (3-5)
- (cyan): (past tense feeling) I/we loved (3+)
- (blue): loved ones – loved it – loved the sound – we loved (1)
- (red): (n., obj.) plant lover – gave them love, met the love of their life, all this love,– not my lover, music lovers, jazz lovers (5+)