

Boston Globe 2018:

1. According to Word2Vec, 'bullying' and 'unaccountable' were the 4th and 5th most similar words to 'black'.
2. Doc2Vec found that 'young' is the 3rd most similar word to 'black'.
3. Also according to Doc2Vec, 'disputing' and 'autistic' were the 4th and 5th most similar words to 'jamaican'.
4. Word2Vec found that 'emigrated' is the 3rd most similar word to 'dominican' and the 2nd most similar word to 'republic'.
5. Word2Vec also found that 'outcountry', 'oozing' and 'duality' are the 1st, 2nd and 4th most similar words to 'mano guay' (a Spanish slang).
6. Another potentially interesting find from Word2Vec is that 'Sumbul' (a Muslim name) and 'Mogadishu' (the capital of Somalia) are the 1st and 2nd most similar words to 'emigrated'.
7. It was interesting to see that when using the Gensim Phrases package to find the most common phrases from the list of formatted documents, and train the Word2Vec model on those, the most similar words to 'black' were just other races. This potentially hints at a lack of bias, perhaps, seeing as other races seem to be used in similar contexts, most notably since 'white' seems to be the most similar to 'black' out of all other races. Perhaps, in the future, other races could be excluded to uncover implicit bias towards the black community.
8. Yet another interesting find using the Gensim Phrases package trained Word2Vec model were the least similar words. However, here, some of the least similar words to 'black' were
 - a. 'regulations'
 - b. 'corp'
 - c. 'leased'
 - d. 'transportation secretary'
 - e. 'renovations'
 - f. 'regulatory commission'

Each of these words hint at potential domains where the black community isn't adequately represented, at least in the media.