

Episode Jaccard Similarity

Mapping South Park

by Fred

Step 1: Collect Underpants

The Data

We found our data on a post from [some generous soul on Kaggle](#). We hoped to find some distinct set of text which was sufficiently binned. The set has

- Season
- Episode
- Character
- Lines

We present one grouping, Episode by Season, and their similarities.

Step 2: ???

Jaccard Simliarity

$$J = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Below we map the Jaccard Similarity across episodes of South Park in two different ways. We first map episodes with stop words left in then proceed to re run our analysis with stop words removed.

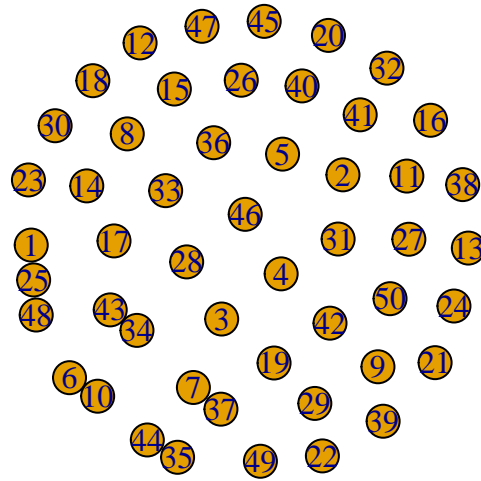
A note on both analyses

We used a `doc_proportion_max` of 0.05

From the documentation: “maximum proportion of documents which should contain term.”

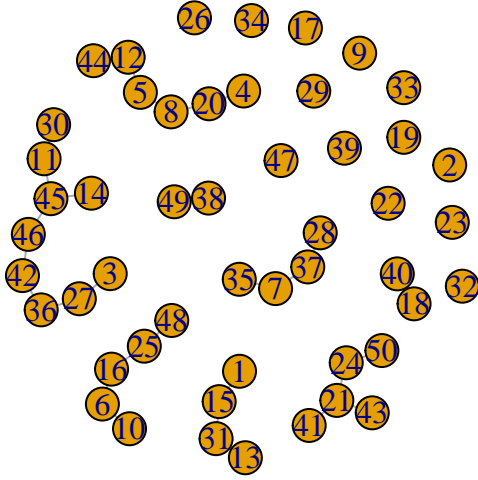
Stop Words Not Removed

Epsilon = 0.985



Stop Words Removed

Where $\epsilon = 0.9825$

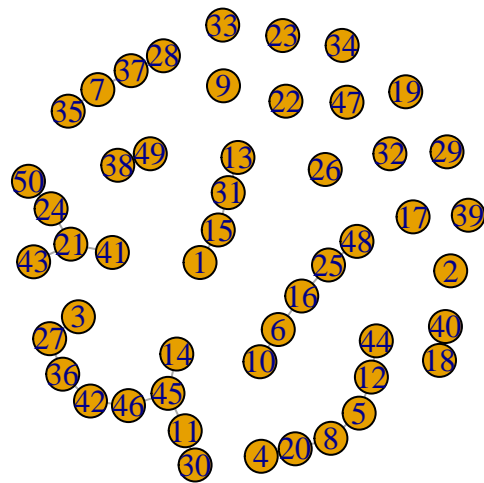


A Quick Thought

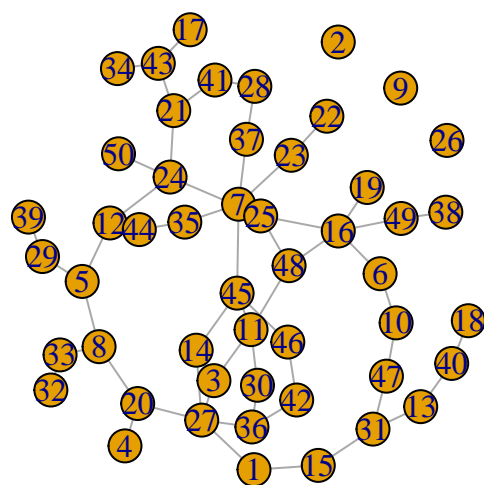
We notice in the two above charts that the connections between points 7-37, 10-6, and 25-48 persist. Further investigation needs to be done to determine why these sets of episodes are paired.

As ϵ gets bigger

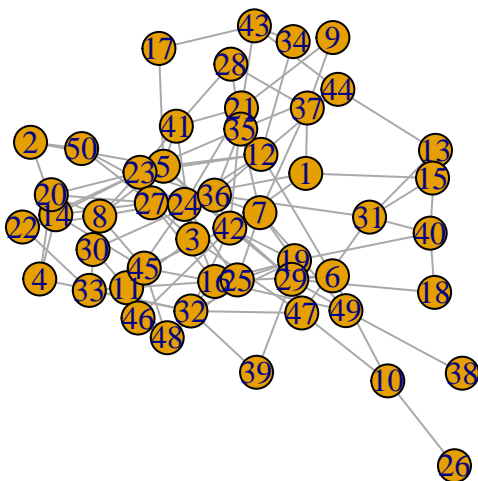
Epsilon of 0.9825



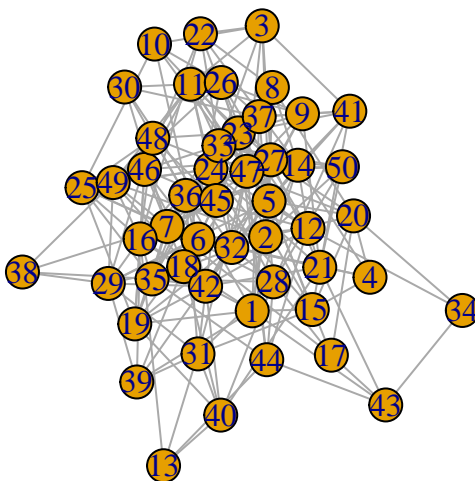
Epsilon of 0.985



Epsilon of 0.9875

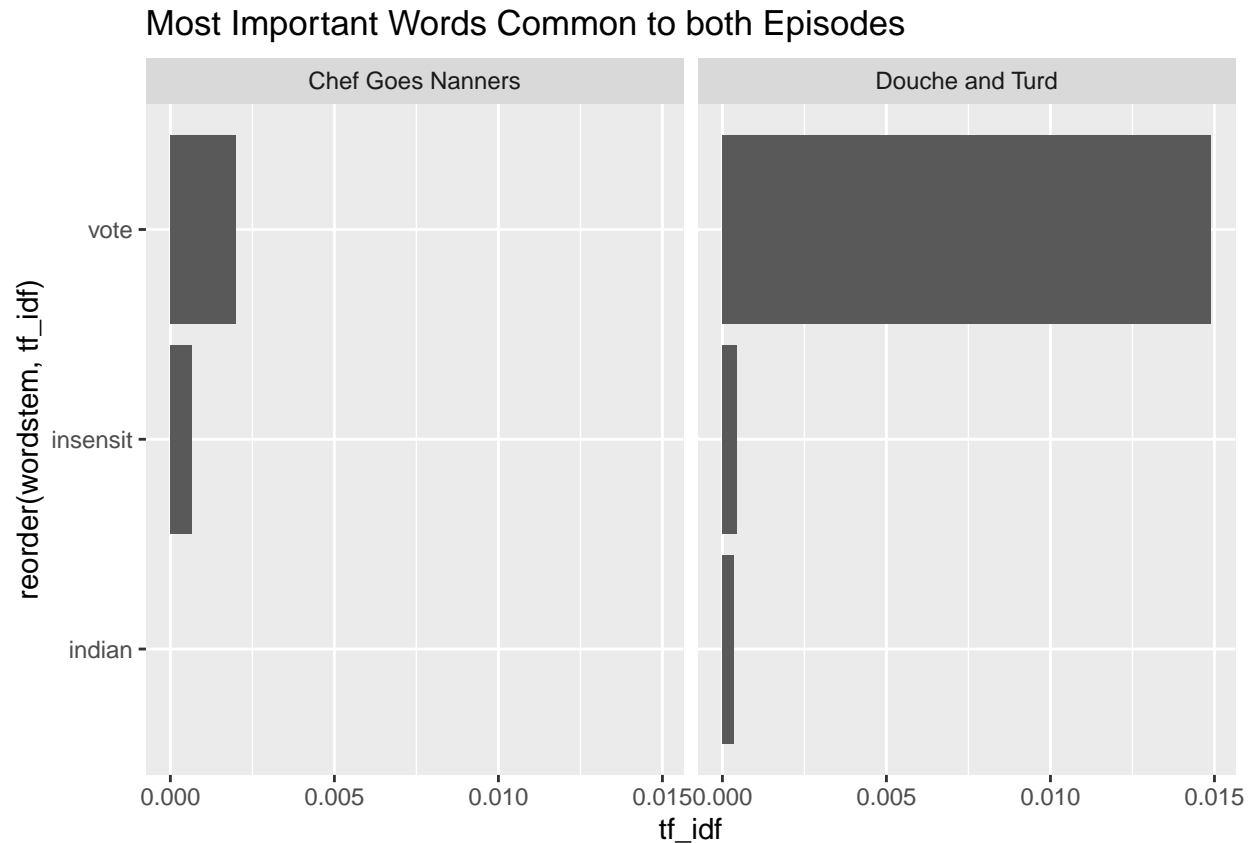


Epsilon of 0.99



Step 3: Profit

We filtered out words that only occurred in one episode, then plotted their tfidf score to find the most important words in common to both episodes.



Season 4, Episode 7: Chef Goes Nanners

IMDB Description

Chef and Jimbo are debating whether the flag of South Park, which depicts four white figures lynching a black figure, should be kept. Jimbo wants to keep the flag out of tradition, while Chef feels it's a racist symbol. As the debate spreads through the town, the mayor backs out of making a judgment, saying the issue will be decided by a debate held by the schoolchildren. This pits Wendy and Cartman, who want it changed, against Stan, Kyle and Kenny, who don't see anything wrong with the flag as it is.

Season 8, Episode 8: Douche and Turd

IMDB Description

Stan refuses to vote in the school mascot election because his choices are a giant douche and a turd sandwich. Unfortunately, his views on voting are seen as un-American, and he is ultimately banished from the town.

insensit
indian
offensplenti
citi allow mackei
natur
vote

offens
allow indian
mackei citinatur
plenti
insensit
vote