**Cultural evolution of emotions in 50 years of pop song lyrics**

A. Acerbi, C. Brand, A. Mesoudi

## Preliminary analysis

We analysed the emotional content of two datasets of english language song lyrics. The dataset “billboard” contains the lyrics of the songs included in the annual Billboard top-100 from 1965 to 2015, both included (N=4,913 songs). The dataset “mxm” contains the lyrics of the English language songs present in the musixmatch.com website, the world's largest lyrics platform where users can search and share lyrics, from 1965 to 2010 (years for which we had more than 500 songs) both included (N=159,015 songs). Full descriptions of the datasets and of the sentiment analysis tool are in the document “Data preparation”.

In both datasets, we found a decrease in the usage of *positive* emotions-related words (Figure 1), coupled with an increase in the usage of *negative* emotions-related words (Figure 2).

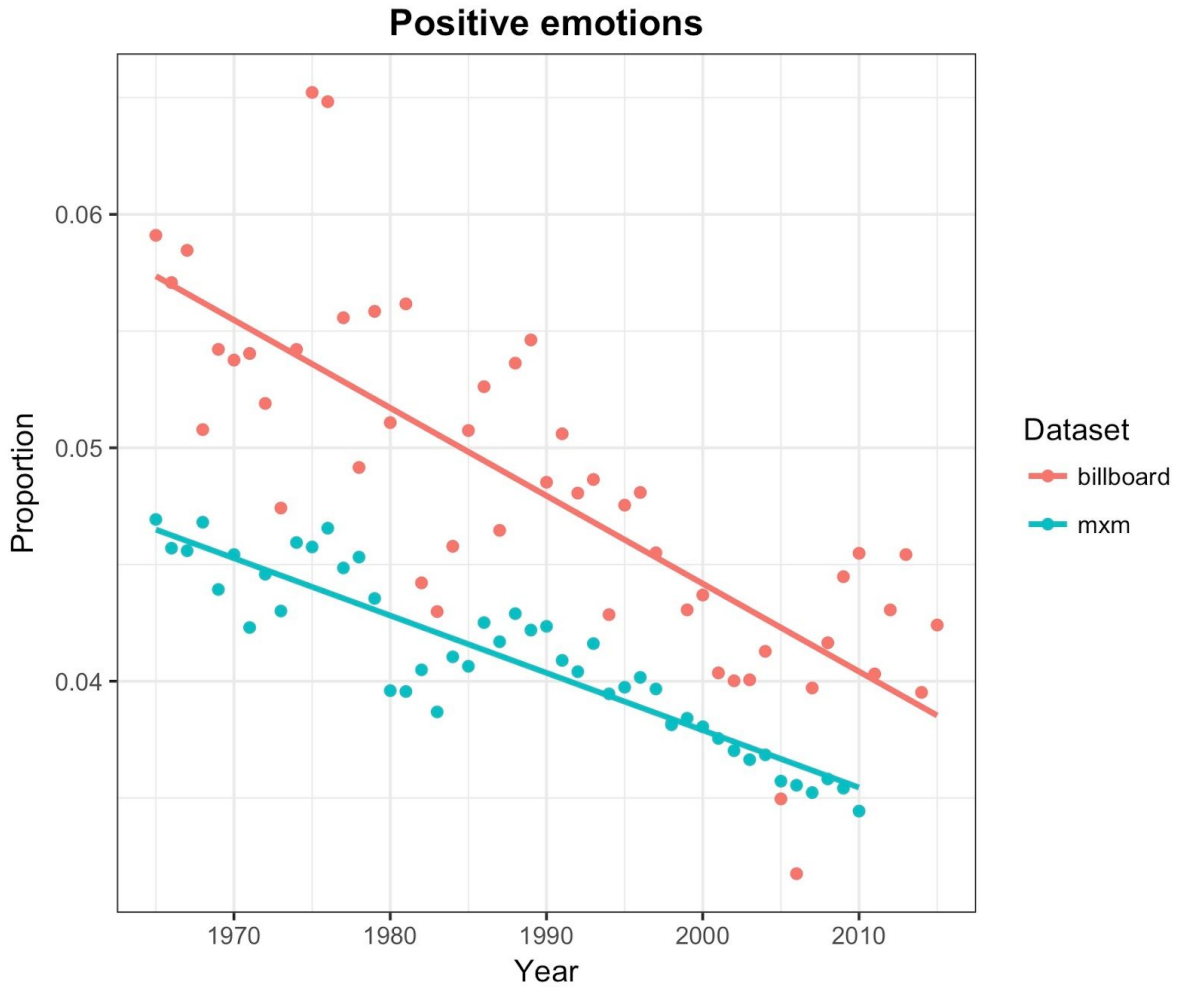


**Figure 1:** Proportion of positive emotions-related words, on the total number of words for all songs, by year.

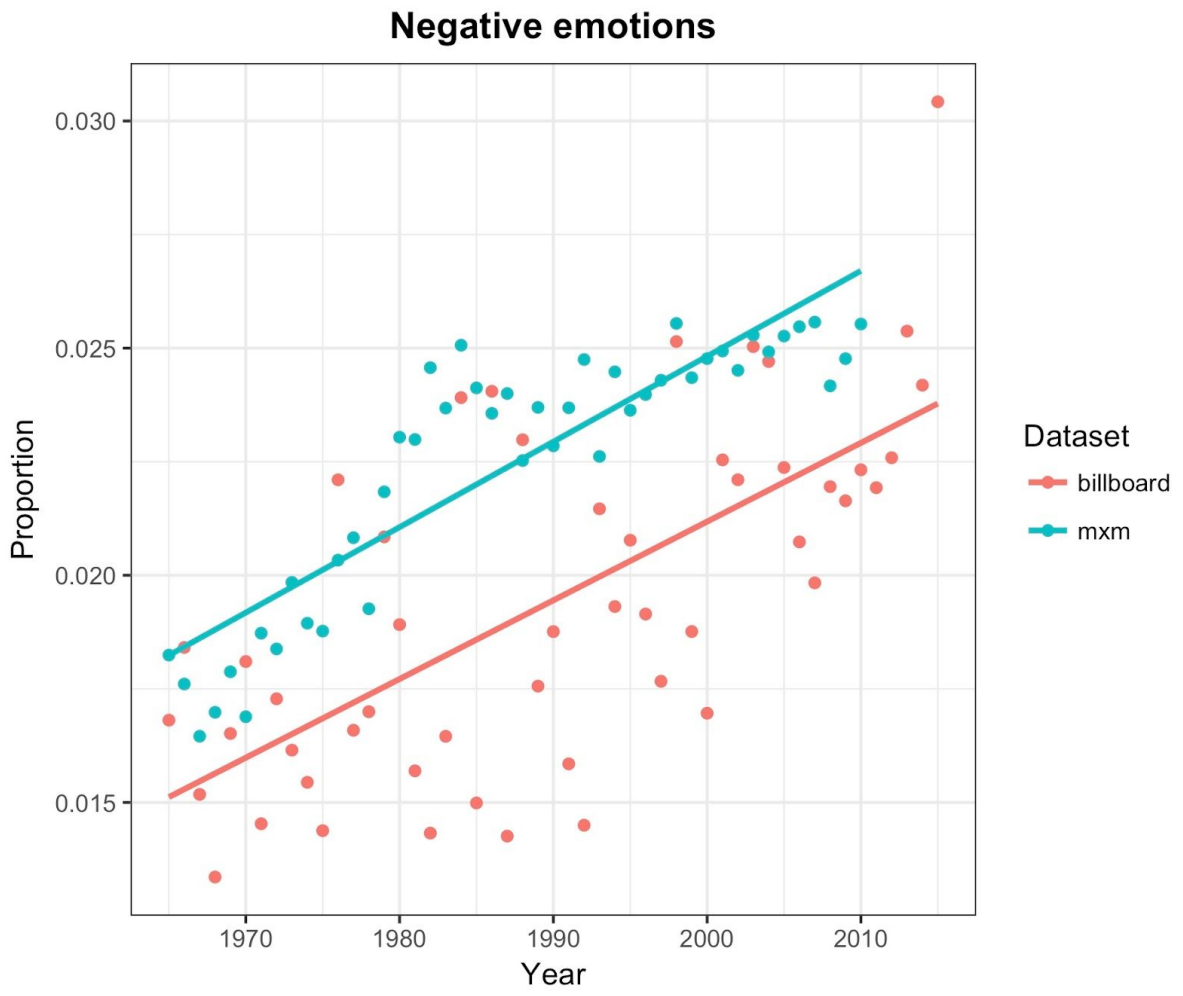


**Figure 2:** Proportion of negative emotions-related words, on the total number of words for all songs, by year.

These results are broadly consistent with previous analyses (e.g. Dodds & Danforth 2010), that found that pop music lyrics got sadder through time. A similar result has been found in literary fiction, where an analysis of two centuries of different English language corpora pointed to a decrease in positive emotion-related content, while the negative content remained stable, or slightly increased (Morin & Acerbi 2017).

Single words show considerable change in usage during the time frame considered. For example, the usage of the term “love” more than halved in both datasets (see Figure S1).

We also analysed the emotional words that contributed more to the trends. The relative contribution is calculated as the absolute value of the difference between the frequency in the last year of the dataset and the frequency in the first year of the dataset. Figure 3 shows the 20 most contributing positive and negative words for the “billboard” dataset, and figure 4 shows the same for the “mxm” dataset. Notice the words are stemmed (see “Data preparation” document).

The words mostly contributing to the decrease of positive emotion content are largely overlapping in the two datasets. The words mostly contributing to the increase of negative emotion are instead, in part, different, with the “mxm” dataset showing a larger effect of slang/swear words.

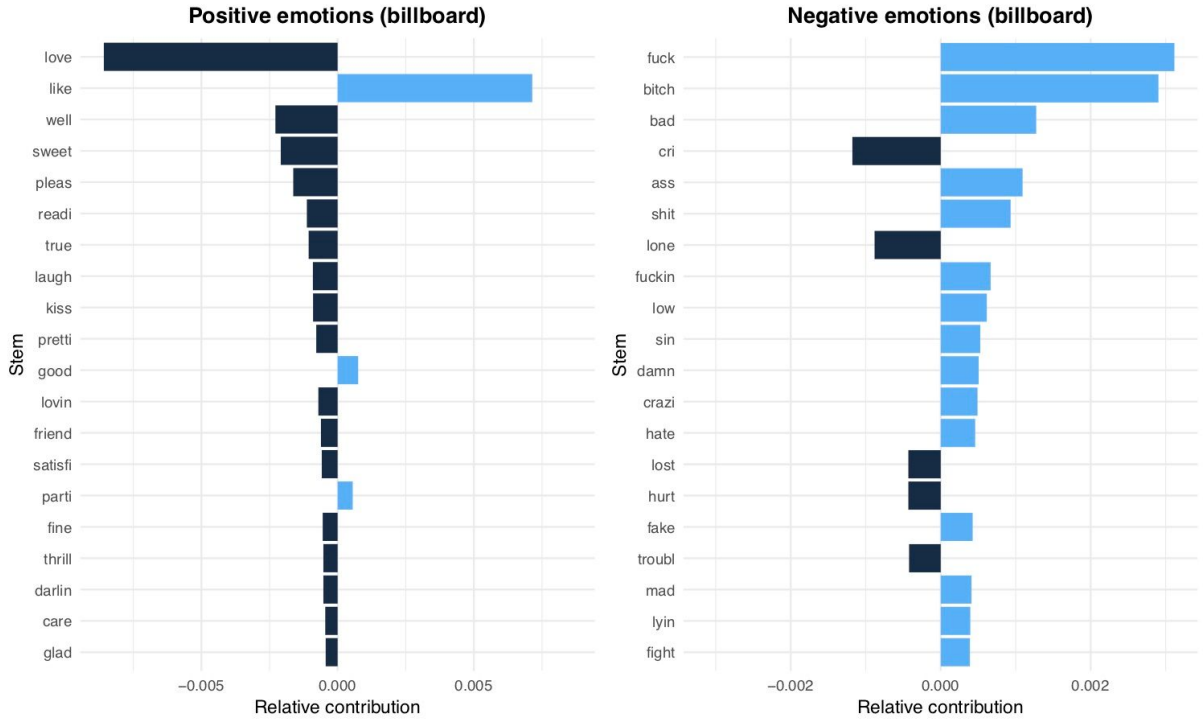


**Figure 3:** Relative contribution of the most contributing 20 positive and negative emotions-related stems for the dataset “billboard”.

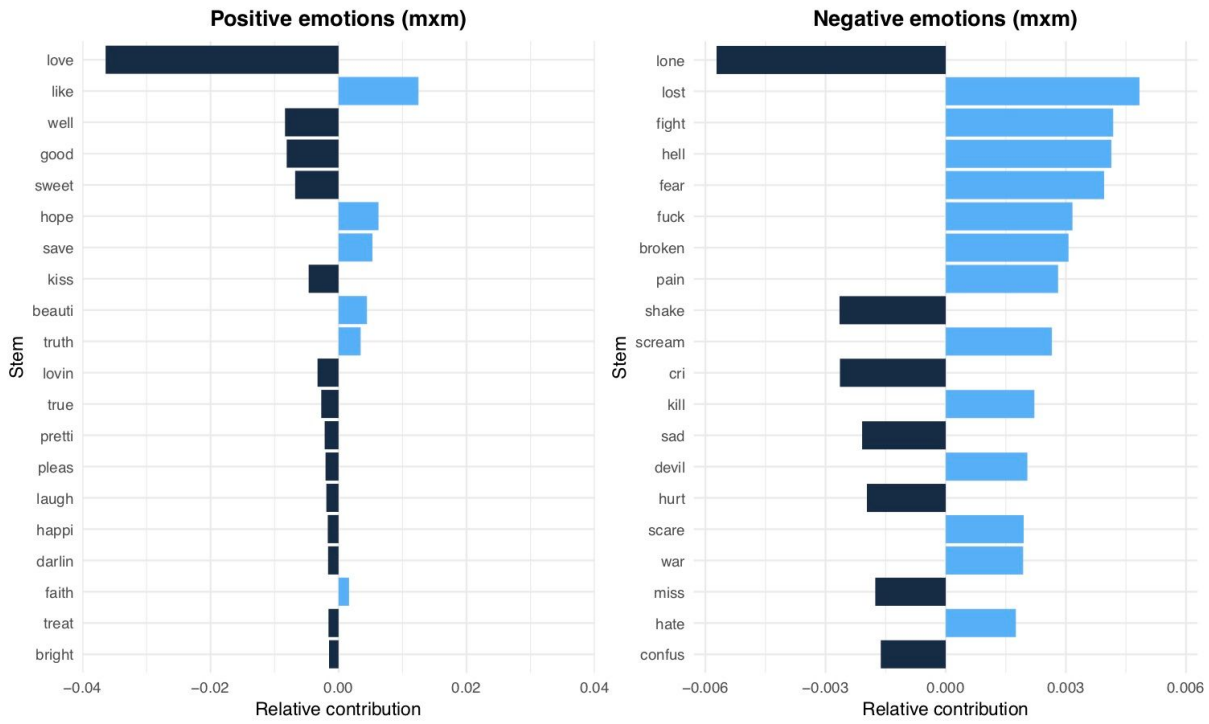


**Figure 4:** Relative contribution of the most contributing 20 positive and negative emotions-related stems for the dataset “mxm”.

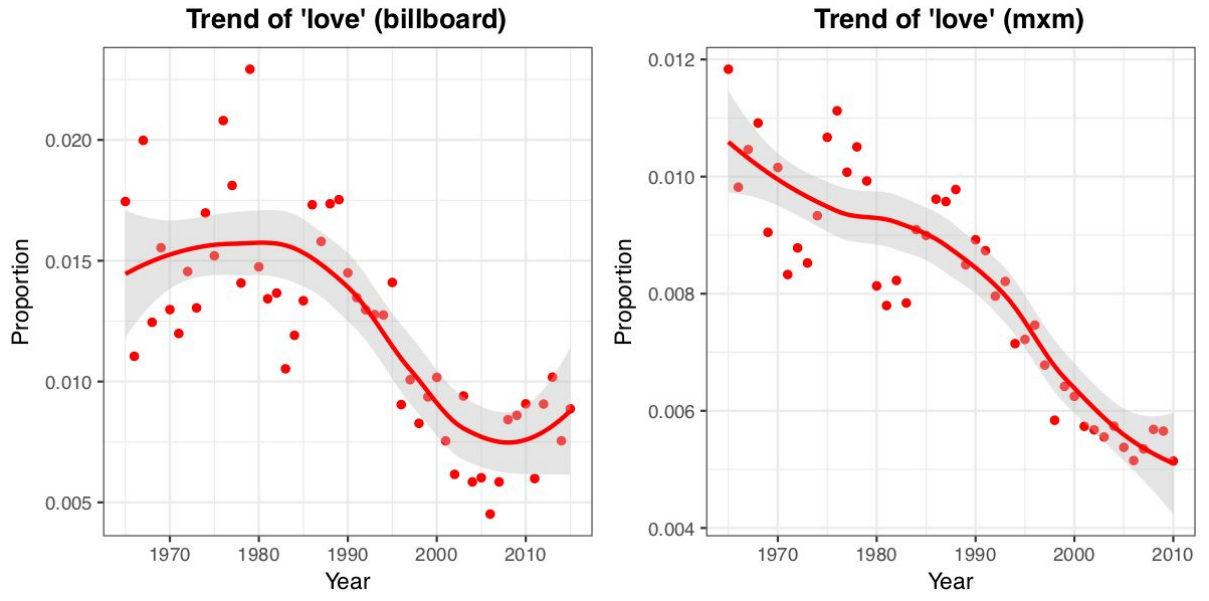


**Figure S1:** Frequency of the term “love” in the two datasets.