DV FINAL PRESENTATION PREP

Martina: 86

Lars:113

Hugo: 176

Raph: 150

Jim: 86

Fengan: 90

Total: 11 min 40 sec

Intro: **Martina**

**Actual Time: 86s**

Suggestion for Intro

Good morning,

We are Group 3 and this is our project “A Lyric Space Odyssey”. We will briefly introduce you to the data sets and then we will show a set of visualisations we have realized experimenting three different methods and tools.

The main dataset we used is The Million Song DataSet which is a freely available collection of audio features and metadata for one million contemporary popular music tracks.

We also used two complementary data sets, The MusixMatch Data set which contains the lyrics collection of the Million Song Data Set, including a set of 5000 unique words,

and the TagTraum Data Set which contains the genre annotations.

In fact we are working with a subset of these data sets:

1st Constraint: We work with Songs with Lyrics with Genre Annotations. (MSD has instrumental songs).

2nd Constraint: We consider only English lyrics.

3rd Constraint: We analyze songs from 1980 – 2010.

What we want to do is visualizing the song lyrics in order to find patterns and encouraging the user to explore the lyrics. We will take you through an exploration of this dataset starting from a more general level down to individual lyrics and titles.

**Chap 1: Lars**

**Actual Time: 66s + +47s = 113s**

Graph 1:

When starting to explore a dataset, a **network graph** is a good place to **start**. This type of **visualization** can often reveal **subtle** **patterns** and **connections**, difficult to **pick** **out** from the **data**, but easily **visible** to the human **eye**.

In this **first** case, we set out to examine whether there were any **patterns** in songs that **share** **similar** lyrics. Using Gephi, we input **connections** for each song that shared at least **20** **%** **unique** **words** in the dataset, outside of stop words. This is the result. Here the colors indicate genre. You can see a cluster of orange nodes drifting off to the side. Apparently, in this dataset, rap songs have a different vocabulary.

Next we wanted to look into the connections between the words themselves à

Graph 2:

On this chart, each connection corresponds to two words cooccurring in a song in the dataset, while the size of the node and text shows how frequent a word is cooccurring. Seeking to grapple with the actual connections between words in this dataset, this graph shows us what we are up against. It makes for a pretty dense graphn becoming more of a wordcloud than a network graph. It became clear that to really dive down into the connections between song lyrics, we would have to dive deeper.

**Chap 2: Hugo**

**Actual Time: 61s + 56s + +49s**

**[Fengan]**

So, in order to dive deeper, we provide a tool that calculates and visualises individual word connections.

Here in this bar you can type words of your interest and discover their connections. The bubbles here present only the relevant connections - and by “relevant” I mean a good balance between popularity and uniqueness. In other words, the connected words here are neither too frequent, nor too infrequent. The results are conveniently categorised by genre.

Let’s try “journey”, for instance. We can see that “journey” is found most frequently in New Age songs. And here you can see some interesting connections. Bigger bubbles display more unique connections like “ocean”, “dark”, “heaven”. And small bubbles represent those overall very popular but relatively less relevant words, like “promise”, “twice”, “fallen”.

**[Hugo]**

The next stop in this *odyssey* is about word co-occurrences within a song

[Change to 2.2, Lyrics Overview]

...in the Lyrics Overview dashboard, you can *explore* a representative song of each genre. Bigger bubbles indicate relevance within lyrics. So, in this case, “translation”

[hover over translation]

...is one of the words that best characterize this New Age song.

Now, let’s jump to *another dimension* with the Ego Position Chart. Here, a *red star* situates this song in the “You – Others” quadrant. Which means that pronouns like “me” and “I” are less dominant than pronouns like “they” or “you”. Then, we can say that this song has low egocentricity.

To learn more about the Ego Position lets goto the next dashboard,…

[Change to 2.3, Yearly Ego]

…here we see that New Age is not clustered. This indicates a *broad* *spectrum* in the use of pronouns. And if we focus in other genres,

[ Remove New Age ]

..we can unveil a tendency towards more egocentric songs. Going back in time

[ Adjust years: 1980-1990 ]

...we see that Pop and Rock used to be centered, closer to less egocentric quadrants.

**Chap 3: Raph**

**Actual Time: 150**

After the exploration section, let move onto the creation part.

“Create your own lyric” is an exploration environment and a creation tool as well.

The concept is to try making your own lyric by diving, voyaging within the million song dataset.

[hover over the **mega** bubble chart]

You can explore around the mega bubble chart. This is the same as the word communities visualized in the Gephi Chart, only without the edges.

[click on one bubble of the **mega** bubble chart]

**(! Click 1) By selecting a word**, you have the interconnected words visualized in another bubble chart on your right.

[**click** on another bubble in the **connect** bubble chart, you can do it more than one time]

Selecting another word brings you to the connection of that word. Your chosen word is always at the centre of the regenerated bubble chart.

Or You can search a word to begin:

[search spa] (select space afterwards)

Let’s see what insight the word “space” bring about:

To help you to create your own lyric, you can switch around different parts of speech

[switch to POS noun]

noun: “serenity”, “oblivion”, “solitude”, “wasteland”

[switch to POS Adj/Adv]

adj/adv: “endless”, “distant”, “vast”, “desolate”

Like before, the size of the bubbles represent the frequency (or popularity) of a word. A large bubble represents a popular word. A small bubble represents a unique word.

[switch the Popularity-Uniqueness toggle]

You have the freedom to choose to see more popular words or unique words. With the scale of the bubble, you can compare the relative popularity or uniqueness of your chosen word to the connected word.

[hover over genre link]

You may also choose to explore the connection of the word in different genres.

Once you are happy, click the “add button” to add that word to the lyric composer.

[add a word]

We also offer a pronoun toolbox to help you to create your customized lyric

[add a pronoun]

Next, we are going to talking about song title

**Chap 4: Jim (90 seconds flat)**

**Actual Time: 86s**

What would you name the next hit song? Let’s take inspiration from the hottest existing songs.

By starting from one word and following the arcs to the next and the next, we can build new song titles. Some examples here: if Kesha had this chart, she could have named her song something more descriptive than “tik tok” which is most commonly the name of a video sharing app. Likewise, we can find alternate reality names for No Scrubs and Country Roads.

You’ll notice that some common words are not here, the most noticeable example being the word “Love.” We’ll see on the next chart why.

>>>GOTO 4.2<<<

This is the chart of the worst songs on record. “Love” is one of the most used, or overused words, in every bottom-ranked song. Great love songs describe the theme, they don’t need to put it directly in the title. Bad songs hit you over the head with “here’s my message, don’t miss it!”

Stepping away from love, we can find other examples of song titles not to use. “I Don’t Do It”? There are so many interpretations of this, all of them bad.

**Conclusion: Fengan**

**Actual Time: 119s**

(????? - 90 sec)

To wrap up, our team visualised English song lyrics in 13 genres from the period of 1980 to 2010 from 4 different perspectives.

One of the main challenges with our dataset is to avoid trivial connections. On the one hand, we have to tackle words that are used in almost all songs, including for example love, baby, and all kinds of stop words. And on the other hand, we also have words that appear in too few songs, such as symbols, names, onomato’poeias. In the end, I believe that we managed to present words with a relatively good balance between popularity and uniqueness.

Insight wise, we noticed that while each genre has its distinct word preference, RAP songs really stand out for a very very unique field of words. English songs are becoming increasingly egocentric in most genres. In terms of song titles, it’s revealed that hot titles use uncommon word pairs to create descriptive and memorable titles, while poorly-ranked songs use more common words.

In addition, we invite the user to explore songs, words, part of speech, genres, as well as their popularity by herself. With this autonomy, the user is able to generate her own insights far beyond us.

**Concrete Insights:**

Jim: Hot titles use uncommon word pairs to create descriptive and memorable titles. Poorly-ranked songs use more common words.

Feedback:

elaborate, interaction well thought out

stronger insights

Katrien: implementation?

Diego: big data, lots of tools (d3 js)