

1 Introduction

- **Group members:** Enrico Borba, Claire Goeckner-Wald
- **Team name:** Papa Mart's Mini Gary - The Comeback
- **Division of labour:** Enrico Borba: Programming, ideas, report visualization. Claire Goeckner-Wald: Programming, ideas, report assembly.

2 Pre-processing

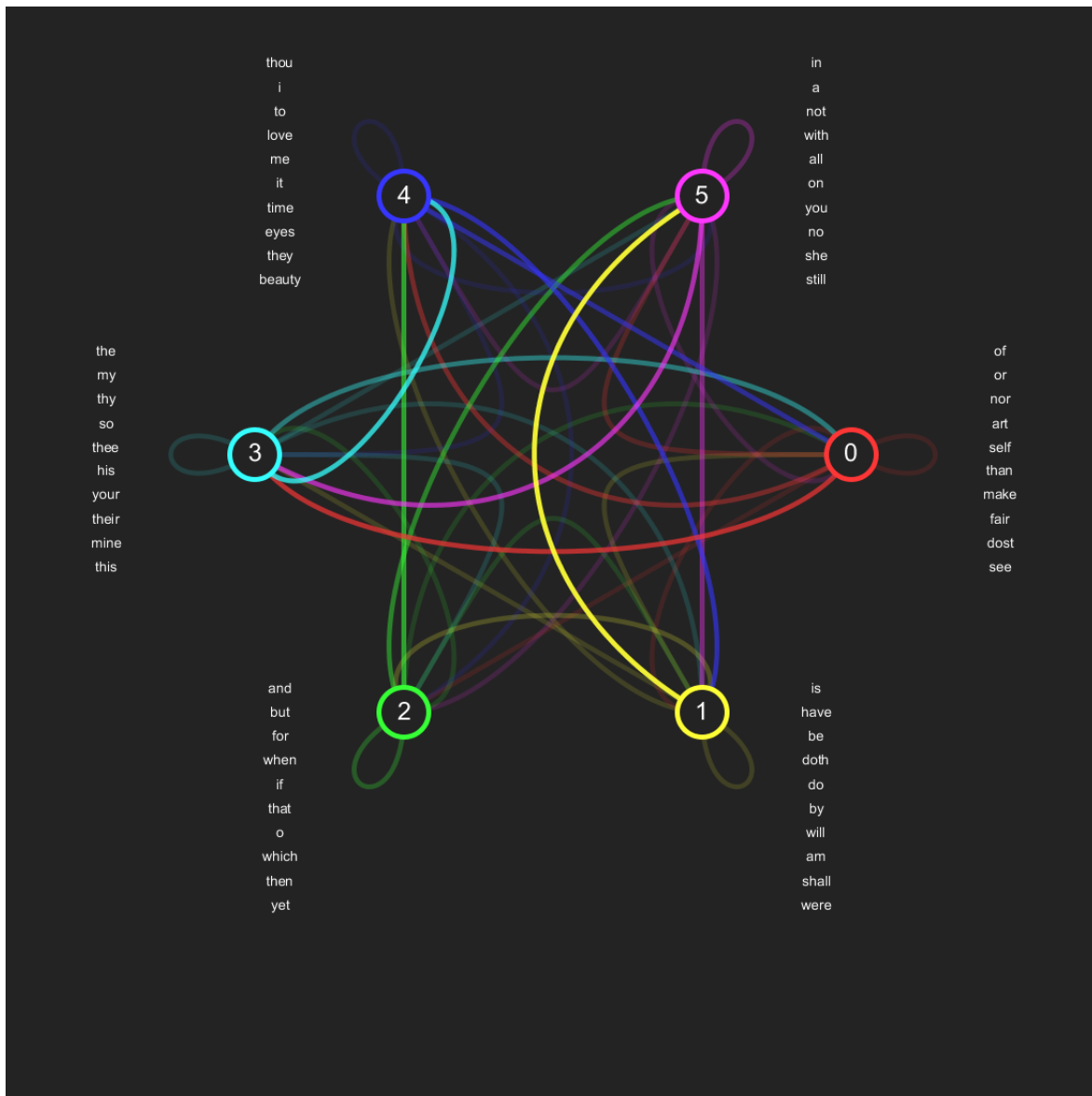
- **Handling the dataset**
 - **Quatrains versus couplets:** Initially, we thought that we should train two different models, one on lines from the quatrains, and one on lines from the couplets. This way, we could more accurately capture the “shift in tone” that Shakespeare often performs during his sonnets. However, since we decided to ‘force’ rhyming using a rhyming dictionary garnered from the dataset, we ended up combining quatrains and couplets into one model. We
- **The dictionary**
 - **Reason for use:** We used a dictionary to assign each word a unique number. We had to set all word to lowercase first, to avoid using “The” in the middle of a sentence, when we would rather have “the”.
 - **Unexpected trouble:** The use of the dictionary is also one reason why we decided to treat lines from quatrains and couplets equally. Initially, we had one large dictionary that covered all words Shakespeare used in the dataset. As expected, some words were used only in the couplets, or only in the quatrains. However, we ran into errors when training two separate Hidden Markov Models, most likely because the model expected that if we gave it words (represented by numbers) $\{3, 15, 23, 14, 194\}$, that there would be $(1 - 194)$ states available. However, this was not the case.

3 Unsupervised Learning

- **Topic**
 - Subtopic:
 - Subtopic:
- **Topic**
 - Subtopic:
 - Subtopic:

4 Visualization & Interpretation

- Topic
 - Subtopic:
 - Subtopic:



5 Poetry Generation

- Topic

- Subtopic:
- Subtopic:

6 Additional Goals

- Topic
 - Subtopic:
 - Subtopic:

7 Extra Credit: Recurrent Neural Networks

- Topic
 - Subtopic:
 - Subtopic: