# Word2Vec

- Since we're using the keywords from songs to retrieve images with tags associated with them, we need to find a way to compare them
- Most modern ML algorithms don't deal with strings directly
- This is where word2vec comes in: it takes words and converts them to a point in a n-dimensional vector space
- The model we train can be used to determine how similar the keyword is to the tags of the images so that we can pick the "best" (relevant) image
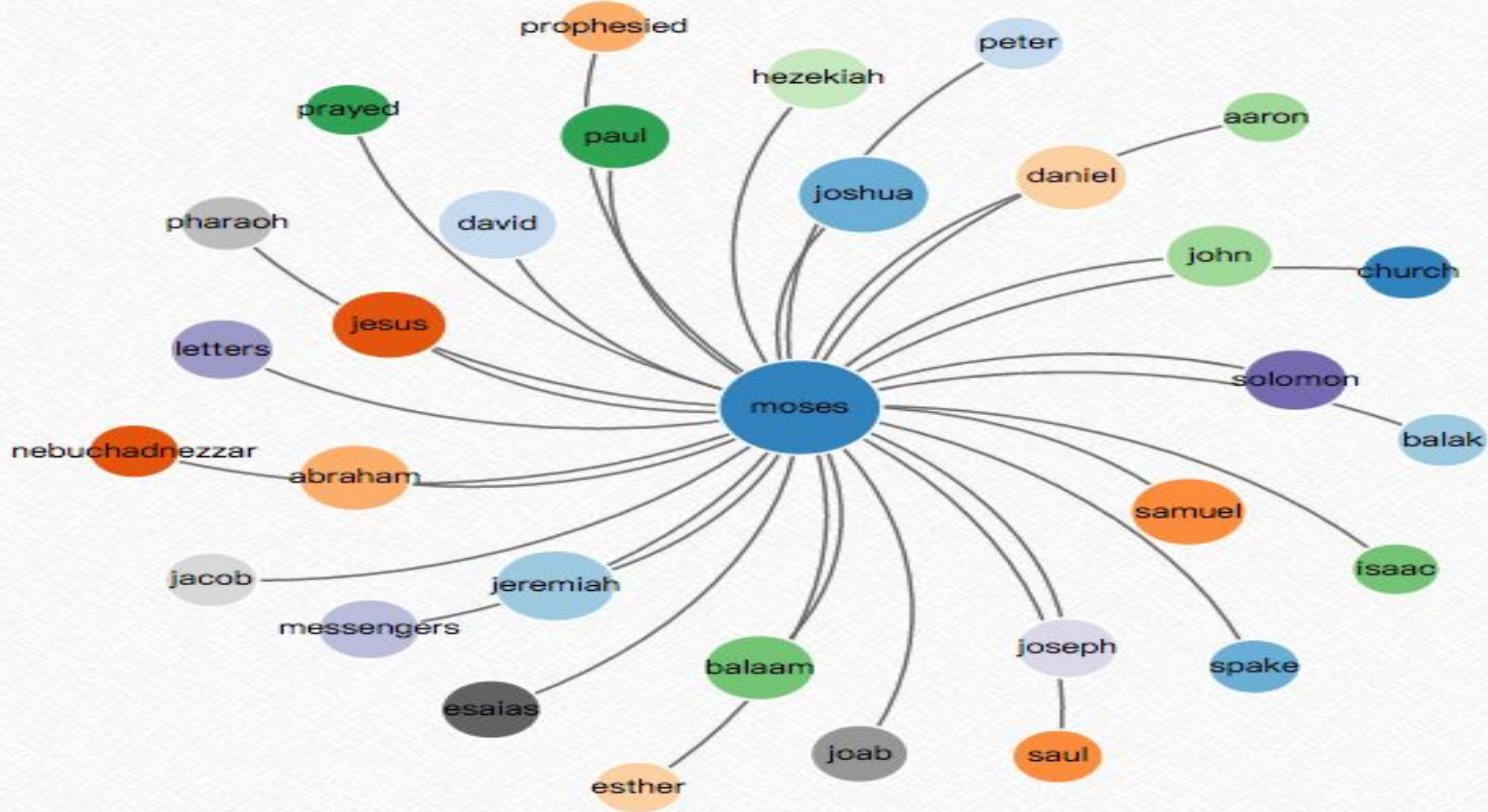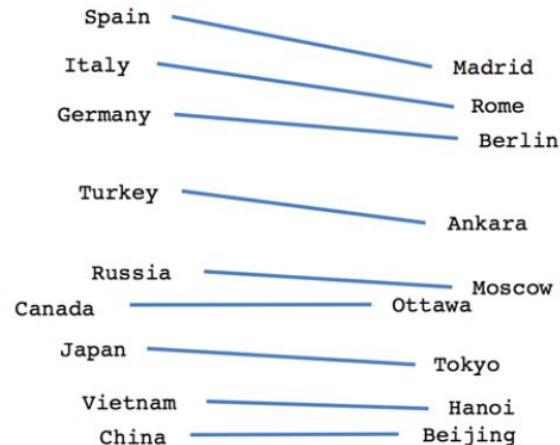
- We can see which words are close to each other
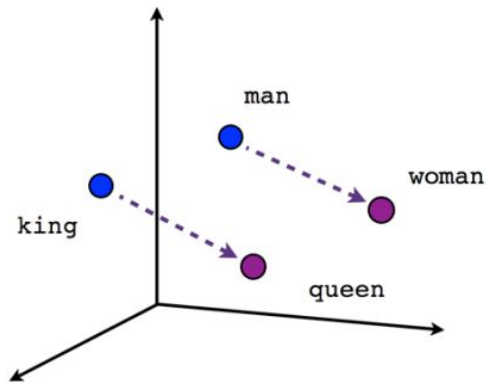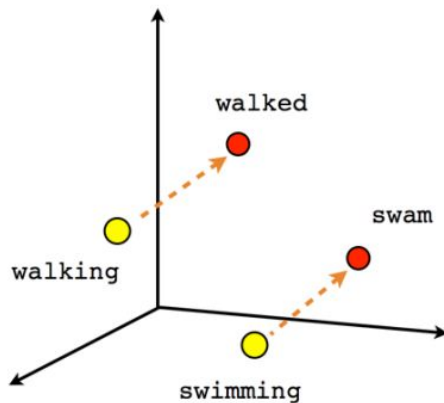
Male-Female      Verb tense      Country-Capital

- This is essentially what it's doing: words with similar relationships (king and queen or man and woman) have similar euclidean distances between them

# Development Updates

- We were able to get a running algorithm to find the most relevant image for each paragraph
- Ran it through a few times to find errors
- We've done the manual search using pixabay to see how well it's doing

# Manual Search vs Algorithmic Version

Manual: https://docs.google.com/document/d/10HvHlIFRJy6NmU9uxNbhjY92KNjwS3qWGhpdtbY_GE4/edit?pli=1

Algorithmic: https://github.com/michael1albq/DataStories

# Challenges & Next Steps & Outline

- Dealing with any more potential errors through testing and debugging
- Creating the second algorithm (where we have a keyword mapped to an image and the song maps to all the images)
- Need to change the dataset we use since it can only obtain a limited number of pictures for many keywords and none for many other keywords → so we're losing a lot of information