UCSB Python Network Science: Word2Vec Graphing

Skyler Chandos, Mason Lender, Casmali Lopez and Jake Trimble
Learning Python (Thanks Guido)

Our introduction to the class began by learning how the coding language Python operates and transferring our prior knowledge of Java to a new language.

Lots of practice problems, some from Project Euler!

def hello3():
    for i in range(3):
        print "Hello"

def greaterThan(num1, num2):
    if num1 == num2:
        print "Are Same"
    elif num1 > num2:
        print "{} is greater than {}".format(num1, num2)
    else:
        print "{} is greater than {}".format(num2, num1)
Graph Theory!

Is a field of mathematics that describes the complex relationships in a network of connected nodes, where said nodes can represent a multitude of objects or ideas. We used this theory throughout our time working on the project as the basis of our program. A simple example of graph theory being implemented is in a representation of the system that transfers power across the country, as shown in the bottom-right graph (Color is determined by degree).
We learned two methods of search in order to find the connections between nodes in a network of nodes. DFS (Depth First Search) uses a stack to store nodes as it searches through the graph. Therefore, it reads from the most recent node added to the stack. BFS (Breadth First Search), on the other hand, uses a queue and reads from the first node added to the queue.
Machine Learning

- The process by which computers interpret and learn from data
- There are two types of machine learning, classification and reinforcement learning. We worked with a library that used classification learning.
- In classification learning a program is “trained” by feeding it data and telling it how to classify the input. It can be either supervised or unsupervised, meaning that the program can take user input while training or train based upon its algorithm alone.
- Once trained, the computer can make decisions and predictions based on the input without requiring additional user input.
- For example:
  - A program is fed images and told what objects it contains
  - After the training period is complete, the program can be given an image and it will tell the user what objects the image contains
Word2Vec from Gensim

An unsupervised machine learning library that takes as an input a large corpus and returns a vector space wherein words with similar linguistic contexts have similar vectors.
The Program

```python
# Determine which nodes belong in which communities
partition = com.best_partition(fullGraph)

# Returns the number of communities
maxComm = max(partition.itervalues())

# An empty 2D list for the communities
comms = [[] for i in range(maxComm+1)]

# Add nodes into their specific communities
for k,v in partition.items():
    comms[v].append(k)

# Select all the communities with more than three words
res = []
for i in range(len(comms)):
    if len(comms[i]) > 3:
        res.append(comms[i])

# Turn all the communities into subgraphs of fullGraph
sG = []
for i in res:
    sG.append(fullGraph.subgraph(i))
```
The Santa Barbara Municipal Code (SBMC from hereon), Graphed (Sample 500 Words)
Word Communities in the SBMC
Isolated Word Community in SBMC
Isolated Word Community in SBMC
Isolated Word Community in SBMC
Thank you!