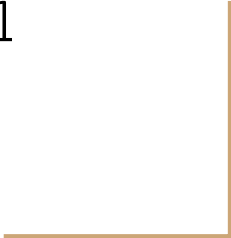# Programming, Problem Solving, and Algorithms

CPSC203, 2019 W1

Announcements

Final Exam: 12/11, 12p, DMP301

OH next week:

Today:

More NLP

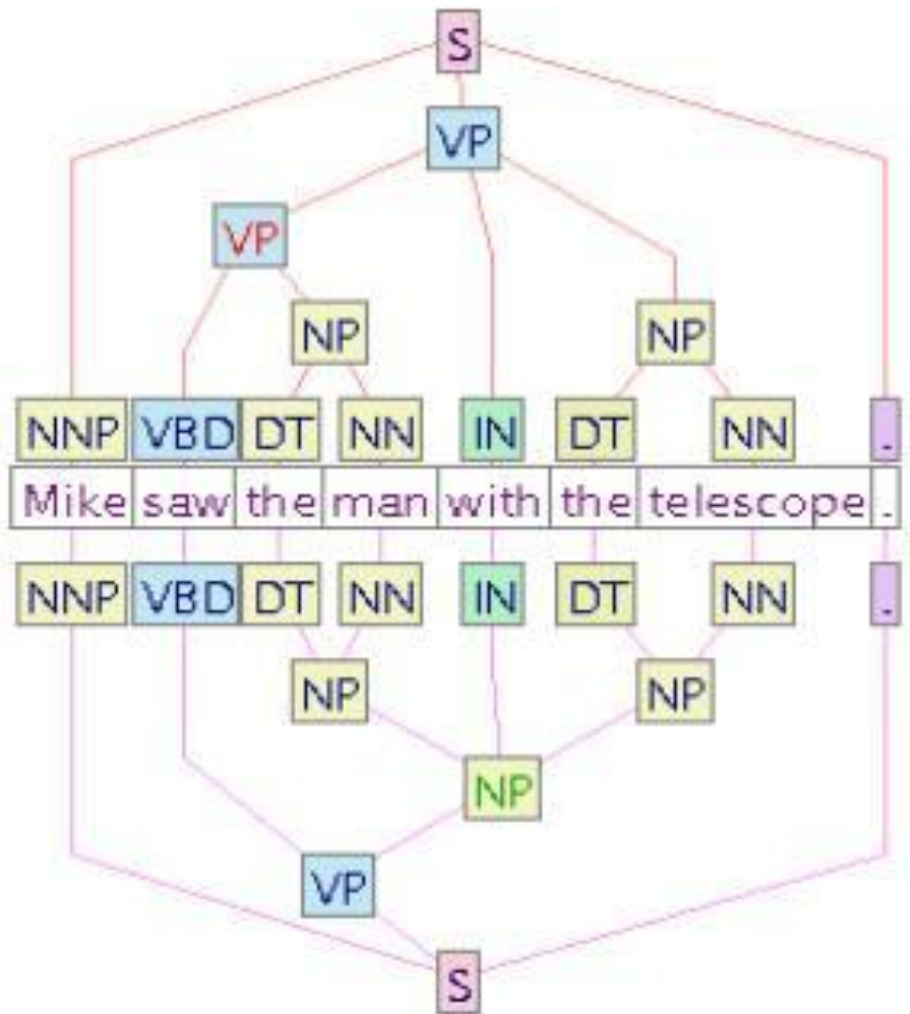Graphs in Python

Harry's social network

POS and NER via nltk

NER used for inferring relationships between entities.

```
PERSON *lives* LOCATION

LOCATION *has* ORGANIZATION
```

Named Entity Recognition (NER)

1. Underline all of the proper nouns (named entities) in this text..

Mr. and Mrs. Dursley, of number four, Privet Drive, were proud to say that they were perfectly normal, thank you very much. They were the last people you'd expect to be involved in anything strange or mysterious, because they just didn't hold with such nonsense.

Mr. Dursley was the director of a firm called Grunnings, which made drills. He was a big, beefy man with hardly any neck, although he did have a very large mustache. Mrs. Dursley was thin and blonde and had nearly twice the usual amount of neck, which came in very useful as she spent so much of her time craning over garden fences, spying on the neighbors. The Dursleys had a small son called Dudley and in their opinion there was no finer boy anywhere.

Typical categories of entities are PERSON, LOCATION, ORGANIZATION. Think about how you might discover each of the entities using a program.

NLTK NER discovery...

2. Repo LecHP contains a file called test.py. Modify and execute this file to answer the following questions. In each case, sketch an example of the output, and explain it briefly in English.

a. if `textRaw` is the string above, what is the result of

```
sents = sent_tokenize(text)
```

b. if `sents` is the result of part a, what is the result of

```
sentWords = [word_tokenize(s) for s in sents if s]
```

NLTK NER discovery...

c. if `sentWords` is the result of part b, what is the result of

```
sentWordsPOS = [pos_tag(s) for s in sentWords]
```

d. if `sentWordsPOS` is the result of part c, what is the result of

```
sentWordsNER = [ne_chunk(s) for s in sentWordsPOS]
```

e. if `sentWordsNER` is the result of part d, what is the result of

```
subtrees = [chunk.subtrees() for chunk in sentWordsNER]
```

NLTK NER discovery...

f. if `subtrees` is the result of part e, what is the result of

```
entities = [[ s for s in st if s.label() == "PERSON"] for st in subtrees]
```

g. if `entities` is the result of part f, what is the result of

```
entities = [[ ' '.join(c[0] for c in s.leaves()) for s in st] for st in entities]
```

3. Write python code that would extract all the verbs from the text above. The answer to problem 2c will help you!

4. (challenge) Write a function `personVerbs(person, text)`, that returns a list of all the verbs that occur in sentences that also contain `person`.

One last project...

Repo LecHP's file test.py does more than just find the characters. It also

_____.

Suppose we want to know more: We'd like to understand the bonds between pairs of people!

1. How can we infer connections?
2. How can we draw a graph?
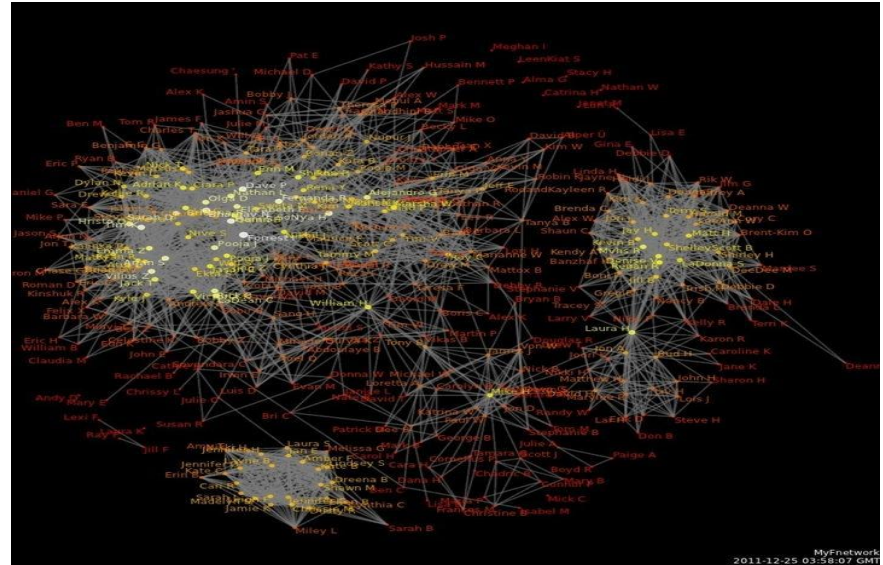
Friends?

Given the text from a novel, how can we infer interaction or connections between characters? Discuss this question, and write down your ideas.

_____

_____

_____

_____

_____

_____

We are not the first...

Extracting Social Networks from Literary Fiction

David K. Elson
Dept. of Computer Science
Columbia University
delson@cs.columbia.edu

Nicholas Dames
English Department
Columbia University
nd122@columbia.edu

Kathleen R. McKeown
Dept. of Computer Science
Columbia University
kathy@cs.columbia.edu

Abstract

We present a method for extracting social networks from literature, namely, nineteenth-century British novels and serials. We derive the networks from dialogue interactions, and thus our method depends on the ability to determine when two characters are in conversation. Our

We present a method to automatically construct a network based on dialogue interactions between characters in a novel. Our approach includes components for finding instances of quoted speech, attributing each quote to a character, and identifying when certain characters are in conversation. We then construct a network where characters are vertices and edges signify an amount of bilateral conversation between those charac-

https://nicholasdames.org/wp-content/uploads/2013/06/ACL2010-ElsonDamesMcKeown.pdf

Starting from the goal

The social network graph will have…

Vertices:

Edges:

Edges

We could consider every pair of people and check every paragraph for their presence. Do you like this plan?

OR, we could...

1.
2.

Edges

"I've heard of his family," said Ron darkly. "They were some of the first to come back to our side after You-Know-Who disappeared. Said they'd been bewitched. My dad doesn't believe it. He says Malfoy's father didn't need an excuse to go over to the Dark Side." He turned to Hermione. "Can we help you with something?"

What names appear?

What pairs should be tallied?

General observations:

Designing from the middle

If I gave you [[RW, HG, HP], [RW, AD], [H, HP, HG]]...

1. could you create [RW, HG, HP, AD, H]?

2. and {(RW,HG):1, (RW,HP):1, (HG,HP):2, (RW,AD):1, (H,HP):1, (H,HG):1}?

Demo

LecHP updated to include all the new pieces.

We'll be in touch!!