Benjamin @benjamin_draves

4 Dec • 16 tweets • benjamin_draves/status/1335008268129869826

Hello. Yes - I'd like to report my Q4 network pictures. A thread.

Over the past few days, I've constructed a co-occurrence multiplex network for the characters in Harry Potter.

I started by grabbing the text from a pre-scraped github page. Here's the first page of Book 1. (Source:

formcept/whiteboard

FORMCEPT Whiteboard used by developers for developers - formcept/whiteboard

https://github.com/formcept/whiteboard/tree/master/nbviewer/notebooks/data/harrypotter

)

```
THE BOY WHO LIVED

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 Dursley s had a small son
called Dudley and in their opinion there was no finer
boy anywhere.

The Dursleys had everything they wanted, but they
also had a secret, and their greatest fear was that
somebody would discover it. They didn't think they
could bear it if anyone found out about the Potters.
Mrs. Potter was Mrs. Dursley's sister, but they hadn't

Page | 2 Harry Potter and the Philosophers Stone - J.K. Rowling
```

Next up was grabbing the character names. Wikipedia has an ~insanely~ comprehensive list. A little beautiful soup magic and bam: character list.

List of *Harry Potter* characters

From Wikipedia, the free encyclopedia

This article **does not cite any sources**. Please help improve this art

Find sources: "List of Harry Potter characters" – news · newspapers · books · sch

This is a **list of characters in the *Harry Potter* series**. They are all characters who have appeared in a *Hari*

Contents [hide]

1 Characters by surname
2 Characters with no or unknown surname
3 External links

Characters by surname [edit]

Contents

A · B · C · D · E · F · G · H · I · J · K · L · M · N · O · P · Q · R · S · T · U · V · W · X · Y · Z

A [edit]

- Hannah Abbott – Hufflepuff student in Harry Potter's year. Prefect and member of Dumbledore's Army.

B [edit]

- Ludo Bagman – Head of the Department of Magical Games and Sports within the Ministry of Magic.
- Bathilda Bagshot – Author of *A History of Magic*, and the great aunt of Gellert Grindelwald.
- Katie Bell – Gryffindor Quidditch Chaser one year above Harry Potter. Member of Dumbledore's Army.

```
['Aberforth Dumbledore',
 'Alastor (Mad-Eye) Moody',
 'Albert Runcorn',
 'Albus Dumbledore',
 'Albus Severus Potter',
 'Alecto Carrow',
 'Alice and Frank Longbottom',
 'Alicia Spinnet',
 'Amelia Bones',
 'Amos Diggory',
 'Amycus Carrow',
 'Andromeda Tonks',
 'Angelina Johnson',
 'Anthony Goldstein',
 'Antioch, Cadmus, and Ignotus Peverell',
 'Antonin Dolohov',
 'Arabella Figg',
 'Aragog',
 'Argus Filch',
```

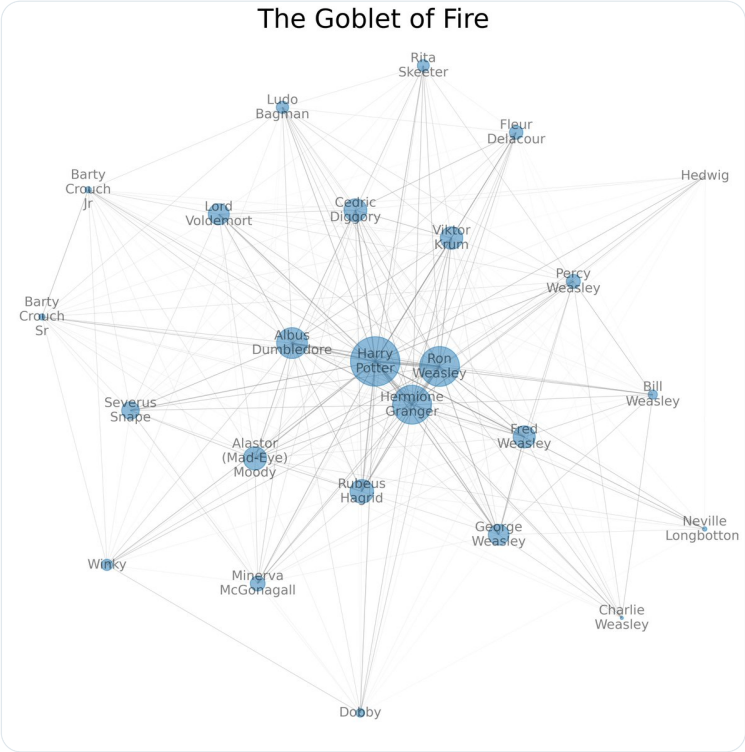
Since most of these names were very formal, I decided to build a synonym dictionary for each name. e.g. { 'Albus Dumbledore': [Albus Dumbledore', 'Albus', 'Dumbledore'] }

```
'Marvolo Gaunt': ['Marvolo Gaunt', 'Marvolo '],
'Mary Cattermole': ['Mary Cattermole', 'Mary '],
'Mary Riddle': ['Mary Riddle', 'Mary '],
'Merope Gaunt': ['Merope Gaunt', 'Merope '],
'Michael Corner': ['Michael Corner', 'Michael '],
'Millicent Bulstrode': ['Millicent Bulstrode', 'Millicent '],
'Minerva McGonagall': ['Minerva McGonagall', 'Minerva ', 'McGonagall
'Molly Weasley': ['Molly Weasley', 'Molly '],
'Morfin Gaunt': ['Morfin Gaunt', 'Morfin '],
'Mrs. Cole': ['Cole'],
'Mundungus Fletcher': ['Mundungus Fletcher', 'Mundungus '],
'Myrtle Warren': ['Myrtle Warren', 'Myrtle '],
'Nagini': ['Nagini', 'Nagini '],
```

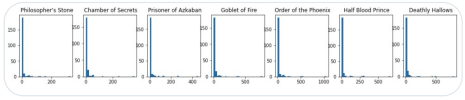
This step was all rather subjective and will eventually double count some characters with similar names but 🤷 we do our best.

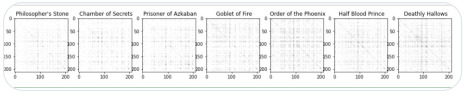
Edge A <--> B exist when A and B's name are on same page in the same book. Edges are weighted by # of co-occurrences. Each layer in the multiplex refers to a different book.

The resulting networks are shown in tweet #1. Here's a zoom in of the of Book 4: The Goblet of Fire (my personal favorite.)



Expectantly, the networks are really sparse and have heavy degree skew. The adjacencies and degree distributions are given here:



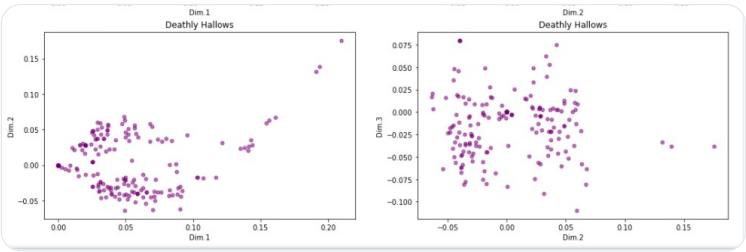


Using the degree distribution you can order the character's by # of appearances across different books. Here's the top 10 for each book.

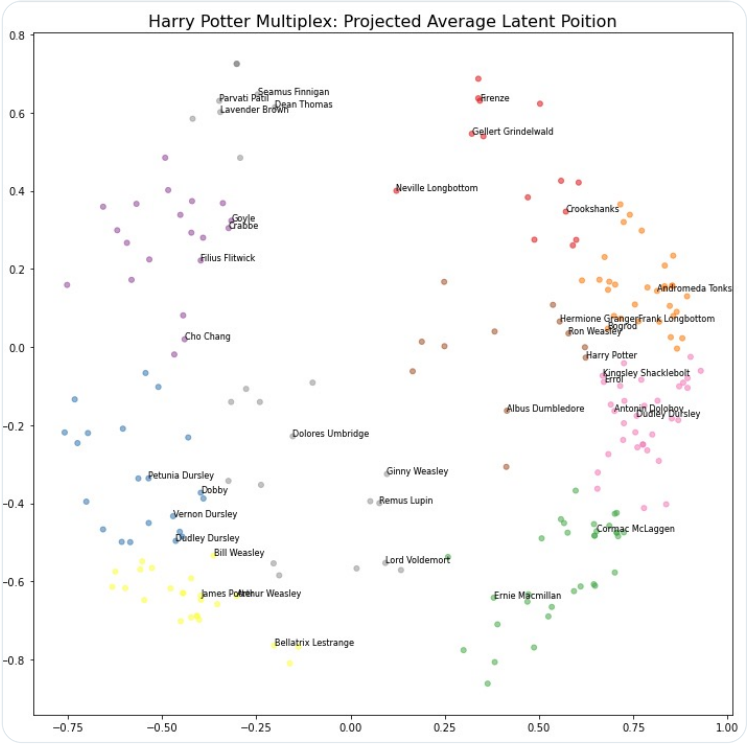
	Philosopher's Stone	Chamber of Secrets	Prisoner of Azkaban	Goblet of Fire	Order of the Phoenix	Half Blood Prince	Deathly Hallows
0	Harry Potter	Harry Potter	Harry Potter	Harry Potter	Harry Potter	Harry Potter	Harry Potter
1	Ron Weasley	Ron Weasley	Ron Weasley	Ron Weasley	Hermione Granger	Albus Dumbledore	Hermione Granger
2	Hermione Granger	Hermione Granger	Hermione Granger	Hermione Granger	Ron Weasley	Ron Weasley	Ron Weasley
3	Rubeus Hagrid	Albus Dumbledore	Remus Lupin	Albus Dumbledore	Albus Dumbledore	Hermione Granger	Albus Dumbledore
4	Albus Dumbledore	Fred Weasley	Sirius Black	Rubeus Hagrid	Dolores Umbridge	Horace Slughorn	Lord Voldemort
5	Severus Snape	George Weasley	Severus Snape	Alastor (Mad-Eye) Moody	Fred Weasley	Severus Snape	Severus Snape
6	Minerva McGonagall	Rubeus Hagrid	Rubeus Hagrid	Cedric Diggory	George Weasley	Ginny Weasley	Rubeus Hagrid
7	Dudley Dursley	Ginny Weasley	Albus Dumbledore	Viktor Krum	Ginny Weasley	Lord Voldemort	Ginny Weasley
8	Neville Longbottom	Minerva McGonagall	Minerva McGonagall	Fred Weasley	Rubeus Hagrid	Rubeus Hagrid	Remus Lupin
9	Vernon Dursley	Severus Snape	Crookshanks	George Weasley	Neville Longbottom	Draco Malfoy	Luna Lovegood

Next I tried to find a core component shared across layers to jointly embed (my dissertation topic 😎). I defined this component as any character that occurred on average over 3 times in a book. After rearranging the adjacencies, the core - periphery looks something like this.

Using some omnibus embedding magic (arXiv:1705.09355), the characters are mapped into \mathbb{R}^3. The characters not in the core component are then embedded using an out-of-sample method. Here's the embedded points for Book 7.



I then averaged node embeddings across layers and projected them onto the unit sphere (to correct for degree heterogeneity). The characters here are colored by a naive k-means with K = 10.



In conclusion - network science is dope and can be used for so many cool things. #networkscience

Code for all of these figures can be found here:

dravesb/OmniAnalysis
Data Analysis Completed with the Omnibus Embedding - dravesb/OmniAnalysis
https://github.com/dravesb/OmniAnalysis/tree/master/analysis/Harry_Potter

Also 🙂 Trans rights are human rights. So if you're able please consider donating:

Donate - Transgender Law Center
https://transgenderlawcenter.org/donate

• • •