# DOCTOR WHO FANDOM TEXT ANALYSIS



LAURIE CAGNEY
REGIS UNIVERSITY



# PRESENTATION OVERVIEW

- What is fandom?
- Introduce Dataset
- Cleaning Methods
- LDA Topic Model
- Final Thoughts
- Appendix



# WHAT IS FANDOM?

Fan-dom

noun

the state or condition of being a fan of someone or something.

"my 17 years of sports fandom"

the fans of a particular person, team, fictional series, etc. regarded collectively as a community or subculture.

"the Breaking Bad fandom"



Me with a Tardis at DragonCon 2019

# THE DOCTOR WHO FANDOM

- Doctor Who is a British TV show that has been airing since the 1960s, depicting the adventures of a time traveling alien and their companions they pick up along the way
- Fans are called "Whovians"



• This project focuses on one specific aspect of fandom: Fanfiction



# **FANFICTION**

- Fanfiction = Fan created works of fiction!
- Diverse and multifaceted
- Legality? (infamously: Anne Rice)













# THE QUESTION...

What topics are fanfiction authors writing about in the Doctor Who fandom that resonate with readers?



# INTRODUCING THE DATASET



# DATASET ORIGIN

- Scraped from ArchiveOfOurOwn.org
- Sorted by the ones with the top kudos (highest score)
- Completed only
- Rated Teen and below
- English only

1400 chapter one's of Doctor Who (out of 46,000 total)

### \*NOTE\*

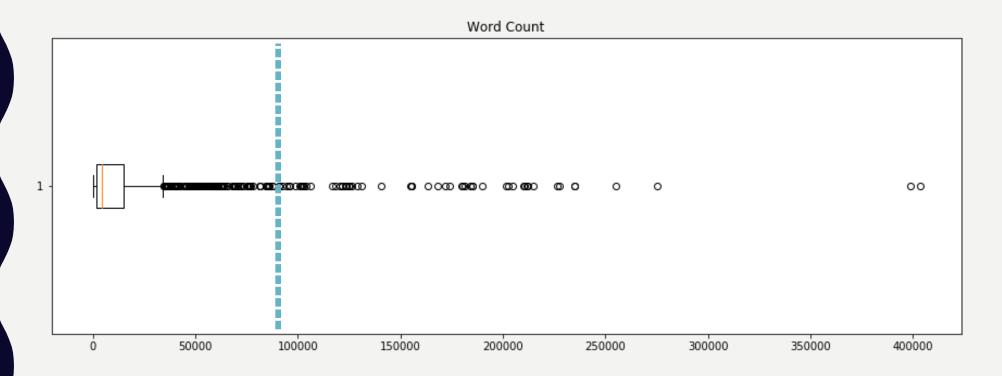
I used a prebuilt scraper that adhered to AO3's TOS and didn't collect any identifiable data from profiles nor anything that required logged in authentication



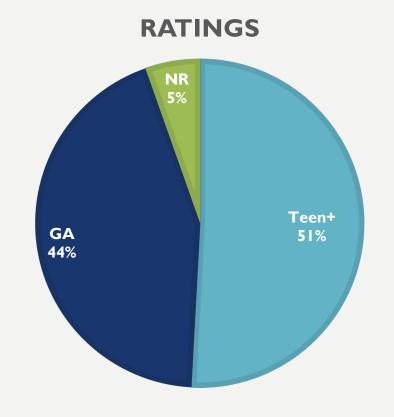
# HOW DO THESE STORIES COMPARE TO PUBLISHED NOVELS?

The average 200 page novel has about ~90,000 words...

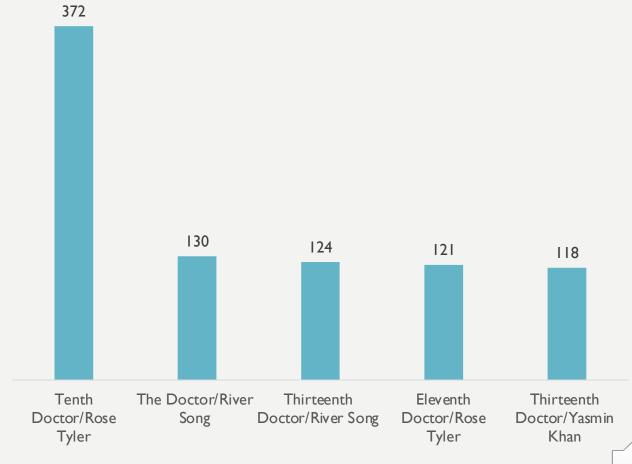
57 of these stories had more than that...



# OTHER DISTRIBUTIONS



### **RELATIONSHIP PAIRINGS**







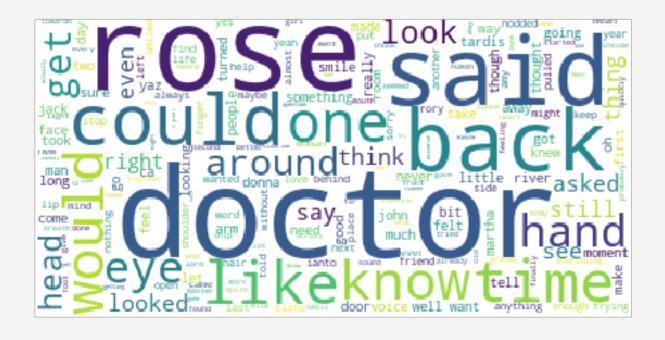
# **METHODS**

Tried Lemmatizing, Stemming, and Noun Only....





# WORD CLOUD





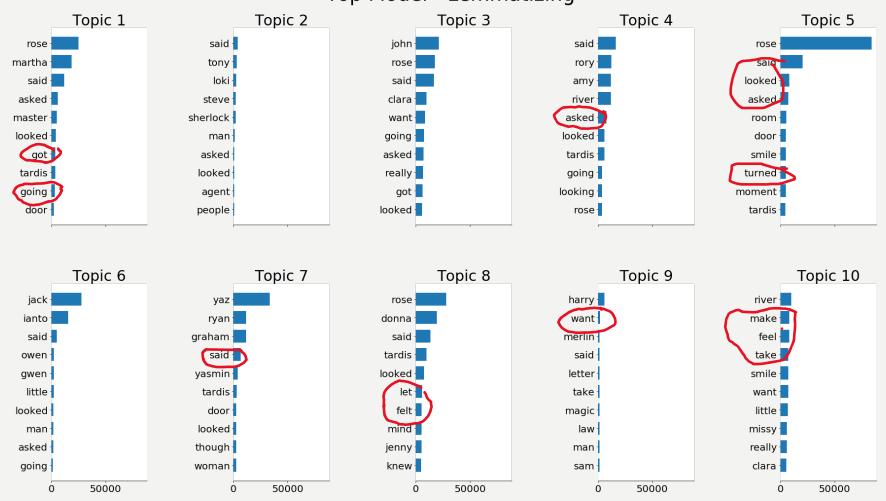
# TOPIC DELING

LATENT DIRICHLET ALLOCATION



# FIRST ITERATION

### Top Model - Lemmatizing





# UNDERSTANDING MAX\_DF

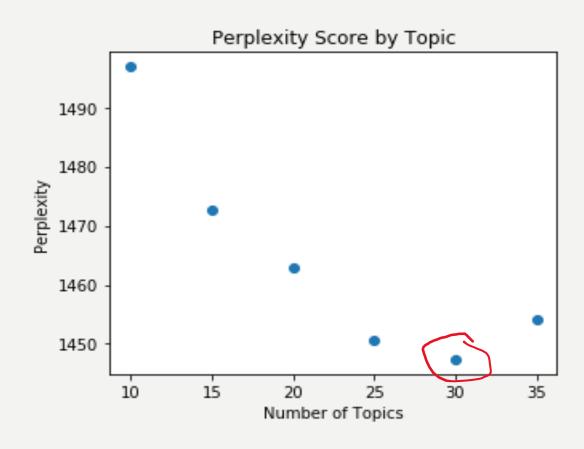
Document Frequency Based on Top Terms from ngrams/model I





# PERPLEXITY SCORE

### **DETERMINE NUMBER OF TOPICS**



Inputs:

 $Min_DF = 25$ 

 $Max_DF = 70\%$ 

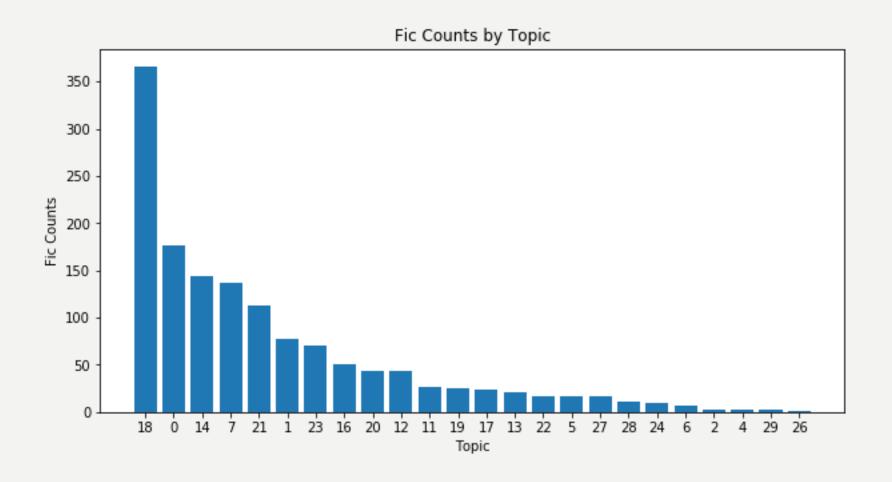
Word Length > =3

N\_Components tested:

[10,15,20,25,30,35]

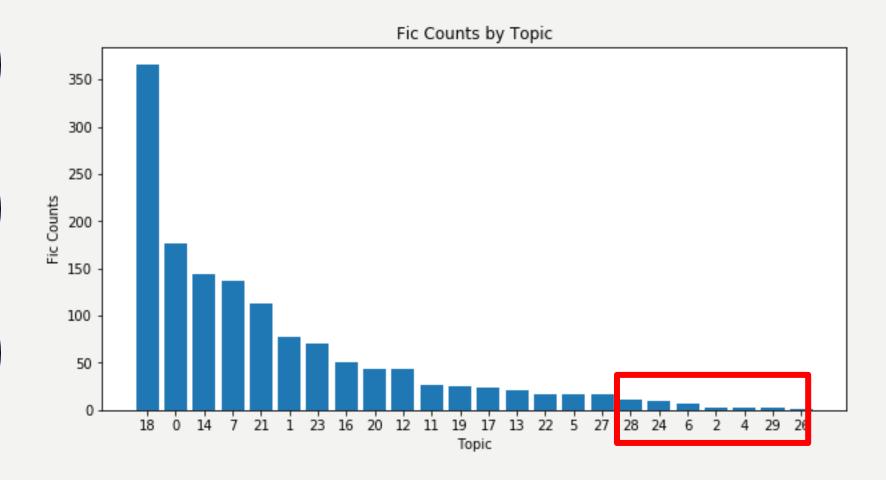
# EXPLORING TOPICS

# FIC COUNTS BY TOPIC



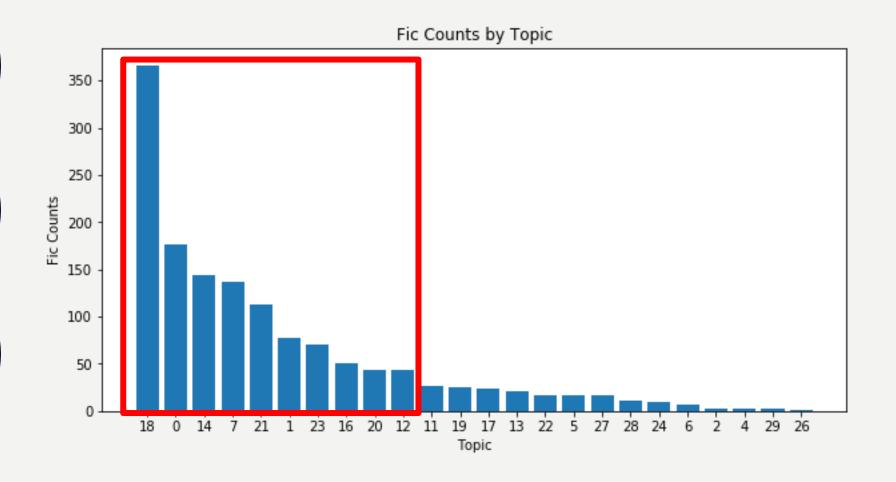


# FIC COUNTS BY TOPIC





# FIC COUNTS BY TOPIC





# TOPIC TO FIC MAPPING

### Every topic had words associated to it, each fic was assigned a topic

Topic 18 words				
body	0			
foot	1			
question	2			
memory	3			
help	4			
ship	5			
mouth	6			
sort	7			
feel	8			
planet	9			
air	10			
lord	11			
world	12			
hour	13			
floor	14			

topic	author	rk_id title	work_id	
18	['Riv_ika']	2116 My Fandom Needs Me	12292116	3
18	['Try2CatchMe']	9625 Past Tense	609625	4
18	['fingersfallingupwards']	8785 Connect-the-Dots	6218785	7
18	['esama']	8043 Fantastic Humanity	9448043	9
18	['dls']	7144 Stars Spell Out Your Name	10707144	13
18	['NoPondInTheForest']	1638 The Consequences of Dreaming	1241638	30
18	['LicieOIC']	2555 Heartsmith	1402555	39
18	['esama']	7304 Bargain	17287304	40
18	['morningsound15']	No Rose Without a Thorn	16371371	55
18	['NoPondInTheForest']	7700 The Aftermath of Forever	3587700	64



# TOP 10 TOPIC THEMES

### **BASED ON DOMINANT TOPIC SCORE**

Торіс	Theme
18	General Regeneration
0	13th Doctor Regeneration
14	TenToo AU
7	Rose as Bad Wolf
1	River Song's Husband

Topic	Theme
21	Falling in Love
13	Heaven and Hell Collide
16	The Family Pond
20	Time Lords
12	Friends of The Doctor

# **PULLED IN A NEW FIC**

### TITLE

Earthly Companion

### **SUMMARY**

Following the departure of Amy and Rory Pond, The Doctor's grief compels him to make a terribly irrational decision: use the Chameleon Arch to become human, and forget the pain of all that he's lost.

It's been decades since Rose Tyler discovered that she can't age. Through the power of Bad Wolf, she finds herself torn through the void, and pulled back to her original universe. She lands in Victorian London, picks up a job as a governess, and befriends a local recluse named Matthew Pond. He's a bit odd and far too young, but she can't help but feel drawn to him.

### **MY OPINION**

TOPIC "ROSE AS BAD WOLF" OR "TENTOO AU"



# **APPLY MODEL**

### #Clean fic

```
new_Fic.body = html.unescape(new_Fic.body)
new_Fic.body = new_Fic.body.replace('\n', ", regex=True)
new_fic_lem = new_Fic.body.apply(clean_text_lemmatizer)
new_fic_nouns = new_fic_lem.apply(noun_only)
```

### #Apply Model

```
new_fic_tf = cv.transform(new_fic_nouns)
new_fic_topics = model.transform(new_fic_tf)
```

### #Topic this fic fits in

new\_fic\_topics.argmax()



TOPIC 7
Rose as Bad Wolf







# **NEXT STEPS...**

- Subtopic modeling within the larger topics
- Understanding the fics in the smaller topics (seems to be crossovers)
- More comprehensive cleaning merging bigrams
- See if other topic model paradigms work better
- Better dataset: full stories, remove crossovers etc.

# FINAL THOUGHTS

- Required extensive domain knowledge
- Freeform text is difficult to process
- Interpretable topics are hard to create
- Lots of manual work, not easily scalable to other fandoms



# THANK YOU!





ALL THE EXTRAS

# LINKS

Full Topics with Weights

<u>GitHub</u>

Ao3 Scraper

Archive Of Our Own

Anne Rice (and others) vs. Fanfic Authors

# FANDOM DEFINITIONS

Term	Definition		
AU	Alternate Universe – timeline of the authors creation that is not canon compliant		
Canon	Material that is accepted to be an officially in-universe, there were no liberties taken to expand or change the source material.		
Companion	The Doctor's chosen partners while traveling time and space.		
Regeneration	When the Doctor is mortally wounded, they can heal themselves by regenerating into a new version (in other words, how the role gets recast)		
Tardis	The Doctor's means of traveling through time and space. Stands for Time and Relative Dimensions In Space.		
TenToo	Fandom name for when the Tenth Doctor partially regenerated into a copy of himself		
The Master	The doctor's arch nemesis		
Time Lords	The alien species that The Doctor belongs to; he is considered the last time lord along with The Master		

# FULL THEME LIST

Theme	Count
13th Doctor Regeneration	176
Falling in Love	77
Human AU	3
Omegaverse	2
Sarah Jane Adventures	16
White Collar Crossover	7
Rose as Bad Wolf	137
Avengers Crossover	26
Friends of the Doctor	43
Good Omens Crossover	21
TenToo AU	144
The Family Pond	51
Domestic Fluff	24
General Regeneration	366
Harry Potter Crossover	25
Time Lords	43
River Song's Husband	113
Sherlock Crossover	16
Heaven and Hell Collide	71
Royalty AU	9
Marriage Law	1
???	16
Camelot AU	11
Torchwood Sherlock	2
	I3th Doctor Regeneration Falling in Love Human AU Omegaverse Sarah Jane Adventures White Collar Crossover Rose as Bad Wolf Avengers Crossover Friends of the Doctor Good Omens Crossover TenToo AU The Family Pond Domestic Fluff General Regeneration Harry Potter Crossover Time Lords River Song's Husband Sherlock Crossover Heaven and Hell Collide Royalty AU Marriage Law ??? Camelot AU

No stories had the dominant topic for topics 3, 8, 9, 10, 15, 25

# QUICK STATS MEAN VALUES

Words

17,571

**Comments** 

80

Kudos

502

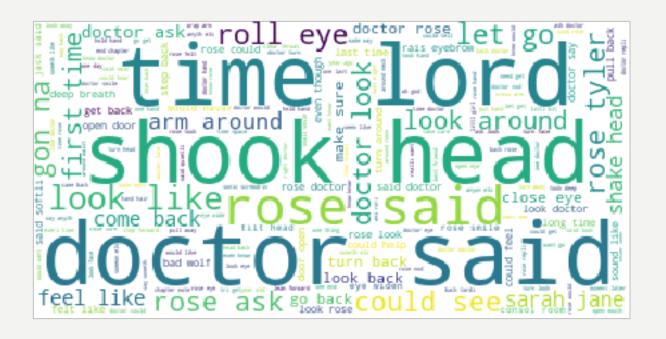
**Bookmarks** 

70

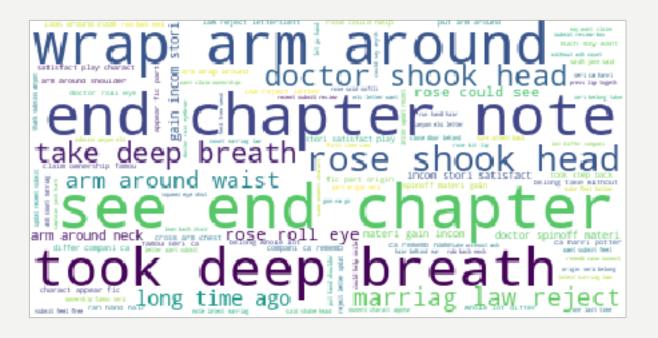
Hits

6,015

# **BIGRAMS**



# **TRIGRAMS**



# POPULAR FICS OVER TIME

