

“Everyone is Gay and Nothing Hurts”: The Presentation of Queer Relationships in Fanfiction

Michael Miller Yoder
Luke Breitfeller



Carnegie Mellon University
Language Technologies Institute

NLP and LGBTQ Issues

- Existing focus is on identifying and quarantining hate speech in “mainstream” corpora (Dinakar et al. 2012, Djuric et al. 2015, Waseem 2016)
- Less effort into studying queer language within queer-friendly spaces.

A Different Corpus: Archive Of Our Own

Archive Of Our Own (AO3) is a noncommercial database for fanfiction and transformative works, gaining popularity as a source of data in the digital humanities (Milli and Bamman 2016)

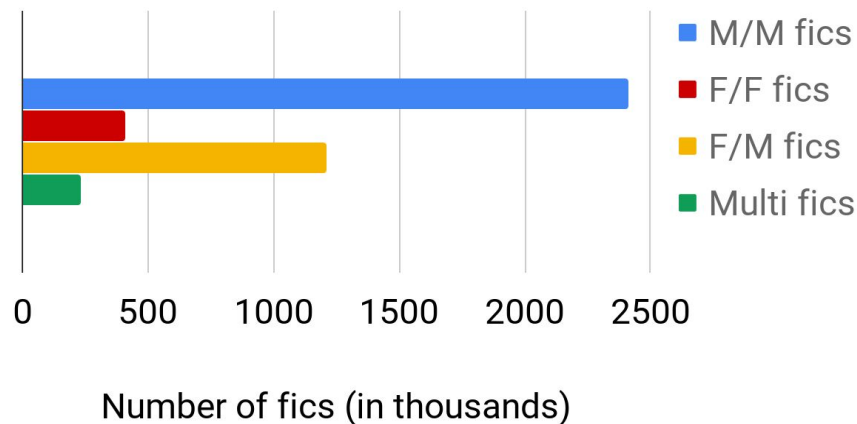
- Large--almost 5 million works, more each day
- “Average” writers, contemporary
- Metadata: author annotation in form of authors’ notes and “tags” describing the work in great detail

AO3: A Queer Space

Fanfiction writers are often queer and write queer stories to fix a lack of queer content in traditional media (Lothian et al. 2007)

Community mores are very pro-queer (cishet stories often framed as objectionable), with few open queerphobes.

Relationship types in AO3



there's a straight scene sorry (1)

I cannot believe I wrote something with an heterosexual couple in it (1)

theres like 1 straight couple im sorry but theyre cool (1)

Not Every Character is Queer but the Writer Is Not Overly Interested In Being Polite or Heterosexual (1)

The Question

Queerness is present in AO3, but there's still debate over whether this is “good” representation of queer issues.

- Ex. little representation for queer women, fetishization, queer relationships without using explicit queer labels

The problem: AO3 is a context-rich corpus, and traditional NLP techniques are designed to work in a lack of context. Traditional NLP techniques lead to naive conclusions.

Methods: Projecting words onto “semantic axes”

- Goal: measure and inspect how LGBTQ characters and concepts are presented differently across corpora

Methods: Projecting words onto “semantic axes”

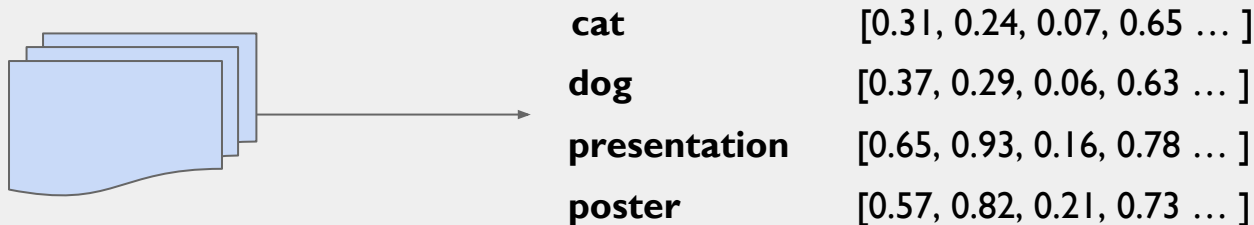
- Goal: measure and inspect how LGBTQ characters and concepts are presented differently across corpora
- Semantic word vectors (Mikolov et al. 2013, Bojanowski et al. 2016)

Methods: Projecting words onto “semantic axes”

- Goal: measure and inspect how LGBTQ characters and concepts are presented differently across corpora
- Semantic word vectors (Mikolov et al. 2013, Bojanowski et al. 2016)
 - Words that occur in similar contexts have similar vectors

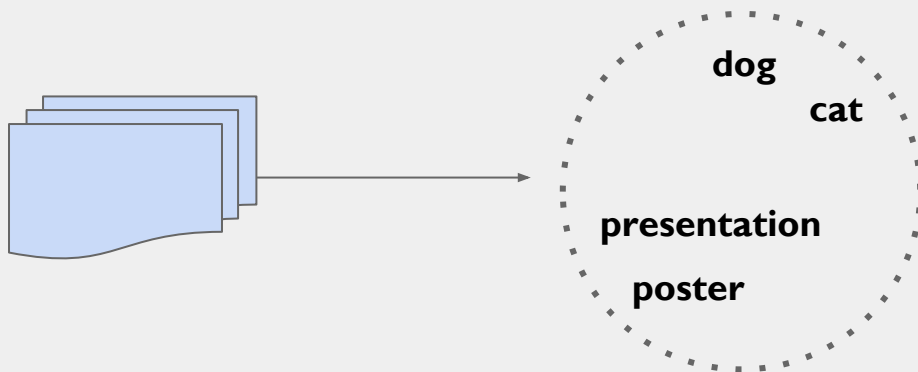
Methods: Projecting words onto “semantic axes”

- Goal: measure and inspect how LGBTQ characters and concepts are presented differently across corpora
- Semantic word vectors (Mikolov et al. 2013, Bojanowski et al. 2016)
 - Words that occur in similar contexts have similar vectors



Methods: Projecting words onto “semantic axes”

- Goal: measure and inspect how LGBTQ characters and concepts are presented differently across corpora
- Semantic word vectors (Mikolov et al. 2013, Bojanowski et al. 2016)
 - Words that occur in similar contexts have similar vectors

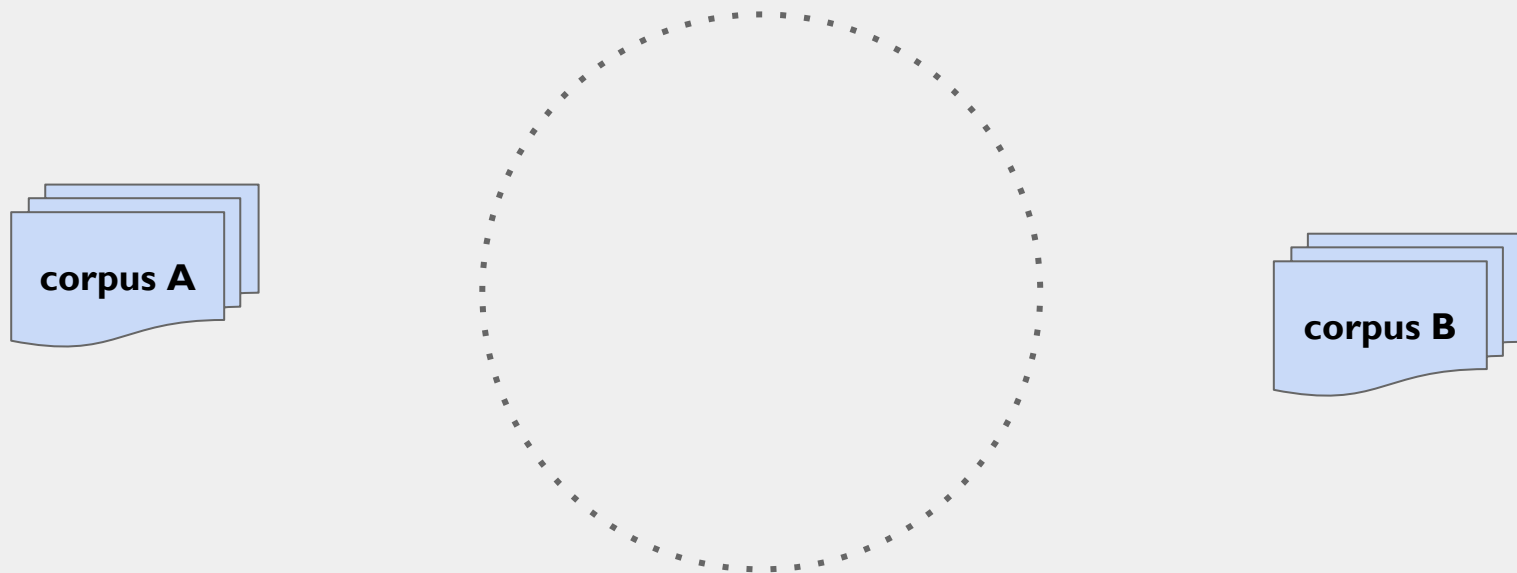


Methods: Projecting words onto “semantic axes”

- Project word vectors adapted to different corpora on semantic “axes” of interest (An et al. 2018)

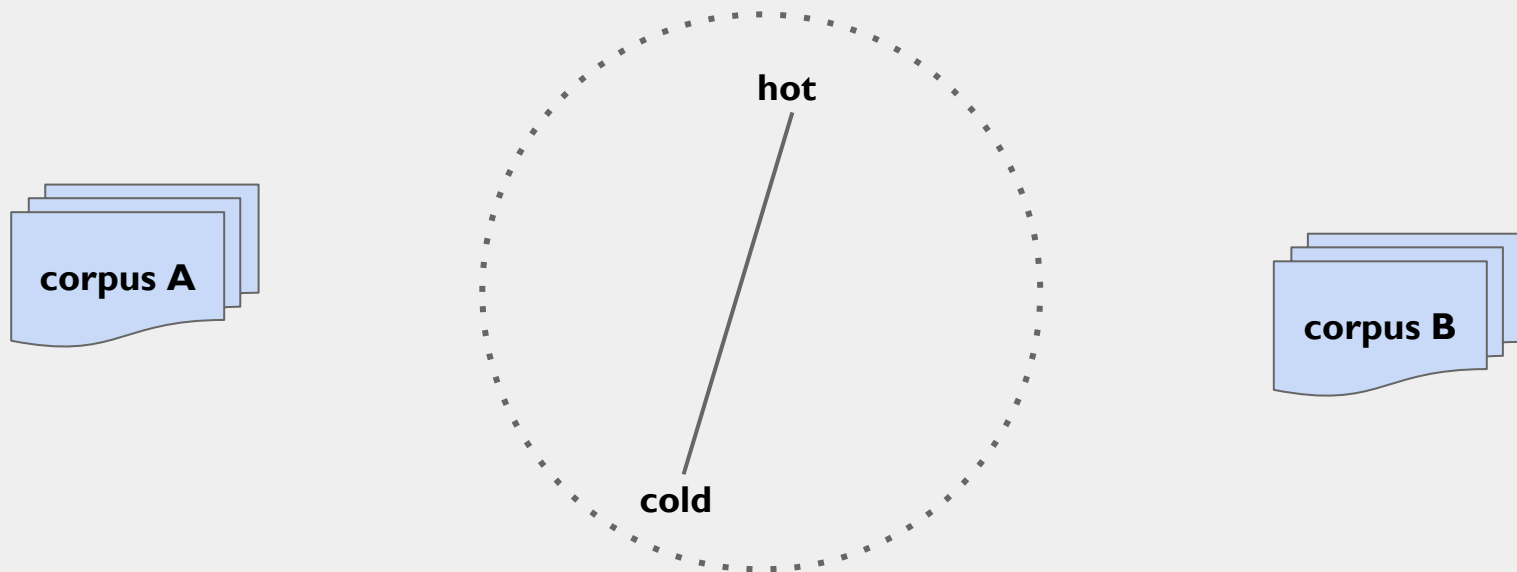
Methods: Projecting words onto “semantic axes”

- Project word vectors adapted to different corpora on semantic “axes” of interest (An et al. 2018)



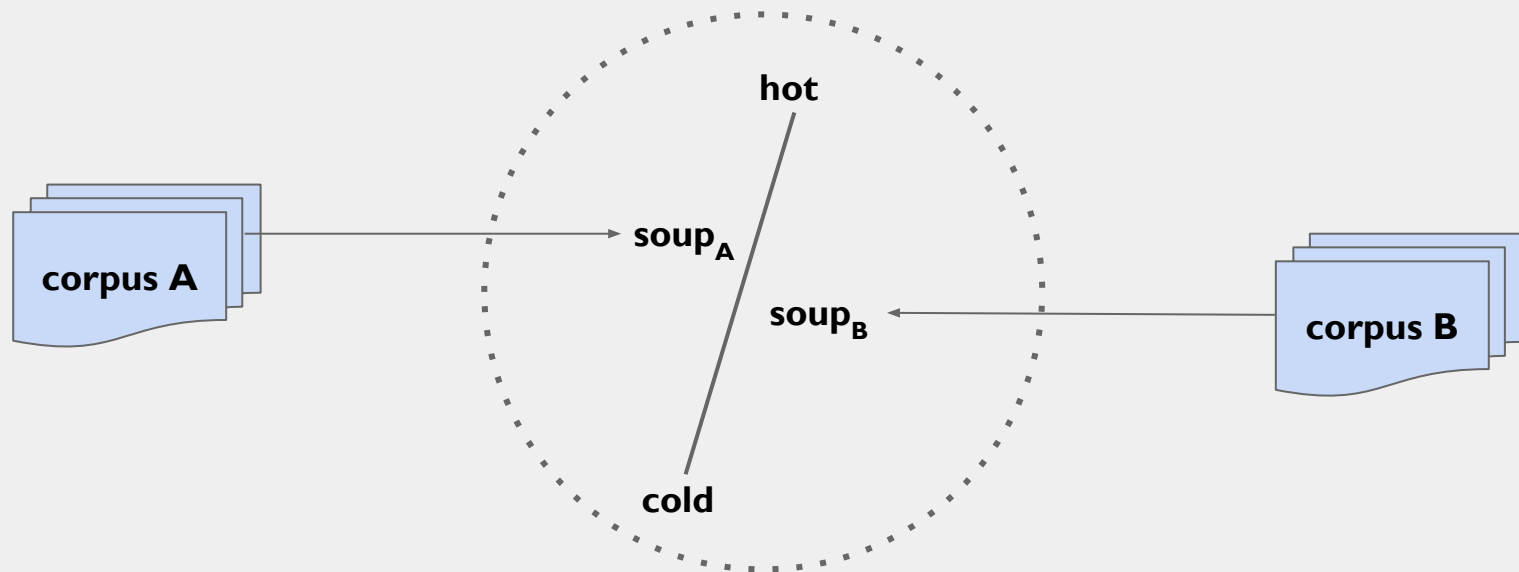
Methods: Projecting words onto “semantic axes”

- Project word vectors adapted to different corpora on semantic “axes” of interest (An et al. 2018)



Methods: Projecting words onto “semantic axes”

- Project word vectors adapted to different corpora on semantic “axes” of interest (An et al. 2018)



Methods: Projecting words onto “semantic axes”

- Which semantic dimensions are relevant to the presentation of identity?

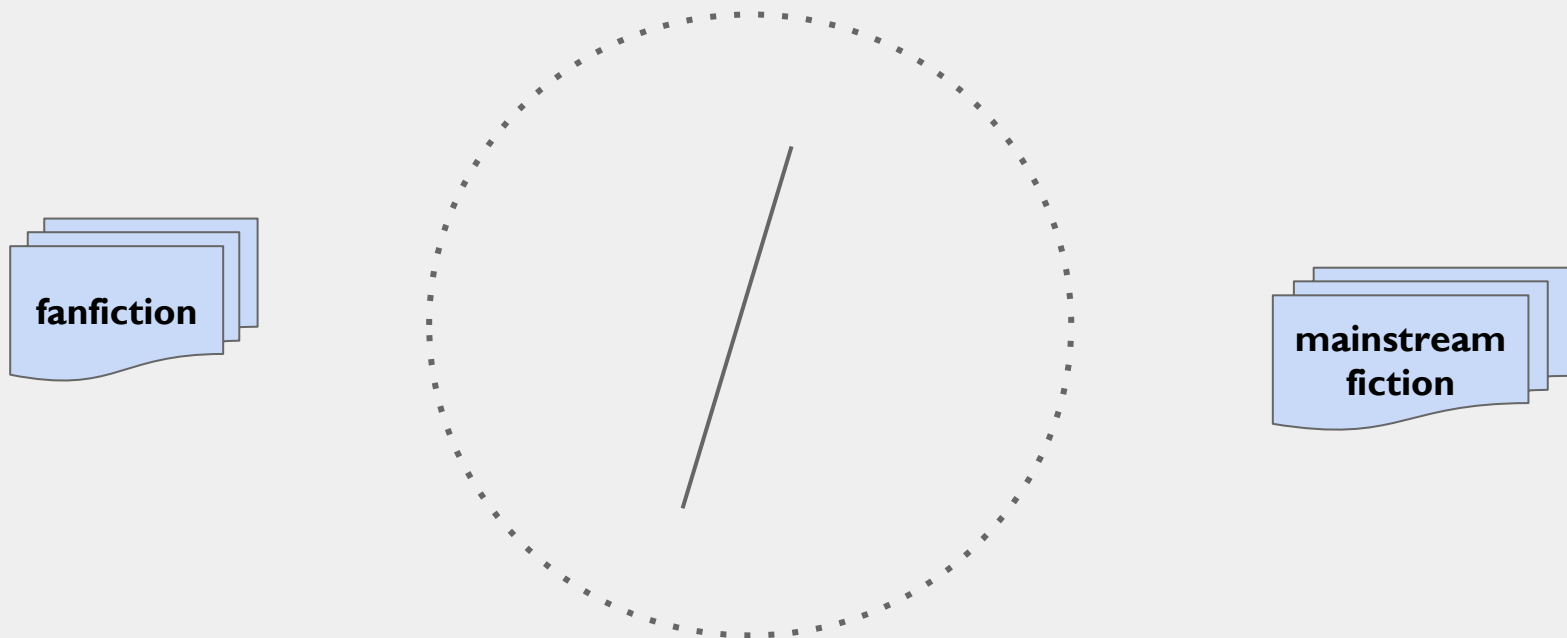
Methods: Projecting words onto “semantic axes”

- Which semantic dimensions are relevant to the presentation of identity?
- *Theorizing Identity in Language and Sexuality Research*, Bucholtz and Hall (2004)

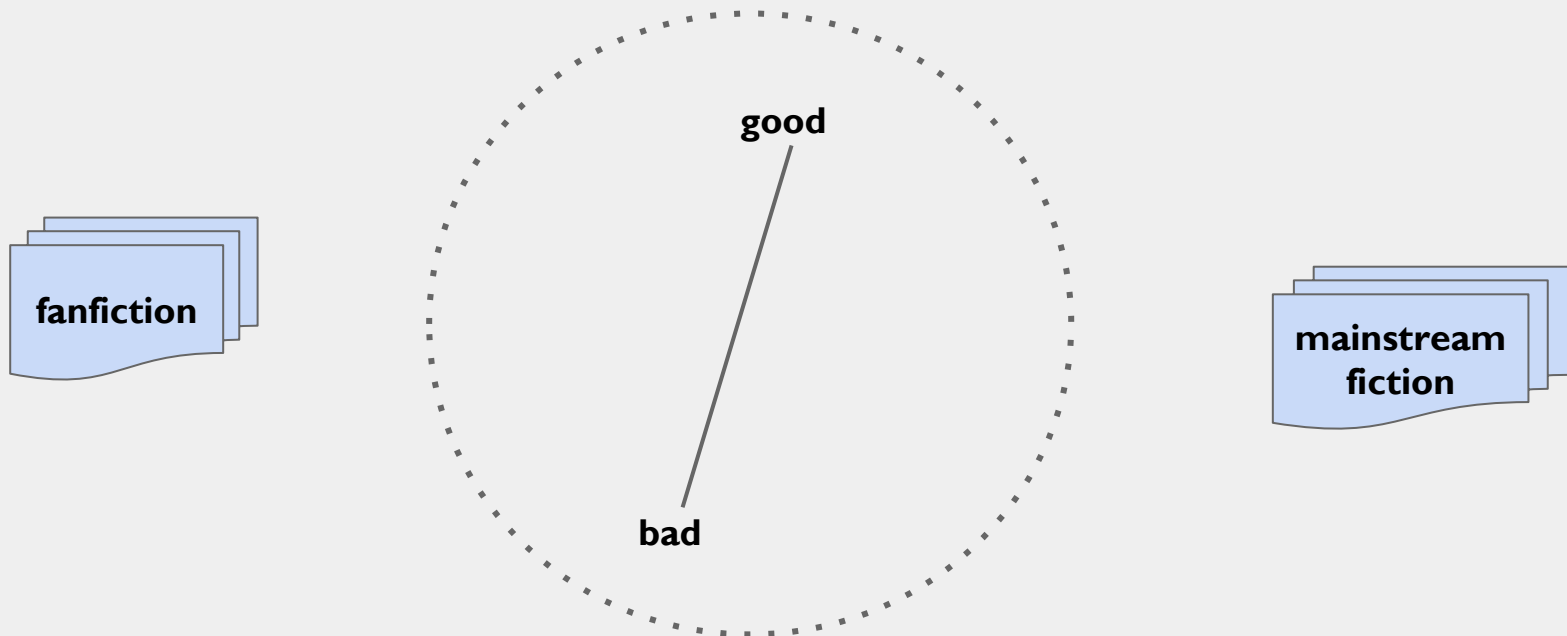
Methods: Projecting words onto “semantic axes”

- Which semantic dimensions are relevant to the presentation of identity?
- *Theorizing Identity in Language and Sexuality Research*, Bucholtz and Hall (2004)
 - Recognition vs illegitimation (*good/bad*)
 - Sameness vs difference (*same/different*)

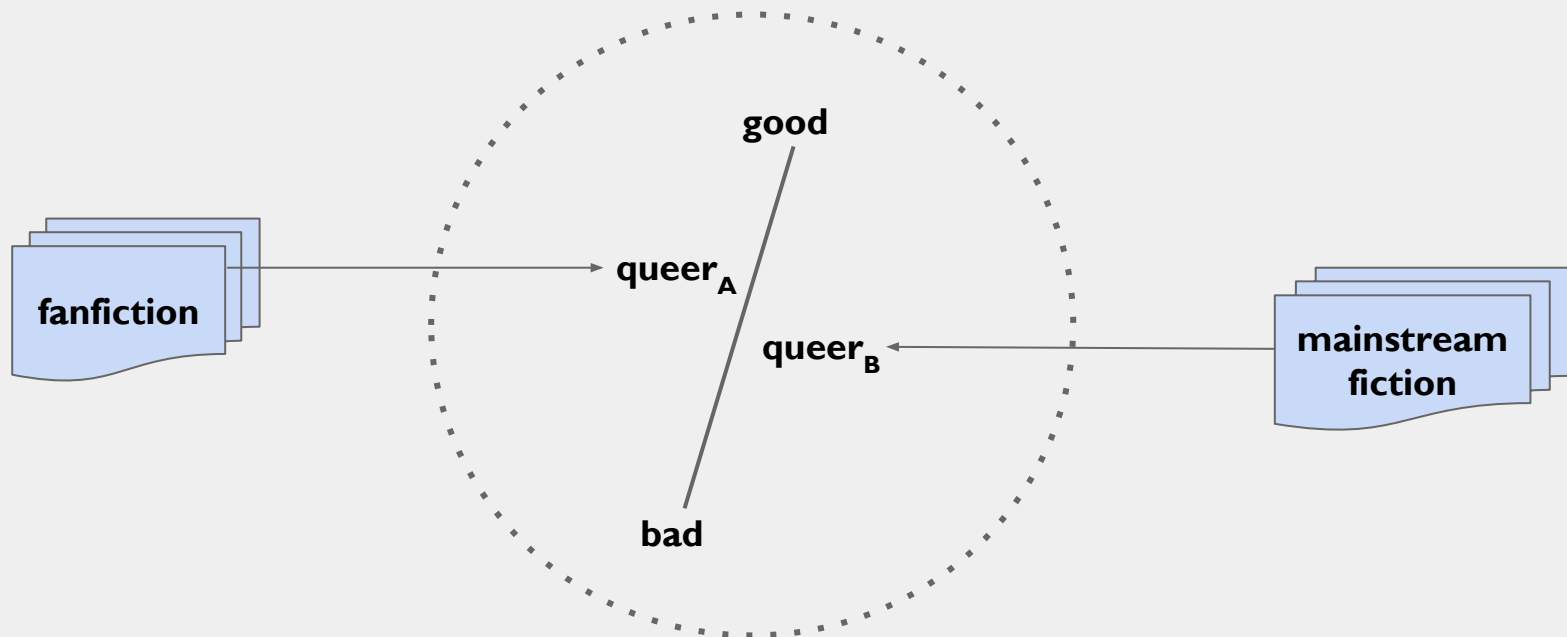
Methods: Projecting words onto “semantic axes”



Methods: Projecting words onto “semantic axes”




Methods: Projecting words onto “semantic axes”



Fanfiction data

- 7 most popular fandoms on Archive Of Our Own
 - Marvel
 - Supernatural
 - DC
 - Sherlock Holmes
 - Teen Wolf
 - Star Wars
- 21,000 stories
- 150 million words

 Archive of Our Own *beta*

Fandoms Browse Search About

Rating: Teen And Up Audiences

Archive Warning: No Archive Warnings Apply

Category: M/M

Fandom: Harry Potter - J. K. Rowling

Relationships: Draco Malfoy/Harry Potter, Harry Potter/Ginny Weasley

Characters: Draco Malfoy, Harry Potter, Astoria Greengrass, Daphne Greengrass, G
James Sirius Potter, Lily Luna Potter, Hermione Granger, Ron Weasley,

Additional Tags: Minor Character Death, Emotional/Psychological Abuse, Physical Abuse,
Harry Potter, Number Twelve Grimmauld Place, Draco Malfoy & Harry Pot
Scorpius Malfoy & Albus Severus Potter Friendship, Abusive Relationships

Language: English

Stats: Published: 2019-05-08 Updated: 2019-05-08 Words: 8324 Chapters

How Love Hurts

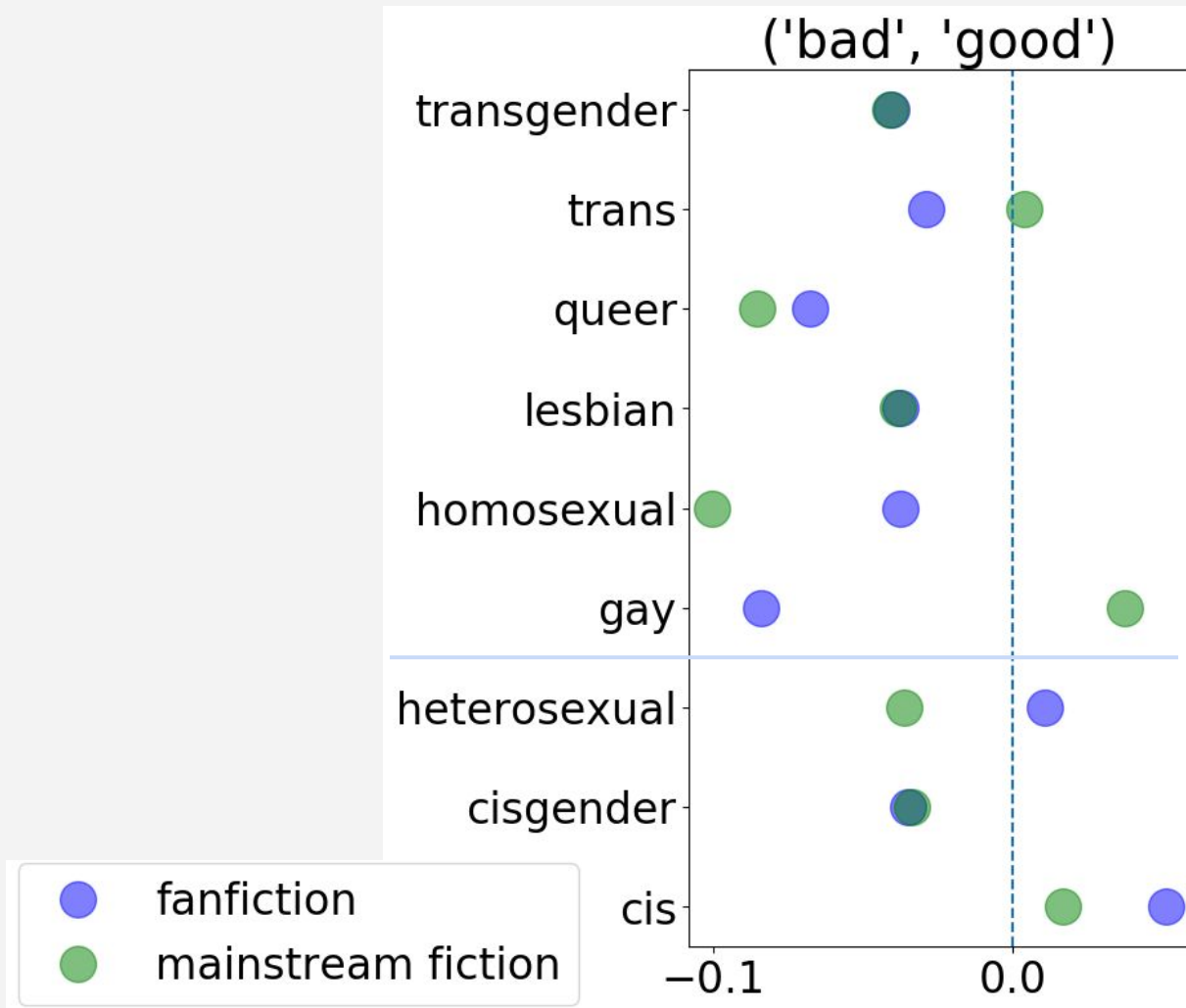
LindyA1985

Summary:

Mainstream fiction data

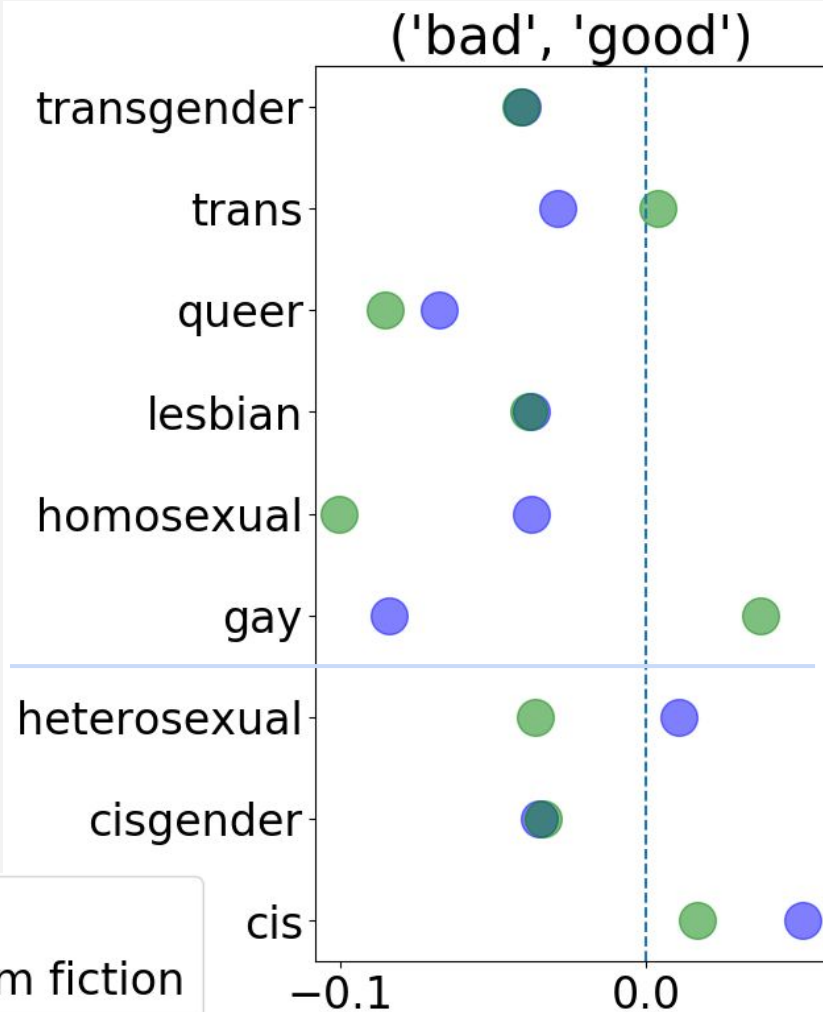
- Corpus of Contemporary American English (COCA)
 - 1990-2017
- Fiction section includes
 - short stories and plays from literary magazines
 - children's magazines
 - popular magazines
 - first chapters of first edition books
 - movie scripts
- 128 million words

Results



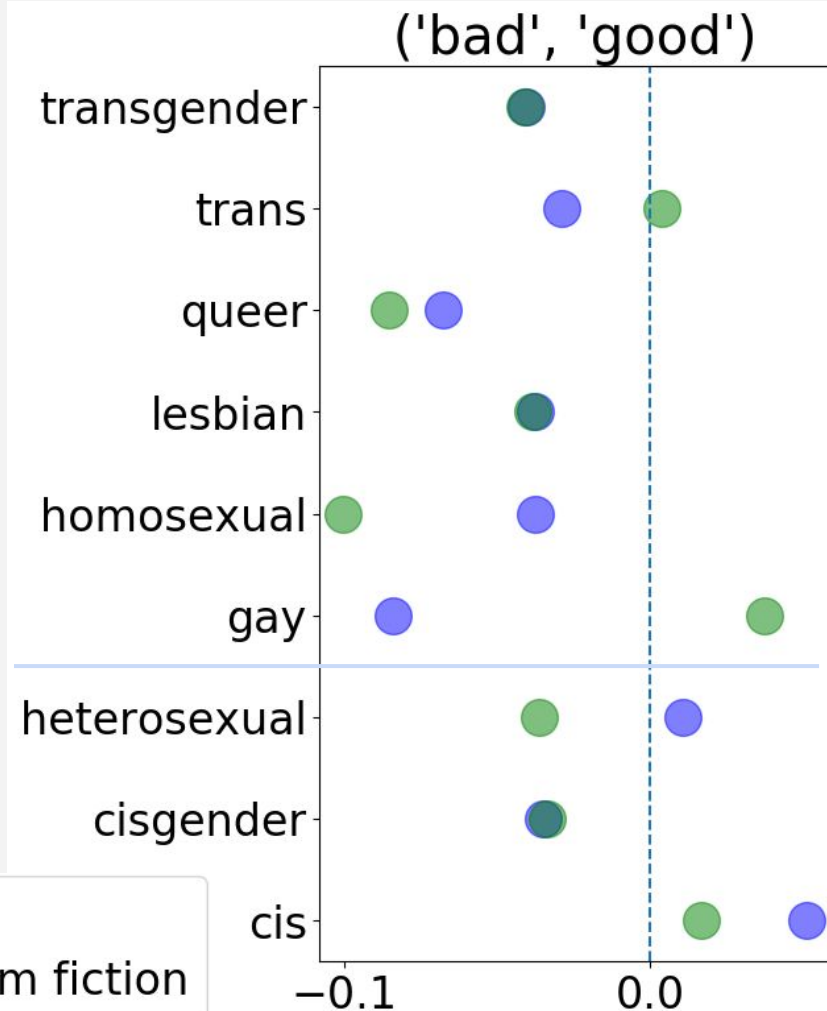
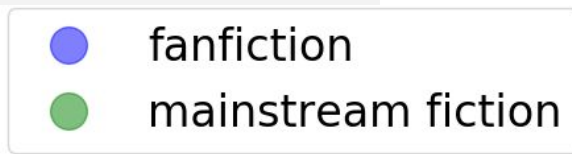
Results

- Sometimes expected: 'queer' more toward *good* in fanfiction



Results

- Sometimes expected: 'queer' more toward *good* in fanfiction
- Otherwise not: 'gay' much more toward *good* in mainstream fiction



The Paradigm Shift

Using only direct linguistic observation ignores social motivation of fanfiction.

- Writing on AO3 skews towards negative subjects overall (“angst”)

As a next step, we utilize a social examination of the AO3 author group and the significant advantage of AO3--its user-annotated metadata--to create a new paradigm for studying this space.

Social Analysis

Authors have specific, varied motivations for writing fanfic.

The “vent fic”: stories by queer authors about discrimination, written for emotional catharsis.

“Aspirational” fics: stories by queer authors fantasizing about worlds without discrimination.

(Mostly) identifiable by metadata tags.

Methodology: Sub-Corpora

We isolate two sub-corpora from AO3 based on user-generated metadata tags, “challenges” versus “aspirations”.

Idea: use these sub-corpora as more nuanced baselines.

Everybody's Gay!, everybodys gay, Everybodys gay and you can fight me, everyoen else is gay, everyone
fic is gay fight me, everyone in this series is gay ok, everyone in this show is homosexua
ie is a little gay in this, everyone is a tad gay, everyone is also gay, everyone is at least a little gay, Everyo
everyone is gay !, Everyone Is Gay (mostly), Everyone Is Gay . . . kind of, EVERYONE IS GAY AF, everyone is
sorry, everyone is gay and im not sorry, everyone is gay and it is okay, everyone is gay and its great, ever
ause my boys deserve to be happy, everyone is gay and nothing hurts, everyone is gay and thats just how
AS everyone is gay as , everyone is gay at some point, Everyone is Gay AU, everyone is gay bas
h it, everyone is gay dont @ me, everyone is gay fight me, Everyone is gay for some reason, everyone is c
make the rules, everyone is gay in chapter 4 and probably other future updates, everyone is gay in some w
ven funny, Everyone is gay lmao, everyone is gay noone is straight, everyone is gay one way or another In
one is gay too, everyone is gay!, everyone is GAY!!!, Everyone Is Gay...cuz i wanna, Everyone Is Gay.mp3,

Methodology: Corpora Selection

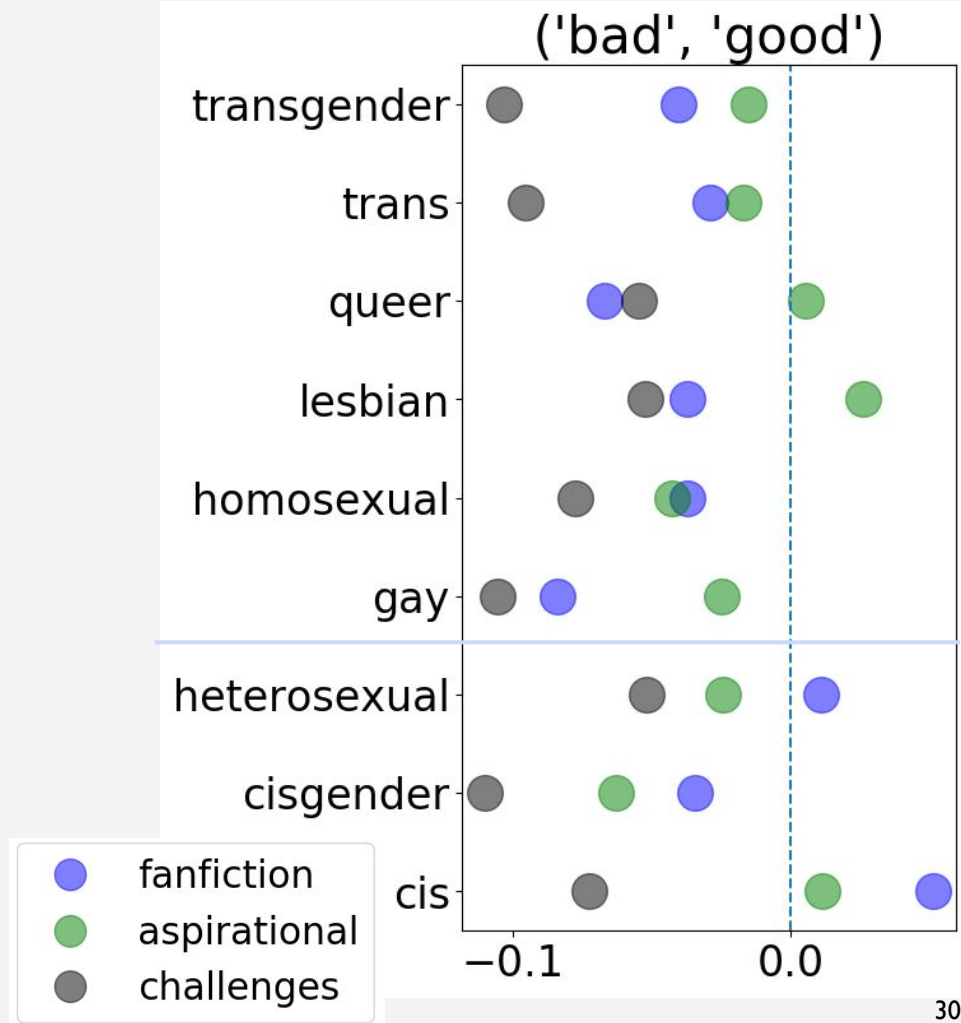
“Challenge fic” (6k total):

- “Homophobia”
- “Transphobia”
- “Biphobia”
- “Closeted Character”/“Outing”
- “Slurs”/“Hate Crimes”
- “Internalized Homophobia”
- “Outing”

“Aspiration fic” (7k total):

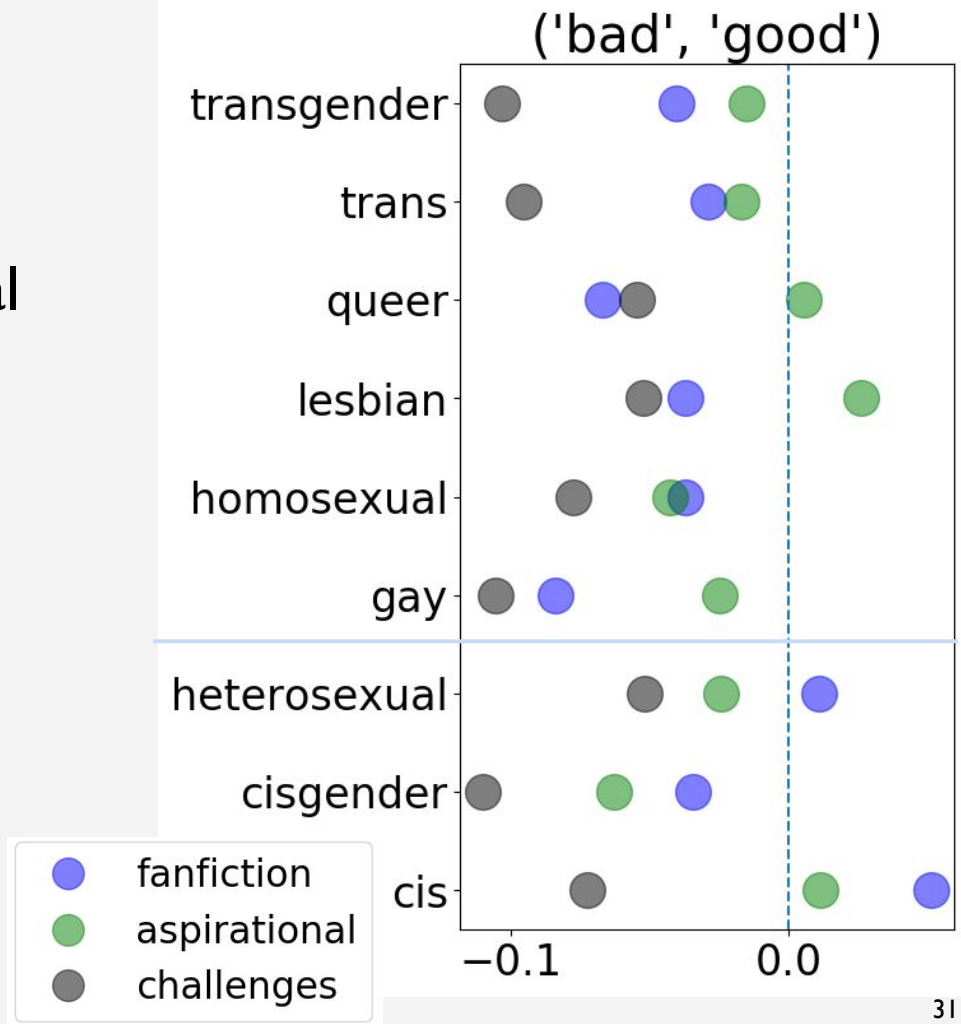
- “LGBTQ Themes”
- “Coming Out”
- “Gay Character”
- “Lesbian Character”
- “Bi Character”
- “Trans Character”

Results



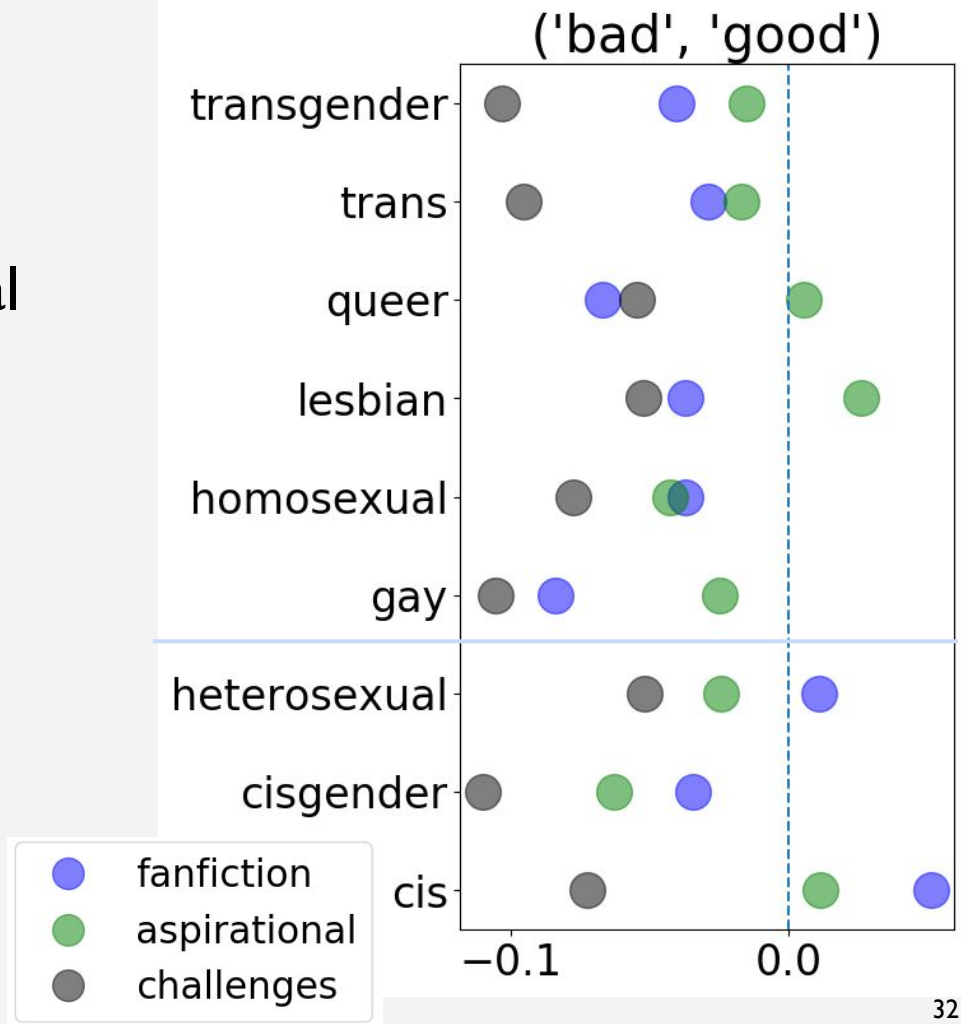
Results

- General trend that aspirational fics have queer terms closest to *good* overall, fits intuition



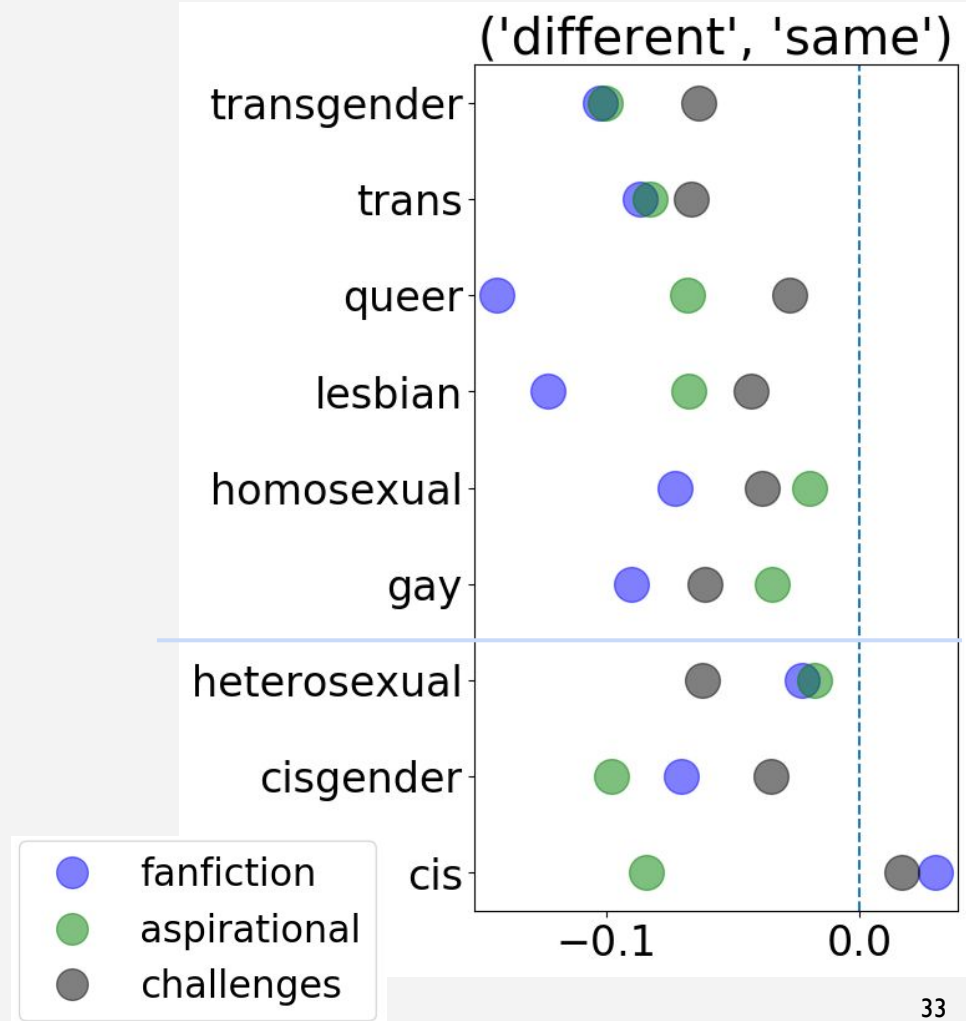
Results

- General trend that aspirational fics have queer terms closest to *good* overall, fits intuition
- “Normative” terms closer to *bad* in aspirational fics than random selection of fanfic, even closer in challenge fics



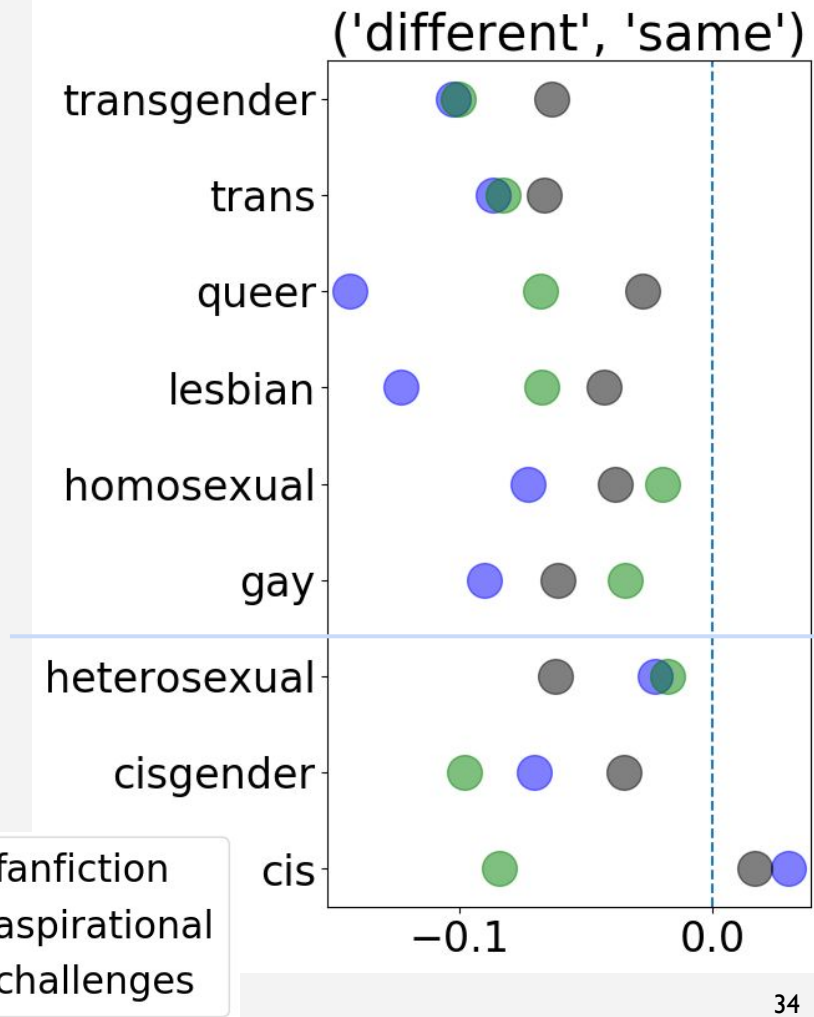
Results

- Results less clear for this axis



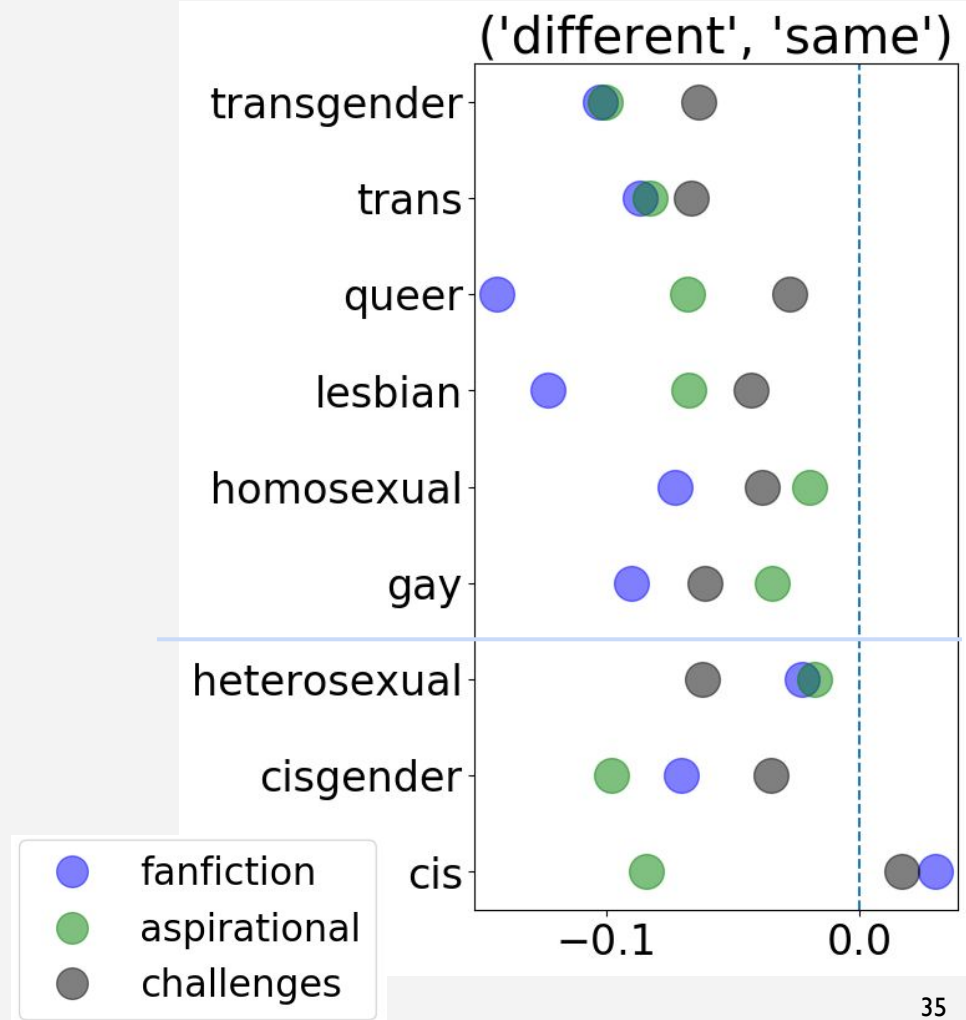
Results

- Results less clear for this axis
- In general, terms in challenges closer to *same* than aspirational



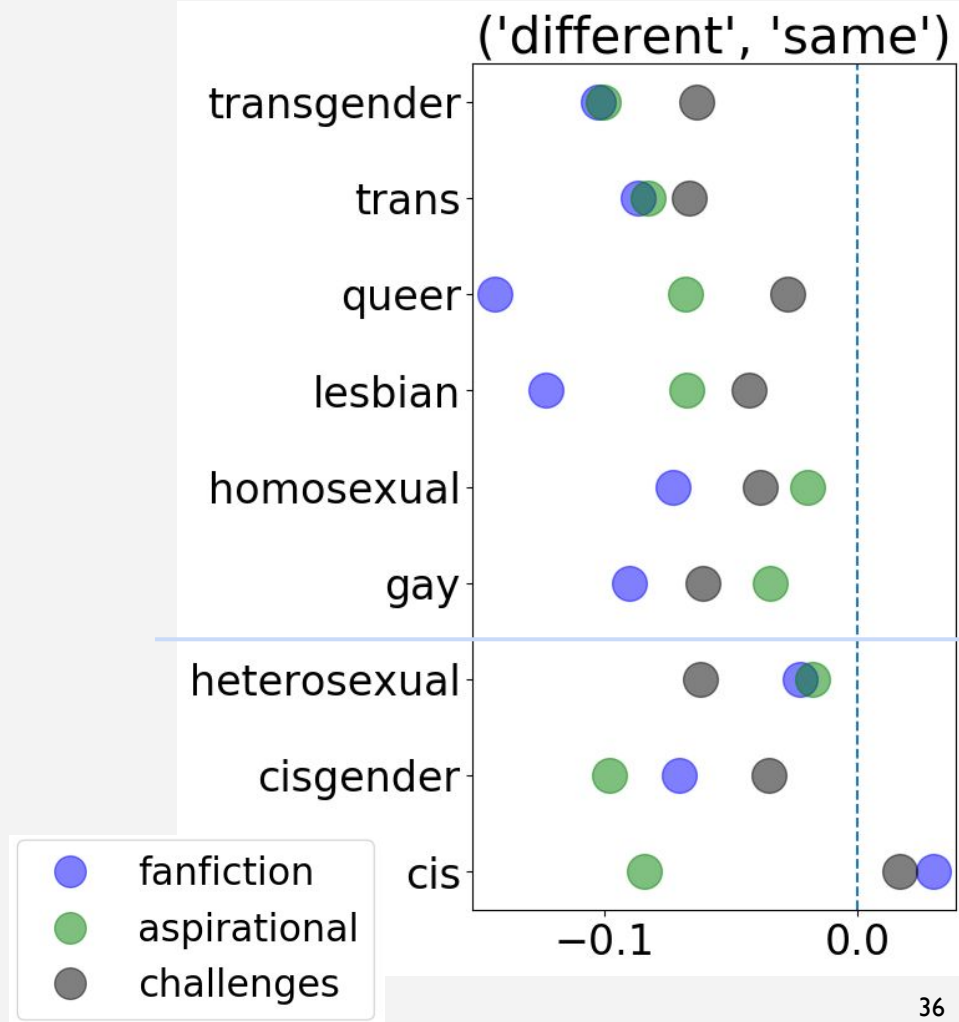
Results

- Terms in challenges closer to *same* than aspirational:



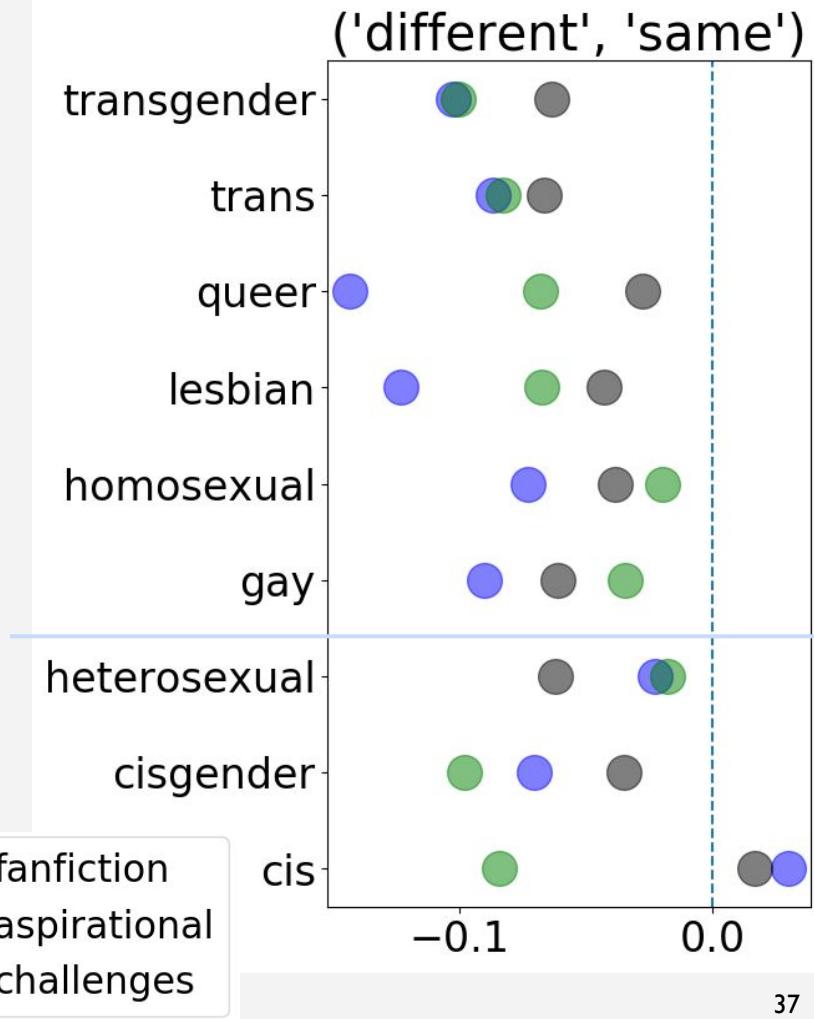
Results

- Terms in challenges closer to *same* than aspirational:
 - “I’m not a queer.”



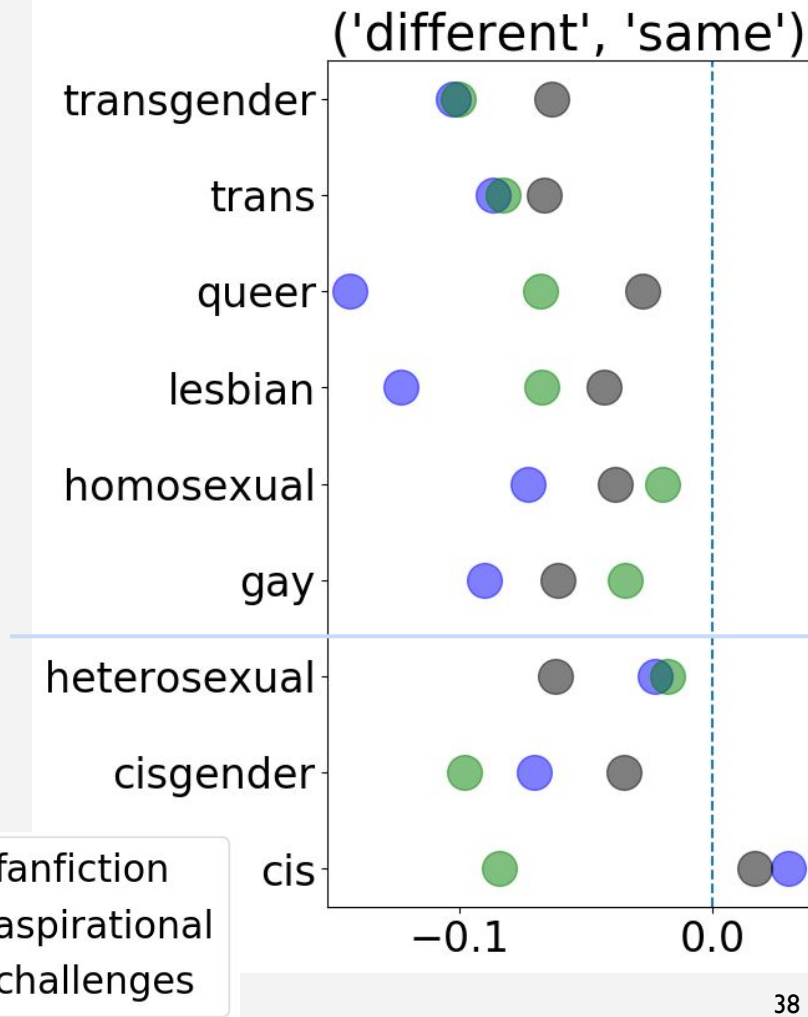
Results

- Terms in challenges closer to *same* than aspirational:
 - “Im not a queer.”
 - ‘I’ close to *same* in challenge fics



Results

- Terms in challenges closer to *same* than aspirational:
 - “Im not a queer.”
 - ‘I’ close to *same* in challenge fics.
 - Challenge fics often come from personal experience, “same” as author.



Conclusion

- Language lets us indirectly measure treatment and acceptance of marginalized groups
- Word vectors alone ignore social context unique to marginalized groups
- Metadata allows us to shape input in a way that captures nuance missing from straightforward NLP

Future work

- Manually locate better corpus splits for other semantic axes
- Automatically identify meaningful corpus splits based on word vector variation
- Indirect associations with gender and sexuality made where stories contain queer relationships, but do not mention queer identity labels

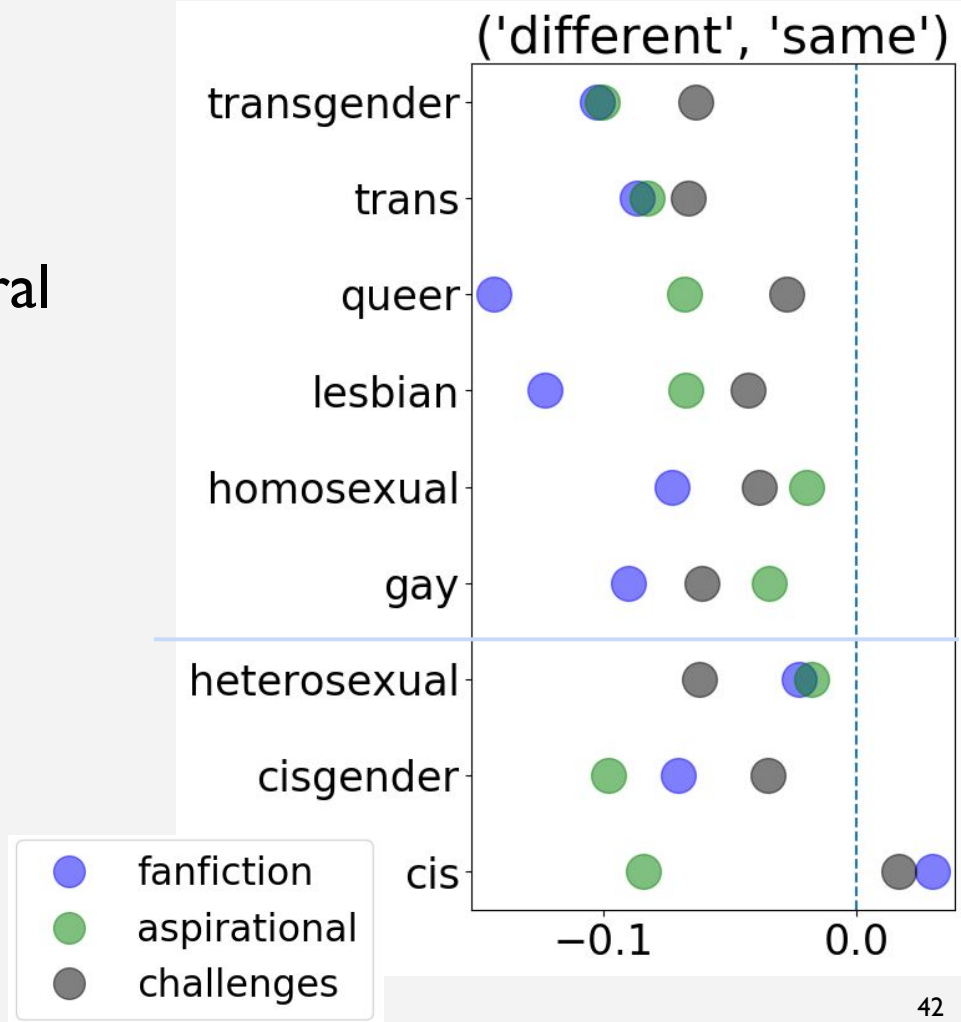
Thank you!

Luke Breitfeller mbreitfe@andrew.cmu.edu

Michael Miller Yoder yoder@cs.cmu.edu

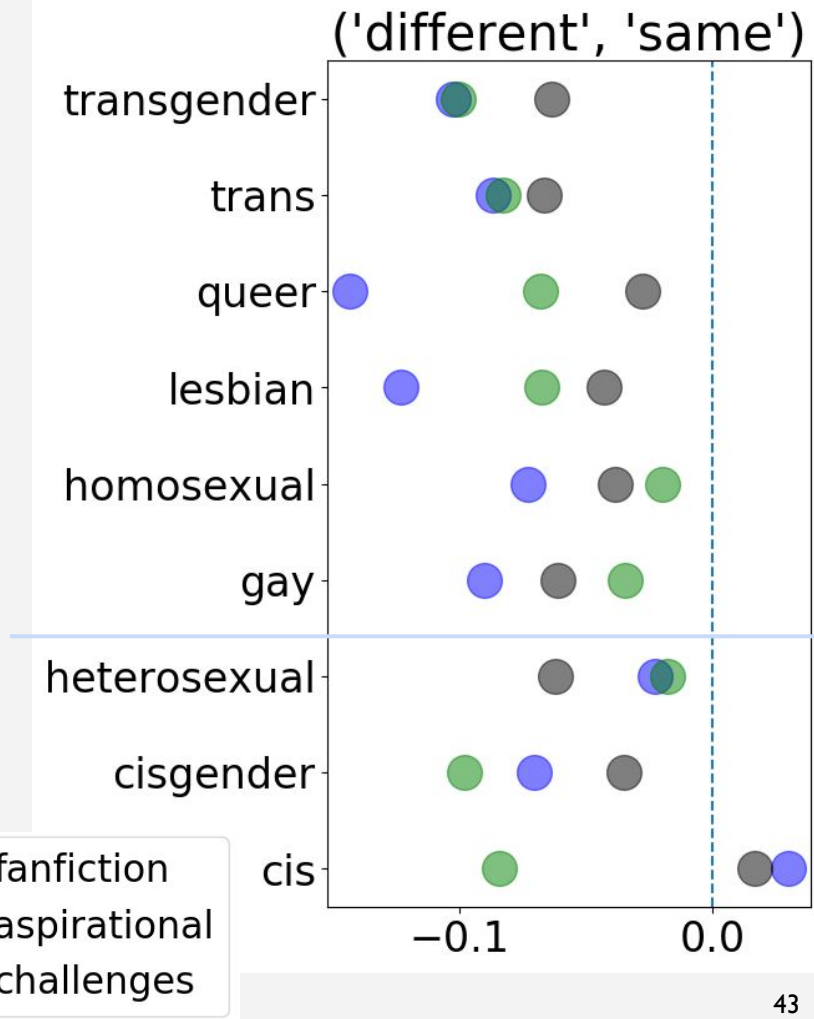
Results

- 'Queer', 'lesbian', 'gay' in general fanfic toward *different*

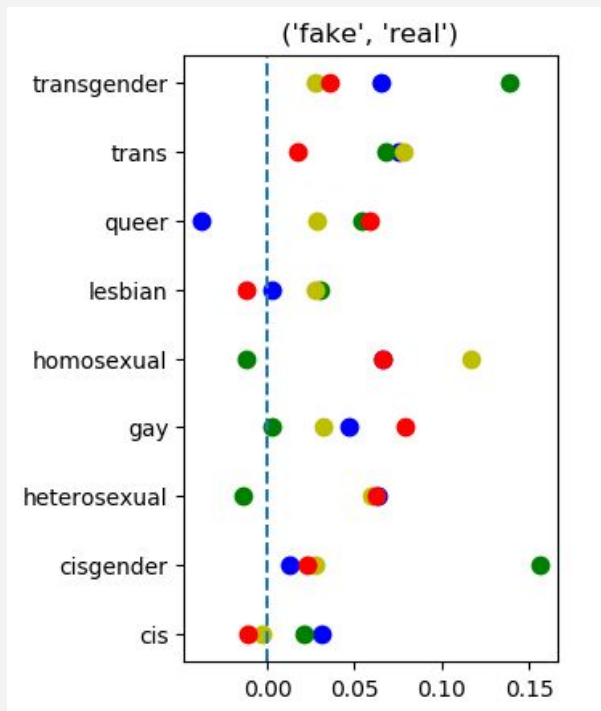


Results

- ‘Queer’, ‘lesbian’, ‘gay’ in general fanfic toward *different*
 - Identity labels often in sharper, negative contexts: “Was it so awful to have a lesbian for a daughter?”



Results



- 'Transgender' unexpected aspirational closer to *fake* than challenges corpus
- Could expect terms to be more "real" in challenges since occur in more serious contexts, or could expect negative attitudes around them which might tend things toward *fake*
 - Challenges corpus more real than aspirational for 'transgender', 'queer', 'gay'
 - Challenges corpus more fake than aspirational for 'trans', 'lesbian', 'homosexual'
- Hegemonic terms more toward fake than non-hegemonic terms are (not sure)