

## **The Evolution of LGBTQ Words in American News**

### **Introduction**

Several news outlets started publishing their first major articles about LGBT issues in the 1950s, specifically about gay men and lesbian women. Many of the top outlets used the term “perverts” as a synonym for gay and lesbian, for example, the New York Times published an article called “Perverts Called Government Peril” and “Federal Vigilance on Perverts Asked.”<sup>1</sup> Over time news outlets used different words and tone to discuss LGBTQ issues and people, often reflecting the cultural norms of society, and oftentimes shaping those norms. Over time, cultural acceptance, norms and media narratives around LGBTQ identities and issues have evolved rapidly.

In this paper I seek to explore how lesbian, gay, bisexual, transgender and queer issues and communities have been discussed by news outlets, and how such discussions have evolved since 1990. The goals of my project are:

1. Study how the description of LGBTQ groups have changed over the last 20 years in literature and the broader culture,
2. Determine where various words related to sexual orientation and gender identity are located on the cultural dimensions of race, class and gender, and
3. Detect linguistic shifts in the meaning of LGBTQ words over the last 20 years

---

<sup>1</sup> Streitmatter, Rodger. From “Perverts” to “Fab Five”, The Media’s Changing Depiction of Gay Men and Lesbians. 2009.

## Data

I use the The Corpus of Contemporary American English (COCA) in this study, specifically relevant news articles from 1990 to 2015. COCA is one of the most widely-used corpora in English and offers unparalleled insight into variation in English.<sup>2</sup> I specifically include in my corpus only all news articles using the words lesbian, gay, bisexual, transgender or queer, which amounts to a total of 3,648 documents.

## Methodology

We then begin to illustrate them with word networks that we can analyze to understand the structure of how words connect with one another, and the dynamics of how their meanings flow through a discursive system. For example, we can define links between words as a function of their co-presence within a document, chapter, paragraph, sentence, noun phrase or continuous bag of words. We can also define links as a function of words that rely on one another within a directed dependency parse, or links between extracted Subjects, Verbs and Objects, or nouns and the adjectives that modify them (or verbs and the adverbs that modify them). Rendering words linked as a network or discrete topology allows us to take advantage of the network analysis metrics like centrality and community.

In this paper I used various methods discussed in class, homeworks and assigned readings. The detailed methods for each task I complete in the paper are:

- To calculate basic descriptive statistics about the make-up of my corpus, the the frequency of documents using lesbian, gay, bisexual and transgender, using techniques discussed in homework 4.

---

<sup>2</sup> <https://www.english-corpora.org/coca/>

- To understand how *transgender* is related to the other LGBTQ words, I use semantic network analysis and investigate subnetworks and nodes around *transgender*. The approach allows me to define links between transgender and other words as a function of their co-presence within each news article, and investigate whether other LGBTQ words or issues have This approach was discussed in homework 5.
- To identify how the context of LGBTQ words have changed over time in the news, I use the word2vec word embedding models, created for each of my time periods, and compare the most similar to the LGBTQ words across the time periods. This approach was discussed in homework 7.
- To understand how LGBTQ words are related to the dimensions of gender, race, and class, I project the words onto three semantic spaces defined by words related to those dimensions. For example gender is defined using the words “man,” “he,” “him,” and on the other end, “woman,” “she,” and “her.” This approach was also discussed in homework 7.
- To identify the semantic changes of LGBTQ words across time periods, I use an approach piloted at Stanford by William Hamilton, Daniel Jurafsky and Jure Leskovec to align the dimensions of the embeddings from the five time periods and identify semantic changes as the word vectors change their loadings on the LGBTQ words. This approach was discussed in homework 7.
- To quantify whether the news discussed lesbian, gay, bisexual and transgender issues in a positive or negative light, and the sentiment of the tone in relation to each other, I used the BERT model on specific sentences used in articles in my corpus. The BERT model is sensitive to the context of the sentence, which makes it appropriate in sentiment analysis

about LGBTQ words. This approach is used in homework 8. Note that these results come directly from my homework 8 notebook, in a section labeled “Used for Final Project.”

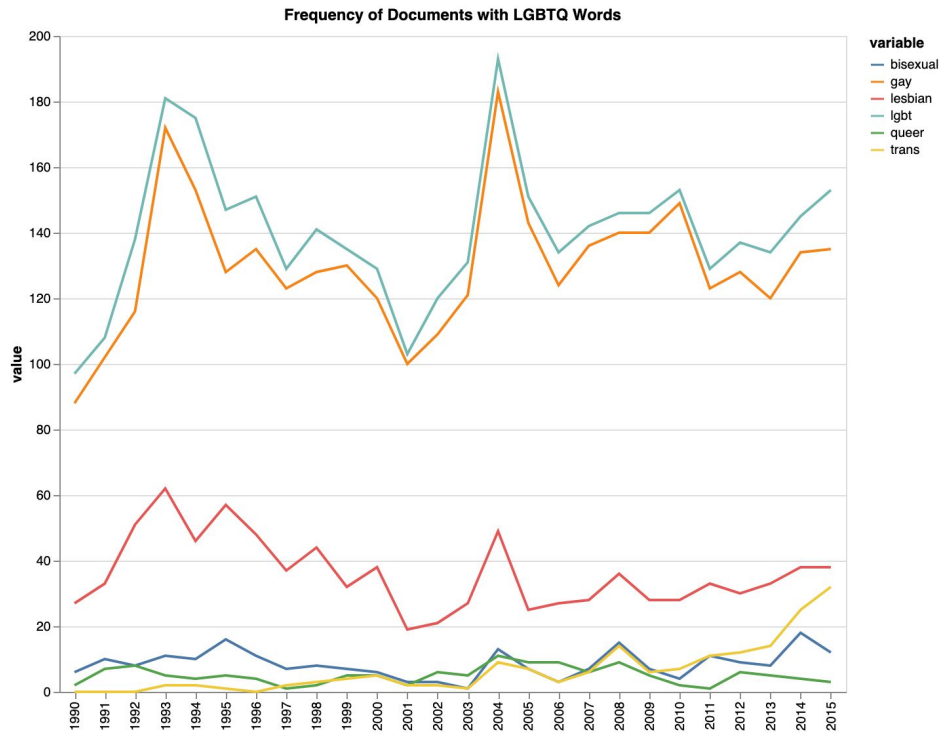
## **Descriptive Data, Analysis, and Interpretation**

### **Word and document frequencies**

The document frequency using lesbian, gay, bisexual, transgender and queer over time, displayed in Table 1, shows the following:

- The usage of gay makes up majority of LGBT mentions in the news over the whole time period
- Big spike in new articles with “gay” in it during the early 1990s, due to the AIDS epidemic
- Also a spike in 2004, when same-sex marriage first became legal in Massachusetts and new outlets across the country started discussing the ethics of same-sex marriage
- The usage of transgender has steadily increased since 2009
- The usage of “lesbian” peaked in 1992 and has slowly been declining since then, with a slight resurgence since 2005
- “Queer” is the least used word in the news

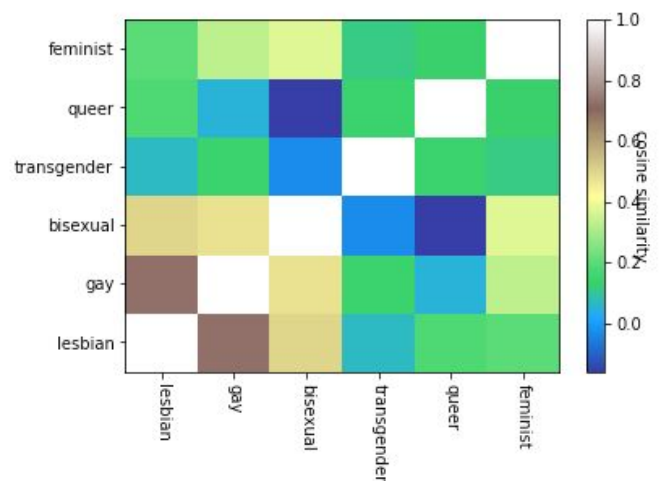
Table 1



## Relationship of LGBT words to each other

News outlets use the term “gay” far more than lesbian, bisexual, transgender and queer. One possible explanation is news articles use it as a synonym or umbrella term for other LGBT terms. To test this hypothesis I looked at how the different documents relate to each other within the

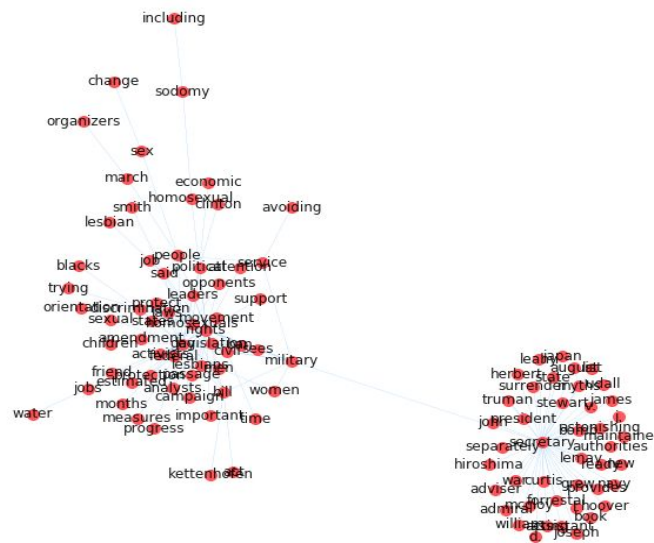
Table 2 - Doc2Vec Similarity



space using the Word2Vec algorithm, shown in a heatmap in Table 2. I found that queer and transgender are the most different from all other words in the space. I also included “feminist” in the doc2vec heatmap, which shows that the word feminist is more closely related to all the LGBTQ words than transgender and queer is to gay, lesbian and bisexual. This is an interesting finding, and one I explore in more depth through time.

**Table 3**

**Transgender Semantic Network (1990-1995)**



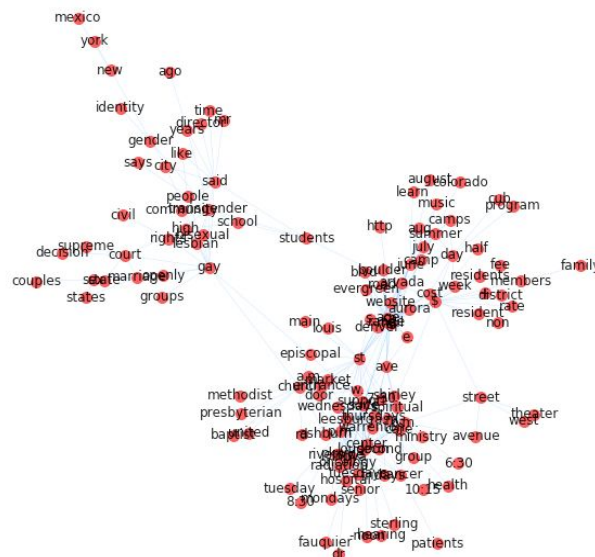
I explore how the word transgender relates to lesbian, gay, bisexual and queer in more detail by examining the semantic network of the documents that include the word “transgender.” To understand how the relation between these words have evolved over time I compare the semantic network analysis of transgender in the time period 1990-1995 to the period from 2010-2015. In the first period, only five articles used “transgender” and the topics covered were about trans folks in

the military and trans rights. Table 3 above shows the semantic network of documents including transgender from 1990-1995.

In contrast, in the period from 2010-2015, 101 articles used the word “transgender,” and the network is far more diverse than from 1990-1995. The network is clustered roughly around 4 topics:

1. Health Care
2. Religion
3. Civil Rights
4. Gender identity in locations (schools and cities)

**Table 4**  
**Semantic Network of Transgender Articles (2010-2015)**



The increase in coverage and topics about transgender issues in the news reflects the fact that trans issues are more included and advocated for by the larger LGBTQ community than ever before. The cluster around health care reflects that fact that trans communities suffer the most from mental

health issues and are the least likely to receive adequate health care in the United States, and the media is finally discussing more of these issues. Table 4 above shows the semantic network of documents with transgender from 2010-2015.

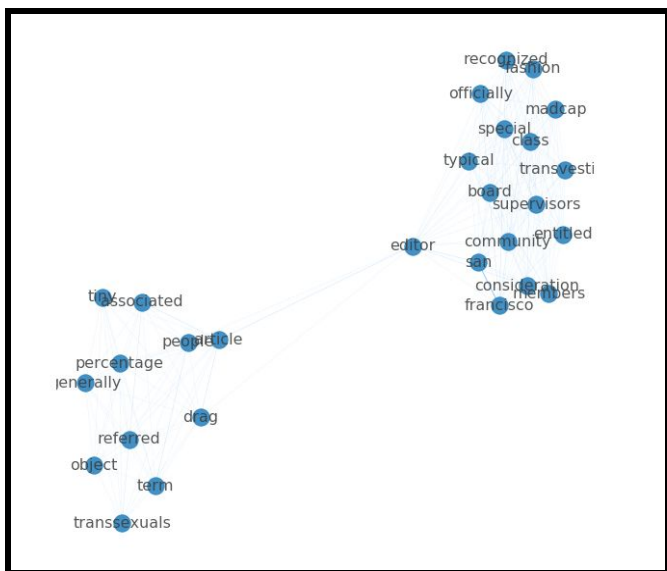
Additionally I looked closer at the nodes within two network steps of transgender in 1990-1995, and from 2010-2015. The graphs in Table 5 below show that lesbian, gay, bisexual and queer are not within 2 nodes of transgender in 1990-1995, but are in 2010-2015, in addition to “rights.”

This could be because there has been a long history of marginalization of trans issues and communities by the lesbian and gay communities. Lesbian and gay activism in the 1990s often ignored trans issues at best, and some groups were actively advocated against trans inclusion in the community and trans rights. News coverage of trans issues was also minimal until the early 2000s.

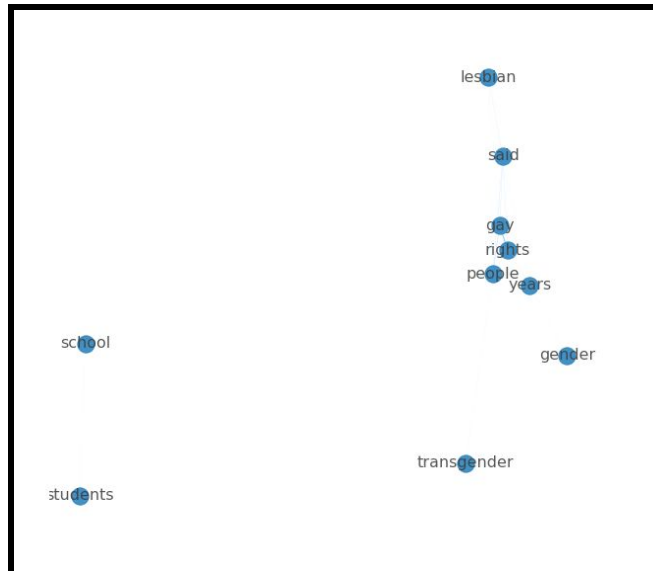
**Table 5**

**Nodes within 2 network steps of “transgender”**

**1990-1995**



**2010-2015**





## **Evolution of the Meaning and Use of Lesbian, Gay, Bisexual and Transgender**

I created five word2vec word embedding models for the five five year periods in my corpus (1990-2015). I compared each model in various ways to explore the evolution of the meaning of lesbian, gay, bisexual and transgender over time. To begin, I compared similar words to gay in the first period, from 1990-1995, to the last period, from 2010-2015.

Table 6 below shows the words most similar to “gay” in Period 1 and Period 5. The tables have many similarities, but in 2010-2015, the words closest to gay in the vector space include “equality,” “licenses” and religious. It also does not include lesbian or black, which were closer to “gay” in 1990-1995. These patterns reflect many changing attitudes and trends towards LGBT people over the years, and within the LGBT community. For example, same-sex marriage became one of the most discussed and advocated for “civil right” for gay and lesbian people by the mid-2000s. However this was not the case in the 1990s, when the community was ravaged by the AIDS epidemic, and facing harsher discrimination in all realms of life leading to economic hardship and homelessness. Many communities continue to face these obstacles - poor healthcare, discrimination at work, poverty, mental illness and even violence. However the “mainstream,” often meaning White, LGBT movement instead chose to make same-sex marriage the most prominent issue to advocate for by the early 2000s. This is reflected in the different words similar to gay in the two time periods.

Another interesting difference is the word “abortion” shows up in the early 1990s and not in 2010-2015. One could assume this is because abortion rights were more threatened at this time,

espeically under the Bush Administration. An interesting next step would be to run this same analysis on the time period from 2015-2020 to see if the LGBT community is again talking about abortion as these rights are facing renewed threats under the Trump administration.

**Table 6**  
**Words Most Similar to Gay Across Two Time Periods**

1990-1995	2010-2015
Lesbian	Sex
Black	Civil
Civil	Human
Rights	Equality
Anti	Licenses
Human	Religious
Men	Lgbt
Community	Bisexual
Abortion	Gays
Women	Interracial

In Table 6.2 below, I compare the words most similar to lesbian in 1990-1995 and in 2010-2015.

Doing so reveals very interesting patterns that reflect the increasing inclusion of other identities into the lesbian community, and the willingness of the community to advocate for issues beyond simply women and lesbian specific issues. In 1990-1995 the only other identity associated with lesbian was gay and bisexual. However in 2010-2015, transgender is also included.

**Table 6.2**  
**Words Most Similar to Lesbian Across Two Time Periods**

1990-1995	2010-2015
Anti	Bisexual
Rights	Openly
Black	Rights
Community	Anti
Marlinton	Marriages
Gay	Men
Civil	Lgbt
Men	Couples
Human	Transgender
Abortion	Weddings

In support of previous findings above, the most words most similar to queer have almost nothing to do with LGBT issues or communities. From 2010-2015, the words most similar to queer include words like “digital” and “rabbis,” which have not connection to other LGBTQ issues. “Feminism” and “1960s” are also included, indicating that the news media does not use the word queer to

describe LGBTQ issues, but does associate it with some feminist movements and perhaps art and music. The most dissimilar word in the word to vec models between all LGBTQ words is queer, which makes sense given the context the news articles use “queer” in.

The difference between the words above begs the question, when did the meaning and context of “gay” change the most between 1990 and 2015? In order to do this I use an approach piloted at Stanford by William Hamilton, Daniel Jurafsky and Jure Leskovec to align the dimensions of the embeddings from the five time periods and identify semantic changes as the word vectors change their loadings on the LGBTQ words.

The heatmaps describing the changes over time (defined as 1-cosine similarity, with a range of 0-2, so low values mean little change, and high values mean lots of change) are below. Table 7 shows that the meaning of gay changed the most in epoch 5, and actually changed very little between 1990-2000. Again this may be attributed to the evolution of LGBTQ issues discussed from the AIDS epidemic in the 1990s to gay marriage in the 2000s. The meaning of lesbian also changed quite a bit over time, though slightly less than gay. Lesbian changed more from 2005-2010 relatively than the word gay. Interestingly, the meaning of *bisexual* changed very little over time. This could be due to relative infrequency of the word, and the erasure of bisexual specific issues in the community. Note the differing scales in the two heat maps.

Table 7

Semantic Change of *Gay* and *Lesbian* Over Time

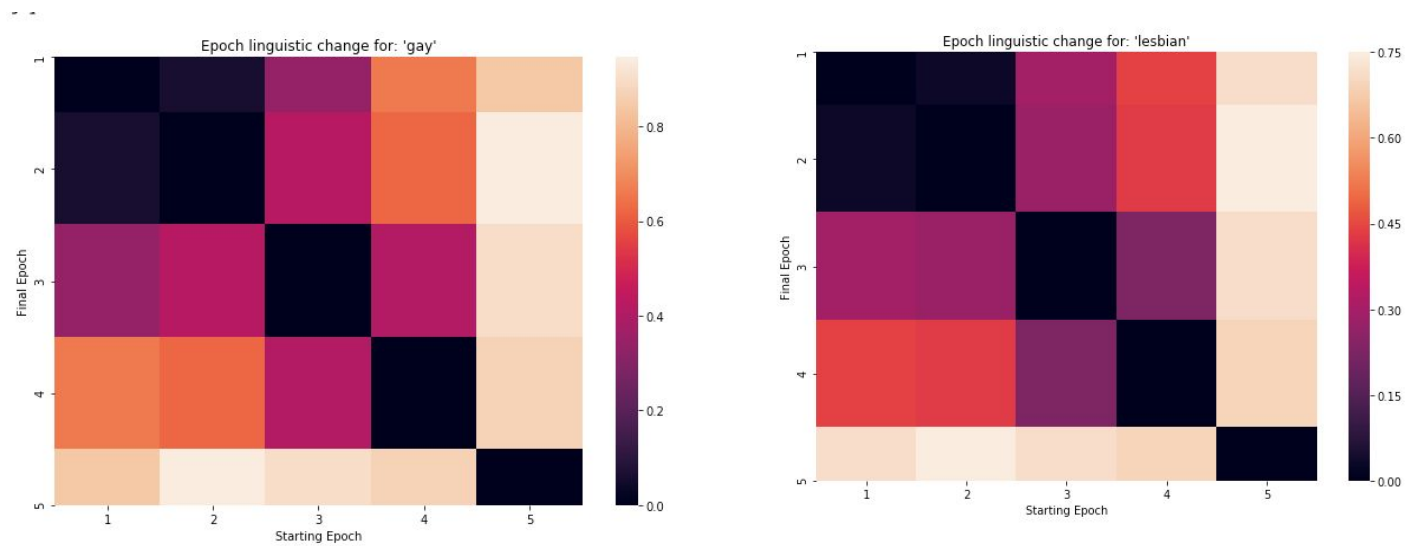
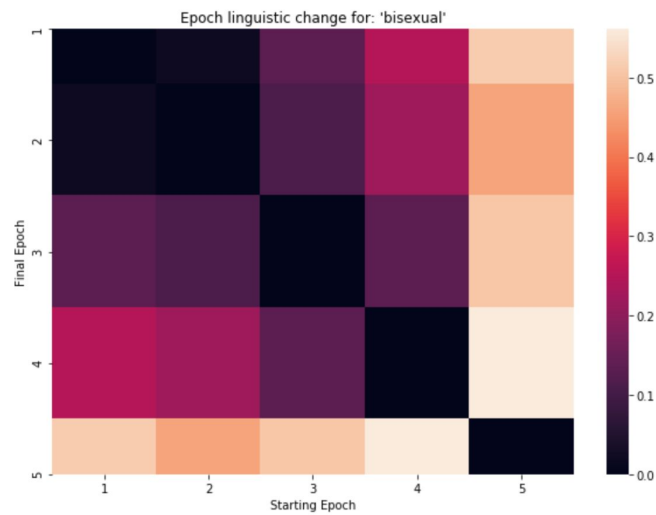


Table 8

Semantic Change of Bisexual Overtime



## **LGBTQ Words Projected on the Dimensions of Race, Class, and Gender**

Lesbian, gay, bisexual, transgender and queer communities, issues and the labels itself have a complicated place in the dimensions of race, class, and gender. The identities themselves are incredibly diverse and nuanced, with countless subgroups and identities within each label. The news however has typically portrayed such identities in a much less nuanced way, and often focused on issues pertaining most to the most privileged sectors of the community, and marginalizing or flattening other parts of the community. For this reason, I projected the words onto the dimensions of race, class and gender. The results are shown in Table 9 below. The most notable finding include:

- Gay is the most feminine word in the LGBT lexicon, more so than “woman.” This is an interesting finding, but it is difficult to conclude why this may be. Typically stereotypes of gay men in media depict them as feminine, flamboyant, and put them in the role of “gay best friend.” If many news articles discuss gay men and their depiction in film and TV, then one can imagine outlets associating men with more feminine adjectives and with woman. Exploring the question further could be a good next step to expand this project.
- Transgender is a close second - this is interesting because transgender is a gender neutral word referring to transmen, transwomen, or gender non-binary folkx. The fact that transgender is associated very closely with women could be attributed to the media discussing trans women more than trans men or non-binary folks
- All LGBT words are slightly closer to black than white. Again this is difficult to say anything conclusively about. I would have expected words like “stud” that have typically been used by the black community to be closer to black, while others that are used more broadly like gay

to be used by both. The absence of this tells me that perhaps the dimension of race is more nuanced or not well captured by simply labeling one end “white” and the other “black.” The races are not a binary or a dichotomy, many people identify as both, and this scale ignores Latinx, Asian, Native American and other races.

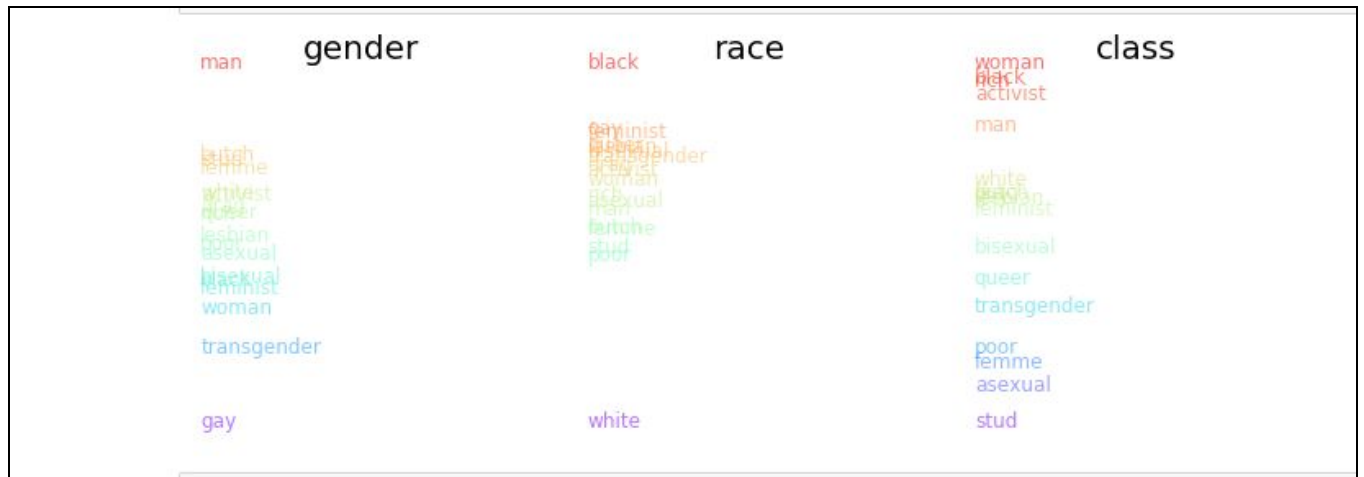
- Bisexual, queer, and transgender are all closer to poor than rich, while gay and lesbian are closer to rich. Typically bisexual and transgender people face greater mental health risks and receive poorer healthcare than other LGBTQ identities. Transgender folkx especially still face job discrimination and a higher threat of violence that causes economic hardship not faced by non-transgender people.<sup>3</sup>
- I also chose to project “feminist” on these dimensions, because the feminist movement has always had an interesting and sometimes contrary relationship to the LGBTQ movement, and especially the transgender movement. Interestingly, feminist is more closely related to rich than poor.

**Table 9**

**LGBTQ Labels Projected on the Dimensions of Gender, Race, and Class**

---

<sup>3</sup> “Economics and the Transgender Community.” Engage TU.  
<http://www.engagetu.com/2018/04/12/economics-and-the-transgender-community/>.



### Brief Sentiment Analysis of LGBTQ words generally

I also train the last layers BERT deep neural net model on my full corpus and then analyze the sentiment of specific sentences in articles about LGBTQ civil rights. I was particularly interested in investigating whether changing the LGBTQ identifier in a specific sentence would change the sentiment of a sentence. I used the following sentence from my corpus: “Judge Walker had suggested that the group lacked standing because he found no harm caused by granting gay people the right to marry.”

I then switched out the word gay for queer in the sentence, and analyzed how the sentiment of the sentence and the meaning changed. I found that the sentence using queer lies far away in semantic space to the same sentence using gay. I believe this implies that the words changed the meaning of the sentence quite dramatically. Using the word queer also made the sentiment of the sentence slightly more positive.



I also thought it would be interesting to look at how the BERT model could identify the sentiment of a sentence if many 'negative' words were used in it, even if it had a positive meaning overall. I used words/sentences within the LGBTQ lexicon to investigate.

I worked with this sentence, also found in my corpus: "The show was conceived, said Gonzalo Casals, the museum director, as an intimate platform for artwork that holds up a mirror to the lives of sex workers in the L.G.B.T.Q. community, some pieces by artists who themselves worked as prostitutes."

BERT classifies the words prostitutes and sex workers as highly negative, but identified the sentence as being positive.

## **Conclusion**

With the various methods described above, I come to the following conclusions about the evolution of LGBTQ words in American news:

1. Out of all LGBTQ issues, the media discusses "gay" issues the most
  - There was a spike in the usage of gay and lesbian in the early 90s with the AIDS epidemic, and all LGBTQ words in the mid 2000s with greater discussion of same-sex marriage
2. Queer is the most different from the other LGBTQ words. News outlets do not seem to use queer as a synonym for other LGBTQ words in relation to other LGBTQ issues. The LGBTQ community however largely considers queer to be an umbrella term for other identities. For a future project it would be interesting to compare the usage of queer in blog posts, magazines, and social media to demonstrate how the word is used differently by

different communities, and much more frequently by LGBTQ folks more broadly.

3. After queer, transgender is the most different from lesbian, gay, and bisexual. When looking at the semantic network surrounding transgender from 1990-1995, the word only appeared in 5 articles (less than .01% of my corpus) and did not co-occur with gay, lesbian or bisexual. This reflects the fact that trans issues were largely excluded from lesbian and gay movements in the 1990s, and not discussed in the media at the time.
4. The change in the semantic network surrounding *transgender* towards the late 2000s indicates more inclusion by the media of trans issues as part of LGBT issues, and generally reflects more court cases and media coverage of transgender issues.
5. The words in semantic space located near lesbian, gay, and bisexual have shifted over time from being more about civil liberties to more of equality and marriage. Again this reflects changes in the issues mainstream LGBTQ groups have advocated for over time, from being more about the AIDS epidemic and economic hardship, to same-sex marriage in the 2000s.
6. The semantic meaning of gay has changed the most of any LGBTQ words over time, while lesbian and bisexual have changed very little. This could be because the news tends to use gay as an umbrella term at least for lesbian. Again it also reflects the vast difference between issues facing the gay male community in the 1990s to the 2000s. The scarcity of transgender and queer in the 1990s makes it difficult to apply the methods of aligning word embedding models across time periods.

## References


Ellen Blast. “The Economics of the Transgender Community.” Regional Economic Studies Institute. Engage TU. April 2018.

<http://www.engagetu.com/2018/04/12/economics-and-the-transgender-community/>.

Computational Content Analysis Course Material.

<https://github.com/Computational-Content-Analysis-2020/Content-Analysis-2020>

Kozlowski, Austin, Matt Taddy, James Evans. “The Geometry of Culture: Analyzing Meaning through Word Embeddings.” 2019, American Sociological Review.

Manning and Schütze. 1999. Foundations of Statistical Natural Language Processing. MIT Press: selections from Chapter 5 (“Collocations” ): 151-163, 172-176, 183-186.

Nikhil Garg, Londa Schiebinger, Dan Jurafsky and James Zou. “Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes.” 2017. PNAS.

Gentzkow, Matthew & Jesse M. Shapiro. 2007. [“What Drives Media Slant? Evidence from U.S. Daily Newspapers.” \(Links to an external site.\)](#) *Econometrica* 78(1): 35–71.

Streitmatter, Rodger. From “Perverts” to “Fab Five”, The Media’s Changing Depiction of Gay Men and Lesbians. 2009.

William L. Hamilton, Jure Leskovec, Dan Jurafsky. “Diachronic Word Embeddings Reveal Statistical Laws Of Semantic Change.”