After the Vote: Mapping Online Trans Discourse

Mikael Poli*

*mikael.poli@studenti.unipd.it

Abstract—This project explores the topics of discussion around transgender identities on Reddit after Donald Trump's 2024 re-election. A dataset of 15,819 posts and comments from trans-focused communities was collected and grouped into 13 topics using the Leiden algorithm. Sentiment and agency were analyzed with LIWC-22 and BERTAgent. While political topics tended to be more negative, the overall tone was slightly positive, especially in conversations about support and resources. The results provide an example of how network science methods can help make sense of online discussions during times of political pressure.

Keywords—*Network Science, Semantic Network Analysis, Reddit, Transgender*

1. Introduction

Transgender people face stigma and discrimination, leading to minority stress: a chronic strain linked to negative mental and physical health outcomes [1]. It realistic to think this stress can be intensified by political attacks on trans rights, especially efforts to restrict access to gender-affirming care and fuel the demonization of trans people. Donald Trump's 2024 Agenda 47 campaign included several such proposals, framing them as protection from so-called "gender insanity" [2].

2. Aim

The overarching goal of this project is to capture a snapshot of online discourse within trans-oriented Reddit communities in the months following Trump's re-election. The goal is divided into two objectives:

1. Cluster online posts into topics in order to understand what people are discussing within the broader transgender discourse.
2. Perform sentiment analysis on each topic to get an overview of the emotional tone and affect expressed across discussions.

3. Process

All code was written in Python 3.11.

3.1. Phase 1: Building the Network

3.1.1. Data Collection

The first step was to identify the top subreddits and their comments. This was done by querying Reddit's official API to retrieve the names of the most popular subreddits for the keyword "trans". These were r/trans, r/transgender, and r/asktransgender. Data was collected on May 17, 2025. Since Reddit's API does not allow custom date-based filtering, posts were collected in bulk by requesting top posts withing the whole year, and were then filtered post-hoc by indicating December 6, 2024, as the start date. This yielded 256 posts and 18085 comments, for a total of 18341 documents and 6247 words, which were then deep cleaned and POS-tagged (allowed POS: ADJ, ADV, PRON, NOUN, PROPN, VERB).

3.1.2. Graph Construction

The next step was to build the words-documents bipartite network (using TF-IDF scores), which was then projected onto the documents. In order to make performance more efficient, the document-projected network was pruned based on the Network Community Profile (NCP), which suggested a 99-percentile threshold, keeping therefore only the top 1% strongest connections. The final undirected weighted graph was then obtained by extracting its giant connected component. A Maximum Likelihood estimation of the scaling exponent indicated $\gamma = 2.36$. The graph is composed of 15,819 nodes and 924,976 edges. It has a small diameter of 6 and an average path length of around 2.51, meaning most nodes are only a few steps apart. The average degree is 116.94, while the median is 50, suggesting that a few highly connected nodes (hubs) are pulling the average up. Despite the network's size, the density is low (0.00739): the network is sparse overall, but still efficiently connected.

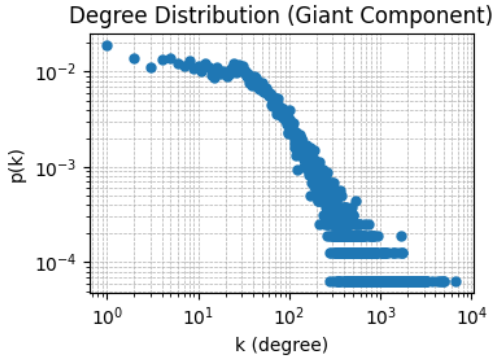

Figure 1. Degree distribution of the final document network.

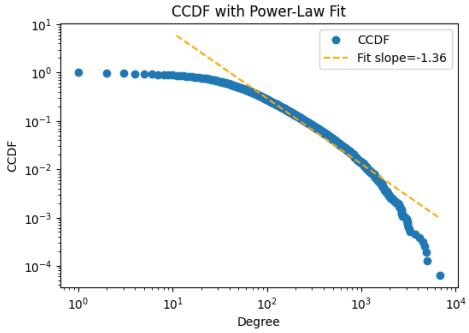

Figure 2. Complementary cumulative distribution function of the final document network.

3.2. Phase 2: Topic Detection and Interpretation

3.2.1. Topic Detection

Four topic modeling methods (BERTopic, BERTopic with outlier reduction, Louvain, Leiden) were ran on the graph to cluster the documents into topics. BERTopic was ran on raw text. Their performance was evaluated using Modularity, Ncut, and NMI, defined as:

- Modularity
$$Q = \sum_t (P_{tt} - p_t^2)$$
- Normalized Cut
$$\text{Ncut} = 1 - \frac{1}{n}\sum_{j=1}^{n} \frac{(P_{tt})_{jj}}{p_j}$$
- Normalized Mutual Information
$$\text{NMI} = \frac{I(W;T)}{H(T)} = \frac{\sum_{w,t} P_{wt} \log \frac{P_{wt}}{p_w p_t}}{-\sum_t p_t \log p_t}$$

where

- P_{tt} represents the topic-topic probability matrix
- p_t represents the probability vector of topics
- P_{wt} the joint probability matrix of words and topics.

The results were as follows:

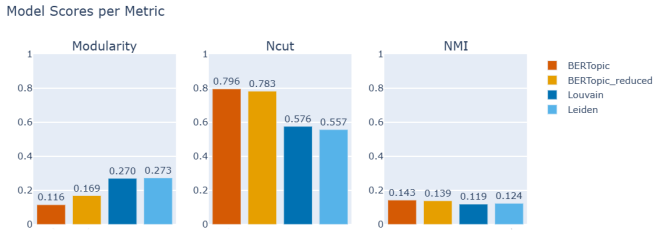

Figure 3. Topic detection models comparison.

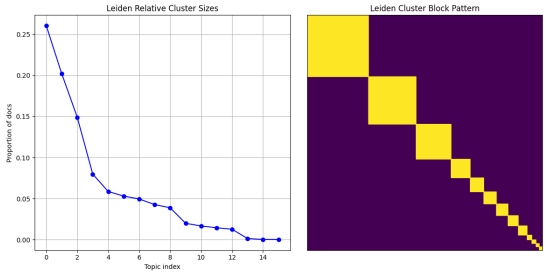

Figure 4. Leiden topic distribution.

The Leiden algorithm yielded the best results, and its output was selected for topic interpretation.

3.2.2. Topic Interpretation

All 16 resulting topics were inspected manually by extracting their top 10 representative documents, selected as those with the highest weighted degree within the their respective topic sub-graph. These were then fed to Chat-GPT to output a single short label and description for each topic, which were then once again refined manually. This resulted in 13 topics, and 3 outlier groups:

Size	Label
26.60%	Erosion of Trans Rights in the U.S.
20.61%	Daily Challenges for Trans People
15.22%	Comprehensive Care for Trans Youth
8.16%	Christianity and Trans Identity Conflict
5.99%	HRT Physiological and Anatomical Changes
5.42%	Cultural Sensitivity in Name Choice
5.06%	Hypersexualization of Trans People
4.36%	Legal Document Changes
3.97%	LGBTQ+ Representation and Controversies in Media
2.04%	Gender-Neutrality of Language
1.70%	Trans Resources and Legal Context
1.49%	Sex and Gender in Nature vs. Norms
1.30%	Misrepresentation of School Shooter Narratives
0.14%	Outlier
0.03%	Outlier
0.03%	Outlier
100.00%	

Table 1. Document distribution by Leiden topic.

The short descriptions of all topics can be found in the Appendix section in Table 2.

3.2.3. Topic Graph

The undirected weighted topic-level graph was built by projecting the document graph onto the topics.

3.3. Phase 3: Sentiment Analysis

Sentiment analysis was carried out through LIWC-22 [3] using the categories *tone_pos, tone_neg, emo_pos, emo_neg, politic, health, mental*), and calculating two composite measures:

- Tone (*tone_pos* – *tone_neg*);
- Emotion (*emo_pos* – *emo_neg*);

Agency perception analysis was conducted using BERTAgent [4] (*BATot*). Due to the size of the documents, BERTAgent was run on the top 15 documents per topic, superficially cleaned and split into sentences of maximum 50 characters each. The *BAtot* score was then extracted for each sentence.

Each LIWC-22 score and the BAtot score was then averaged by topic. The six resulting attributes appended to each node in the topic graph were:

1. Tone
2. Emotion
3. LIWC-Politic
4. LIWC-Health
5. LIWC-Mental
6. Agency

The graph was exported into .graphml format and loaded to Gephi for visualization.

The Tone, Emotion, LIWC-Health, LIWC-Mental, and Agency scores were plotted as a function of LIWC-Politic to identify any trend in their values, and a final overall measure of each score was computed for the entire network by taking the average value of each one weighted by the nodes' PageRank score.

4. Results

The final network was visualized using the Force Atlas 2 layout in Gephi. Nodes were resized according to their PageRank score, and colored based on their tone (blue = positive, vermillion = negative). The result is shown in Figure 5. Bar plots for each topic's sentiment scores can be found in the Appendix section in Figure 8.

The network-level results of the sentiment and agency analysis are reported in Figure 6.

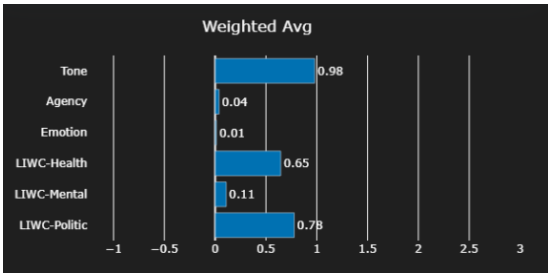

Figure 6. Network-level sentiment analysis results.

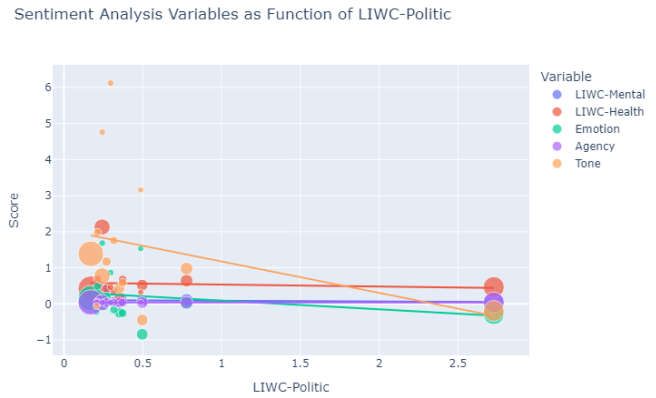

Figure 7. Behavior of Tone, Emotion, LIWC-Health, LIWC-Mental, and Agency, as a function of LIWC-Politic.

5. Discussion

5.1. Measures

Louvain and Leiden significantly outperformed BERTopic variants in terms of Modularity and Ncut. This suggests that graph-based clustering yielded more coherent communities in terms of internal connectivity, and that they provided cleaner inter-cluster separation in topic space. All models, however, yielded very low NMI scores, despite the meaningful nature of the top 10 words in each topic. One hypothesis is that the tail of overlapping words pulls down the score even though the core topic-defining words differ.

5.2. Topics

The central theme in the network is *Daily Challenges for Trans People*, which delves into the everyday difficulties trans individuals may face and strive to overcome across the various levels of Bronfenbrenner's ecological systems theory [5]. These include the individual (personal), microsystem, mesosystem, exosystem, and macrosystem levels, ranging from immediate relationships to broader social norms, political and economic systems, and cultural contexts.

The network's main concerns closely reflect Agenda 47's priorities regarding LGBTQ+ issues, particularly *Erosion of Trans Rights in the U.S.* and *Comprehensive Care for Trans Youth*. Another key theme is the historical and ongoing tension between certain religious traditions and non-cis-heteronormative identities, as captured in *Christianity and Trans Identity Conflict*. Notably, documents within the *Comprehensive Care for Trans Youth* category place more emphasis on both physical and mental health than those in other topics.

The overall tone across the network leans slightly positive. This could be influenced by the supportive nature of discussions found in *Trans Resources and Legal Context* and *Gender-Neutrality of Language*. Tone appears to shift more negative in politically charged topics, though it's important to note that this study does not establish a causal relationship between political content and tone.

Overall affect and perceived agency remained mostly neutral. The difference between tone and affect may stem from how LIWC-22 categorizes language: tone includes words that convey not just direct emotion but also those indirectly tied to it (e.g., "birthday," "funeral"). The notably low agency scores detected by BERTAgent might reflect both the tool's limitations (discussed in the next section) and the nature of the documents themselves. Many are long, and sections explicitly expressing agency may be too sparse to stand out.

Some topic-specific sentiment patterns are worth highlighting. The theme *Christianity and Trans Identity Conflict* is associated with the strongest negative tone and affect. Interestingly, *Sex and Gender in Nature vs. Norms* also trends negative, despite dealing with biology and non-human species. This may be because the documents often shift into cultural commentary, which tends to describe societal norms as harmful or exclusionary toward non-cis-hetero-normative identities.

In an unintentionally dark humorous twist, *Misrepresentation of School Shooter Narratives* is categorized as having neutral tone and highly positive emotion. This result likely reflects a limitation of

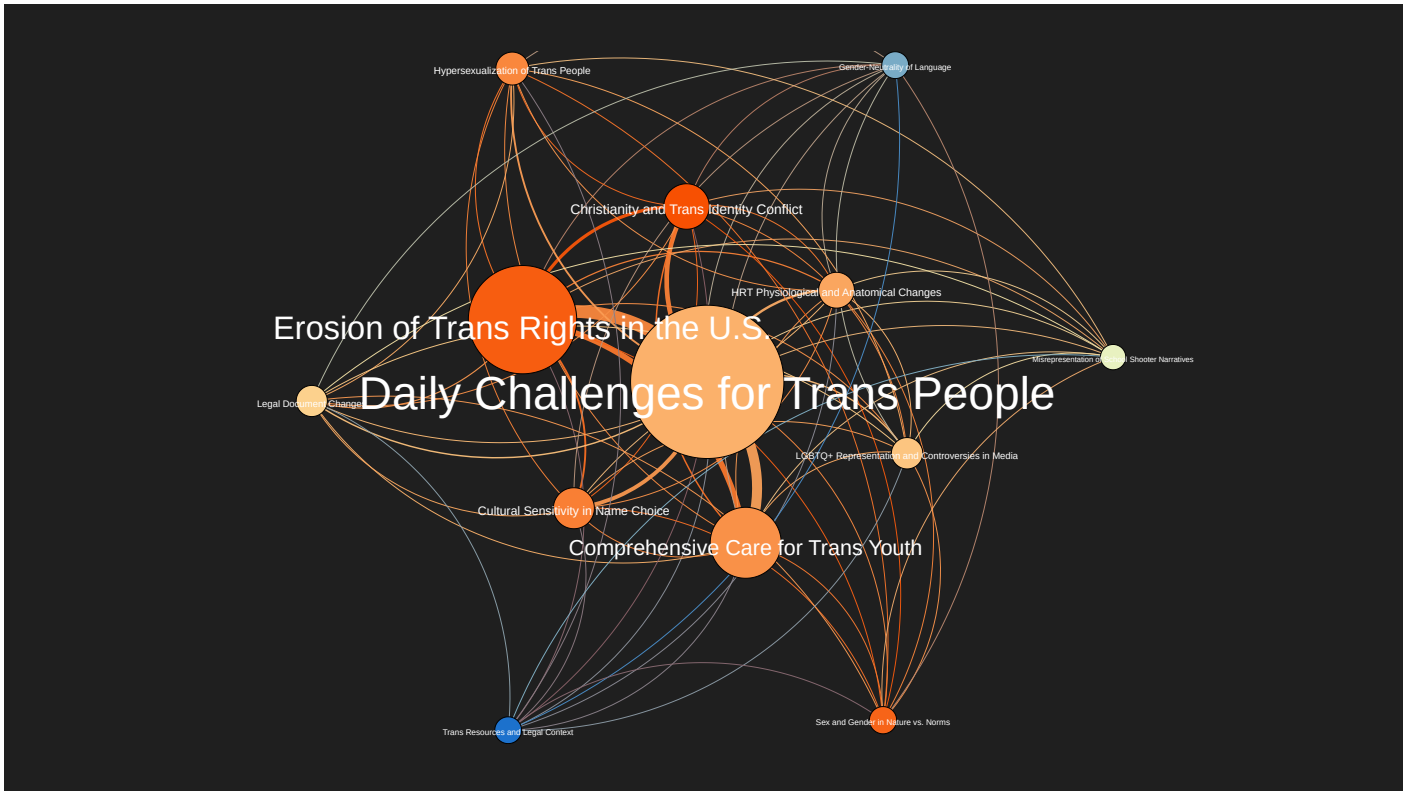

Figure 5. Topic Network. Size = PageRank score; Color = Tone, with blue indicating positive tone, and vermillion indicating negative tone.

LIWC-22, which is explored further in the next section.

5.3. Limitations

This project has several limitations that are important to acknowledge.

First, the data collection relied on Reddit's "top posts" filter, which does not support custom date-based queries. Because of this, post-hoc filtering was used to narrow down the dataset. This approach may have introduced some bias, as it is influenced by Reddit's own ranking system and may overrepresent posts with higher engagement or older content that had more time to accumulate votes.

Second, while the network analysis was useful in identifying topic clusters, it did not fully account for overlapping communities. Many documents are relevant to more than one topic, and standard community detection methods do not always capture this complexity. Using methods specifically designed for overlapping structures could improve the accuracy and depth of future analyses.

Third, LIWC-22, which was used to assess tone and affect, has some known limitations. The tool may struggle to interpret sarcasm, irony, or more subtle forms of language. As a result, the emotional tone it assigns may not always reflect the intended meaning of a post, especially in a context where informal and Internet-specific language is common (e.g., in *Misrepresentation of School Shooter Narratives*).

Fourth, the use of BERTAgent for measuring perceived agency presented practical challenges. Reddit posts are often quite long, and the model requires inputs to be shortened or segmented. This preprocessing step may have removed context that is important for detecting agency-related expressions.

Overall, these limitations suggest that while the project provides useful insights, the findings should be viewed as preliminary. Future work could benefit from improved filtering options, more sophisticated modeling techniques, and tools that are better suited to handling complex and informal online discourse.

6. Software

The code, data, figures, and documentation for the project can be found on GitHub.

References

[1] J. Hunter, C. Butler, and K. Cooper, "Gender minority stress in trans and gender diverse adolescents and young people", *Clinical Child Psychology and Psychiatry*, vol. 26, no. 4, pp. 1182–1195, Jul. 2021. DOI: 10.1177/13591045211033187. [Online]. Available: https://doi.org/10.1177/13591045211033187.

[2] *President Trump's Plan to Protect Children from Left-Wing Gender Insanity | Donald J. Trump For President 2024*. [Online]. Available: https://www.donaldjtrump.com/agenda47/president-trumps-plan-to-protect-children-from-left-wing-gender-insanity.

[3] R. L. Boyd, A. Ashokkumar, S. Seraj, and J. W. Pennebaker, "The development and psychometric properties of liwc-22", *Austin, TX: University of Texas at Austin*, vol. 10, pp. 1–47, 2022.

[4] J. Nikadon, C. Suitner, T. Erseghe, L. Džanko, and M. Formanowicz, "BERTAgent: The development of a novel tool to quantify agency in textual data.", *Journal of Experimental Psychology General*, May 2025. DOI: 10.1037/xge0001740. [Online]. Available: https://doi.org/10.1037/xge0001740.

[5] U. Bronfenbrenner, *Ecological systems theory*. Jan. 2000, pp. 129–133. DOI: 10.1037/10518-046. [Online]. Available: https://doi.org/10.1037/10518-046.

7. Appendix

Topic	Description
Erosion of Trans Rights in the U.S.	Examines the escalating threats to transgender rights and healthcare in the U.S., driven by conservative policy agendas, judicial shifts, and inconsistent political support. Underscores the growing disillusionment with the political system and call for urgent advocacy to counter bipartisan failures and safeguard LGBTQ+ protections.
Daily Challenges for Trans People	Examines the personal, social, and systemic challenges faced by trans and intersex individuals, including identity formation, medical transition, discrimination, and the quest for acceptance and support amid cultural and institutional resistance.
Comprehensive Care for Trans Youth	Addresses the medical, psychological, ethical, and social dimensions of gender-affirming care for transgender minors, highlighting the need for informed, individualized treatment supported by parents and professionals while confronting misinformation and societal challenges.
Christianity and Trans Identity Conflict	Examines the complex intersection of Christian beliefs and transgender identities, revealing internal religious conflicts, varying interpretations of scripture, and the political and social tensions that shape acceptance, prejudice, and activism.
HRT Physiological and Anatomical Changes	Discussion of individuals' varying changes associated with hormone replacement therapy (HRT), esp. regarding temperature sensitivity, hunger, and the possible consequences of anatomical changes (e.g., at TSA security screenings).
Cultural Sensitivity in Name Choice	Explores perspectives on name choice, esp. on the adoption of a non-culturally-coherent name, focusing on cultural appropriation, racial identity, and intent.
Hypersexualization of Trans People	Discussion around the social perception of fetishes within and outside trans communities. The documents emphasize rejecting harmful generalizations and stereotypes that see trans people as driven by sexuality, while advocating for nuanced understanding of individual expression and sexuality.
Legal Document Changes	Discussion of challenges and legal complexities in updating gender markers and names on state-issued IDs, birth certificates, passports, and Social Security records, highlighting varied state policies, potential federal restrictions, and strategic advice for navigating these changes.
LGBTQ+ Representation and Controversies in Media	Discussion of debates around LGBTQ+ inclusion and censorship in media, focusing on controversies involving J.K. Rowling's views on trans people, Disney's handling of queer characters and stories, and broader issues of representation, corporate motives, and cultural reception of LGBTQ+ narratives in film and animation.
Gender-Neutrality of Language	Discussion of how certain informal terms like (e.g., "dude", "guys") vary in perceived gender neutrality depending on context and generational usage.
Trans Resources and Legal Context	Provides extensive resources and guidance for transgender individuals and their families on understanding gender identity, accessing support, and navigating medical options, with discussion of recent U.S. Supreme Court justices' stances on LGBTQ+ legal cases and their ongoing judicial impacts on transgender rights.
Sex and Gender in Nature vs. Norms	Explores the complexity and variability of biological sex beyond the male-female binary in the animal kingdom, while distinguishing these biological phenomena from human gender and its social policing.
Misrepresentation of School Shooter Narratives	Discussion on how school shooters are often mischaracterized through scapegoating, particularly targeting transgender individuals, while ignoring broader patterns and systemic issues like gun access and societal failures. Highlights media tendencies to politicize shooter identities to fuel fear and prejudice.

Table 2. Topics and Descriptions of Trans-Related Issues

Figure 8. Sentiment analysis results by topic, ordered by each topic's PageRank score in descending order from top left to bottom right.