Pitch Your Project

In the three cells immediately following, describe \*\*WHAT\*\* you are planning to analyze for your final project (i.e., texts, contexts and the social game, world and actors you intend to learn about through your analysis) (<200 words), \*\*WHY\*\* you are going to do it (i.e., why would theory and/or the average person benefit from knowing the results of your investigation) (<200 words), and \*\*HOW\*\* you plan to investigate it (i.e., what are the approaches and operations you plan to perform, in sequence, to yield this insight) (<400 words).

What?

I plan to analyze Reddit data from two main subreddits: r/Adoption and r/Adopted. These subreddits both discuss the topic of adoption, that is in the legal sense of the word where an individual (most often a child) is taken from their biological or legal guardians and placed into the care of an adoptive family. (Definition adapted from: https://dcfs.illinois.gov/loving-homes/adoption.html). The difference between the two subs are their main goals. r/Adoption is a general purpose subreddit where they welcome anyone to post questions or engage in discussion about adoption related topics. Though all are welcome, most the posters tend to be from the adoption triad (adoptee, adoptive parent, or birth parent). This sub is quite a bit larger than r/Adopted as well, so it has far more posts in it. On the other hand, r/Adopted explicitly states that their community is meant for adoptees only; the community is constructed as a safe space for adoptees to engage with other adoptees. Though adoptee posts are dominant in r/Adoption as well, my hunch is that adoptees who post in r/Adopted are much more likely to display negative sentiment towards adoption than those who post in r/Adoption.

Why?

Fisher (2003) notes that adoption is a stigmatized status, thus learning more about how adoptees discuss adoption might provide valuable insights into the complexities of what adoption means in regards to the identity of adoptees. Furthermore, this research adds to the theory regarding social identity construction in online spaces. What it means for individual adoptees as well as adoptees as a collective identity (r/Adopted as whole) will be explored through this research. Thus I hope to explore the differences in individual and collective identity as well.

How?

I plan to conduct a semi supervised technique to label posts from both communities as having been written by adoptees or not. After labeling all data, I also wanted to perform a topic modeling to discern the topics that these discussion boards participate in. Though both are generally linked to adoption, the specific topics may be different from one another. After topic modeling. I also wish to perform a sentiment analysis to discern how adoptees and non adoptees feel about these different topics in these two subreddits.

Pitch Your Sample

In the cell immediately following, describe the rationale behind your proposed sample design for your final project. What is the social game, social work, or social actors you about whom you are seeking to make inferences? What are its virtues with respect to your research questions? What are its limitations? What are alternatives? What would be a reasonable path to "scale up" your sample for further analysis (i.e., high-profile publication) beyond this class?

The sample of textual data will gather the posts and comments from r/Adoption and r/Adopted. The posts from r/Adoption come from two sources: an archive for large subreddits last updated in 2022 and a supplementary set of posts scrapped using a personal scraper (due to limitations the scraper only takes the 1000 most recent posts). For r/Adopted, these text data come from the personal scraper as well, so it is much more limited. The social game I am interested in understanding are the complex interactions/discussions between adoptees and other persons related to adoption. Furthermore I hope to reveal insights into how adoptees in online spaces . The virtues for this data set are its relatively large amount of example text. The main limitations is that the population of individuals who post in r/Adoption and r/Adopted are not necessarily representative (nor should they be assumed to be) of the general population of adoptees and other members of the adoption triad. Like many other discussion boards, there are some people who frequently post in the subreddit, and then those that go on a burner and post a single question. Alternatives would include Facebook groups; however, I ruled out using these groups due to prominent ethical considerations. Reddit is a public discussions site, whereas these Facebook groups are often private and contain a lot more personally identifiable information. Other social media sites could also provide some insight but Reddit seems to be the most commonly used and openly available. To scale up this sample, I would ideally want to analyze all posts from r/Adopted, but this would require a better scraper or access as a researcher to the Reddit API.

## \*\*\*Which (words)?\*\*\*

<300 words

Pitches will include one (potentially very) short paragraph describing your question(s); one short paragraph describing your "data" (e.g., texts, corpus, labels, etc.); and a final short paragraph describing the three methods you expect to use.

Questions:

Main question:

How do the conversational dynamics and post sentiments within online adoption communities differ between general-purpose adoption subreddits like r/Adoption and adoptee-exclusive spaces like r/Adopted?

Supplementary questions to answer main:

Are the topics discussed in r/Adoption and r/Adopted distinctly different from one another? If no, it is not detrimental. But I have a hunch that even though both talk about adoption, they talk about different aspects of it. Since more adopters post in r/Adoption, there are more discussions about the process of adopting. Whereas in r/Adopted there might be more discussion about trauma.

Are the posts from adoptees distinctly different from posts not made by adoptees in r/Adopted and r/Adoption? If the answer is no, then the main question becomes much harder to answer.

Data:

My data comes from the two largest subreddits dealing with adoption: r/Adoption and r/Adopted. The data from the former contains archived posts since its inception as a subreddit as well as more recently scraped data. r/Adopted contains only scrapes the top 1000 most recent posts (due to the limitations of my scraper; I could properly configure selenium to dynamically scrape the website, so I used an older version of reddit to gather conversations by pages). This data contains posts and comments with corresponding metadata like user, flairs, etc.

Methods:

In order to answer these questions there are three distinct methods that I will need to employ. I will need to use ML, specifically neural networks to classify posts made by adoptees and those not made by adoptees. I will incorporate topic modeling to understand the different topics discussed in each subreddit. And then I will perform a sentiment analysis on these topics to discern whether these subreddits are more positive or negative.