INFO 5601 Final Project
Analyzing Reddit Commenter's Opinions on Net Neutrality

Luke Meszar

1   Introduction

Over the last few years, discussions around net neutrality have become more prevalent on the internet, including Reddit. The project aimed to study the public opinion of net neutrality on Reddit. In particular, to understand how opinions differed across different subreddits. In particular, subreddits that likely have different types of users. I choose a thread from five different subreddits and performed both a qualitative coding and sentiment analysis on the comments in each of these threads.

2   Net Neutrality

Net neutrality has been a hotly debated topic recently with recent policy changes by the FCC. This section will provide background on what net neutrality is and the issues surrounding it. The debate stems from how ISPs and their broadband networks should be classified. The two options are as an information service under Title I of the 1934 Communications Act of as a Title II telecommunication service. The key difference is Title II gives the FCC the ability to regulate ISPs in an effort to protect consumers and online businesses against discrimination [Han18]. Since Congress has not yet written a bill deciding this matter, it is up to the FCC to make these decisions. In 2015, the FCC, under the Obama administration ruled to classify the internet as a Title II service. Then, in December 2017, the FCC, now under the Trump administration, ruled to reverse this decision. Clearly there has been disagreement on how the FCC should view the internet and net neutrality, particularly among party lines. So, what exactly is net neutrality?

Roughly speaking, net neutrality says that ISPs should not be allowed to treat any discriminate any data that passes through their networks. This means that ISPs shouldn't be allowed to provide fast lanes for content they want to prioritize. To truly understand this, some of the economics of the internet need to be discussed.

First, the players need to be explained. There are the ISPs like Comcast and TimeWarner. Then there are the content providers like Amazon and Google. Finally, there are the consumers of thsi content, often average people. The first fact to realize is the ISPs are fundamentally operating in a two sided market. One side is the content providers, and the other is the content consumers. In two sided markets, it is important for there to be enough actors from both sides. For the internet, content providers wouldn't create content if there were no consumers to view it and no consumers would be online if there was no content for them to view. Clearly, ISPs need to work with both sides if they want any success.

Many of the largest content providers do not have to pay ISPs for their use of the internet as a means to reach other users [LW09]. This is because, in practice, the internet does distinguish between content providers and content consumers. This obviously leads to ISPs feeling like they are not being payed what they are owed. Being allowed to charge content providers for the services they use and provide "fast lanes" they argue will provide a better outcome for consumers. First, the money obtained by ISPs by companies like Google and Amazon could be used to fund the construction and maintenance of better infrastructure. Providing fast lanes would allow important content to be prioritized. It could even be as simple as charging more for peak hours which is often done in other areas like electricity and tolls [HW06].

It is very easy to imagine a situation where this backfires on consumers. The first is ISPs decide that content providers they support will always get the fastest treatment and every other content provider is hindered. In this

scenario, even companies like Google wouldn't be able to access the fast lane and the content on the internet becomes very fractured. The other option is just purely a huge financial cost on accessing the fast lanes. This means that only big companies like Google and Netflix would be able to pay for this. In either scenario, small content providers would never be able to succeed since they will never be able to afford the fast lanes.

A lot of the fear of this reality is due to the monopolistic natures of ISPs. Due to the high infrastructure start up cost and many local governments limiting competition, it can be impossible for competition to exist. With any monopoly, companies have no incentive to offer fair pricing for anything. One of the solutions [HW06] proposes involves making it illegal for local governments to create monopolies.

3   Methodology

The first choice that had to be made was which subreddits and which threads to consider. A few factors went into choosing this. First, a thread had to have a decent number of comments. A thread with only a handful of comments might not be the most representative of a subreddits overall opinions due to small sample size. The subreddits chosen to pick threads from was based on getting a diverse set of subreddits.

The five that were chosen were /r/askreddit, /r/technology, /r/books, /r/the_donald, and /r/democrats. /r/askreddit was chosen because it is one of the largest subreddits, and likely has one of the most diverse user bases of any subreddit. The thread chosen was titled "*[Serious] Americans that *DO NOT* support net neutrality, why?*" This article in particular was chosen because it asks what I assumed to be for the opposite opinion than the majority of Reddit users hold. I thought this would make for an interesting discussion from both sides. /r/technology was chosen because net neutrality is directly related to topic of the subreddit. The article chosen was "*Net neutrality day of action update: Twitter, Soundcloud, and Medium, have joined. Reddit, This could be as big as SOPA*" primarily because it was one of the threads with the most comments. /r/books was chosen because it was a seemingly neutral subreddit. It is neither directly political or technology related. The thread chosen was "*Join the Battle for Net Neutrality!! We need to stop them from allowing ISPs to charge us extra fees to access ebooks, games or anything else!*" The subreddits /r/the_donald and /r/democrats were chosen for their clear political leanings. I realized they aren't exactly a fair comparison since one is about a specific president and the other is about a whole political party. However, because the thread chose for /r/democrats, "*This is President Barack Obama. He did not sell Americans out to the telecom lobby, but instead called upon on the FCC to take up the strongest possible rules to protect net neutrality, which they did at his instruction in 2015*" is very much about Obama. The thread chosen from /r/the_donald was "*One Year Later, 'Net Neutrality' Zealots Proved Dead Wrong. 'net neutrality' zealots warned that its repeal would spell doom for a 'free and open" internet.*'"

Once the threads were chosen, comment data was collected using the Reddit API and the Python wrapper praw. After this, open coding was done on 50 comments, with 10 coming from each post. Two types of codes where chosen to generally represent people's opinions on net neutrality: their stance on net neutrality and their reason for their stance. The following tables list the codes with a short description:

Codes for Stance on Net Neutrality

Code	Description
For NN	It is clear that the comment is for NN
Against NN	It is clear that the comment is for NN
Probably for NN	The comment's stance on NN is ambiguous but context says they are likely for NN
Probably against NN	The comment's stance on NN is ambiguous but context says they are likely against NN
Ambiguous	It is entirely unclear the comment's position on NN

Code for Reason for Stance

Code	Description
Political Reasons	Primary reason stems from opinions of politicians, or general political views
Economic Reasons	Primary reason is driven by economic ideas like markets, monopolies
Personal Reasons	Primary reason is how it affects their lives individually
Other	The other categories don't fit well

I originally tried a form of topic modeling known as Latent Dirichlet Allocation to find which comments were on

topic. However, this proved ineffective since all the topics were too broad. I decided to write my own crude system to determine which comments were on topic. I created a dictionary containing terms like "net neutrality, Comcast, ISPs" etc. Using this dictionary, I filtered out comments that contained these words in them.

Then, I built a simple Python script that allowed me to go through each comment, choose whether to accept or reject it, and then choose a numerical code for each of the two codes presented above. For all the threads except /r/democrats, I coded 41 comments. For /r/democrats, I only coded 38 because so many of the comments where off-topic. Using these codes, I computed simple percentage statistics for all the codes and combinations of codes in each of the threads which is the basis of my results. Finally, I ran the pretrained sentiment analysis known as VADER [HG14] on all the comments I coded. For each comment, a compound score was given in the range $(-1, 1)$ where a more negative number is a negative sentiment and a larger positive number is a more positive comment. I considered any comment with a compound score greater than 0.25 to be positive, anything with a compound score less than -0.25 negative, and everything else is considered neutral. From here, percentages for each category and average compound scores were calculated. I also ran the sentiment analysis on the thread titles.

4 Results

The complete results of all the percentages from the coding are long so they can be found in section 9. Here, only the top two codes from each category for each thread will be shown.

/r/askreddit		Stance		Reason	
	Probably against NN	22.0%	Economic Reasons	68.3%
	Probably for NN	22.0%	Political Reasons	24.4%
/r/technology					
	For NN	48.8%	Political Reasons	31.7%
	Probably for NN	36.6%	Economic Reasons	29.3%
/r/books					
	Probably for NN	48.8%	Political Reasons	34.1%
	For NN	29.3%	Other	26.8%
/r/democrats					
	Probably for NN	47.2%	Political Reasons	66.7%
	Ambiguous	33.3%	Economic Reasons	16.7%
/r/the_donald					
	Probably against NN	65.9%	Political Reasons	63.4%
	Against NN	26.8%	Other	19.5%

I am choosing not to report the most popular combinations of stance and reasons since they can almost always be inferred from the topic reason and stance separately. The sentiment analysis for the titles resulted in:

	Compound Score	Sentiment
/r/askreddit	0.402	Positive
/r/technology	0.000	Neutral
/r/books	-0.539	Negative
/r/democrats	0.807	Positive
/r/the_donald	-0.889	Negative

The results for the sentiment analysis are as follows:

	Positive	Neutral	Negative
/r/askreddit	53.7%	17.1%	29.3%
/r/technology	46.3%	24.4%	29.3%
/r/books	46.3%	19.5%	34.1%
/r/democrats	33.3%	11.1%	55.6%
/r/the_donald	29.3%	26.8%	43.9%

The average sentiment analysis for the comments in each thread are:

	Average Compound Score
/r/askreddit	0.140
/r/technology	0.132
/r/books	0.102
/r/democrats	-0.217
/r/the_donald	-0.101

5   Analysis

The results for the splits based on subreddits is predominantly what I would have expected. This is especially true when it comes to the stance. For /r/technology, having 48.8% for net neutrality and 36.6% probably for net neutrality means that over 85% of comments on this thread were leaning for net neutrality. Considering [Pea18], most people in the technology industry left-leaning politically. With some accuracy, this assumption this should extend to the technology subreddit. As we will see when discussing the two political subreddits, liberals tend to be in favor of net neutrality.

The two political subreddits' stances are definitely as expected. The liberals in /r/democrats almost have a majority of the comments as probably being for net neutrality while /r/the_donald have over 90% of comments on the side of being against net neutrality. This is to be expected considering the Obama administration was for net neutrality whereas the Trump administration is against it.

Since /r/books was chosen to be the most "neutral" subreddit as a way to account for the average Reddit population, it is interesting that it is over 75% on the side of pro net neutrality. I think this reflects that Reddit as a whole likely leans left.

I think the most interesting result was /r/askreddit's. There was an even split between for and against net neutrality in the top two stance categories. The full results for /r/askreddit's stance categories are

Stance	Percentage %
Probably against NN	22.0
Probably for NN	22.0
For NN	22.0
Ambiguous	19.5
Against NN	14.6

By looking at all the results, generally for net neutrality wins out, but not by much. This is almost certainly due to the particular question asked in this thread which asks for people who are against net neutrality to explain their reasoning. A question was something like "What is your opinion on net neutrality?" would probably have led to

results more along the lines of /r/books.

When looking at the results for the reasons for peoples stance, it is interesting that threads have a higher percentage of political motivations behind their stance. This is to be expected from the two political subreddits. For both /r/technology and /r/books, political reasoning doesn't dominate by too much. I would have expected /r/technology to be dominated by economic reasoning instead of political reasoning though.

It is hard to draw any clear conclusions using the sentiment analysis portion of the data. Computing the absolute difference between the average comment compound score and the compound score gives the following result

	Absolute Difference
/r/askreddit	0.262
/r/technology	0.132
/r/books	0.641
/r/democrats	1.024
/r/the_donald	0.788

The only conclusion is there were large differences between the sentiment of the tile and the sentiment of the comments in /r/books, /r/democrats, and /r/the_donald. Whereas there were relatively small differences in /r/askreddit and /r/technology. I wonder if the large difference in the political subreddits is caused by the general tendencies for political discussions to become so heated. It is hard to draw any conclusion from these two data points especially since the thread titles are almost completely opposite ends of the positive and negative spectrum.

6   Potential Biases

There are a number of things with the data used and how it was analyzed that likely led to incomplete results. The first is the small number of threads. It would have been better if there were more threads from the same subreddit, more subreddits, or both. This would give a more representative view of the opinions of people on Reddit. Two subreddits in particular to add would be /r/republicans and /r/obama as opposites of /r/democrats and /r/the_donald respectively.

Adding more "neutral" subreddits than just /r/books would have likely given a better baseline of Reddit's average position. It is likely that /r/books isn't the most neutral subreddit. I imagine that reading in a person's spare time is often a sign of better education and higher income. This probably in turn correlates higher with liberal views, at least in the United States. This means /r/books might have a higher liberal bias than subreddits like /r/pics or /r/gifs.

Another issue was all the thread titles were different. Most obviously, the /r/askreddit thread asking for reasons of people who don't support net neutrality. Clearly this will lead to certain types of responses. On the other hand the /r/the_donald and the /r/democrats thread titles were about Trump and Obama respectively. This definitely led to certain kinds of discussions of net neutrality which were very focused around a specific political figure.

There was also the inherent difference between times of the threads and the state of net neutrality in the government. The /r/askreddit was posted on 11/21/17, /r/technology on 07/22/17, /r/books on 11/21/17, /r/democrats on 12/01/17 and /r/the_donald on 01/08/19. The first four were before or right around the time of the FCC's repeal of the net neutrality rules set in place by the Obama administration. The /r/the_donald post is much more recent and meant to be a reflection on the state of the internet after the repeal.

A huge potential source of bias is the inherent bias I have since so much of the work of this project was based on my subjective opinion. Every time I coded a comment, I was making a judgment based on my ideology and perceptions about what other people were thinking. There may have times where I would assume too quickly that someone supported net neutrality just because they expressed other liberal views or vice versa with conservative views. Similar things happened when deciding between politician reason versus economic reasons for their stance. For instance, is discussing government created monopolies a political reason or an economic reason? I am not sure I made decisions on questions like these consistently which could have skewed the results. This is likely where having

more than one person on this project would have been valuable to give a more neutral view.

A last potential source of bias is the way the comments were scraped from the threads. The way the Reddit API's Python wrapper works, the hierarchy of comments is not preserved since a depth first traversal is employed when traversing the comment tree. This meant that comments didn't necessarily follow from their parents as they do when browsing Reddit normally. This removes any interaction and context between two different comments since it wasn't always possible to tell which comments were replying to another. On one hand, this could have removed some biases introduced by associating context of surrounding comments. Instead, I generally treated each comment in a vacuum within the thread. This same lack of context also could cause issues where a misunderstanding of someone's point causes me to miss code their comment. It is not clear what the effect of this loose comment ordering is. Maybe it would have been better to entirely randomize the comments so there is no discernible order.

7 Further Work

The first thing to do would be to gather more data to get a more complete picture. It would be interesting if possible to be able to control for when the threads were posted. It would be interesting to see if people's opinions shifted significantly over time. Especially during the Obama administration and the Trump administration.

What would probably make this work better would to conduct surveys in each of the subreddits. Something with as neutral of a prompt as possible like "What is Your Opinion on Net Neutrality?" Then having a few simple questions that would fall into the categories above like stance and general reasoning. Finally, there could be a open ended response section. This would probably give a more accurate view of what each subreddits population think of net neutrality.

8 Conclusion

This project looked to answer if different subreddits had generally different opinions on net neutrality. The answer was as expected. The generally left-leaning Reddit population was generally in support of net neutrality. This was seen primarily on the subreddits /r/technology, /r/books, and /r/democrats. Due to the nature of the question asked, /r/askreddit's thread was the most balanced one. The heavily conservative /r/the_donald very strongly leaned toward being against net neutrality.

References

[Han18] Mina J Hanna, *Net neutrality: A brief overview of the policy and the fcc's ruling to upend it*, Computer **4** (2018), 78–81.

[HG14] Clayton J Hutto and Eric Gilbert, *Vader: A parsimonious rule-based model for sentiment analysis of social media text*, Eighth international aaai conference on weblogs and social media, 2014.

[HW06] Robert W Hahn and Scott Wallsten, *The economics of net neutrality*, The Economists' Voice **3** (2006), no. 6.

[LW09] Robin S Lee and Tim Wu, *Subsidizing creativity through network design: Zero-pricing and net neutrality*, Journal of Economic Perspectives **23** (2009), no. 3, 61–76.

[Pea18] Joanna Pearlstein, *It's true: Tech workers overwhelmingly support democrats in 2018*, Wired, 2018.

9 Appendix

Full results:
   /r/askreddit:
   Stances:
   Percentage for Probably against NN: 0.21951219512195122
   Percentage for Probably for NN: 0.21951219512195122
   Percentage for For NN: 0.21951219512195122
   Percentage for Ambigious: 0.1951219512195122

Percentage for Against NN: 0.14634146341463414
Reasons:
Percentage for Economic Reasons: 0.6829268292682927
Percentage for Political Reasons: 0.24390243902439024
Percentage for Other: 0.07317073170731707
Percentage for Personal Reasons: 0.0
Combined Stance and Reason:
Percentage for Probably for NN and Economic Reasons: 0.1951219512195122
Percentage for For NN and Economic Reasons: 0.17073170731707318
Percentage for Probably against NN and Economic Reasons: 0.14634146341463414
Percentage for Ambigious and Economic Reasons: 0.14634146341463414
Percentage for Against NN and Political Reasons: 0.12195121951219512
Percentage for Probably against NN and Political Reasons: 0.04878048780487805
Percentage for Probably for NN and Political Reasons: 0.024390243902439025
Percentage for Probably against NN and Other: 0.024390243902439025
Percentage for For NN and Political Reasons: 0.024390243902439025
Percentage for For NN and Other: 0.024390243902439025
Percentage for Ambigious and Political Reasons: 0.024390243902439025
Percentage for Ambigious and Other: 0.024390243902439025
Percentage for Against NN and Economic Reasons: 0.024390243902439025
Percentage for Probably for NN and Personal Reasons: 0.0
Percentage for Probably for NN and Other: 0.0
Percentage for Probably against NN and Personal Reasons: 0.0
Percentage for For NN and Personal Reasons: 0.0
Percentage for Ambigious and Personal Reasons: 0.0
Percentage for Against NN and Personal Reasons: 0.0
Percentage for Against NN and Other: 0.0
/r/technology:
Stances:
Percentage for For NN: 0.4878048780487805
Percentage for Probably for NN: 0.36585365853658536
Percentage for Probably against NN: 0.07317073170731707
Percentage for Against NN: 0.04878048780487805
Percentage for Ambigious: 0.024390243902439025
Reasons:
Percentage for Political Reasons: 0.3170731707317073
Percentage for Economic Reasons: 0.2926829268292683
Percentage for Other: 0.2682926829268293
Percentage for Personal Reasons: 0.12195121951219512
Combined Stance and Reason:
Percentage for Probably for NN and Economic Reasons: 0.14634146341463414
Percentage for For NN and Other: 0.14634146341463414
Percentage for For NN and Economic Reasons: 0.14634146341463414
Percentage for For NN and Political Reasons: 0.12195121951219512
Percentage for Probably for NN and Political Reasons: 0.0975609756097561
Percentage for Probably for NN and Other: 0.07317073170731707
Percentage for For NN and Personal Reasons: 0.07317073170731707
Percentage for Probably for NN and Personal Reasons: 0.04878048780487805
Percentage for Probably against NN and Political Reasons: 0.04878048780487805

Percentage for Against NN and Political Reasons: 0.04878048780487805
Percentage for Probably against NN and Other: 0.024390243902439025
Percentage for Ambigious and Other: 0.024390243902439025
Percentage for Probably against NN and Personal Reasons: 0.0
Percentage for Probably against NN and Economic Reasons: 0.0
Percentage for Ambigious and Political Reasons: 0.0
Percentage for Ambigious and Personal Reasons: 0.0
Percentage for Ambigious and Economic Reasons: 0.0
Percentage for Against NN and Personal Reasons: 0.0
Percentage for Against NN and Other: 0.0
Percentage for Against NN and Economic Reasons: 0.0

/r/books:
Stances:
Percentage for Probably for NN: 0.4878048780487805
Percentage for For NN: 0.2926829268292683
Percentage for Ambigious: 0.17073170731707318
Percentage for Probably against NN: 0.04878048780487805
Percentage for Against NN: 0.0

Reasons:
Percentage for Political Reasons: 0.34146341463414637
Percentage for Other: 0.2682926829268293
Percentage for Economic Reasons: 0.24390243902439024
Percentage for Personal Reasons: 0.14634146341463414

Combined Stance and Reason:
Percentage for Probably for NN and Political Reasons: 0.17073170731707318
Percentage for Probably for NN and Other: 0.12195121951219512
Percentage for Probably for NN and Personal Reasons: 0.0975609756097561
Percentage for Probably for NN and Economic Reasons: 0.0975609756097561
Percentage for For NN and Political Reasons: 0.0975609756097561
Percentage for For NN and Economic Reasons: 0.0975609756097561
Percentage for For NN and Other: 0.07317073170731707
Percentage for Ambigious and Other: 0.07317073170731707
Percentage for Probably against NN and Political Reasons: 0.04878048780487805
Percentage for Ambigious and Economic Reasons: 0.04878048780487805
Percentage for For NN and Personal Reasons: 0.024390243902439025
Percentage for Ambigious and Political Reasons: 0.024390243902439025
Percentage for Ambigious and Personal Reasons: 0.024390243902439025
Percentage for Probably against NN and Personal Reasons: 0.0
Percentage for Probably against NN and Other: 0.0
Percentage for Probably against NN and Economic Reasons: 0.0
Percentage for Against NN and Political Reasons: 0.0
Percentage for Against NN and Personal Reasons: 0.0
Percentage for Against NN and Other: 0.0
Percentage for Against NN and Economic Reasons: 0.0

/r/democrats:
Stances:
Percentage for Probably for NN: 0.4722222222222222
Percentage for Ambigious: 0.3333333333333333
Percentage for Against NN: 0.111111111111111

Percentage for Probably against NN: 0.05555555555555555
Percentage for For NN: 0.027777777777777776
Reasons:
Percentage for Political Reasons: 0.6666666666666666
Percentage for Economic Reasons: 0.16666666666666666
Percentage for Other: 0.1388888888888889
Percentage for Personal Reasons: 0.027777777777777776
Combined Stance and Reason:
Percentage for Probably for NN and Political Reasons: 0.3888888888888889
Percentage for Ambigious and Political Reasons: 0.16666666666666666
Percentage for Ambigious and Other: 0.08333333333333333
Percentage for Probably for NN and Economic Reasons: 0.05555555555555555
Percentage for Ambigious and Economic Reasons: 0.05555555555555555
Percentage for Against NN and Political Reasons: 0.05555555555555555
Percentage for Probably for NN and Other: 0.027777777777777776
Percentage for Probably against NN and Political Reasons: 0.027777777777777776
Percentage for Probably against NN and Economic Reasons: 0.027777777777777776
Percentage for For NN and Political Reasons: 0.027777777777777776
Percentage for Ambigious and Personal Reasons: 0.027777777777777776
Percentage for Against NN and Other: 0.027777777777777776
Percentage for Against NN and Economic Reasons: 0.027777777777777776
Percentage for Probably for NN and Personal Reasons: 0.0
Percentage for Probably against NN and Personal Reasons: 0.0
Percentage for Probably against NN and Other: 0.0
Percentage for For NN and Personal Reasons: 0.0
Percentage for For NN and Other: 0.0
Percentage for For NN and Economic Reasons: 0.0
Percentage for Against NN and Personal Reasons: 0.0
/r/the_donald:
Stances:
Percentage for Probably against NN: 0.6585365853658537
Percentage for Against NN: 0.2682926829268293
Percentage for Ambigious: 0.07317073170731707
Percentage for Probably for NN: 0.0
Percentage for For NN: 0.0
Reasons:
Percentage for Political Reasons: 0.6341463414634146
Percentage for Other: 0.1951219512195122
Percentage for Economic Reasons: 0.14634146341463414
Percentage for Personal Reasons: 0.024390243902439025
Combined Stance and Reason:
Percentage for Probably against NN and Political Reasons: 0.3902439024390244
Percentage for Against NN and Political Reasons: 0.21951219512195122
Percentage for Probably against NN and Other: 0.14634146341463414
Percentage for Probably against NN and Economic Reasons: 0.12195121951219512
Percentage for Against NN and Other: 0.04878048780487805
Percentage for Ambigious and Political Reasons: 0.024390243902439025
Percentage for Ambigious and Personal Reasons: 0.024390243902439025
Percentage for Ambigious and Economic Reasons: 0.024390243902439025

Percentage for Probably for NN and Political Reasons: 0.0
Percentage for Probably for NN and Personal Reasons: 0.0
Percentage for Probably for NN and Other: 0.0
Percentage for Probably for NN and Economic Reasons: 0.0
Percentage for Probably against NN and Personal Reasons: 0.0
Percentage for For NN and Political Reasons: 0.0
Percentage for For NN and Personal Reasons: 0.0
Percentage for For NN and Other: 0.0
Percentage for For NN and Economic Reasons: 0.0
Percentage for Ambigious and Other: 0.0
Percentage for Against NN and Personal Reasons: 0.0
Percentage for Against NN and Economic Reasons: 0.0