Exploring How Online Communication Relates to Personality Types: A Reddit MBTI Community Analysis

Lukas Wanzeck[a,1]**, Andreas Raaskov**[a,1]**, and Clair Mutebi**[a,1]

This manuscript was compiled on March 22, 2025

Personality tests are very common in both corporate settings and universities, with over 2 million individuals taking these tests annually to gain insights into their personality traits. Understanding different personality types allows for deeper understanding of each member's work approach and communication style.

Our primary objective is to figure out how personality types defined by Myers and Briggs relate to online communication in the Reddit forum MBTI. An analysis of data is performed shedding light on the overrepresented usage of online communication of intuitive personality types. Simultaneously, we investigate tendencies of personalities to cluster within specific communities or bridge across diverse groups using modularity analysis and the Louvain algorithm. The best partition showed temporal structures and a few meaningful groupings of various, but mainly introverted, personality types. Additionally, our study employs sentiment analysis to explore the linguistic nuances within posts for each personality type. Individuals who either are feeling, judging or extroverted showed a more positive sentiment when compared to the opposite traits.

| Reddit | MBTI | Network analysis | Personality type insights | Personality type behavior|

Significance Statement

Communication is a key element in every society. Combined with the reality that every human can be classified into one of 16 personality types using the Myers-Briggs Type Indicator, it is important to understand how these character traits influence communication. The understanding of such has not only the ability to optimize group work in a business setting but reaches far into private life as well. If aware, an individual will be able to shape their communication and comprehend the communicating partner. Moreover, this analysis is performed on online communication taking place on Reddit. This is particularly relevant as a substantial amount of communication happens online nowadays. Also, Reddit in particular offers the possibility to communicate freely, as users keep their anonymity.

I n today's world, personalities are not just about how one acts in person, they also come through in how we behave and communicate online. Universities and workplaces are starting to realize how much someone's personality impacts teamwork and use personality tests to look at team dynamics. The influence of personality types extends beyond the confines of workplaces and universities. An emerging trend sees individuals openly showcasing their personality type, often adopting it as a distinctive identifier within various online platforms. Displaying one's personality type serves to not only build communities but also foster a richer understanding of individual personality traits. The system that is often employed is Myers-Briggs Type Indicator (MBTI), which identifies personality patterns by answering a series of questions.

Our aim is to uncover the connection between the MBTI classification and the dynamic of the personality types' interactions in the online forum MBTI subreddit during the years 2010 to 2022. Moreover, this research seeks to explore the online behaviour of different personality types and understand their communication styles, preferred interaction partners and variance in expressions. This paper revolves around the exploration of the following distinct research problems

- Is the data representative?

- Is there a connection between personality types or traits and social structures?

- Do personality types influence the sentiment of communication?

As the analysis of these problems resolves around the MBTI system, it allows for a comparative analysis into how the description of the 16 personality types is correlated with the results of this exploration.

Author affiliations: [a]DTU Compute, Richard Petersens Plads, Build. 324 DK-2800 Kgs. Lyngby

Myers Briggs defines 16 personality types by categorizing people by four aspects. Hence, all personality types can be named with four letters. The first letter represents how one interacts with others where one gets labelled as introvert ("I - - -") or extrovert ("E - - -"). The next aspect determines how one sees the world and processes information. One is then either observant ("- S - -") or intuitive ("- N - -"). Also how one makes decisions and copes with emotions is taken into account. One can be thinking ("- - T -") or feeling ("- - F -"). Finally, how the approach to work, planning and decision-making is handled also plays a role. One will either be labelled as judging ("- - - J") or prospecting ("- - - P") (1).

Results

Is the data representative? A solid basis for every analysis is to check whether or not the data is representative. In our case, this means whether or not the distribution of personality types in the Reddit forum is a reflection of the distribution of personality types in the real world. As not every person has taken such a test, we will have to go with an estimate given by (2) on the distribution in the real world.

Looking at the direct comparison in figure 1 it becomes apparent that the distributions are not similar. A few personality types share similar representations, one such is ENFJ. On the other hand, there are many personality types that are overrepresented extremely such as INTP, INFP, INFJ, INTJ and ENTP. Also, the opposite can be seen with ESTP, ESTJ, ISFJ, ISTJ, ESTP and ESFJ which are greatly underrepresented. We further notice that the four introverted intuitive personality types are the most represented ones in the forum.
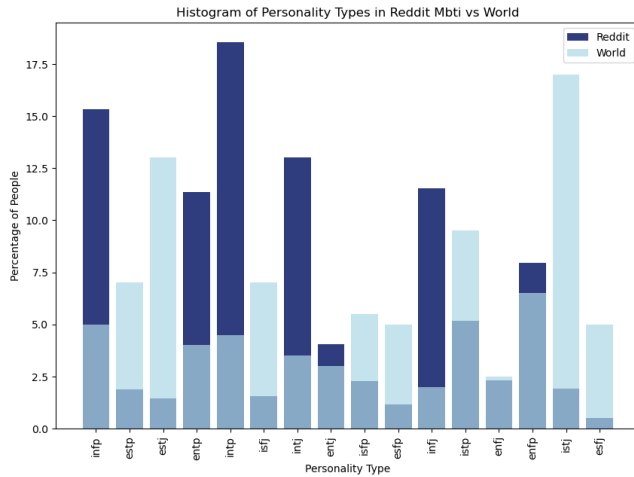

Fig. 1. Histogram of the different personality types both for the Reddit forum MBTI and an according estimate of the distribution in the world given by (2). *The code for this figure can be found in 3_Stats.ipynb*

We head on to investigate whether or not there exists a system in the personality types' communication unveiling community structures.

Is there a connection between personality types or traits and social structures? To evaluate communities we partition

the data set and compute the modularity of each partition. In total, there are 6 different partitions. Below is a list of these partitions including their modularity in parenthesis. We partition

- Randomly (0.008)
- Into personality types (0.078)
- Introverts versus extroverts (0.039)
- Intuitive versus observant (0.029)
- Thinking versus feeling (0.048)
- Judging versus prospecting (0.042)
- Using the Louvain algorithm (0.518)

First of all, we notice that as we would expect a random partition has a modularity close to zero. Moreover, a partition into personality types is better compared to partitioning into character traits. Comparing only partitions into character traits there is no significantly better partition.

It is evident that the partition achieved by running the Louvain algorithm yields the highest modularity. As this is the best partition we will investigate this one further. To be more precise, we would like to find out how many communities there are, how large they are and what personality types are grouped together.
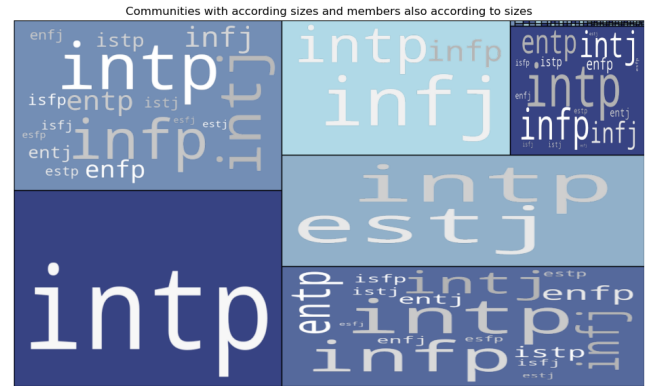

Fig. 2. Result of the Louvain algorithm where the sizes of the communities are proportional to the members. Each square also contains a word cloud with the distribution of the member's personality types. *The code for this figure can be found in 5_Communities.ipynb*

This is exactly what figure 2 shows. There are six significant communities, ranging from 10198 members down to 3371 members. In contrast, the other 87 communities have 5 or less members. Most noticeable is the fact that the largest community has members only from INTP. The second, third and sixth largest communities exhibit a similar distribution pattern as the overall forum, see figure 1. Additionally, the are two large communities consisting of INTP and ESTJ and one of INTP, INFP and INFJ.

As it is now clear which personality types prefer to be in communication with which personality types, we would like to investigate how the tone in these conversations is.

Do personality types influence the sentiment of communication? To find out how the sentiment throughout the personality types' communication is dispersed, we performed a sentiment analysis based on all the posts from each personality type. As seen in figure 3, a clear pattern with a few exceptions has formed. Below are patterns including the p-value of a one-sided welsh-t-test.

- Feelers have a higher sentiment than thinkers ($p < 0.0001$).
- Judgers have a higher sentiment than prospecting ($p < 0.0001$).
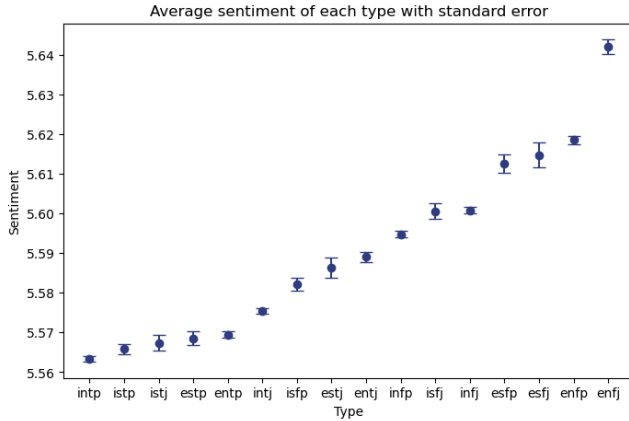- Extroverts have higher sentiment than introverts ($p < 0.0001$).

Fig. 3. Average sentiment of each personality type with standard error. *The code for this figure can be found in 4_text_analysis_graph*

Of interest is not only with which tone these personality types generally communicate but also if they have preferred communication partners or not. Figure 4 shows the amount of posts addressed to the personality types. It is obvious that personality types generally often post on things coming from their own personality types and on posts from large represented personality types.

Even more interesting than knowing with which sentiment personality types communicate and who they like to address would be knowing whether personality types communicate extra positively or negatively to another personality type. This could disclose information about which personality types work well together and which do not. In figure 5 the mean sentiment each personality type uses when commenting on another personality type's post can be seen as well as the sentiment for submissions. The pattern we see here is that feelers are writing more positive comments than thinkers, but also that feelers receive way more positive comments in return.

Discussion

In regards to the analysis concerning the representability of our data set, the conclusion must be that the results found in this paper are not representative of the world's population. It will however still give valuable indications about the communication behavior of the different personality types. Just the fact that intuitive personality types are overrepresented and observant is underrepresented displays the fact that intuitive individuals like to discuss matters

Fig. 4. A graph on matrix form that shows which personality types talk to each other. Each column is how many posts a type has made and each row is what type was commented on or if the post was a submission (post that is shown on the Reddit page). The weights is how many times each type of post was made. *The code for this figure can be found in 4_1_sentiment_analysis.ipynb*
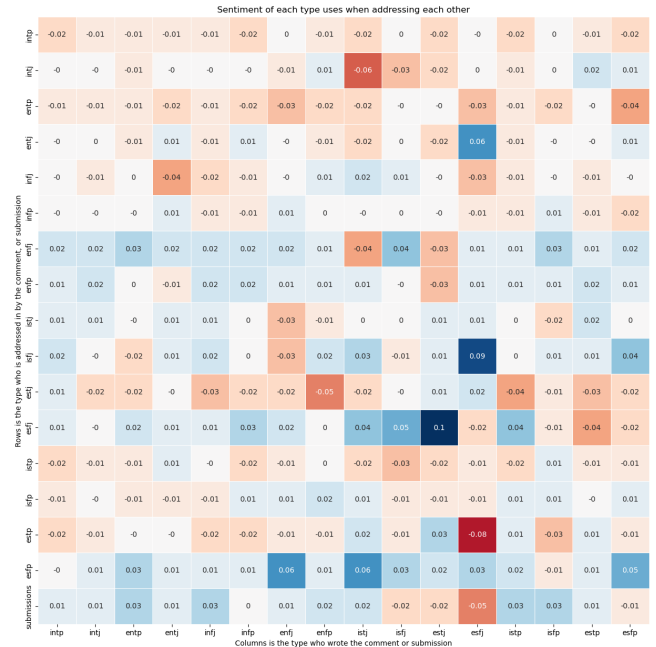

Fig. 5. A graph on matrix from. It has the same structure as the graph in figure 4 however the weights are now the sentiment of each type of post, in order to highlight the difference in communication the mean from each type seen in figure 3 was substrate from each row corresponding to normalising on each nodes outgoing edges. *The code for this figure can be found in 4_1_sentiment_analysis.ipynb*

more often than observant individuals do ([3](#)). As Reddit is a platform giving people the opportunity to discuss it makes sense that these personality types are overrepresented. It should also be noted that all personality types on Reddit are self-reported and that we have no way to confirm if people reported their personality correctly.

Our findings are limited by the choice of the data set. This raises the question of whether another platform as a source could have been preferable. In social media dynamics, the correlation between personality types and platform preferences sparks an interest for investigation. For instance, contrasting platforms such as Instagram, known for it's visually-centric approach emphasising on visual storytelling and self-presentation, tend to resonate with traits associated with extroversion. Conversely, Twitter thrives on conciseness and public nature. Meanwhile, Reddit emphasises anonymity, community-driven content and discussion, where users value depth and community engagement.

Analyzing the personality types within these platforms can say something about individual personality tendencies, user engagement, communication and language. For instance, recent studies have highlighted trends, revealing that INFJs are the dominating personality type on Twitter ([4](#)), while INTPs are dominating personality type on Reddit. Understanding the characteristics of these personalities sheds light on their likemindedness towards these platforms. The INFJ's nuanced expression and empathetic communication might naturally resonate with Twitter's platform dynamics, whereas the INTP's analytical and introspective nature could resonate with the anonymous, in-depth discussions on Reddit. Due to time constraints, an exploration into this aspect remain unfulfilled.

From figure [2](#) we saw that there were 6 large communities containing 99.7% of the total users. We suspect this community structure to be present due to temporal structures independent of personality types. The MBTI Reddit forum experienced a surge in new subscribers and posts, especially in the past couple of years ([5](#)). It is possible that new users are active for some time and then tend to be inactive. This leaves redundant users and could be a reason that three large communities with a similar personality type distribution exist. Simultaneously, as the temporal aspect seems to partially control the community structures, it is not astounding that three of the large communities reflect the overall personality type distribution. To determine the exact reasons for this community structure further research would be needed. One could investigate the time of activity within a community and do a qualitative assassination of the connectivity between members.

The fact that the largest community only consists of INTP can be explained by the fact that INTP dominates the MBTI forum as seen in figure [1](#) and that INTP mostly writes to other INTP's as seen in figure [4](#). The community of INTP and ESTJ exists possibly because ESTJ writes more with INTP than with any other personality type, however this is also true for almost all observant personality types according to figure [4](#). The community of INTP, INFJ, INFP also makes fine sense since they are all introverts and intuitive that according to figure [1](#) is common on the internet but rare in the general population. Thus it makes sense that they meet in their own community, however according to this theory INTJ should also be present in this community.

Using Clique Percolation to find communities would have presented an interesting angle on the topic of communities. This would have allowed to unveil structures over more than one character trait. This would be a method for further research.

From the results of the sentiment analysis we see that feelers tend to express themselves, and receive comments, more positively compared to thinkers. The result regarding expression is confirmed by Dr. A.J. Drenth ([6](#)). Feelers naturally incline towards addressing emotional needs, whereas thinkers lean towards logical thinking ([7](#)). This inclination could explain why feelers are more expressive while thinkers may display less variance in expression as shown in figure [3](#). We also found that extroverts have a more positive sentiment than introverts do indicating that they are more sociable. Judging personality types have a higher sentiment than prospecting ones, this may be because judgers are more inclined to follow rules of general politeness.

Materials and Methods

Data and Preprocessing All data was downloaded from https://the-eye.eu/redarcs/ in two files one containing all submissions and one containing all the comments. The data-cleaning process involved removing all posts without an author (a post stayed even when a Reddit profile was deleted but the author is no longer shown). Posts where the author had not indicated a flair (their personality type) were removed as well and posts not containing text were also removed (happens if a post is just an image or a link).
All text was preprocessed by removing all "'s" and all non-alphabetic characters. The whole text was converted to lowercase. The text was then tokenized and had its English stopwords removed. Furthermore, we lemmatized the text with WordNetLemmatizer, all provided by the NLTK library ([8](#)).

Network The Graph is structured such that each comment is an edge and each Reddit user is a node. This means that each node has a personality, and each edge has some text. The graph is directed, where an edge goes from the author of a comment to the person whose post was commented on. This also means that we have a lot of multi-edges where two people engage in a thorough conversation. We also had to delete a lot of edges without a recipient node for cases where the post commented on was deleted.
The final network has 45044 nodes and 880399 edges, of which 381805 are multi-edges.
The power law exponent of the network was calculated using the power-law package ([9](#)).
- The exponent for in-degree is: 2.84
- The exponent for out-degree is: 3.11

We see that the in-degree distribution is less than 3 indicating scale free properties, while the out degree is more than 3 indicating a random network([10](#)). This also indicates that a few single users start debates that a lot of people comment

on while users each user is limited in how many comments they can make in commenting.

In order to perform the sentiment analysis we needed a graph on how many times personality types comment on each other so we made a graph with 16 nodes (one for each personality type). Each edge was then weighted by how many times a person of a personality type had commented on another personality type this graph has been visualised as matrix in figure 4. Each node in this graph also has a number for how many submissions this personality type made.

Communities For visualisation and computational purposes, we exclude 332 users who only post but do not comment in this part of the analysis.

Modularity. (11) Modularity is a measure aimed at determining, whether a partition of nodes in a graph forms a set of communities or not. It works only on the assumption that communities are more tightly linked than a random graph would be (10, H2 in section 9.2). Modularity compares the number of links in a proposed community to the expected number of links if the graph was randomly wired. The modularity of a single community is calculated as follows

$$M_c = \frac{L_c}{L} - \left(\frac{k_c}{2L} \right)^2$$

where L_c is the total number of links in the proposed community, whereas L is the total number of nodes in the whole graph. k_c is the total degree within the proposed partition of nodes. The modularity for the whole graph is simply the sum of all the modularities for all partitions.

The intuition behind this formula is quite simple. It is a difference between the normalized amount of links in the community with the expected amount of links in a random graph with the same degree sequence. In other words, if there are more links in the partition than expected it is more likely to be a community.

Hence, if the difference and therefore the modularity is positive, it is more likely to be a community. If the modularity is zero, then it has exactly the number of links as we expect a random graph to have and is not a community. In the case that the modularity is negative, we have isolation.

The Louvain Algorithm. We used the implementation community_louvain from the community package. Louvain's algorithm does two things. It starts by finding local communities through modularity optimization. When the communities are found it aggregates each community as one node. This is repeated iteratively until a maximum of modularity is reached (11)(12). Furthermore, to take the randomness in the algorithm into account we ran it 100-times and then took the partition with the median number of communities.

Sentiment analysis The sentiment analysis was made using the LabMT wordlist that consists of 5000 words rated between 1 to 9 by mechanical Turks to give a robust sentiment per word (13). To calculate sentiment for a personality, each word was then given a sentiment score by the LabMT wordlist or ignored if it was not in the list. The mean and standard error for each type over all their text can be seen in 3.

To see if the findings from figure 3 were significant we sorted all personality types after the character traits and performed a one-sided welsh-t-test (14). To figure out how different personality types talk to each other we performed a sentiment analysis on each edge of the graph shown in figure 4. We then subtracted the sentiment mean of each personality type in order to make the graph in figure 5.

Conclusion

We headed out to explore how MBTI personality types communicate on the Reddit forum MBTI and whether we could draw parallels to the character traits described by (1). The analysis revealed the personality types generally acted as expected. We saw that introverts prefer anonymous online communication, that several personality types tend to be in closer contact with other personality types sharing similar character traits and that emotional individuals overall use a more positive language.

These insights are of great value and can be used to improve the quality of communication in many different settings. Though, many more insights can be gained by considering different data sets, finding more complex community structures (Clique Percolation) or performing a qualitative analysis of the exchanged texts used to do the sentiment analysis. That this research still has great potential is also shown in the fact that the major limitation was time and else only lied in the choices of data and research methods.

Contributions

In the following the authors are abbreviated the following: Lukas Wanzeck (LW), Andreas Raaskov (AR) and Clair Mutebi (CM).

- Abstract & Significance Statement (LW,CM)
- Introduction (LW,CM)
- Results (LW,AR)
- Discussion (LW,AR,CM)
- Materials and Methods (LW,AR)
- Conclusion (LW)

References

1. 16personalities theory (https://www.16personalities.com/articles/our-theory) (2023) Accessed: 2023-12-01.
2. Mbti world population & mbti by country statistics (https://typologytriad.wordpress.com/mbti-population-by-country/) (Oct. 2020) Accessed: 2023-11-29.
3. Mind: Intuitive (n) vs. observant (s) (https://www.16personalities.com/articles/mind-intuitive-vs-observant) (2023) Accessed: 2023-11-29.
4. Y Wang, Understanding Personality through Social Media. (year?).
5. r/mbti reddit statistics (https://subredditstats.com/r/mbti) (2023) Accessed: 2023-12-03.
6. How Thinkers (T) vs. Feelers (F) Deal with Emotions. (year?).
7. Thinkers and Feelers: Bridging the Personality Divide (https://www.aim.com.au/blog/thinkers-and-feelers-bridging-the-personality-divide) (year?).
8. NLTK :: Natural Language Toolkit (https://www.nltk.org/) (year?).
9. A Clauset, C Rohilla Shalizi, ME J Newman, POWER-LAW DISTRIBUTIONS IN EMPIRICAL DATA. (2009).
10. AL Barabási, M Pósfai, *Network science.* (Cambridge University Press, Cambridge), (2016).
11. Assignment 2 02805 (2023).
12. Louvain algorithm (https://sites.google.com/site/findcommunities/) (2015) Accessed: 2023-12-03.
13. PS Dodds, KD Harris, IM Kloumann, CA Bliss, CM Danforth, Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. *PLoS ONE* **6**, 26752 (2011).
14. scipy.stats.ttest_ind — SciPy v1.11.4 Manual (https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest$_ind.html)(year?).

Wanzeck, Madsen, Mutebi *et al.*

PNAS — **March 22, 2025** — vol. XXX — no. XX — **5**