THE UNIVERSITY OF CHICAGO

Exploring Personality and Online Social Engagement: An Investigation of MBTI Users on Twitter

By

Partha Kadambi

June 2022

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Prof. Allyson Ettinger
Preceptor: Dr. Sanja Miklin

Exploring Personality and Online Social Engagement: An Investigation of MBTI Users on Twitter

Partha Kadambi

Abstract

Text-based personality analysis by computational models is an emerging field with the potential to significantly improve on key weaknesses of survey-based personality assessment. We[1] investigate 3848 profiles from Twitter with self-labeled Myers-Briggs personality types (MBTI) – a framework closely related to the Five Factor Model of personality – to better understand how text-based digital traces from social media can be used to predict user personality traits and explore relationships between community discourse and personality traits. We leverage machine learning and network analysis approaches to analyze relationships between behavioral traits and various facets of engagement on social media. We find that both social engagement metadata and multiple modalities of text from social profiles significantly correlate with factors of the MBTI system – especially the Intuitive/Sensing trait that corresponds to the dimension of Openness in the Five Factor model of personality (the Big 5). We discuss our findings and their implications for the validity of the MBTI and the lexical hypothesis, a foundational theory underlying the Five Factor Model of personality that links language use and behavior. Our results hold optimistic implications for personality psychologists, computational linguists, and other social scientists seeking to clarify the role of behavioral dimensions underlying semi-structured text-based social engagement and illuminate the links between sociolinguistics and core dispositional traits.

[1] The use of 'we' refers to the sole author.

1. Introduction

The modeling of personality traits offers a powerful approach to understanding the diversity of human behavior at the level of the individual and in theorizing the emergence of complex society-wide phenomena. A complete understanding of several social phenomena – from the formation of specific political beliefs to the diversity of lifestyle and information preferences – lies downstream of an understanding of the nature of core cognitive/psychological preferences of persons. Personality psychometrics attempts to discover these crucial axes of human behavior and devise methods to measure these dimensions.

A growing number of political analysts, social psychologists, and behavioral economists are incorporating models of personality in their studies to account for the effects of individual differences in the decision-making attitudes of people. These differences are important in determining social outcomes[1]. Particular measures of personality differences such as the five-factor model (FFM) have demonstrated significant correlations with several indicators of achievement and social outcomes, including college GPA and job satisfaction[2, 3]. The Myers Briggs Type Indicator (MBTI) [4] – which comprises four scales that all directly correlate with FFM traits[5] – is closely related and enjoys high popularity in public culture and has shown to correspond to qualitative assessments of communication styles and other behavior[6, 7]. Personality psychometrics can serve as a useful tool in understanding the links between specific social outcomes germane to social scientists and individual behavior. They also provide efficient methods to control for behavioral variance in human subject research in the social sciences.

Psychometrics has conventionally employed survey-based personality questionnaires in its methodology. The increasing availability of rich social data online and the development of advanced data science tools – including syntactic parsers capable of analyzing text at scale – has spurred interest in computationally enhanced methods for text-based personality psychometrics. Data science approaches leverage a variety of statistical and machine-learning methods to extract signals of personality traits from text data. These include powerful methods of exploratory analyses that help visualize relationships between personality labels and semi-structured text and 'supervised' methods that explicitly search for and quantify signals between labeled traits and the data of interest.

A computationally enabled, text-based observational research design has certain key advantages over a questionnaire-based design. An observational approach is nonreactive – the subject's behavior is closer to their natural state in this case as compared to when prompted by a survey. Social text is a window into the social behavior of a person: it encapsulates a wide range of behavioral granularities. This allows for a possibly more intimate measurement of personality traits – within the context of online behavior – than survey instruments that use Likert scales. Conversely, it also allows for understanding relationships between core personality traits and a larger and finer range of behavioral outcomes – some of which are marked by only subtle psycholinguistic cues. This stands in contrast to the mainstream paradigm of outcome-oriented personality research that has historically focused largely on macro-outcomes such as health, education, and career success. Lastly, a computationally enabled observational study is vastly cheaper and more scalable than similarly powerful human subject experiments.

Advancements in Natural Language Processing in the past few years, which have significantly improved the ability of computers to consider context when extracting semantic information, have ushered in a new paradigm of digital text processing. While digital text-based psychometrics has been demonstrated as a possibility for longer than a decade, the recent wave of progress in NLP can likely empower significantly improved text-based predictive models of personality. This opens the door to a new class of questions that further probe the links between semantic patterns underlying text-based social communication and personality traits.

While these advancements can be framed as 'new', the principles at the core of this approach are strongly cemented in the foundations of some of the most influential frameworks of personality, such as the five-factor model (FFM). The FFM was originally conceived from the lexical hypothesis[8] – the idea that psychological states and dimensions of behavior are encoded within language. We advance the view that computationally leveraged large-scale text analyses can provide crucial validation of some of the fundamental theories underlying dominant models of personality. In our view, text-based approaches are not principally sub-standard to survey-based approaches to personality psychometrics (which are often termed as the gold standard for ground truth); rather, we believe that text-based approaches directly investigate empirical claims underlying the construction of personality traits construed from the lexical hypothesis. Kulkarni et al. (2018)[9] precisely espouse this position in their attempt to distil latent personality traits from unstructured text from discourse on Facebook.

This paper aims to understand how personality traits relate to online social communications through an analysis of social data from the profiles of Twitter users. We investigate Twitter users who self-label their personality 'type' within the MBTI personality framework. This includes exploring the semantic spaces occupied by the social discourse *generated* by different personality types and also how personality traits relate to information *consumption* preferences online. We do this through an analysis of multiple facets of user data on Twitter – tweets, likes, biographies, and user engagement metadata.

Our second objective is to evaluate the validity of key constructions and theories underlying the MBTI system by analyzing the prediction accuracies of machine-learning-based models tasked with predicting user MBTI traits (given user profile data). These models, some of which can extract contextually aware semantic information, may be sufficiently powerful to give us a rough estimate of the true upper bound of prediction accuracy on the task. The ability of these models to label user traits correctly would be indicative of, to some extent, the significance of the trait under question itself. High accuracies would validate the existence of the trait, though the opposite does not necessarily hold if we were to obtain relatively low accuracies.

2. Background and related literature

Text-based approaches are relatively new to personality psychometrics, as methods capable of analyzing text data to significant depth have only been developed recently. Quercia et al. (2011)[10] were amongst the early pioneers of using Twitter data to predict personality. That personality traits are strongly linked to social media interaction patterns was best demonstrated by Kosinski et al. (2013)[11] through a large-scale analysis of Facebook data. The same year, Schwartz et al. (2013)[12] showed that an open-vocabulary approach (a consideration of all tokens appearing in text, as opposed to looking for specific pre-established words/phrases) could succeed in capturing behavioral and demographic attributes. In 2015, Park et al.[13] further explored the possibility of assessing personality through language expressed on social media and found significant correlations between the Big 5 traits and language use online.

While these studies have validated the language-based social media approach to personality assessment, the prediction of personality traits using machine learning and Natural Language Processing is still an active and relatively unexplored research area. This paper aims to improve prior models by adopting sophisticated NLP methods that have been developed in the last two years and using the largest dataset of its kind for the Twitter platform. This would serve the purpose of validating a text-based approach to predict MBTI traits and also possibly support the validity of the MBTI framework itself. We also aim to understand how multiple aspects of social engagement – self-descriptions, information consumption (the kind of information consumed by the user), and information generation (original messages shared by the user) – are related to personality traits. This would investigate how various *kinds* of text-based digital trace, particularly on Twitter, relate to personality traits - a question relatively untouched by prior research.

2.1 MBTI and text-based predictive approaches

The MBTI is a commercially and publicly popular psychometric framework that classifies persons into one of 16 'types'. It measures four traits (Introversion/Extraversion, Sensing/Intuition, Thinking/Feeling, and Judging/Perceiving) and assigns a letter label to each trait dichotomy to produce a 4-letter personality 'type'. The four dimensions are briefly described here:

• Extraversion/Introversion – preference of directing energy and time to external goals and other people or towards personal ideas and emotions; whether people are energized by spending time with other people or by time alone (correlates with FFM extraversion)

• Intuition/Sensing – the kind of information people pay attention to; whether you prefer an abstract and open-ended worldview or one oriented towards details and concrete facts (correlates with FFM openness)

• Feeling/Thinking - preference for decision-making strategy; prioritizing people-oriented, emotionally sensitive decision-making or decision-making oriented by objective principles guided by logic (correlates with FFM agreeableness)

• Judging /Perceiving – whether you prefer clear goal-oriented scheduling or are more flexible and prospecting for new experiences (correlates with FFM conscientiousness)

While this rigid classification of dispositional preferences into two categories has attracted strong criticism[14], it should be noted that the MBTI is a close cousin of the FFM – all four dimensions correlate significantly with four FFM scales[5]. It has been pointed out that the FFM is a superior alternative to the MBTI – however, this is only partly true. While the FFM is a generally better accepted and well-validated tool, it is important to note that the underlying theories behind the MBTI and the FFM are very different. While these frameworks are ostensibly comparable, the interpretational structure underlying the MBTI is more complex than that underlying the FFM – extensive qualitative analysis and anecdotal observations are built into the theory underlying the MBTI. MBTI is based on investigative work by Carl Jung[15] while the FFM is based on the lexical hypothesis and derived from factor analysis. The MBTI simplifies the theory of psychological functions advanced by Carl Jung by framing it as a typology - by doing so, it distorts the concepts it was originally built on (for example, Carl Jung's psychological functions are not dichotomous to the degree espoused by the MBTI, nor are they directly indicative of individual letters in the 4-letter MBTI code). It is noteworthy that the FFM and the MBTI are highly similar – given innumerable possibilities of personality frameworks – despite being developed by two entirely different paradigms. This should give us more confidence that some common behavioral patterns are indeed being captured by these frameworks.

There are two primary justifications for the choice of using MBTI (over the FFM) as the personality framework for purposes of understanding the link between personality traits and social communications online:

1) **Availability of data**: MBTI personality data of users on online social networks is far more accessible than FFM personality data. The MBTI is a highly popular personality theory and large communities of persons who self-identify their MBTI personality type can be found on most large social media websites – including Twitter, Reddit, Facebook, and YouTube. No such communities exist for any of the models that fall under the umbrella of FFM (OCEAN, Big 5, etc.). This makes the FFM unsuitable for large-scale observational studies of social media-based as questionnaire-based approaches will likely be expensive and drastically increase the time needed to collect similar scales of data. On

the other hand, web APIs can be leveraged to search for users who self-label their personality and collect their social engagement information if publicly accessible; automatic digital data collection is far cheaper in terms of both time and money. This project analyzes more than 1M tweets corresponding to about 3500 self-labeled Twitter profiles; running a personality survey (required for measuring FFM dimensions) for this many profiles would be a costly endeavor.

2) **Validating the MBTI:** While the FFM has typically been shown to demonstrate better test-retest validity than the MBTI instrument, this may not reflect the validity of the theory underlying MBTI; rather, this may be because the MBTI survey instrument is itself unfaithful to its underlying theory or is a significantly error-prone approach to analyzing personality. Critical comparisons between the theories underlying the FFM (the lexical hypothesis) and the MBTI (Carl Jung's psychoanalytic theory) are few and far between in existing literature. The theoretical differences between these models of personality create problems for direct comparison of these models. There is some early empirical evidence – from neuroscience – suggesting the existence of neural correlates to Carl Jung's theories of type[16]. Hence, there is sufficient principled cause to consider the MBTI as a distinct instrument from the FFM. This serves to justify the standalone use of the MBTI instrument for this study. Note that validating the presence of MBTI traits would be consistent with the finding that the MBTI attitudes correlate with FFM dimensions. Moreover, as we adopt a purely text-based approach to understand MBTI dimensions, any predictive signals of traits would also speak to the validity of the lexical hypothesis underlying the FFM.

2.2 Review of related literature

Text-based prediction of MBTI labels is currently an active research area and seems to attract more research interest than text-based FFM trait prediction, primarily due to the greater availability of MBTI data. In 2015, Plank and Hovy[17] analyzed 1.2M tweets annotated by MBTI type and found that only the Introvert/Extrovert and Thinking/Feeling traits were predictable with a chance above random chance. However, they used features and techniques that, by today's standards, are fairly basic – they used a logistic regression classifier on bigrams, gender, and meta-features such as tweet counts, etc. In 2016, Verhoeven et al.[18] used word and character n-grams from labeled Twitter posts from users of six languages to predict MBTI trait labels. Only a small fraction of the classifiers (across six languages) they test perform better than the majority baseline. They report that the classifiers that do perform better than random are for the Introvert/Extravert and Thinking/Feeling dimensions; this aligns with the findings of Plank and Hovy.

With the introduction of BERT (Bidirectional Encoder Representations from Transformers) – a deep-learning-based contextual embedding model – by Devlin et al. in late 2018[19], research incorporating BERT to predict MBTI traits started to emerge. Mehta et al. (2020)[20] used an ensemble model composed of BERT and a multi-layer perceptron to predict MBTI labels from a Kaggle dataset composed of posts by users from the personality forum *personalitycafe.com* with self-identified MBTI labels, reporting significant prediction performance for the Thinking/Feeling and Perceiving/Judging dimensions and also SOTA results for the open dataset they used. A similar dataset was analyzed by Keh et al. (2019)[21] using BERT; they found that BERT performed better at I/E and T/F dimensions.

Results from these studies, taken together, suggest that while MBTI personality prediction is not an easy task for NLP-based ML models, some MBTI traits display linguistic signals. However, results based on *personalitycafe.com* posts must be interpreted with skepticism. Firstly, as the forum is dedicated to discussing personality, posts by users are often heavily laden with explicit and implicit references to their (own) personality type, leading to a form of data leakage. Secondly, the discourse captured by these posts reflects a highly specific form of contextual engagement that few persons would ever take part in. As this form of engagement itself is not generalizable to general social media engagement, the scope of models built on such data is highly limited. These are serious issues that, in our view, are grounds for significant skepticism of claims that performance on such datasets reflects the ability of ML models to capture personality traits from text or validates the system of traits in question. Even researchers working on Twitter data express pessimism: in a review of automatic text-based methods to predict MBTI personality, Štanjer and Yenikent (2020)[22] conclude with a rather grim note – they argue that '*Twitter data simply do not make for a good dataset for the MBTI personality detection, and even more, that purely textual data do not exhibit clear linguistic (either content-based or stylistic) signals for it.*'

As we will see in the results section, this view is at least partially incorrect; the authors have likely framed their argument in the context of some Twitter studies that report poor prediction accuracy when attempting to classify MBTI traits using profile data. This paper will argue that improvements in the quality of the (Twitter) data used and the prediction techniques obviate concerns that Twitter data is fundamentally incomplete or inappropriate for predicting MBTI traits. We do this by building a new Twitter dataset of users who self-label their MBTI type and using powerful text-embedding techniques that are capable of contextually aware syntactic parsing to predict type.

3. Methods and Results

We obtain our dataset by mining Twitter data of profiles who identify their MBTI type in their biographies. We apply a range of methods across a comprehensive array of data – user text data and metadata – to understand how discourse and information preferences align with MBTI personality traits. Our analysis begins with a basic regression-based analysis of profile metadata which includes count-based features. We then train machine learning models based on Naïve Bayes, basic neural networks, and a deep-learning-based NLP model (BERT) to analyze text data from profiles. Finally, we perform exploratory analyses of text using network analyses and visualizations of text embeddings, to gain a clearer understanding of how semantic structures align with personality traits.

3.1 Data and discussion

We mine Twitter data of a total of 3848 profiles. These profiles display MBTI type - a four-letter code – in their biographies. The data we mine include statuses (which include retweets), tweets liked by the user, user biography, and metadata on counts of profiles followed, number of statuses, and number of tweets liked. Users with an abnormally high number of follows (>2500), statuses (>20000), or likes (>20000) were excluded due to a higher likelihood that the profiles are fake/controlled by bots. We clean all user text of MBTI labels to ensure no contamination of training data with labels. Text data is cleaned through the removal of stop words and lemmatized depending on the requirements of the analytical methods used downstream. All mentions (tokens starting with '@') and URLs are removed to further sanitize our text. We are left with a total of 3501 profiles after removing profiles with non-typical metadata values. This corresponds to about ~1M tweets in total.

The personality labels for our profiles are user-provided; hence, these labels are prone to erroneous identification. We expect that most users will not have arrived at these labels by undertaking a well-validated test administered by professionals. We also expect some users to have declared personality traits that they *aspire* to have rather than ones that reflect true behavior. While both these concerns are valid, we think the issue of label accuracy is partially mitigated as these users are a self-selected group of personality theory enthusiasts that are sufficiently confident in the veracity of their results to share their 'type' publicly. Even if there are errors in self-identification, we expect some signal related to true MBTI type (if it exists) to remain. In the worst case, we expect that the results of this study generalize only to *self-identified* personality. Finally, note that due to the presence of these issues, the prediction accuracies reported by our analysis are a *lower bound* of the true prediction accuracies (that is if the labels were generated by a perfect instrument). Hence, any significant predictive performance generated by our models has optimistic implications for this line of research.

Our data also suffers from some degree of imbalance in the distribution of Intuitive and Sensing types in the data (see table 1) – Intuitive types are disproportionately dominant in our dataset. They are also overrepresented: They appear about *four times* more frequently in our data than in the general population, based on evidence from population-level sampling[23]. This leads to an unbalanced dataset, but we correct this downstream through subsampling. More importantly, this finding is in line with the theory underlying intuitive/sensing preferences: persons who create profiles on Twitter are more likely to be seeking new ideas and conversations (high openness) than concrete experiences and facts (low openness). It is also possible that self-selection affects those who choose to *report* their MBTI preferences as well; however, it is not entirely clear from the standpoint of theory that persons with Intuitive preferences be more likely to self-report their personality traits. Either way, self-selection effects serve as our first indicator that *some* degree of validity exists in our data – that the Intuitive/Sensing labels are not merely randomly reported. Other trait classes are roughly equal to population-level estimates.

Personality Trait	**N**
Introverts (I)	**1985**
Extraverts (E)	**1863**
Intuitive (N)	**2279**
Sensing (S)	**1569**
Feeling (F)	**1986**
Thinking (T)	**1862**
Prospecting (P)	**1713**
Judging (J)	**2135**

Table 1: No. of profiles (before cleaning)

3.2 Metadata Analysis

Understanding relationships between key profile metadata features and personality traits is the first step to grounding our subsequent analysis. Our metadata contains valuable information on both the degree to which users engage with the Twitter platform as well as *how* they engage with the platform. The number of likes can be thought of as a proxy for user activity, the number of tweets a proxy for active engagement, and the number of follows a proxy for diversity of information consumption. While these data points do not necessarily track our suggested interpretations, they nonetheless serve to give us a better idea of who our users are. In conjunction with an analysis of user text, they may serve to add crucial context to our results and discussion.

We perform simple linear regression to analyze the relationship between our metadata features and personality traits. We do not perform any balancing of classes for this exercise. In addition to generating regression statistics, we visualize trait-metadata relationships through conditional boxplots.

Metadata Analysis: Results

Effect sizes that are significantly different from null with a probability greater than 95% are shown in boldface. Orange bars indicate statistics with the label present (‘1.0’) and blue bars the absence of the label (‘0.0’).

Trait: Extraversion (similar to extraversion on FFM scale)

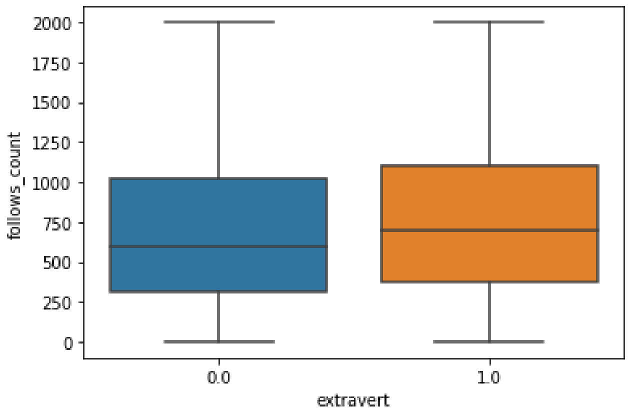

Fig 1. Follow count, Effect: **+70** follows; p-value: **0.02**

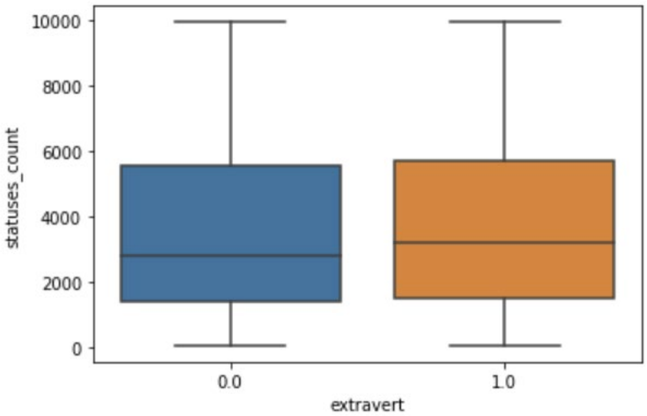

Fig 2. Status count, Effect: +128 tweets; p-value: 0.439

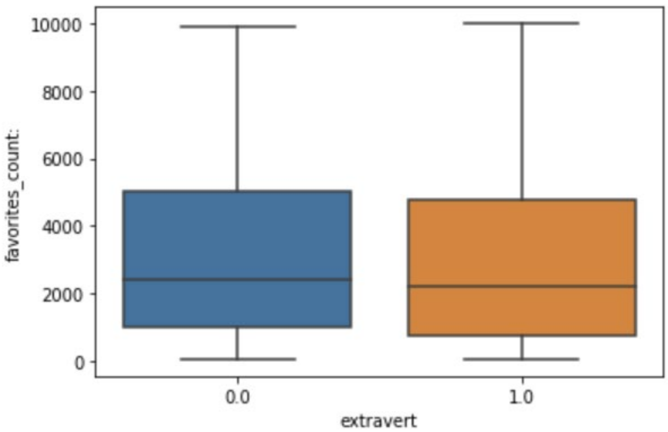

Fig 3. Likes count, Effect: -247 likes; p-value: 0.14

Trait: Intuitive (high openness)

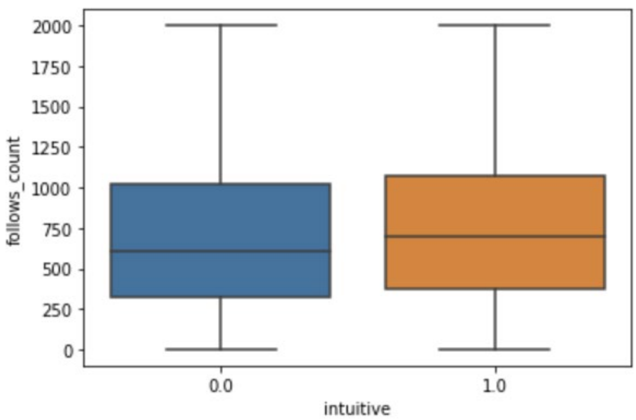

Fig 4. Follow count, Effect: **+62** follows; p-value: **0.05**

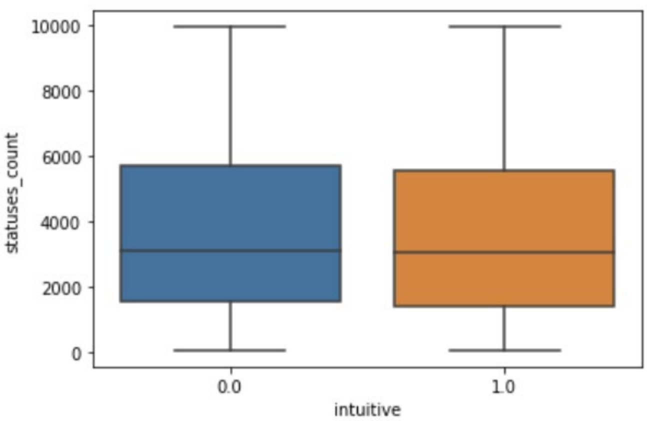

Fig 5. Status count Effect: -43 tweets; p-value: 0.794

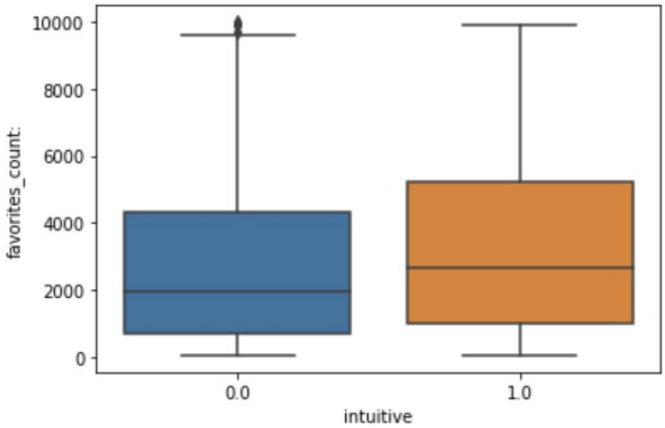

Fig 6. Likes count, Effect: **+612** likes; p-value: <**0.01**

Trait: Thinking (low agreeableness)

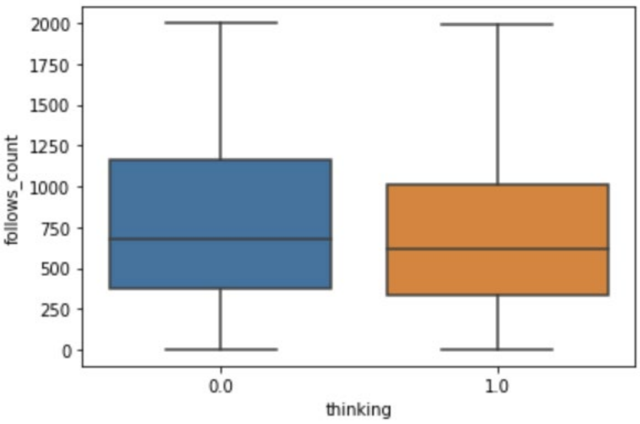

Fig 7. Follow count, Effect: **-76** follows; p-value: **0.01**

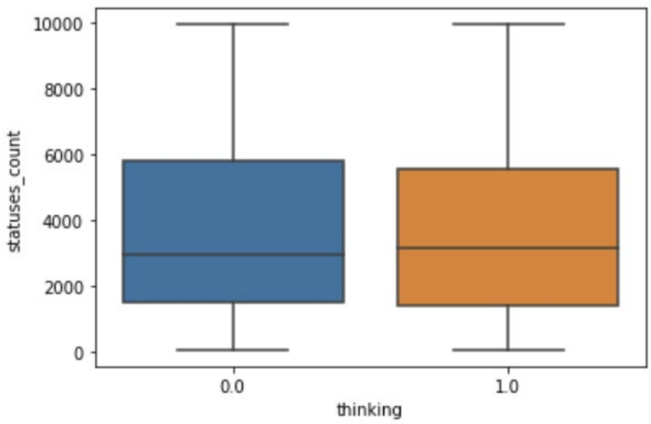

Fig 8. Status count, Effect: +4 tweets; p-value: 0.978

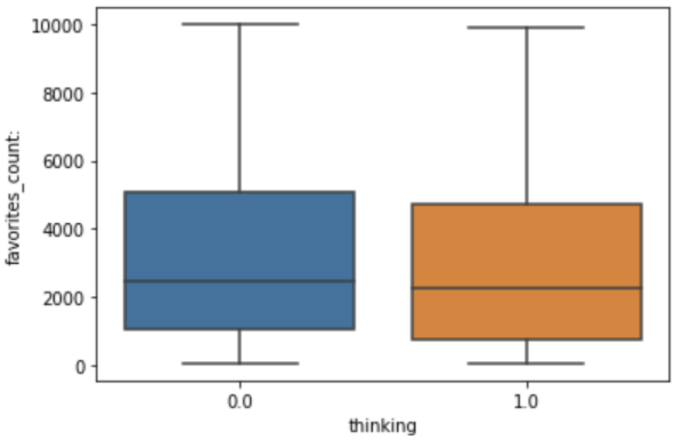

Fig 9. Likes count, Effect: **-322** likes; p-value: **0.05**

Trait: Judging (high conscientiousness

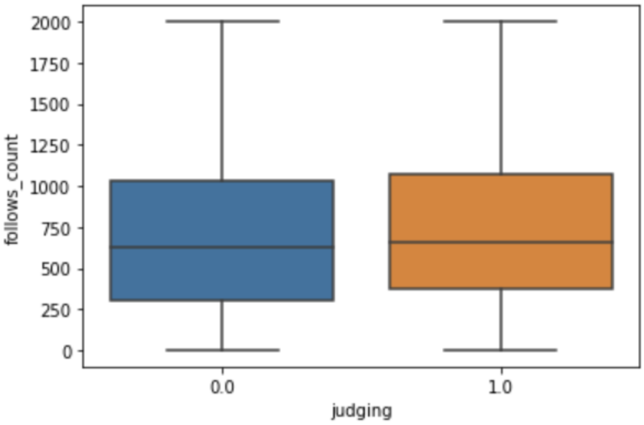

Fig 10. Follow count, Effect: +32 follows; p-value: 0.29

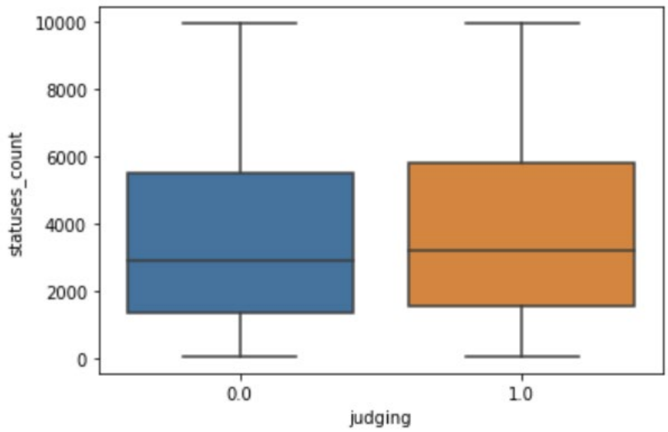

Fig 11. Status count, Effect: +201 tweets; p-value: 0.225

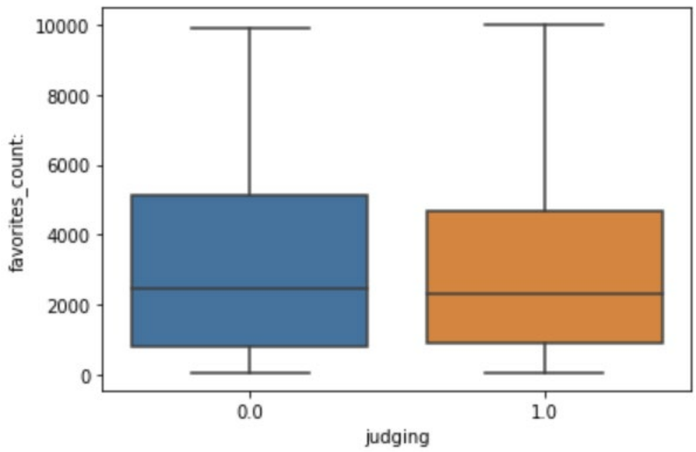

Fig 12. Likes count, Effect: -171 likes; p-value: 0.30

3.3 Classification Baselines (Naïve Bayes and Neural Network classifiers)

A robust baseline for language-based classification tasks is often provided by basic classifiers such as Naïve Bayes and single-layer neural networks. We generate TF-IDF features extracted from text as inputs to these classifiers. For statuses and liked tweets, the text input for each profile corresponds to a concatenated string of tweets. TF-IDF stands for term frequency inverse document frequency – an algorithm that converts documents into vector representations based on word counts. Each document is embedded as a vector wherein each index corresponds to the frequency of a particular word found in the document divided by the logarithm of the frequency of documents containing the word. A Naïve Bayes classifier bases its prediction based on Bayesian estimation of each class being true conditional on the features in the TF-IDF vector. Each index (corresponding to a unique word found in the corpus) is treated as an independent feature. This assumption is flawed but enables highly efficient computation – making it an appropriate choice to set a baseline against which the performance of more sophisticated models can be compared. Our second method to generate baselines is a simple neural network with a single hidden layer containing 200 nodes. We feed TF-IDF features to this classifier as we do for the Naïve Bayes classifier. As the first step in our data processing, we perform a subsampling of our dataset to balance our classes of traits. We then perform a train-test split with a 90:10 ratio for training and testing respectively.

3.4 BERT

Developed by researchers from Google in late 2018, BERT (Bidirectional Encoder Representations from Transformers) represents the latest paradigm in deep-learning approaches to Natural Language Processing. BERT is composed of 12 layers of 'Transformer' encoder-decoders networks that can dynamically allocate weights to relevant words within a sentence, via a method intuitively named 'attention', in the process of language-based inference. BERT improved prior attempts at computerized language analysis significantly and employs 'transfer learning': once it has been trained on language modeling (a task to predict a masked/hidden word in a sentence), it can be further trained downstream to address more specific language tasks, in a process called 'fine-tuning'. The objective of employing BERT is to more accurately estimate the signal between our labels and text – BERT is better able to consider all available information in text than our baseline classifiers, by virtue of its sophistication.

For this project, we build multiple fine-tuned versions of the base model of BERT (BERT-base) as classifiers to predict each of our four MBTI traits We fine-tune three separate models, each using different data as input: liked tweets, own tweets (statuses), or biographies. For liked tweets and statuses, all tweets (per user) are concatenated, and the first 512 tokens are sampled from this body to create the input text. 512 tokens correspond to approximately 35 tweets (an average length of ~15 tokens per tweet). For biographies, up to the first 64 tokens are sampled to create the input text (as biographies are much shorter than either the cumulative sum of liked tweets or statuses). The *BertForSequenceClassification* model from the Transformers Python package, published by Hugging Face, was used for all experiments. This model consists of a pre-trained BERT-base model with an additional classification layer with 2 nodes on top. The final models generated for each of the experiments differed in hyperparameter settings as the hyperparameters were tuned. The learning rate was set to 10^{-5}, the value of epsilon to 10^{-8}, the batch size varied between 4 and 8, and the number of epochs from 3 to 4. We follow the same subsampling and train-test split procedure we employed to create classification baselines to prepare inputs for BERT.

Classification Accuracies

Trait	NB Acc.	NN Acc.	BERT Acc.
Introverts / Extraverts	**58%**	**53%**	**65%**
Intuitive / Sensing	**60%**	**58%**	**73%**
Feeling / Thinking	**52%**	**59%**	**63%**
Prospecting / Judging	**58%**	**54%**	**63%**

Table 2: Accuracies for the Naïve Bayes (NB), Neural Network (NN), and BERT classifiers for biographies, rounded to the nearest integer. All baselines 50%.

Trait	NB Acc.	NN Acc.	BERT Acc.
Introverts / Extraverts	**51%**	**52%**	**64%**
Intuitive / Sensing	**50%**	**49%**	**68%**
Feeling / Thinking	**52%**	**55%**	**63%**
Prospecting / Judging	**51%**	**49%**	**59%**

Table 3: Accuracies for the Naïve Bayes (NB), Neural Network (NN), and BERT classifiers for statuses, rounded to the nearest integer. All baselines 50%.

Trait	NB Acc.	NN Acc.	BERT Acc.
Introverts / Extraverts	**59%**	**55%**	**60%**
Intuitive / Sensing	**57%**	**58%**	**67%**
Feeling / Thinking	**63%**	**55%**	**63%**
Prospecting / Judging	**62%**	**56%**	**59%**

Table 4: Accuracies for the Naïve Bayes (NB), Neural Network (NN), and BERT classifiers for liked tweets, rounded to the nearest integer. All baselines 50%.

3.5 Text embeddings and visualization

We develop visualizations of text embeddings of the statuses, liked tweets, and biographies of users. These visualizations are performed as part of an exploratory analysis of our data, to better understand whether structure emerges from unsupervised transformations of our inputs and whether personality traits relate to that structure. We construct (simplified) representations of the text-space of our data to be able to identify patterns of clustering (along dimensions of personality) in these representations if present.

These visualizations are based on CLS token values, which are high-dimensional embeddings obtained from the BERT model (*before* fine-tuning) that contain aggregate representations of inputs. These embeddings locate our inputs within a global world of text (that which BERT is originally trained on). We then use UMAP (Uniform Manifold Approximation and Projection) - a non-linear dimension reduction algorithm - to project our embeddings to a visualizable lower-dimensional space (2D). Our inputs – biographies, statuses, and liked tweets - for this method are truncated to 512 tokens to accommodate the input length constraints of BERT.

UMAP projection of profile biographies

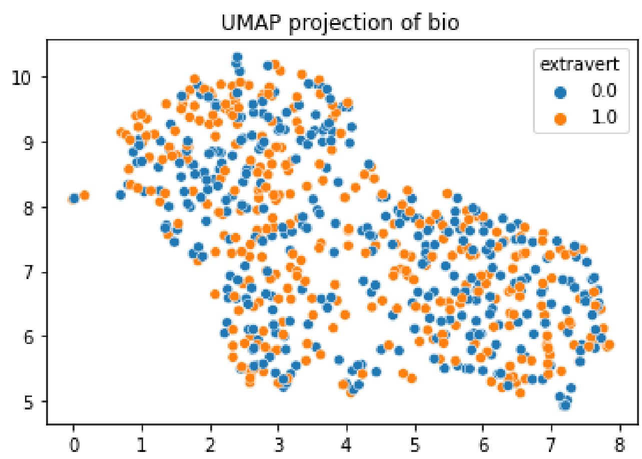

Fig 13.

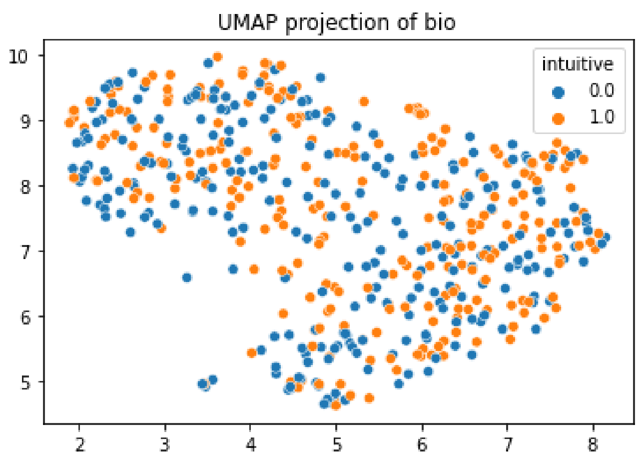

Fig 14.

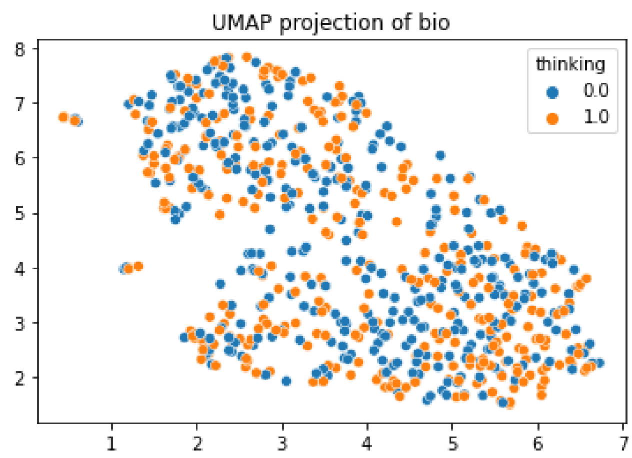

Fig 15.

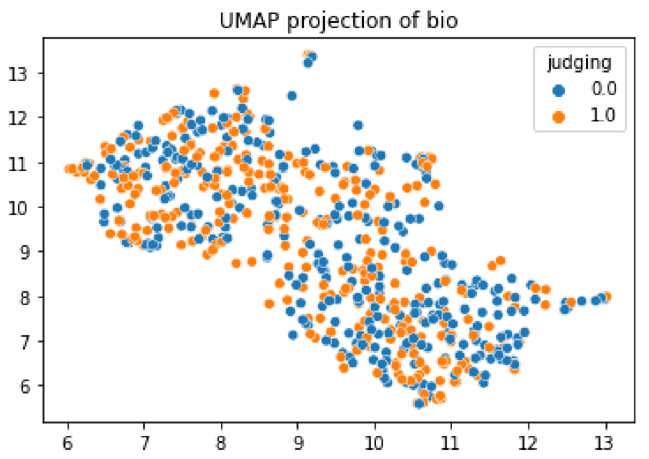

Fig 16.

UMAP projection of profile statuses

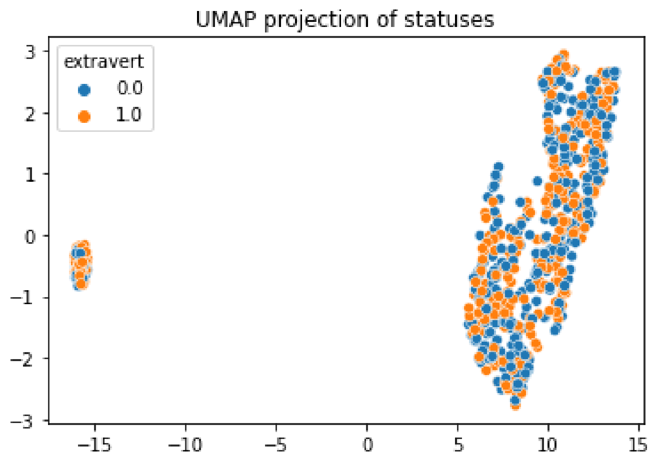

Fig 17.

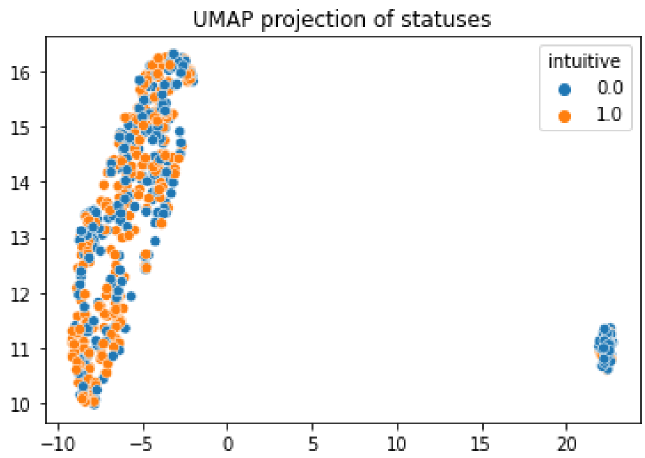

Fig 18.

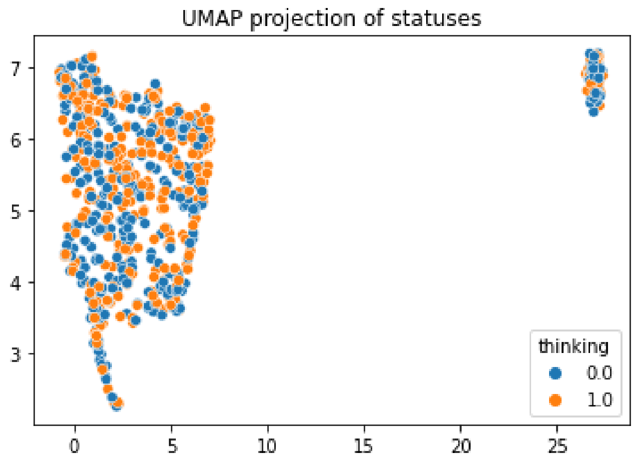

Fig 19.

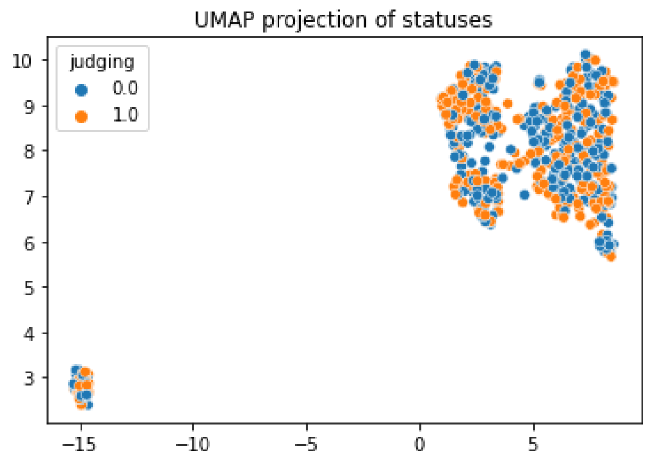

Fig 20.

UMAP projection of profile liked tweets

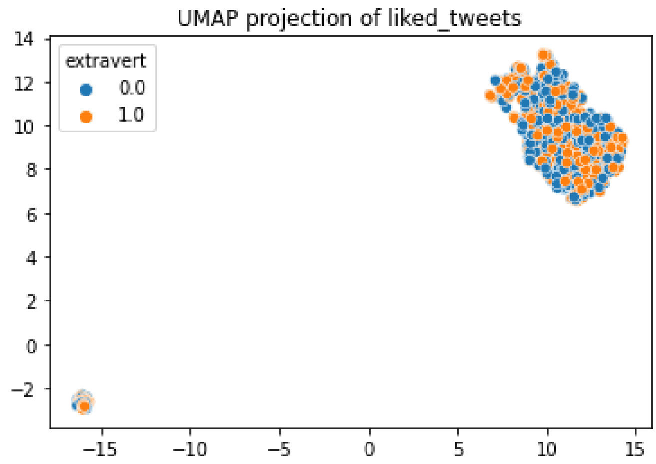

Fig 21.

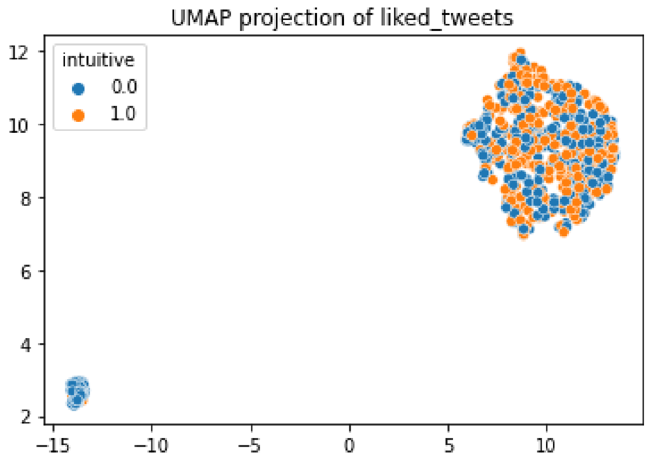

Fig 22.

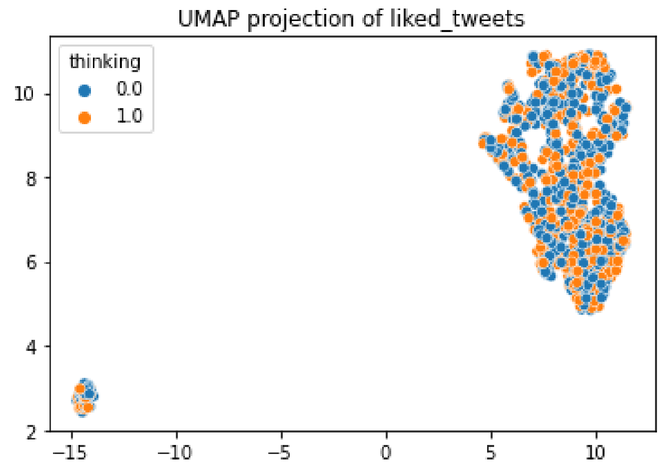

Fig 23.

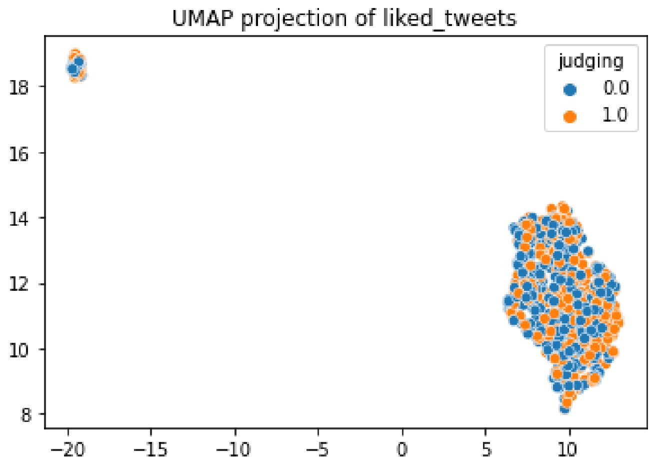

Fig 24.

3.6 Network analysis

A pitfall of relying on visualizations of high-dimensional embeddings for exploratory analysis is that even clearly distinguishable patterns are difficult to interpret – we would not have easy access to understanding the reasons behind clustering patterns or *how* the underlying language relates to personality traits. To account for this weakness and better expose the semantic game underlying our data, we graph semantic networks specific to each personality 'type'.

We build two kinds of networks. The first class of networks is communication networks, wherein nodes are unique documents (profiles) and edge weights are the numbers of shared words between documents. This graph can be understood as a visualization of profiles communicating with one another – the degree of which is determined by the commonality of language between them. Documents closer together in the network share have more language in common; documents with language most representative of the collection tend to appear and cluster together in the center of the network while documents with distinct text would appear at the peripheries. To construct a network for each text source of tweets (statuses and liked tweets), we randomly sample 50 profiles for each trait – a total of 100 per behavioral dimension – and input a concatenated string of tweets for every profile to generate communication networks.

The second class of networks – semantic networks – is based on graphs wherein nodes are unique words and edge weights are defined by the count of the number of times words appear with each other in a single sentence. This network directly showcases semantic connections within the content. It takes as input the total aggregated content generated by profiles under consideration and not individual contributions by profiles. It hence serves as a thematic analysis of the text under analysis. To prepare inputs to build our semantic networks, we randomly sample 500 profiles corresponding to each trait value. As the four dimensions of the MBTI system are binary encodings, we generate a total of eight samples to feed as input to build our networks. We limit the number of tokens per profile to 500 for both statuses ('tweets') and liked tweets to accommodate limited compute resources.

Taken together, these networks: (1) Provide a clearer picture of the *themes* underlying the text we analyze; visualizing semantic networks help engender intuitions of the 'social game' underlying our data. (2) Visually expose homophily and other network patterns with personality traits. These network analyses are hence crucial in building the social interpretability of our data and validating the legitimacy of the results from our machine learning-based modeling exercises.

As our efforts to construct semantic networks yielded results that are noisy and difficult to interpret, we restrict their visualizations and analysis to the appendix.

Profile Status Communication Networks

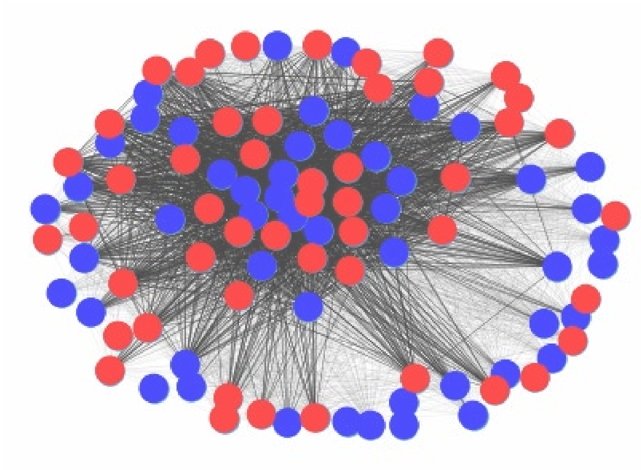

Fig 25. Red: Introverts; Blue: Extraverts

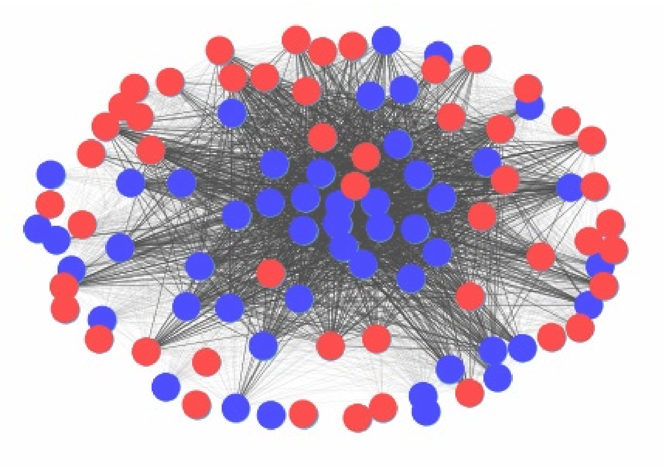

Fig 26. Red: Sensing; Blue: Intuitive

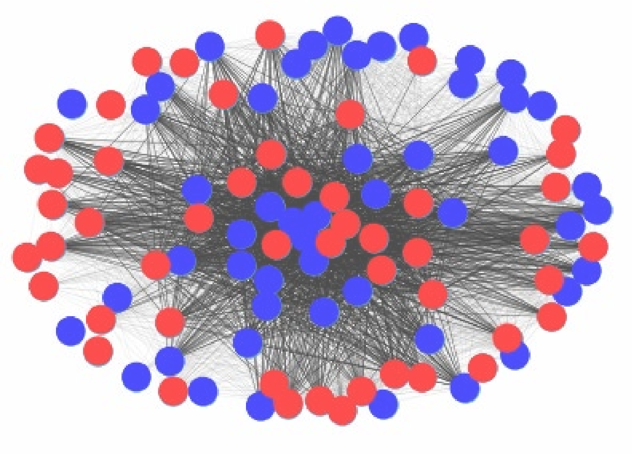

Fig 27. Red: Feeling; Blue: Thinking

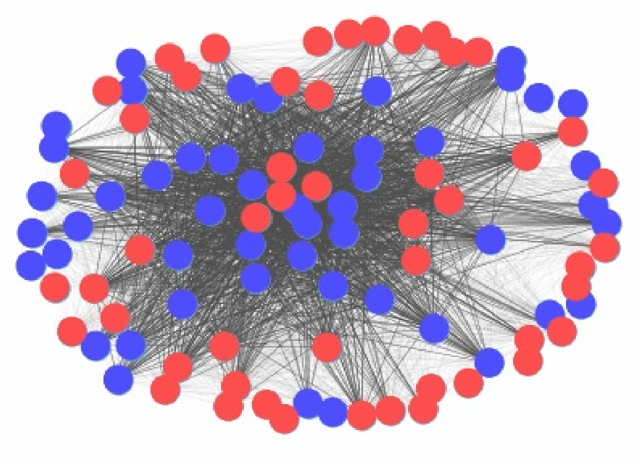

Fig 28. Red: Perceiving; Blue: Judging

Profile Liked Tweets Communication Networks

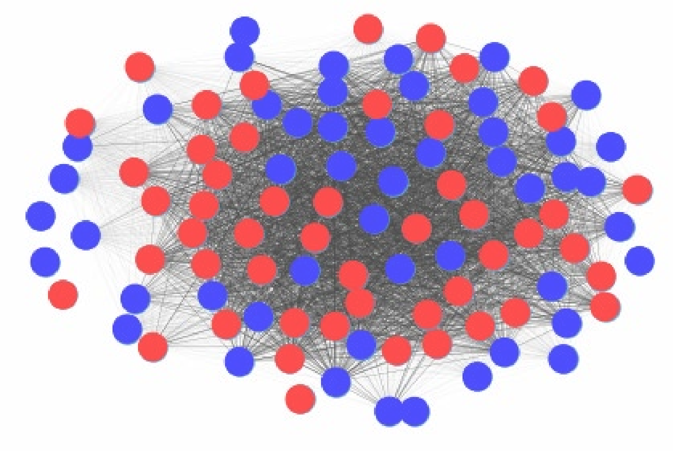

Fig 29. Red: Introverts; Blue: Extraverts

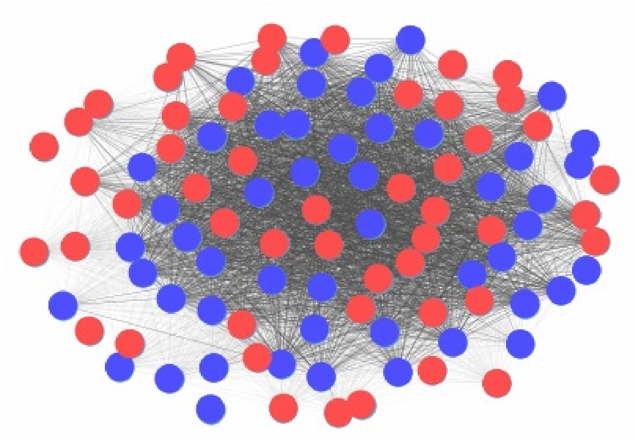

Fig 30. Red: Sensing; Blue: Intuitive

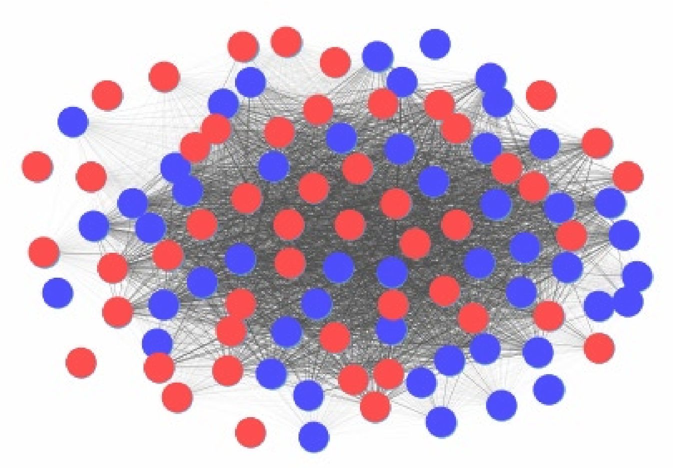

Fig 31. Red: Feeling; Blue: Thinking

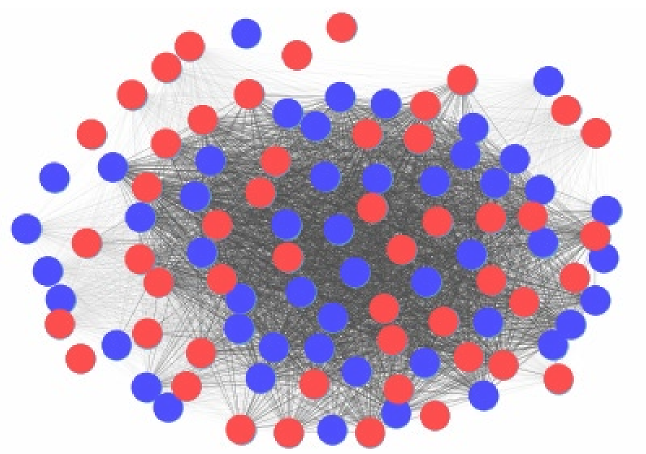

Fig 32. Red: Perceiving; Blue: Judging

4 Discussion

4.1 Metadata Analysis

Our metadata analysis reveals interesting patterns of engagement with the social media platform. Extraverts are more likely to follow a greater number of Twitter accounts than introverts – an effect size of about ~15% more follows than average. This follows general intuitions underlying the theory of extraversion. However, there is no evidence to suggest that they tweet more or like more tweets than introverts. Extraversion seems to be correlated purely with a greater number of follow-connections with other profiles, and not a greater degree of engagement with social media.

It appears that conditional on having a presence on social media, introverts are no less likely to speak out than extroverts.

We also observe that Intuitives (~high Openness) are more likely to both follow more people (~10% more) and *like* more tweets (~30% higher) than Sensing types (~low Openness). Both these effects suggest the same story: Intuitives are more engaged and active on Twitter than Sensing types. The intuition behind this is firmly aligned with our earlier remarks on the overrepresentation of Intuitive types in our data: the Twitter platform is oriented toward ideas, opinions, and fresh conversations (appeal to high Openness users/Intuitives) and not towards concrete experiences or facts (appeal to low Openness users/Sensors). We do not find any evidence that Intuitives tweet more than Sensors, however.

Thinking types were less likely to follow people (~15% lesser) and like fewer tweets (~15% lesser) than Feeling types. Framed the other way, more agreeable persons (~Feeling) tended to follow more people as well be more liberal in liking other tweets than less agreeable persons (~Thinking). This generally corresponds to our folk understanding of agreeableness as a trait. As before, we do find any evidence that either group is more likely to tweet.

Interestingly, we do not see any evidence of correlations between the Judging/Perceiving scale (~high/low conscientiousness) and any metadata feature. This suggests one of three explanations. The first is that there are no significant relationships between J/P and these features. The second is that people are poor at identifying themselves correctly on this scale. The last is that the scale as theorized within the MBTI is itself flawed. Any combination of these explanations is plausible.

4.2 Text Classification

Both our baseline methods and BERT perform better than random on most tasks. Performance differs substantially across traits and features tested. Before we analyze our results, it is important to keep in mind that we cannot interpret classification performance on these tasks on an *absolute* basis. This is because it is not clear what the theoretical maximum performance is. As the labels are self-reported, it is unlikely that they are accurate in the sense that they would match labels generated through standardized testing. Even if users undertook professional tests to determine their types, survey-based measures for personality contain some degree of test-retest error. Even trained humans would be highly unlikely to perform close to perfect on these tasks. Moreover, even if the labels were perfect, we cannot be sure of the degree to which personality traits are indeed predictable from Twitter text; the aim of this analysis is primarily the estimation of this relationship. Hence, as perfect accuracies on these tasks are neither expected nor the aim of our experiments, our classification statistics must be understood solely as (imperfect) measurements of the signal between user text data from Twitter and psychometric traits.

We first analyze model performance on biographies – self-descriptions of the users. An intuitive prior would be that such text contains a strong signal related to personality traits – and the results bear this out. BERT performs with an accuracy of up to 73% on the Intuitive/Sensing (N/S) dimension and nearly 2/3rds on other dimensions. For most traits, we see that Naïve Bayes performance is lower than that of BERT but is substantially better than chance. This indicates that a count-based approach that compares distributions of raw word counts does capture information about personality but also fails to capture all relevant information. Sparsity of data may contribute to underperformance by Naïve Bayes: Biographies on Twitter often contain only singular word/phrasal descriptors of individuals and not fully composed sentences i.e. 'Mother. Engineer. Soccer fan.' Conversely, we could note that BERT does comparatively better with sparser data as it relies on word embeddings generated during pre-training (better representations of words uncommon in our corpus are available as it is already trained on a global corpus). The neural network performs the poorest of the three except on the thinking/feeling dimension.

Moving onto statuses, we see a remarkable pattern – our baseline generation methods essentially perform no better than chance, while BERT performs significantly better with an average accuracy of ~64%. This drastic drop in Naïve Bayes accuracy is despite the increased availability of tokens from a maximum of 64 in biographies to 512 in statuses (which should lead to better estimates). This drop is hence likely due to the feature change itself. Statuses, unlike biographies, are unprompted content generated by users (they are not expected to provide self-descriptive information); they are more indirect self-expressions. Most Twitter content revolves around current affairs: Statuses may contain more generic words which makes it more difficult to distinguish between two lexical distributions, whereas biographies contain more unique tokens – and consequently better distinguishing – on average. We see that N/S stands out as the dimension with the most predictive signal from text – it shows the highest accuracy across all traits for both methods.

The predictive performance of the liked tweets features appears to be comparable to that of statuses. Liked tweets are indicative of the content users find salient: They can be interpreted as proxies for the information consumption preferences of users. Naïve Bayes accuracy significantly improves on liked tweets (as compared to statuses) – comparable to the performance by BERT on every dimension except Intuitive/Sensing. One explanation for the discrepancy between Naïve Bayes performance on statuses and liked tweets is that the average tokens available for liked tweets exceeds that of statuses (people like tweets far more often than they create tweets). This would result in multiple rows with sparsely populated features in our dataset.

In summary, our findings conform to intuitive priors we may have formed before we conducted the experiments. Descriptions of users, text content generated by users, and proxies for information consumption (liked tweets) all carry predictive signals corresponding to each of the personality dimensions we experiment on. Biographies – which are intended as descriptions of persons – carry the most predictive power. Word distributions from statuses appear to carry negligible signal of personality traits – but more complex semantic representations of statuses do appear to capture some signal; for liked tweets both individual tokens as well as higher-order representations carry information.

Through all experiments, we see that the N/S dimension is the easiest to distinguish from text. Note that the N/S dimension demonstrates the strongest correlations to the Five Factor Model of personality (where it is referred to as 'Openness') compared to the other three dimensions. We also observe an overrepresentation of Intuitives in our sample compared to a national average[23]; these findings strongly suggest that the N/S dimension is a significant axis of human behavioral diversity.

4.3 Text Embedding Visualizations

To obtain a better idea of how personality traits relate to the topology of the text-space occupied by users, we visualized CLS-token embeddings obtained from BERT using UMAP. CLS tokens are special tokens that capture an input-level aggregate representation using a 768-dimensional vector. BERT was not trained on the data before we obtained the CLS tokens for visualization. We hoped that an 'unsupervised' visualization of user texts, labeled by personality traits, might reveal interesting patterns of homophily, echo-chambering, and other traces of sociological phenomena.

However, our displays of the reduced two-dimensional representations of the CLS tokens, colored by trait, did not reveal any significant divisions, demarcation, or large-scale patterns between text and any of the personality traits for any of the features we tested. This is not wholly unsurprising as just two components of a 768-dimensional dataspace might not capture enough variation in the data. While some small-scale features are visible in some of our graphs, there is not enough differentiation in general for us to be sure that they are not artefacts from our visualization or otherwise random occurrences. Note that some of the graphs differ between traits for the same text feature – this is because the data for the graphs are sampled again to ensure balanced classes for every trait.

A curious feature of the plots for statuses and liked tweets is the existence of a relatively minor cluster of points distinctly separated from the main cluster body. We attribute this to the existence of unusual regularities in tweets caused by spam or 'junk' tweets generated by promoters, campaigners, or other bot-mediated tweet activity.

4.4 Communication Networks

Most of our eight communication networks – four each for statuses and liked tweets – do not show large-scale patterns of similar activity between profiles as related to personality traits. However, we do see some important patterns.

The network of statuses shaded by intuitive/sensing traits demonstrates a clear pattern of intuitive profiles clustering to the center and sensing towards the periphery. This can be interpreted as Intuitive profiles dominating the 'core' of discourse on the network – with Sensing profiles relegated to peripheral conversations. This fits into our earlier analysis of Intuitive profiles being overrepresented as well as being more likely to engage with the platform. This graph suggests that even after conditioning on these facts, Intuitives are more likely to *produce* content that is more relevant or 'central' to the platform.

The status network shaded by Judging/Perceiving preferences also reveals domination of the core by certain profiles – Judging (high conscientiousness) types. Perceiving types (low conscientiousness) tend to be positioned at the periphery. While this indicated that Judging types tend to dominate the discourse on Twitter, it is not clear from the standpoint of psychometric theory why this should be the case. A possible explanation is that Judging types are more consistent, purposeful, and focused in their communication on the platform as compared to Perceiving types, but we cannot be sure of this interpretation.

In the status network comparing Thinking/Feeling types, we do not see any domination of the center but can observe minor localized patterns at the peripheries. The most straightforward interpretation is that these are local 'communities' that tend to be dominated by either type – perhaps due to alignment between specific community values and the trait of agreeableness.

In communication networks on liked tweets, we see lesser evidence of clustering by behavioral traits. We see small-scale local patterns when we contrast by Extravert/Introvert and Intuitive/Sensing, but these patterns are not very prominent to justify drawing significant conclusions. We see no evidence of clustering when contrasting by Thinking/Feeling or Judging/Perceiving.

Overall, it appears that while tweets generated by Intuitive and Judging types are the most similar to each other and form the 'center' of discourse on the platform, there are not many meaningful distinctions to be made on the kinds of content *consumed* by profiles apart from what appear to be small communities of Introvert/Extravert and Intuitive/Sensing profiles *consuming* similar content.

5 Conclusion

Understanding how core dispositional preferences manifest themselves within social forums may provide direct insight into how human behavioral diversity contributes to discourse at scales far removed from the level of the individual. This project attempts to contribute to this understanding through a detailed analysis of a particularly curious community on social media – a community of personality enthusiasts. Our analyses demonstrate that relationships between online presence and discourse are indeed distinguishable in relation to self-identified personality traits within the Myers-Briggs psychometric framework.

Our quasi-experiments on metadata features indicate that relationships between engagements statistics (number of likes, statuses, follows) and traits indeed generally conform to personality theory from the perspective of both the MBTI system as well as the Five-Factor Model. The most substantial relationships in our metadata analyses were found relating to the Intuitive/Sensing dimension (that corresponds to 'Openness' in the Five Factor model), considering the significance and the effect sizes of the relationships we find. We find that both the number of likes and number of follows are significantly related to (with effect sizes of ~10% and ~30% respectively) this dimension.

We demonstrate that deep-learning-based NLP models can empower observational studies of behavior. As the results bear out, BERT outperforms baseline models with healthy margins on almost all tasks. Classification performance is strongest on the Intuitive/Sensing dimension - an accuracy of up to 73%. The significance of other dimensions in our findings is not far behind. The better-than-random performance of the Naïve Bayes model also lends to the idea that the distributions of lexical items carry intrinsic psychometric signals. The increased performance of BERT suggests that more sophisticated semantic features add additional signal – words themselves don't seem to capture all information relevant to personality traits. We take the performance of these models, which rely solely on text-based prediction, as further validation of recent studies[9,12] that argue that psychometric models based directly on text are principally sound approaches; these results support the lexical hypothesis in so far as psychological attitudes have some form of projection onto everyday language.

We draw limited conclusions from our semantic network visualizations on the thematic differences in tweet content responsible for our classifier performance. However, the communication networks we build from statuses yet again underline our analysis of the Intuitive/Sensing dimension and validate the presence of contrasting patterns of discourse between Intuitive types and Sensing types.

This project demonstrates that some personality traits – especially the Intuitive/Sensing dimension – likely play intrinsic roles in social media discourse. A slew of significant correlations and associations demonstrate that theoretical constructs underlying the Myers-Briggs personality system likely measure meaningful psychometric dimensions at least in part. While many of our results are modest or otherwise limited in their conclusions, it is useful to keep in mind that they represent a lower bound of the true signal strength: They do not account for errors in self-reporting and the unavoidable pitfalls of fake, incomplete, and missing data points, all of which significantly affect data obtained from typical social media sites including Twitter. We thereby suggest an optimistic outlook to further research in this domain.

In addition, many large-scale analyses or applications do not necessarily demand highly accurate classification to achieve their objectives. A study by Matz et al.[25] on personality-based targeting for advertising demonstrated that displaying tailor-made advertisements for persons using a dichotomous coding of the extraversion and openness dimensions – highly similar to the framework we use in this study – results in substantial increases in advertisement interest (up to 50% more purchases!) While they target populations based on Facebook likes, a purely text-based methodology could theoretically function in a similar capacity if validated further. As with text, the accuracy of likes-based identification of personality traits is far from perfect. For such applications and research, a better-than-chance signal is often a sufficient criterion for real-world deployment.

6. Future Work

There are many fronts on which this project can be expanded. While the robustness of all our findings will improve by increasing the size of our available dataset, larger improvements to our prediction accuracies and higher resolutions of our exploratory analyses may be realizable by considering a host of other data sources. These include profile pictures, shared images and GIFs, the tweets of the profiles that the users follow, and, in general, a more extensive accounting of the social network within which the user is situated. This project relies on mostly immediately available information from social profiles and the application of out-of-the-box solutions to classify and explore text; it has likely only scratched the surface of what is possible.

While we discussed the possible sources of predictive power by comparing Naïve Bayes and BERT scores across multiple features, there is a possibility that Naïve Bayes and BERT are using mostly mutually exclusive feature sets to make predictions. If this is so, then an ensemble model composed of multiple classifiers and/or features would outperform BERT noticeably. This possibility has not been investigated in this project but strongly warrants further investigation.

Interpreting 'black box' such as BERT has historically proven challenging; however, the recent introduction of model interpretability toolkits based on game-theoretic feature contribution modeling such as SHAP[26] has opened the door to identify sources of signal through an analysis of trained models. In conjunction with exploratory searches powered by 'unsupervised' machine learning algorithms, modern interpretability approaches can illuminate non-obvious relationships in the data. They will likely be an increasingly key focus of social scientists aiming to further research in this domain.

On a final note, we suggest that similar studies undertaken on other social media platforms are an important next step for this line of research. In theory, relationships between core personality traits and online behavior should hold in principle across comparable media platforms; additional support for the takeaways from this project will be an important demonstration of validity.

References

[1] Roberts, Brent W., Nathan R. Kuncel, Rebecca Shiner, Avshalom Caspi, and Lewis R. Goldberg. "The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes." *Perspectives on Psychological science* 2, no. 4 (2007): 313-345.

[2] Judge, Timothy A., Daniel Heller, and Michael K. Mount. "Five-factor model of personality and job satisfaction: A meta-analysis." *Journal of applied psychology* 87, no. 3 (2002): 530.

[3] Poropat, Arthur E. "A meta-analysis of the five-factor model of personality and academic performance." *Psychological bulletin* 135, no. 2 (2009): 322.

[4] Myers, Isabel Briggs, Mary H. McCaulley, Naomi L. Quenk, and Allen L. Hammer. *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Vol. 3. Palo Alto, CA: Consulting Psychologists Press, 1998.

[5] Furnham, Adrian. "The big five versus the big four: the relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality." *Personality and Individual Differences* 21, no. 2 (1996): 303-307.

[6] Goby, Valerie Priscilla. "Personality and online/offline choices: MBTI profiles and favored communication modes in a Singapore study." *Cyberpsychology & behavior* 9, no. 1 (2006): 5-13.

[7] Harrington, Rick, and Donald A. Loffredo. "MBTI personality type and other factors that relate to preference for online versus face-to-face instruction." *The Internet and Higher Education* 13, no. 1-2 (2010): 89-95.

[8] Goldberg, Lewis R. "The structure of phenotypic personality traits." *American psychologist* 48, no. 1 (1993): 26.

[9] Kulkarni, Vivek, Margaret L. Kern, David Stillwell, Michal Kosinski, Sandra Matz, Lyle Ungar, Steven Skiena, and H. Andrew Schwartz. "Latent human traits in the language of social media: An open-vocabulary approach." *PloS one* 13, no. 11 (2018): e0201703.

[10] Quercia, Daniele, Michal Kosinski, David Stillwell, and Jon Crowcroft. "Our twitter profiles, our selves: Predicting personality with twitter." In 2011 *IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pp. 180-185. IEEE, 2011.

[11] Kosinski, Michal, David Stillwell, and Thore Graepel. "Private traits and attributes are predictable from digital records of human behavior." *Proceedings of the national academy of sciences* 110, no. 15 (2013): 5802-5805.

[12] Schwartz, H. Andrew, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah et al. "Personality, gender, and age in the language of social media: The open-vocabulary approach." *PloS one* 8, no. 9 (2013): e73791

[13] Park, Gregory, H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Michal Kosinski, David J. Stillwell, Lyle H. Ungar, and Martin EP Seligman. "Automatic personality assessment through social media language." *Journal of personality and social psychology* 108, no. 6 (2015): 934.

[14] Pittenger, David J. "Measuring the MBTI… and coming up short." *Journal of Career Planning and Employment* 54, no. 1 (1993): 48-52.

[15] Jung, Carl Gustav. *Psychological types*. Routledge, 2014.

[16] Ciorciari, Joseph, John Gountas, Patrick Johnston, David Crewther, and Matthew Hughes. "A Neuroimaging Study of Personality Traits and Self-Reflection." *Behavioral Sciences* 9, no. 11 (2019): 112.

[17] Plank, Barbara, and Dirk Hovy. "Personality traits on twitter—or—how to get 1,500 personality tests in a week." In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 92-98. 2015.

[18] Verhoeven, Ben, Walter Daelemans, and Barbara Plank. "Twisty: a multilingual twitter stylometry corpus for gender and personality profiling." In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 1632-1637. 2016.

[19] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

[20] Mehta, Yash, Samin Fatehi, Amirmohammad Kazameini, Clemens Stachl, Erik Cambria, and Sauleh Eetemadi. "Bottom-up and top-down: Predicting personality with psycholinguistic and language model features." In *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 1184-1189. IEEE, 2020.

[21] Keh, Sedrick Scott, and I. Cheng. "Myers-Briggs Personality Classification and Personality-Specific Language Generation Using Pre-trained Language Models." *arXiv preprint arXiv:1907.06333* (2019).

[22] Štajner, Sanja, and Seren Yenikent. "A Survey of Automatic Personality Detection from Texts." In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 6284-6295. 2020.

[23] https://www.capt.org/jpt/pdfFiles/Hammer_A_and_Mitchell_W_Vol_37_2_15.pdf

[24] https://huggingface.co/

[25] Matz, Sandra C., Michal Kosinski, Gideon Nave, and David J. Stillwell. "Psychological targeting as an effective approach to digital mass persuasion." Proceedings of the national academy of sciences 114, no. 48 (2017): 12714-12719.

[26] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." *Advances in neural information processing systems* 30 (2017).

Appendix

Semantic Networks: Statuses

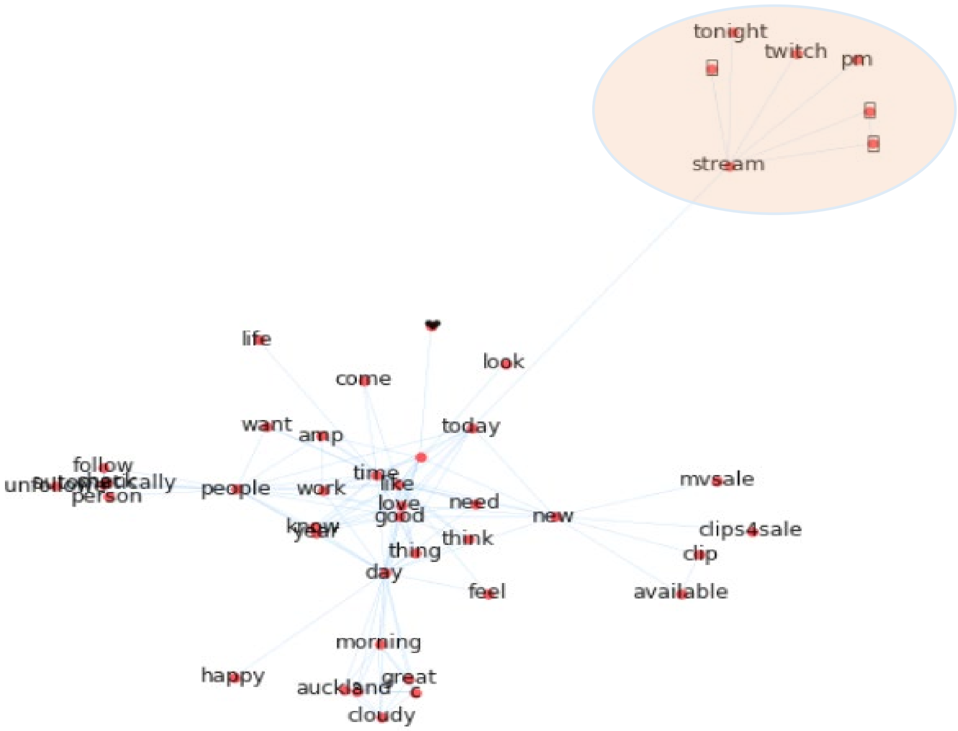

Fig 33**. Introverts**

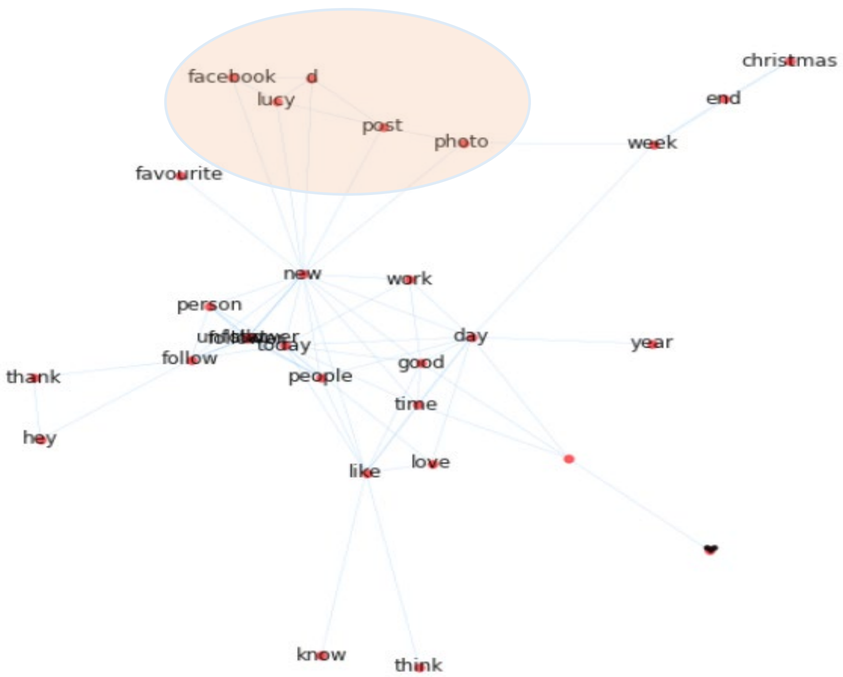

Fig 34. **Extraverts**

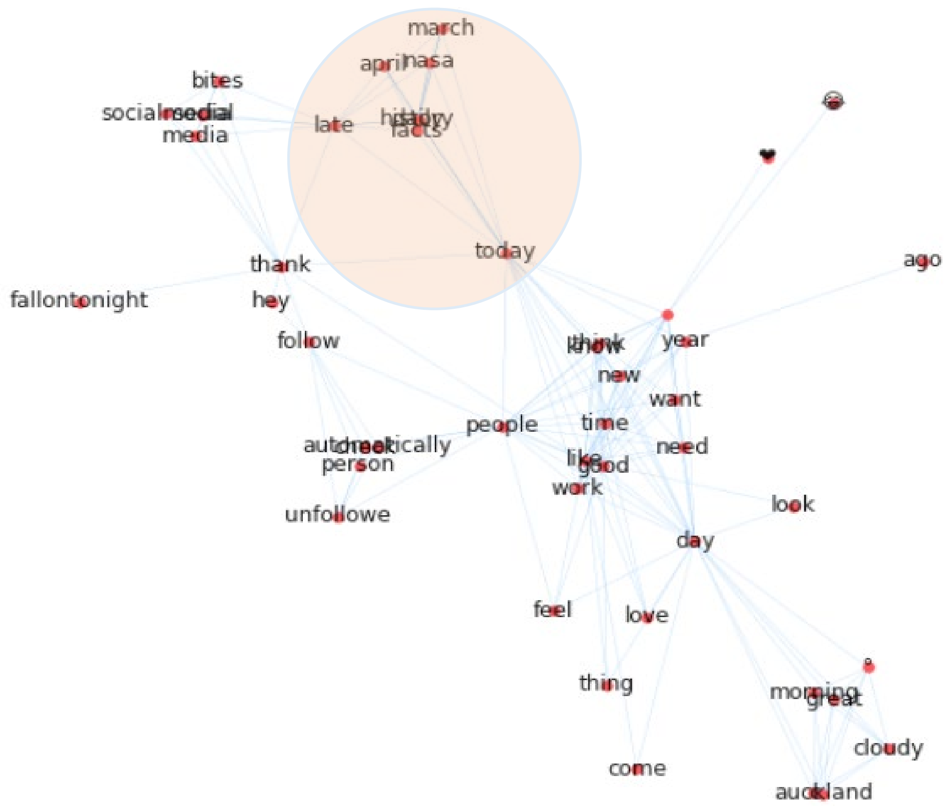

Fig 35. **Intuitive (high Openness)**

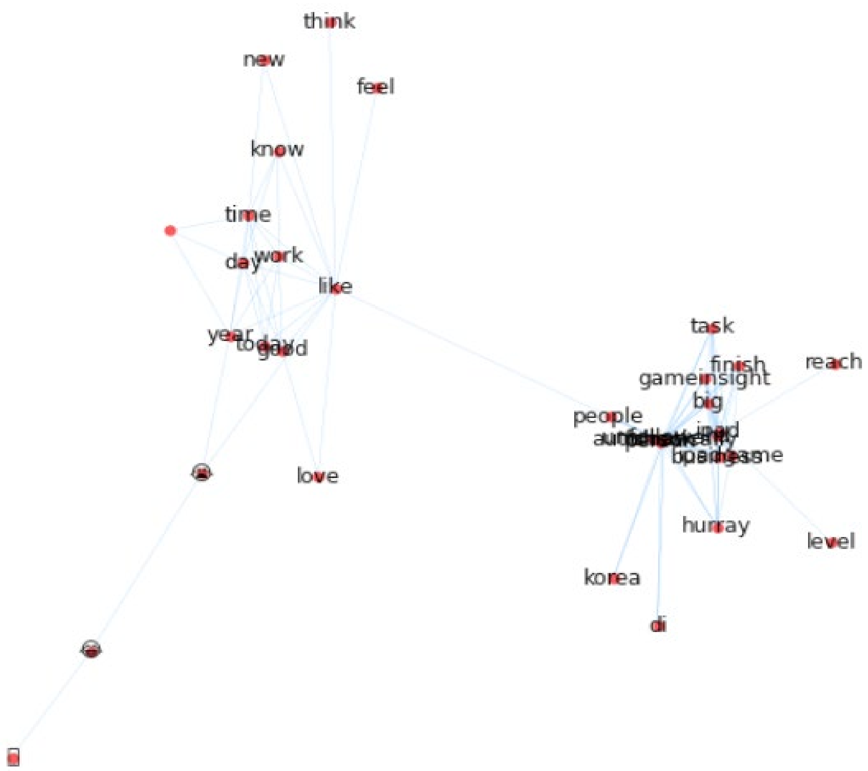

Fig 36. **Sensing (low Openness)**

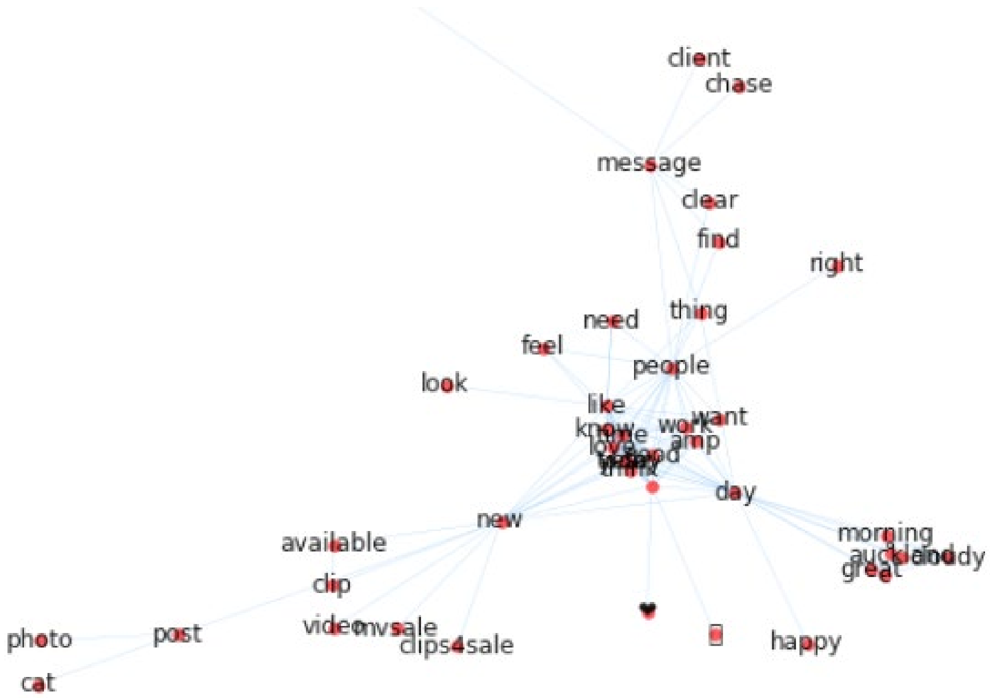

Fig 37. **Feeling (High Agreeableness)**

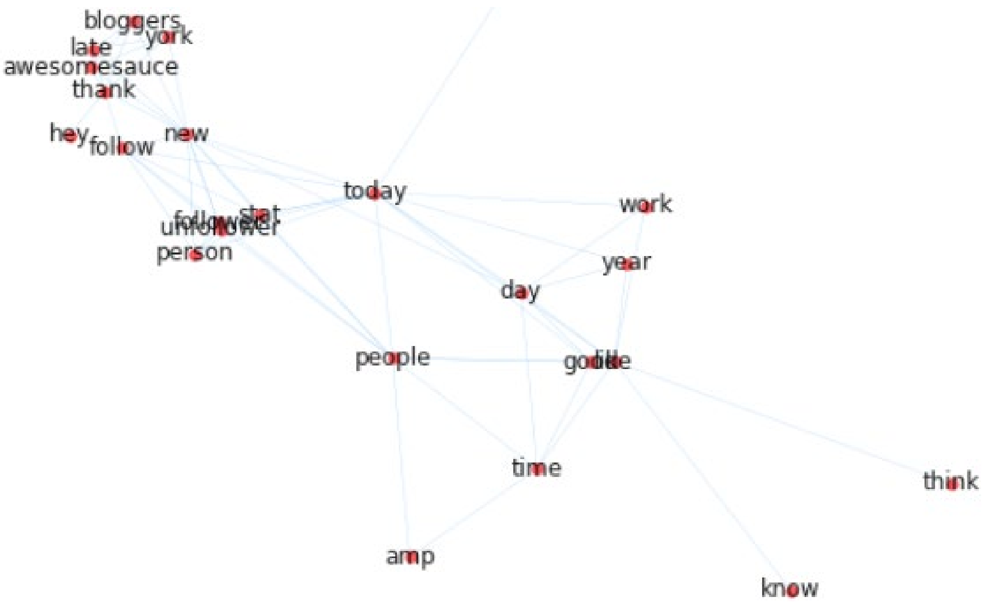

Fig 38. **Thinking (low Agreeableness)**

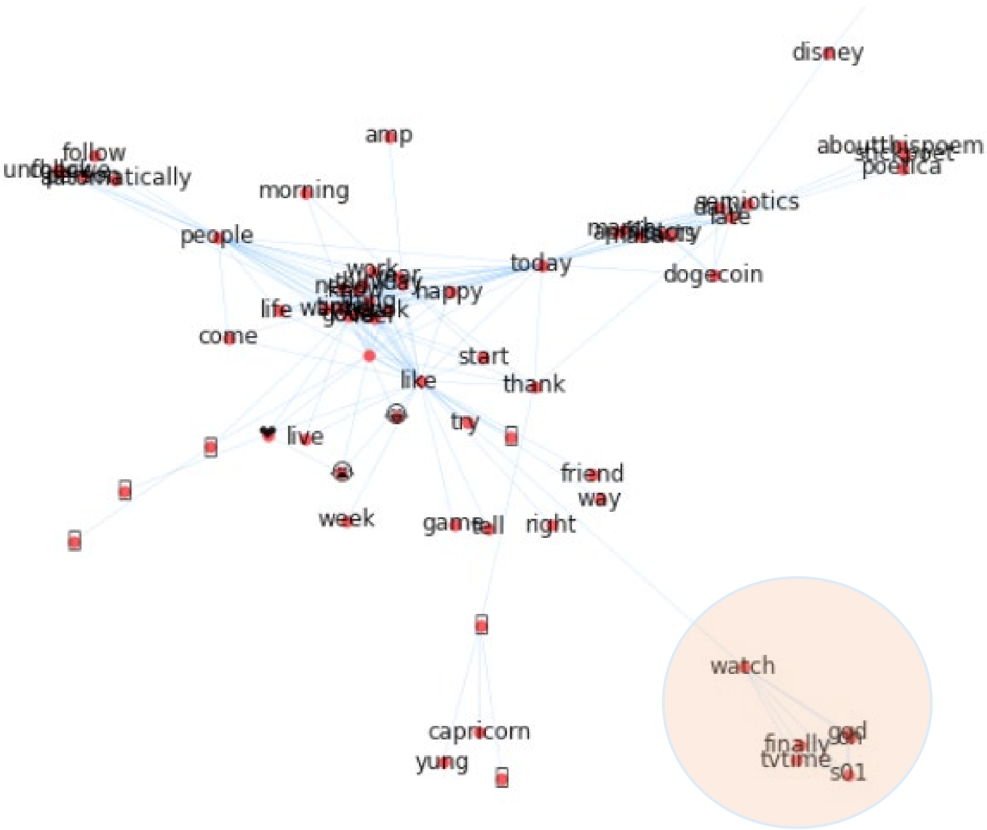

Fig 39. **Prospecting (low Conscientiousness)**

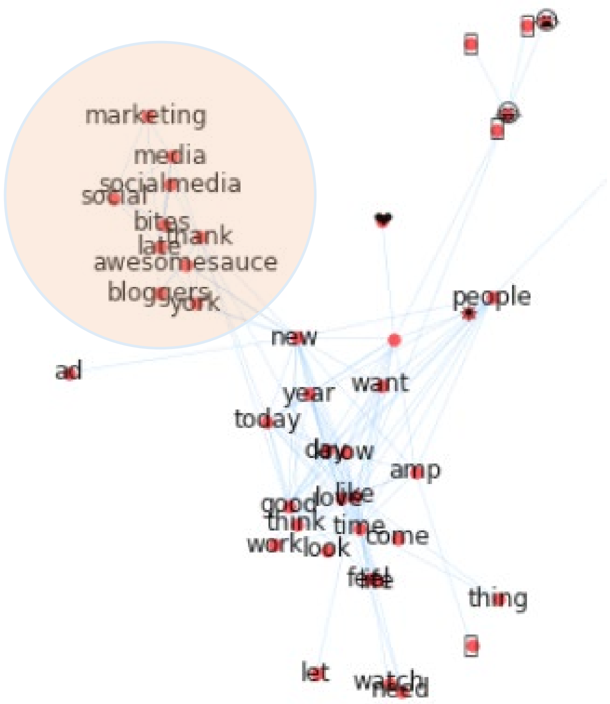

Fig 40. **Judging (high Conscientiousness)**

Semantic Networks: Liked tweets

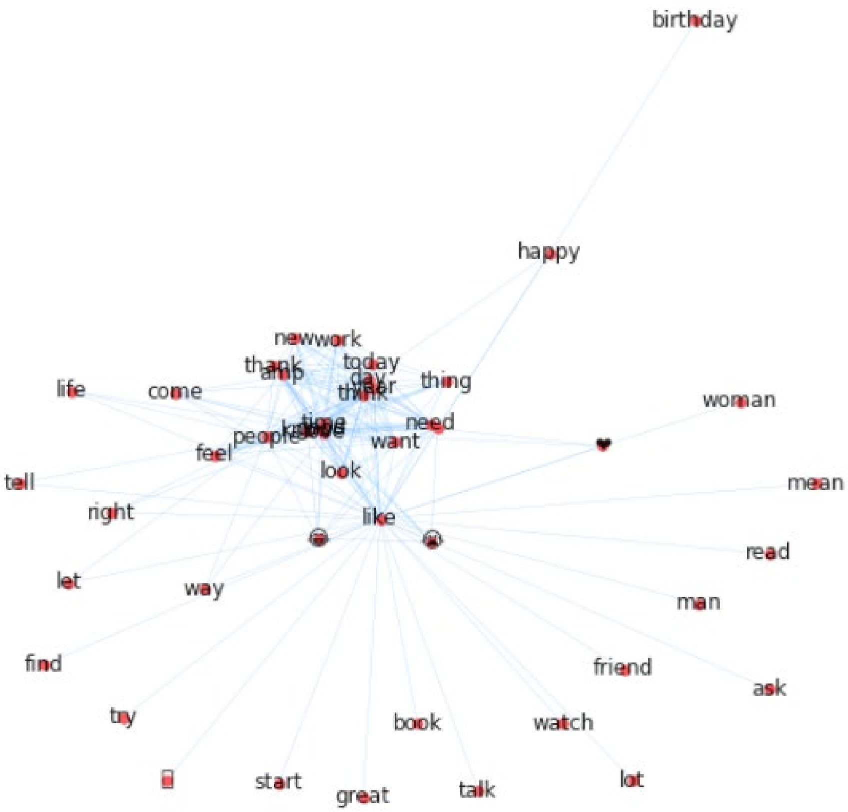

Fig 41. **Introvert**

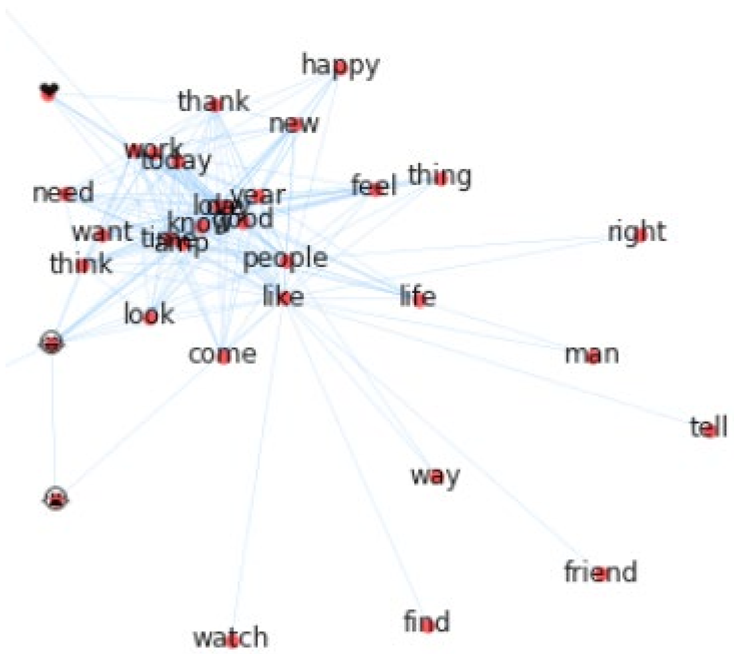

Fig 42. **Extravert**

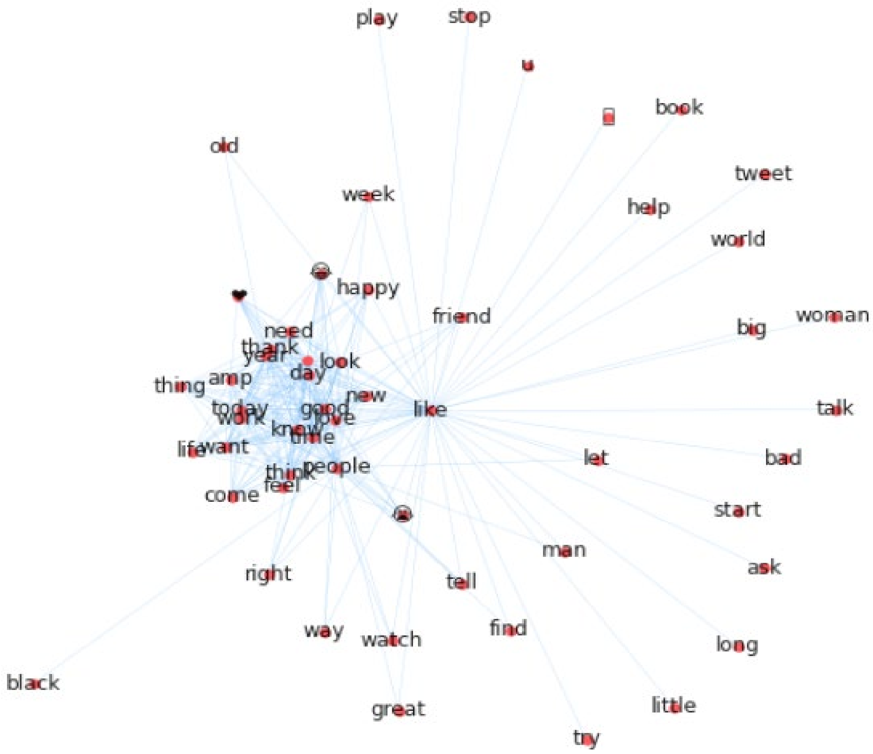

Fig 43. **Intuitive (high Openness)**

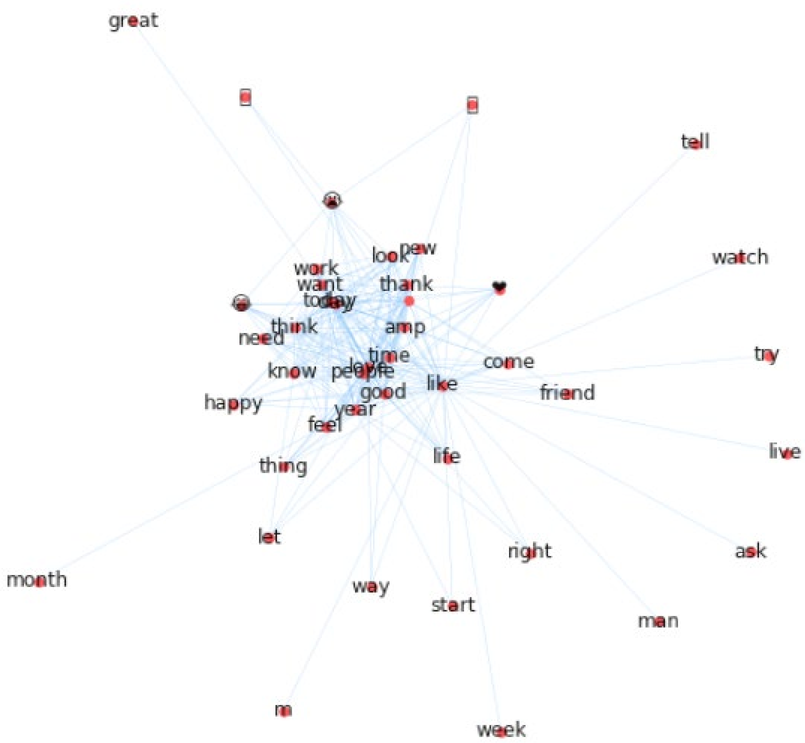

Fig 44. **Sensing (low Openness)**

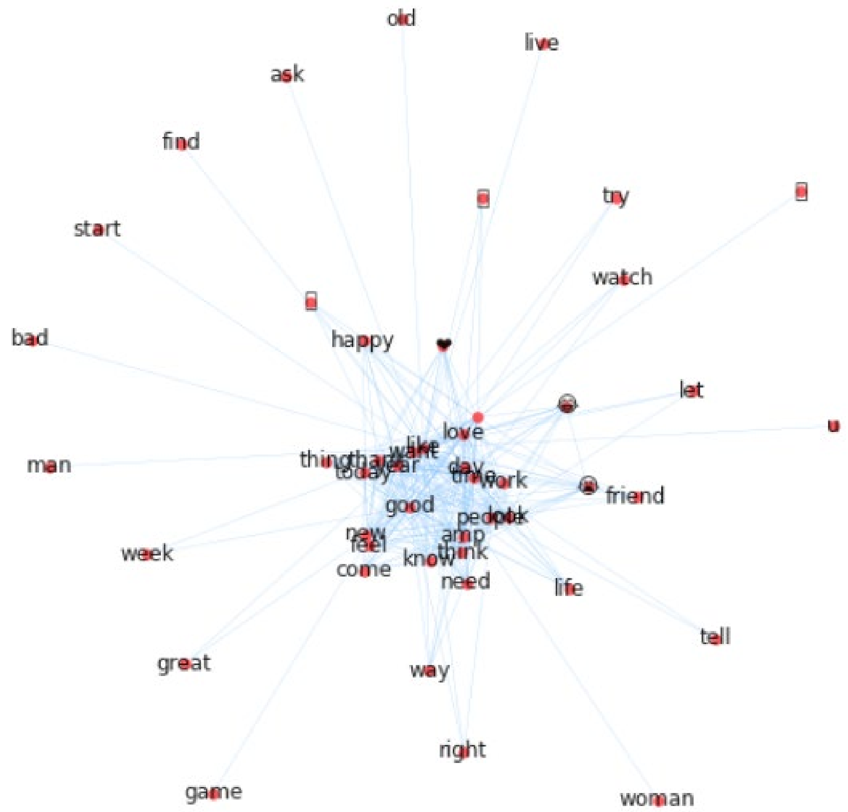

Fig 45. **Thinking (low Agreeableness)**

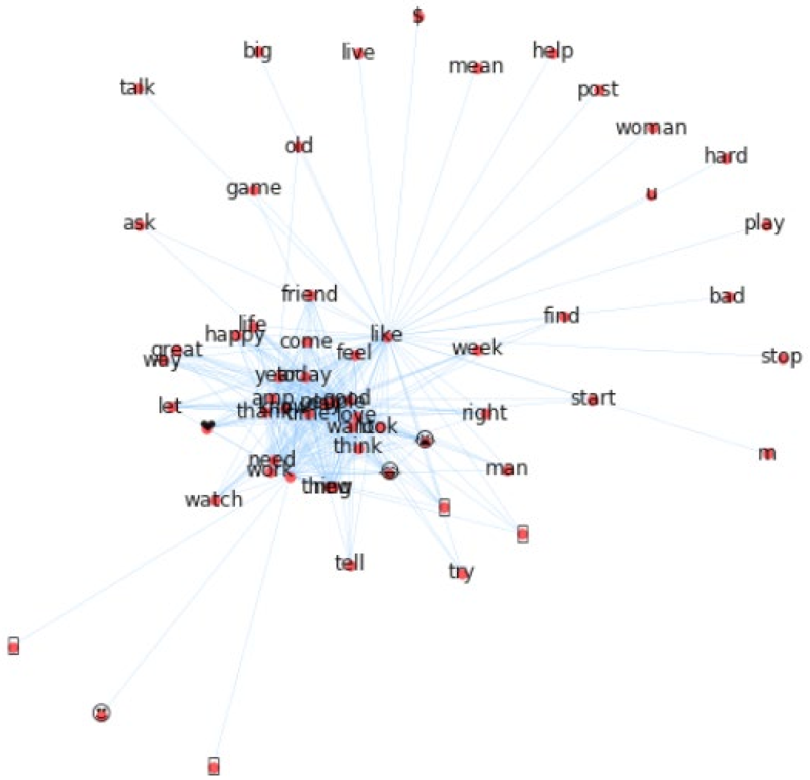

Fig 46. **Feeling (high Agreeableness)**

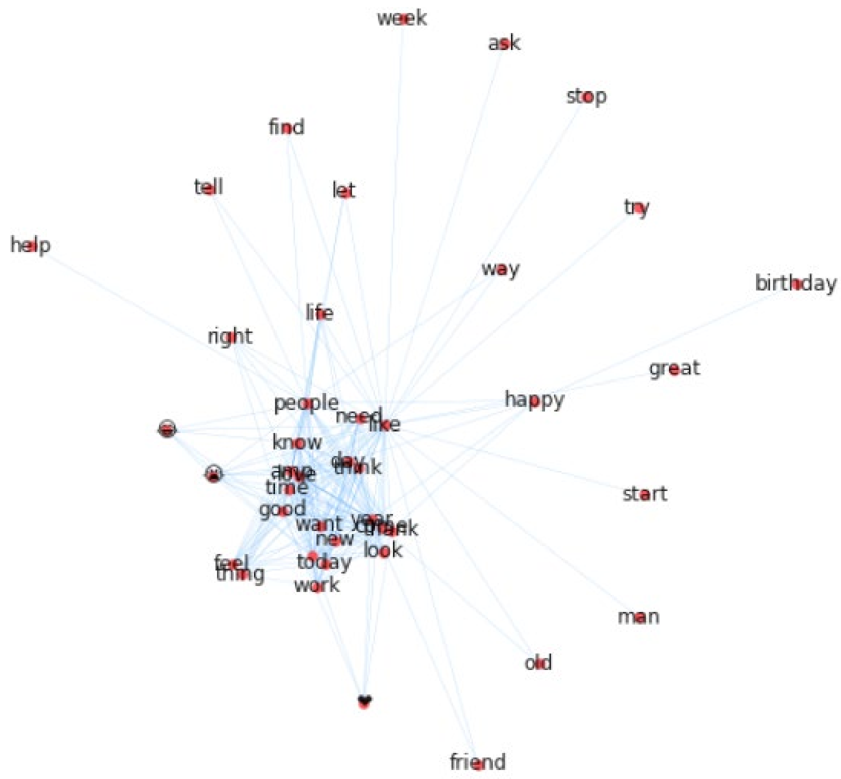

Fig 48. **Judging (high Conscientiousness)**

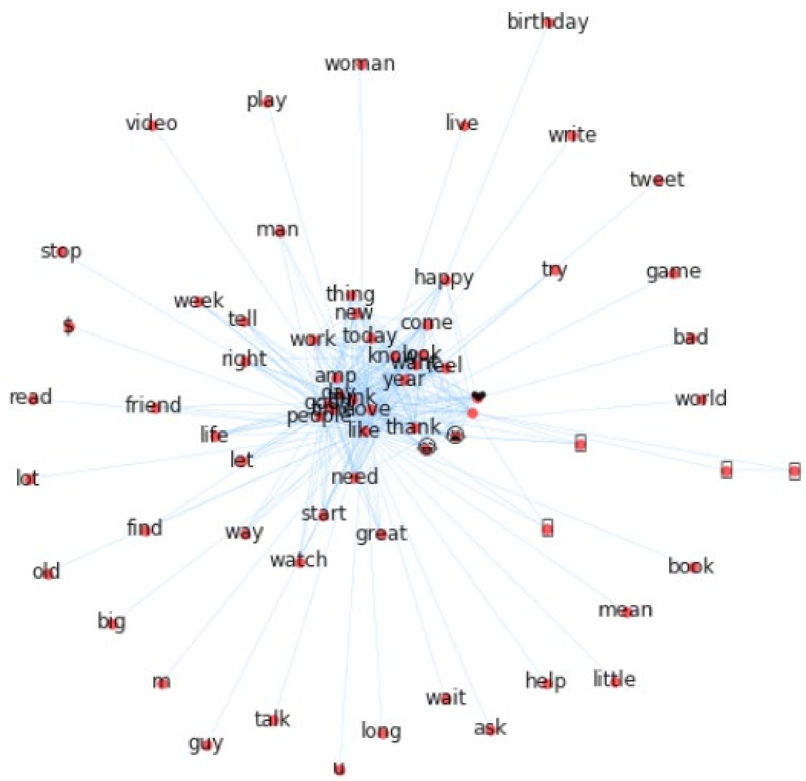

Fig 49. **Perceiving (low Conscientiousness)**

Semantic Networks: Analysis

In all our eight semantic networks – one for each trait – we notice that the core of each network is quite similar. This usually consists of the nodes such as ‘like’, ‘people’, ‘feel’, ‘today’, ‘new’, ‘work’, ‘thank’, ‘day’, and ‘time’, indicating that the bulk of statuses on Twitter are just that: status updates on what and how people are feeling and thinking at the current time. Differences between our networks are usually not found at the core but the peripheries. To draw attention to some interesting differences between each of the networks that may be indications of underlying traits, we highlight relevant areas of our visualizations. In many networks, we discover distant connected components that appear to be ‘junk’ content promoted by bots. In some of the more egregious cases, we only display the largest component within our graph. We also find that promotional content appears often appears within our primary component; unfortunately, it is difficult to remove and clean these instances of noise.

For the introvert/extravert dimension, we find that a subsection of the introvert network corresponds to live streaming of gaming (‘stream’, ‘twitch’), suggesting that introverts may tend to be video game enthusiasts. A theme related to ‘Facebook’, ‘photos’, ‘posts’ was found in the extravert network suggesting an active interest in sharing social information (though we do not find any evidence of higher active engagement with the Twitter platform by extraverts.)

For Intuitive types, we find an interesting subnetwork – a thematic interest in current affairs (‘today’, ‘history’, ‘april’, ‘nasa’). We do not observe any particularly consistent and unique theme within the Sensing type network.

We observe many diverse subnetworks in the Perceiving type network – some select themes include ‘dogecoin’ (cryptocurrency), ‘disney’, and a theme on TV shows (‘watch’, ‘s01’). The network for Judging types is marked by a particularly interesting thematic group on social media-based promotion (‘marketing’, ‘social’, ‘media’, blogger’).

We do not discuss networks on liked tweets in detail as they do not contain significantly differentiated findings: this corroborates our findings from communication networks on liked tweets, where we find that information content consumed by profiles seems to be largely similar across personality types. Visualizations of the same can be found in the appendix.

Overall, while some of our findings on the thematic content within statuses are interesting and may be worth exploring further, we do not find significant evidence of *obvious* relationships between the thematic content of tweets and personality traits.