Analysis of User Temperament and Personality Traits in Social Media through Complex Networks

Matheus Santos, Elaine R. Faria, and Fabíola S. F. Pereira[1]

Faculty of Computing, Federal University of Uberlândia, Brazil
{matheusmgptc,elaine,fabiola.pereira}@ufu.br

Social media has proven to be an important source of data for studies related to user behavior. A gap in the state-of-the-art is the use of visual tools that explore not only the content posted, but also social interactions to understand the personality of users. This work proposes a method to analyze personality traits of Twitter users using complex networks. The networks were created from behavioral and textual data of users. After processing the attributes, it was possible to obtain a network with high modularity and homogeneous clusters. The results show patterns of language and behavior associated with the dimensions of the MBTI, contributing to studies related to the mental health of social media users.

Keywords: complex networks, social media, personality traits

1 Introduction

Since the early 21st century, social media has gained importance due to internet advancements and behavioral shifts, reinforcing its role in social and commercial contexts [5]. These platforms allow global content sharing, capturing user activities, opinions, and emotions [1]. This expansion has drawn interest from mental health and data science experts aiming to uncover behavioral patterns from social data [1].

Social media generates vast volumes of interaction data—likes, comments, and messages—reflecting personal interests and behaviors [1]. According to [2], such data is valuable for identifying personality traits, aiding psychological, business, and clinical studies.

Recent research has focused on classifying user temperament or personality using behavioral and textual features [2, 3, 1, 4]. However, few explore internal patterns within each temperament, limiting insights into behavior profiles [5, 18]. Understanding how users with the same temperament behave on social media can support professionals across multiple fields.

Complex networks can help address this gap by uncovering emerging patterns in data through both local and global relationships [6]. Even when data is not structured as a network, it can be transformed into graph-based models, where vertices represent entities and edges reflect affinity [17]. Applications include modeling communication flows and social interactions.

Their use has grown in fields like machine learning and data mining [6], due to their ability to integrate statistics, graph theory, and dynamics [17]. This makes them a promising tool to explore behavioral structures in online environments.

This work proposes a novel approach: rather than classifying temperament, it aims to analyze and correlate behavioral patterns among users who share similar personality traits using complex networks.

The main contributions include: i) analyzing the relationship between user behavior and temperament/personality traits; ii) providing insights for professionals using such traits in their practice; iii) encouraging new research in the intersection of social media and individual characteristics; iv) presenting a strategy to transform tabular user data into complex networks.

2 Related Works

Several studies explore social media for knowledge extraction, addressing topics like opinion analysis and politics [7]. However, fewer works examine the relationship between social networks and temperament or personality traits. Most focus on classifying users into predefined types, with limited investigation of behavioral patterns within each type [5, 18].

2.1 Works on User Temperament

Various studies investigate personality traits on social media to understand behavioral influence. [2] analyzed MBTI distributions among Twitter users versus the general population, using linguistic and behavioral features. With 1.2 million tweets from 1500 users, they found strong predictive models for E/I and F/T dimensions via logistic regression. Linguistic features proved most informative, revealing language differences aligned with personality.

[3] applied the TECLA framework [1] to identify Keirsey temperaments using Portuguese tweets. Employing TwiSty-PT and classifiers like SVM and Random Forest, they found best performance for artisan and guardian types. [1] proposed a framework mapping MBTI data [2] to Keirsey's model. Algorithms such as Random Forest with LIWC features yielded the highest accuracy.

While most works focus on classification, [5] explored Instagram data to examine behavioral patterns by temperament, using the TEMPS-RIO questionnaire. Findings indicated, for example, that depressed users posted more positive captions than hyperthymic or irritable ones, and anxious users used more emojis. [4] constructed a tweet dataset labeled with personality traits and tested various machine learning models. Using SMOTE for data balancing, Logistic Regression with TF-IDF unigrams achieved the best results.

Except for [5], all studies focused on classification using behavioral and textual features. None employed complex networks.

2.2 Works on Complex Networks

Complex networks are widely used to model real-world systems [6], particularly for identifying patterns and communities. [8] investigated Brazil's political dynamics using networks of deputies connected by voting similarity. Louvain-based community detection revealed evolving coalitions and group isolation. [9] proposed a visual-statistical method for community analysis in social networks across school and hospital datasets. They compared algorithms, finding Louvain and Infomap most effective, depending on context. [10] evaluated eight community detection algorithms on nine diverse datasets. Using modularity as a metric, the multiscale method outperformed others.

Despite its versatility, no prior study has applied complex networks to temperament analysis on social media.

3 Method to extracting patterns in different temperament types/personality traits from social media

The following will present the detailed steps of the proposed method, which can be seen in Figure 1.

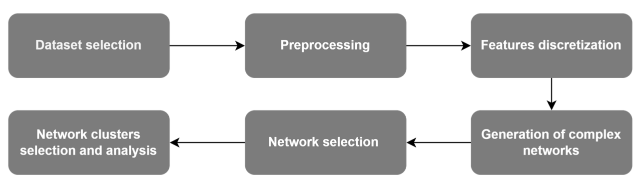

Fig. 1. Overview of the method stages.

3.1 Dataset selection

To apply the proposed method, it is necessary to use social media data that includes user-generated text and social attributes (e.g., followers, engagement metrics). Ideally, demographic variables (such as gender) and information on temperament or personality traits should also be available to enable method validation. Thus, databases meeting these criteria were selected.

There is a scarcity of datasets integrating all these elements, especially MBTI traits. Nonetheless, the method allows adaptation to the variables available.

3.2 Preprocessing

The preprocessing steps included: i) Text cleaning and treatment; ii) Extraction and normalization of LIWC features.

Text processing involved removing stopwords, punctuation, numbers, and special characters, followed by lowercasing and whitespace normalization—ensuring effective LIWC feature extraction.

The LIWC method [11] produced 73 features across linguistic and psychological categories. This study focused on 9 psychological-social attributes: i) *social*, ii) *family*, iii) *friend*, iv) *affect*, v) *posemo*, vi) *negemo*, vii) *anx*, viii) *anger*, and ix) *sad*.

Due to variations in text quantity and size per user, proportions were calculated rather than absolute values. The number of terms in each category was divided by the total word count per user to allow for fairer comparisons.

3.3 Features Discretization

After preprocessing, attributes were normalized to [0,1] to avoid scale bias in Euclidean distance computation (see Section 3.4). However, binary variables dominated due to the concentration of most values in narrow intervals. To mitigate this, three discretization strategies were tested:

- **Transformation 1:** Discretization into 3 groups (0, 0.5, 1) based on attribute min/max values.
- **Transformation 2:** Discretization into 5 groups (0, 0.25, 0.5, 0.75, 1), using the same logic.
- **Transformation 3:** Clustering into 3 groups using K-Means ($K = 3$).

After transformation, each user was represented by their attributes and ID. Euclidean distance matrices were then computed for each dataset, forming the basis for network edge creation.

3.4 Generation of complex networks

Three networks were generated based on each discretization. Nodes represent users; undirected edges were created if their Euclidean distance was 0, forming clusters of users with identical discretized features.

3.5 Network Selection

The best-performing network was selected based on two criteria: modularity [12, 17] and purity. Modularity evaluates the strength of division into communities.

Cluster purity p with respect to a temperament/personality trait x is defined as:

$$p_x = \max Q_x/Q_c \tag{1}$$

Where Q_x is the number of users of trait x in cluster c, and Q_c is the total number in c. A cluster is considered pure if $p_x \geq 70\%$. The selected network showed the highest modularity and most pure clusters across MBTI dichotomies.

3.6 Network clusters selection and analysis

Clusters from the best network were analyzed to uncover patterns related to MBTI dimensions. For each dichotomy, clusters were ranked by purity and user count. Although users in a cluster share identical characteristics (distance = 0), their MBTI labels may vary—hence the need for purity analysis.

Two clusters with the largest number of users were selected for each dimension, among the purest ones, to ensure sample size was sufficient to support pattern interpretation. These clusters also visually stood out in the network layout.

4 Experiments and discussion of results

4.1 Dataset selection

A relevant database was selected, prioritizing social media users with both personality information and social attributes (e.g., number of followers, likes). However, such datasets are scarce, especially those that combine all these features.

The TECLA dataset [2] was chosen. It includes data from 1500 *X* (formerly *Twitter*) users who referenced one of the 16 MBTI personality types along with the terms Briggs or Myers. Each user contributed between 100 and 2000 tweets, totaling 1.2 million messages.

In addition to tweets, variables such as *followers_count*, *statuses_count*, *gender*, *listed_count*, *favourites_count* and *MBTI type* were extracted. Because users had varying text volumes, psychological features were normalized by proportion of terms. Descriptive statistics of the original and LIWC-extracted variables are shown in Tables 1 and 2.

Table 1. Statistics of the original attributes of the TECLA database.

Statistics	num_tweets	followers_count	statuses_count	favorites_count	listed_count	gender
Variable Type	Numeric	Numeric	Numeric	Numeric	Numeric	Binary
Mean	789	1642	16322	4663	36	-
Standard Deviation	550	14631	26363	9650	259	-
Minimum Value	43	3	56	0	0	0
25% Quartile	247	133	2268	211	2	0
50% Quartile	774	309	7372	1241	6	1
75% Quartile	1069	717	18514	4305	17	1
Maximum Value	1994	510168	258427	107889	8776	1

4.2 Preprocessing

To identify the network with the best segmentation across MBTI dimensions, the modularity-based measure from Section 3.5 was applied. Among the three tested transformations, only Transformation 1 yielded clusters with purity $\geq 70\%$.

Table 2. Statistics of the attributes extracted by LIWC in the TECLA database.

Statistics	affect	posemo	negemo	anx	anger	sad	social	family	friend
Variable Type	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
Mean	678	462	211	26	76	43	849	41	41
Standard Deviation	520	375	175	23	77	36	665	44	38
Minimum Value	9	5	1	0	0	0	8	0	0
25% Quartile	191	127	51	7	18	11	237	9	11
50% Quartile	625	413	181	21	53	37	740	31	33
75% Quartile	967	649	308	37	109	62	1192	59	58
Maximum Value	2853	2461	1039	161	526	222	4264	557	362

Numeric variables were discretized into three intervals based on their min and max values. This produced 164 clusters and a modularity of 0.939, which indicates well-defined communities [6].

4.3 Network analysis

From the TECLA network, 201 users were excluded due to lack of connections (distance $\neq 0$). The final network had 1,299 nodes. Clusters were visualized using Gephi and analyzed using iGraph.[1] Figure 2 presents the networks segmented by MBTI dimension.

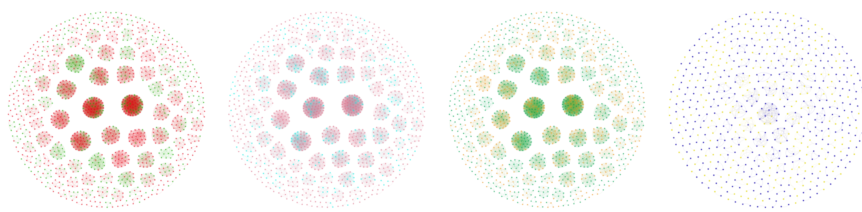

Fig. 2. Complex network with separation of dimensions, for the TECLA database, by means of colors, using the first proposal for transforming attributes. From left to right: E/I, N/S, T/F and J/P, respectively.

In E/I, 63.9% of users were class I; 51.83% of clusters were pure for I, and 18.29% for E. In N/S, class N dominated (74.84%), with 70.73% pure clusters. In T/F, class F represented 58.66%, and 27.44% of clusters were pure. In J/P, class J comprised 59.51% of users, with 31.10% pure clusters. Table 3 summarizes these results.

For each dimension, two clusters per class were selected based on size and purity to support deeper analysis.

[1] https://igraph.org/

Table 3. Network results obtained from the TECLA database.

Dimension	Color Legend	% of users in the network	% pure clusters	Overall average of purity
E/I	green/red	36.10% / 63.90%	18.29% / 51.83%	35.06%
N/S	pink/blue	74.84% / 21.56%	70.73% / 2.44%	36.59%
T/F	orange/green	41.34% / 58.66%	9.15% / 27.44%	18.29%
J/P	blue/yellow	59.51% / 40.49%	31.10% / 8.54%	19.82%

4.4 Analysis of the clusters in TECLA database

In E/I, class E users posted more tweets and used more positive language, while class I showed lower emotional expression. Within class I, female users used fewer negative and social terms than males [13, 14].

In N/S, class N (mostly female) had lower negativity and intermediate positivity. Class S users, especially males, had higher negativity. The contrast reflects the abstract-oriented nature of intuition vs. the concreteness of sensing [13].

In T/F, class T (predominantly male) had low emotional term use, while class F (mostly female) had intermediate levels. This aligns with gendered patterns of emotional expression [15, 16].

In J/P, class J users showed fewer negative and angry terms and more positivity. This aligns with profiles that favor quick, structured decision-making versus exploratory, analytical behavior in class P [13].

5 Conclusion

This work demonstrated the potential of complex networks for analyzing personality traits using social media data.[2] Based on X (formerly Twitter) data labeled with MBTI types, user interactions were modeled through psychological and behavioral similarities from published texts.

Among three tested discretization strategies, Transformation 1 provided the best segmentation, yielding 164 clusters with high modularity (0.939). The methodology—combining network modeling, LIWC-based feature extraction, and normalization—effectively segmented users across the MBTI dimensions.

Results revealed consistent behavioral patterns, particularly in the N/S and E/I dichotomies, where differences emerged in posting frequency, affective language, and social variables, and aligned with prior research on gender influences in the T/F dimension. Despite challenges like class imbalance and limited datasets, the study reinforces the value of complex networks in personality analysis and opens paths for future research.

Acknowledgements: The authors would like to thank the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (Grant no APQ-00226-21) and CNPq (402543/2021-01) for the financial support.

[2] Source code and publicly constructed complex networks available at: https://github.com/mtshnq/asonam2025

References

1. Lima, A. C. E. S., & Castro, L. N. (2019). TECLA: A temperament and psychological type prediction framework from Twitter data. *PLOS ONE*, 14(3), e0212844.
2. Plank, B., & Hovy, D. (2015). 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).
3. Claro, C. F. (2018). Um estudo de caso sobre o modelo de temperamento de Keirsey (Dissertação de Mestrado). Universidade Presbiteriana Mackenzie.
4. Almeida, L. D., & Goya, D. (2023). Detection of Big Five personality traits in Twitter user's profiles based on textual posts. In *Anais do XX Encontro Nacional de Inteligência Artificial e Computacional* (pp. 227–241). SBC.
5. Martins, L. M. (2022). Análises de publicações da rede social Instagram em contraste com o temperamento humano. Dissertação de Mestrado, Universidade Federal de Uberlândia.
6. Carneiro, M. G. (2017). Redes complexas para classificação de dados via conformidade de padrão, caracterização de importância e otimização estrutural. Tese de Doutorado, Universidade de São Paulo.
7. Tabassum, S., Pereira, F. S. F., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *WIREs Data Mining and Knowledge Discovery*, 8(5), e1256. 10.1002/widm.1256
8. Brito, A. C. M., Silva, F. N., & Amancio, D. R. (2020). A complex network approach to political analysis: Application to the Brazilian Chamber of Deputies. *Plos One*, 15(3), e0229928.
9. Linhares, C. D. G., Ponciano, J. R., Pereira, F. S. F., Rocha, L. E. C., Paiva, J. G. S., & Travençolo, B. A. N. (2020). Visual analysis for evaluation of community detection algorithms. *Multimedia Tools and Applications*, 79, 17645–17667.
10. Gul, H., Al-Obeidat, F., Amin, A., Tahir, M., & Moreira, F. (2022). A systematic analysis of community detection in complex networks. *Procedia Computer Science*, 201, 343–350.
11. Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
12. Newman, M. (2018). *Networks*. Oxford University Press. 10.1093/oso/9780198805090.001.0001. Disponível em: https://doi.org/10.1093/oso/9780198805090.001.0001
13. Ribeiro Filho, J. F., de Gusmão Lopes, J. E., Pederneiras, M. M. M., de Almeira, L. B., & de Barros Ribeiro, M. T. J. (2010). Características da personalidade de estudantes de ciências contábeis: Análise do conhecimento baseado no Modelo Myers-Briggs Type Indicator (MBTI). *Contabilidade Gestão e Governança*, 13(2).
14. Nunes, C. H. S. da S. (2005). Construção, normatização e validação das escalas de socialização e extroversão no modelo dos Cinco Grandes Fatores.
15. Martins-Suarez, F. C. & Sousa, J. M. M. (2016). *Homem razão e mulher emoção: Uma análise da relação dicotômica entre homens e mulhere na visão dos assentados(as).*, Ambivalências, vol. 4, n. 7, pp. 288–308.
16. Bourdieu, P. (2010). *A dominação masculina*. In: *A dominação masculina*, pp. 158–158.
17. Barabási, A.-L. (2014). Network science book. *Network Science*, 625.
18. de Alcantara, C., de Faria Paiva, E. R., Barioni, M. C. N., & Junior, L. C. O. (2024). Análise do Temperamento e Percepção do Suporte Social em Redes Sociais Online. *iSys-Brazilian Journal of Information Systems*, 17(1), 11–1.