Received: 16 October 2024 | Revised: 27 November 2024 | Accepted: 16 December 2024 | Published online: 5 February 2025

RESEARCH ARTICLE

Artificial Intelligence and Applications
2025, Vol. 00(00) 1–10
DOI: 10.47852/bonviewAIA52024561

PersonaG: A Quinpartite Graph Convolutional Network for Interpretable Personality Recognition from Text

Mohmad Azhar Teli[1],* **and Manzoor Ahmad Chachoo[1]**

[1]Department of Computer Science, University of Kashmir, India

Abstract: Automatic personality recognition from text has wide-ranging applications in social media analysis, targeted marketing, and personalized user experiences. As such, a lot of researchers have focused on personality recognition in the last two decades. However, existing methods often rely on only shallow semantic features or psycholinguistic features to capture the semantic information in textual data. We propose PersonaG, a novel approach that integrates psycholinguistic categories with WordNet semantics to address these limitations and construct quinpartite graph representations. Our approach combines semantic relationships with the psycholinguistic categories. Classification is performed using a Dynamic Deep Graph Convolutional Network. Our results on the benchmark Essays dataset outperform recent methods, achieving state-of-the-art performance and demonstrating the superiority of our approach. To conclude, the quinpartite graph enables PersonaG to understand the latent personality patterns from text, making it a comprehensive and effective solution for personality recognition.

Keywords: personality computing, APRT, lexical hypothesis, natural language processing, graph convolution network

1. Introduction

Personality is a complex psychological construct encompassing an individual's thoughts, emotions, and behaviors [1]. The study of the personality works to clarify the contrasts in individual conduct. So, the fundamental objective of personality psychology is "to recognize inward properties of the individual from obvious practices and to research the causal connections between them" [2]. The relationship between personality and language has intrigued researchers ever since Theophrastus [3]. The idea gained more impetus with the emergence of the lexical hypothesis, which suggests that significant aspects of personality are encoded in language [4]. With the rise of social media and ever-increasing consumer data availability, it is argued that textual data can be used to train personality computational models to infer traits. These models can be extremely beneficial with widespread applications in social media analytics [5], personalized marketing [6], user experience design [7], and mental health monitoring [8]. As such the last two and a half decades have seen a rising interest in Automatic Personality Recognition from Text (APRT). A wide array of techniques ranging from traditional machine learning methods to sophisticated deep learning and graph-based approaches [9, 10] have been used. Some of these approaches have shown promising results, but there is a need for models that can effectively capture the intricate relationships between personality and natural language. The existing research is filled with benchmark datasets like YouTube blogs [11], MyPersonality [12], Pandora [13], Kaggle [14], Twisty [15], PAN-AP-2015 [16], and Essays [17], we chose Essays for our study as it is the first, yet the most notorious in terms of results [18]. Several psychological scales have been utilized for computational modeling of personality traits, but majority of the researchers follow either MBTI [19] or the BIG 5 [20]. However, we find there is a slight favor for Big 5 scale (Table 1) in [21] and fair amount of underlying correlation between the two scales [22].

Building upon the existing foundation of research, particularly the TrigNet framework [23], we introduce PersonaG. This novel approach extends the Dynamic Deep Graph Convolutional Network (DGCN) [24] with a multi-partite graph structure, dynamic multi-hop (*DmH*) mechanism, and the Learn-to-Connect (L2C) approach. PersonaG combines psycholinguistic categories with WordNet-based semantics to create a comprehensive graph representation. This multi-partite graph is used to capture the relationships among words, sentences, documents, psycholinguistic categories, and semantic associations, producing a highly interpretable model for association between personality and language. This approach offers a more nuanced method for recognizing personality traits from text, addressing a key limitation in previous work, where models often struggled to capture the full spectrum of personality expression in text.

Key contributions of PersonaG include the following:

1) A multi-partite graph structure integrating psycholinguistic and semantic information effectively.
2) Enriched node representation with the integration of WordNet embeddings.
3) A *DmH* mechanism that allows for more flexible information propagation through the graph.
4) Incorporating a L2C approach enables the model to adjust connections between nodes dynamically.

*Corresponding author: Mohmad Azhar Teli, Department of Computer Science, University of Kashmir, India. Email: mohmadazhar.student@kashmiruniversity.net

Table 1
Big 5/OCEAN traits

Trait	Description
Openness (Opn)	Imagination, creativity, curiosity, and openness to experience.
Contentiousness (Con)	Organization, dependability, self-discipline, and goal orientation.
Extraversion (Ext)	Sociability, assertiveness, enthusiasm, and outgoing nature.
Agreeableness (Agr)	Trust, altruism, kindness, and cooperativeness.
Neuroticism (Neu)	Emotional instability, anxiety, moodiness, and irritability.

These features allow PersonaG to interpret the relationship between linguistic features and personality traits and, as such, improve the results achieved by existing models across benchmark datasets. In this paper, we demonstrate how PersonaG combines linguistic and semantic information, with a graph-based approach for personality recognition. We aim to show how our approach offers advantages over existing deep learning and graph-based models in capturing the subtleties of personality expression in text. The rest of the paper is structured as follows: Section 2 builds upon a comprehensive review of related literature, highlighting key advancements and existing gaps. In Section 3, we provide a detailed methodology – including constructing the multi-partite graph, node initialization, and the architecture of the extended DGCN used for the personality classification. The dataset and other experimental setup details form the Section 4. Section 5 includes the results and comparison of PersonaG's performance to existing techniques and discussing its interpretability. Finally, Section 6 summarizes our contributions, addresses limitations, and suggests directions for future research in this rapidly evolving field.

2. Literature Review

Several recent works have focused on personality prediction using text data, leveraging datasets such as Essays, the YouTube dataset, and Big Five personality traits. Tighe et al. [25] used feature reduction techniques such as PCA and information gain on LIWC features. Their logistic regression and SVM models achieved competitive accuracy on the Essay dataset with significantly fewer features than prior methods. In [26], Xue et al. introduced a semantic-enhanced personality recognition neural network, which utilizes context learning to create word-level semantic representations of text. Their model showed good results on the YouTube dataset, with an average accuracy of 70%, but performance on Essays remained below 60%.

Ramezani et al. [27] proposed an ensemble method using various techniques like TF-IDF, Ontologies, and Latent Semantic Analysis for personality prediction on the Essays dataset. Their approach achieved an average accuracy slightly above 60%, with lower accuracy observed for the trait of openness. Wang et al. [28] introduced a Graph Convolution Network-based personality recognition model leveraging a heterogeneous graph structure. While their model outperformed existing methods on the MyPersonality dataset, performance on Essays remained around 60%. In [29], a knowledge graph-enabled model was used for personality trait prediction from Essays, achieving accuracies of up to 71% with various neural network-based classifiers. Additionally, a knowledge graph-based approach proposed in [30] shows improved results using attention-based Graph Neural Networks. Kerz et al. [31] proposed a hybrid approach combining Psycholinguistic and Transformer-based embeddings for personality prediction from Essays. Their best-reported results achieved an accuracy of around 72% for the trait of openness. Roy et al [32] utilized tree-transformers with Graph Attention Network for personality prediction in Essays, reporting an average accuracy of 68%. Zhu et al. [33] introduced the Contrastive Graph Transformer Network, which incorporates LIWC and post-semantic knowledge graph augmentation priors. Their model achieved an average $F1$ score of 75.03% for the Essays dataset. Yang et al. [34] proposed (DeepPerson), a comprehensive model comprising CNN-LSTM, wlpHAN, and SPDFiT components. DeepPerson significantly outperformed existent models on the PANDORA and MyPersonality datasets.

The literature review highlights various techniques and models that have been proposed for APRT, ranging from traditional machine learning methods to more recent deep learning and graph-based approaches. The result trends have promisingly improved over time, thus motivating the need for sophisticated models like our PersonaG. The PersonaG model derives its base from the DeepPerson framework, enhancing the DGCN with a multi-partite graph structure, *DmH* mechanism, and L2C approach. We have incorporated word-, sentence-, and document-level embeddings in addition to LIWC [35] and WordNet categories [36] to enrich node representations, thus improving the model's ability to predict personality traits. We believe our model achieves more accurate personality predictions by dynamically adjusting connections between nodes and leveraging a comprehensive graph structure than existing approaches.

3. Research Methodology

This section explains the complete working of the proposed PersonaG approach. We first describe the construction of the quinpartite graph, followed by the initialization of node representations and the DDGCN architecture used for classification.

3.1. Quinpartite graph construction

We constructed a heterogeneous quinpartite graph for each user, integrating LIWC categories and WordNet embeddings to effectively capture psycholinguistic features. As illustrated in Figure 1, the constructed quinpartite graph comprises five types of nodes:

1) Word-level nodes (V_w): Represent individual words in the text.
2) Sentence-level nodes (V_s): Represent individual sentences in the text.
3) Document-level nodes (V_d): Represent entire documents.
4) LIWC category nodes (V_l): Represent the psycholinguistic categories derived from the LIWC 2015 dictionary.
5) WordNet category nodes (V_m): Represent the semantic relationships between words, derived from WordNet.

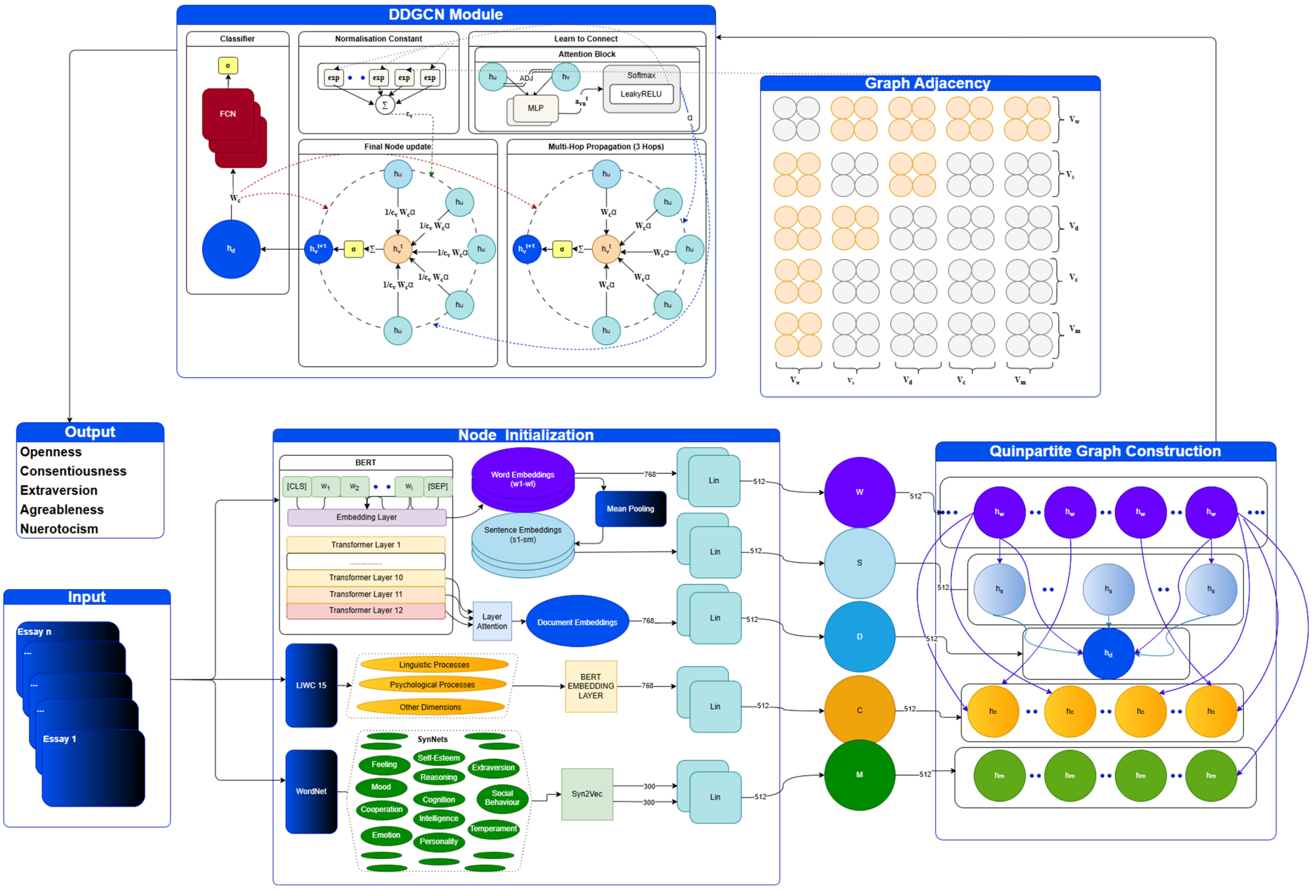
The quinpartite graph can be mathematically represented as:

$$G = (V, E) \tag{1}$$

$$V = V_w \cup V_s \cup V_d \cup V_l \cup V_m \tag{2}$$

where V_w represents the set of word nodes, V_s represents the set of sentence nodes, V_d represents the set of document nodes, V_l represents the set of LIWC category nodes, and V_m represents the set of WordNet category nodes.

Figure 1
Model architecture of PersonaG



3.2. Node initialization

Our quinpartite graph has five node types and each of them is initialized in the following manner.

The word-level vertices V_w and sentence-level vertices V_s initialized using the pretrained BERT [37] model.

3.2.1. V_W: Word embedding (BERT)

$$emb_{word}(w^i) = BERT(w^i) \tag{3}$$

where emb_{word} is the word-level embedding, and for each word w^i, the word embedding is obtained by passing the word through the pretrained BERT.

3.2.2. V_S: Sentence embedding (BERT)

$$emb_{sentence}(s^i) = BERT(s^i) \tag{4}$$

where $emb_{sentense}$ is the sentence-level embedding and for each sentence s^i, the sentence embedding is calculated by passing the sentence through the pretrained BERT model.

3.2.3. V_D: Document embedding (BERT)

The document embeddings are obtained by aggregating the word and sentence embeddings using a mean pooling aggregation function:

$$emb_{document}(d^i) = Agg(w^i) \cup Agg(s^i) \tag{5}$$

where $emb_{document}$ is the document-level embedding and the document embeddings are computed by aggregating all the word embeddings and sentence embeddings for a document d^i. Aggregation is performed using a mean pooling function, which takes the average of the embeddings for words and sentences in the document.

3.2.4. V_L: LIWC categories

The LIWC categories have been obtained for each user from LIWC 2015 dictionary:

$$emb_{LIWC}(l^i) = Embedding\_Matrix(l^i) \tag{6}$$

where $emb_{LIWC}(l^i)$ represents the document-level category scores calculated using LIWC 2015 dictionary, which contains 72 categories covering emotion, cognition, social behavior, personal concerns, biological concerns, and wellbeing.

3.2.5. V_M: WordNet categories

WordNet categories are utilized to capture semantic relationships between words:

$$emb_{\_WordNet}(m^i) = Embedding\_Matrix(m^i) \quad (7)$$

where $emb_{\_WordNet}(m^i)$ denotes the wordnet [] categories (Synsets) that capture specific meanings and relationships between words. Out of 117,000 Synsets, we chose to use 12 primary categories (like emotion, mood, cognition, etc.,) and 72 related secondary categories for capturing the semantic relationships between the words.

3.3. Connections between nodes

The connections between the nodes are what define the structure of the quinpartite graph. These connections enable the model to propagate information between different node types during the graph convolution process. The primary connections in this model include the following:

Word to Sentence Connections: Each word node is connected to the sentence node(s) it belongs to. This connection allows the model to capture the local context of words within a sentence.

Sentence to Document Connections: Each sentence node is connected to the document node, linking individual sentences to the overall document. This relationship allows the model to aggregate sentence-level features into a document-level representation.

Document to Psycholinguistic Category (LIWC) Connections: The document-level node is connected to the LIWC category nodes. This connection allows the model to integrate psychological features from the LIWC categories, enabling the model to predict personality traits based on these psychological dimensions.

Document to WordNet Category Connections: Similarly, the document-level node is connected to WordNet category nodes, allowing the model to integrate semantic knowledge from WordNet into the personality prediction process.

3.4. Dynamic multi-hop structure

Once the node representations are initialized, a *DmH* mechanism is used to propagate information across the quinpartite graph. This approach effectively captures intricate relationships and patterns by iteratively updating node representations based on neighboring nodes' information.

The node representation at time t is represented by h_v^t and is updated at the next time step t+1 by aggregating information from its neighboring nodes $N(v)$, using weights α_{vu}^t that represent the importance of the *neighbors*. The sum of weighted representations of neighbors is passed through a non-linear *sigmoid* activation function σ:

Node Update rule:

$$h_v^{t+1} = \sigma(\sum(u \in N(v))\alpha_{vu}{}^t W^t h_u{}^t) \quad (8)$$

The edge weights between u and v, α_{vu}^t are computed using a *softmax* function applied to the output of a *leaky ReLU* [38], activation, thus creating an attention mechanism that helps model to focus on the most important neighbors:

Node Attention rule:

$$\alpha_{vu}{}^t = softmax(LeakyReLU(a_{vu}{}^t)) \quad (9)$$

These mechanism connections allow information to propagate from word-level nodes to sentence-level nodes, from sentence-level nodes to document-level nodes, and from document-level nodes to LIWC and WordNet category nodes. The connections across different node types enable the model to integrate both semantic (WordNet) and psychological (LIWC) features with the text-level information (word, sentence, and document).

3.5. Learn-to-connect approach

For adjusting the node connections automatically, a L2C mechanism is used. The L2C mechanism helps the model to learn the importance of different connections, enabling it to capture the most relevant relationships for personality recognition.

The importance of a connection between two nodes v and u is learned through a multi-layer perceptron (MLP), which takes the concatenation of the two node embeddings h_v^t and h_u^t as input:

Learned Connection weight:

$$a_{vu}{}^t = MLP([h_v{}^t | h_u{}^t]) \quad (10)$$

This dynamic weighting allows the model to adaptively decide which connections are most relevant for personality trait prediction.

3.6. DDGCN module

The PersonaG model is based on a dual DDGCN. The DDGCN module is responsible for learning node representations within the quinpartite graph structure. It includes the DmH mechanism and L2C approach, enabling the effective propagation of information across the graph and inferring the underlying patterns.

A normalization constant c_v is used to control the scaling of the *neighbors'* contributions. This ensures that the nodes with large neighborhoods do not dominate the update process:

Node Update (with Normalization):

$$h_v{}^{t+1} = \sigma(\sum(u \in N(v))1/c_v W^t h_u{}^t) \quad (11)$$

The *normalization constant* c_v is computed as the sum of the exponentiated *attention coefficients* a_{vu}^t, ensuring that each node's neighborhood influence is properly scaled:

$$c_v = \sum(u \in N(v))\, exp(LeakyReLU(a_{vu}{}^t)) \quad (12)$$

3.7. Classification and training

After obtaining the final node representations from the DDGCN module, we perform classification to predict personality traits. For each personality trait, we employ a separate fully connected layer followed by a sigmoid activation function to obtain the predicted personality:

$$\hat{y} = \sigma(W_c h_d + b_c) \quad (13)$$

where h_d is the final representation of the document node, W_c and b_c are the weights and biases of fully connected layer used for classification. The output is passed through a sigmoid activation function σ to predict the personality trait $\hat{y}$.

4. Experiments

This section describes the experimental setup, including the dataset, evaluation metrics, and implementation details.

Table 2
Statistics of the essays dataset

Statistic	Value
Total Number of Essays	2467
Average Word Count per essay	~650
Total Words in the dataset	~1.6 million
Trait labels	Big 5

4.1. Datasets

We utilize the widely-used Essays dataset, initially collected by Pennebaker and King, for our experiments. This dataset comprises 2,400 stream-of-consciousness essays authored by 1,203 psychology students from the University of Pennsylvania. The essays have been annotated with the Big Five Inventory [39] based on self-reported personality traits, making it a standard benchmark for text-based personality computation. Table 2 illustrates the statistics of Essays dataset.

4.2. Evaluation metrics

The model performance is compared with the baselines using evaluation metrics commonly employed in personality recognition tasks:

1) Accuracy: The fraction of correctly classified labels over the total number of labels.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

where: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives

2) $F1$-score: A balanced measure of performance, calculated as the harmonic mean of recall and precision.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{15}$$

where $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$

4.3. Implementation details

The model is implemented using the PyTorch [40] deep learning framework on a docker [41] based environment. The experiments were implemented and evaluated on NVIDIA DGX A100 AI Server [42], based in Department of Computer Science, University of Kashmir. The server has 320 GB GPU memory, a 3.4 GHz Dual AMD Rome 7742 CPU, 1TB system memory, and a 15 TB Storage.

5. Results

The results of PersonaG model and the baselines on the Essays dataset for personality recognition are presented in Table 3. We report the performance metrics averaged across the five personality traits: Openness (O), Conscientiousness(C), Extroversion(E), Agreeableness(A), and Neuroticism(N). Our proposed model achieves competitive performance across all traits but Extroversion, with an average accuracy of 72.60%.

5.1. Ablation study

The ablation study investigates the impact of different interaction flows within the quinpartite graph on the model's overall performance. We conduct experiments where each interaction flow is individually disabled while rest of the model is kept intact. By comparing the performance of the model with and without each interaction flow, we aim to assess their relative importance in predicting personality traits from text data. Table 4 shows the results of the ablation study, reporting the accuracy across the five personality traits on the Essays dataset.

5.2. Discussion

Our results indicate the effectiveness of PersonaG in identifying personality traits from textual data. The proposed approach outperforms existing methods across all traits and metrics (Figure 2), illustrating the robustness of the quinpartite graph model in capturing linguistic and semantic cues related to personality expression. The combination of psycholinguistic categories from LIWC and semantic relationships from WordNet, along with the word-, sentence-, and document-level embeddings enrich the quinpartite graph enabling a holistic estimation of the underlying relationship between language and personality. The DmH and L2C approaches enable PersonaG to capture intricate patterns and relationships that traditional methods may overlook.

Table 3
Comparison of existing baselines with the proposed model

	Accuracy						***F*1 score**					
Baseline	Opn	Con	Ext	Agr	Nue	Avg.	Opn	Con	Ext	Agr	Nue	Avg.
LIWC	59.5	55.7	56.6	53.3	58.8	56.78	69.0	61.2	54.4	67.5	63.0	63.02
SEPRNN	63.16	57.49	58.91	57.49	59.51	59.31	67.84	63.46	71.5	71.92	62.36	67.42
GCN	64.8	59.1	60.0	57.7	63.0	60.92	67.0	68.0	67.0	69.0	69.0	68.0
KGE	71.4	72.62	73.83	70.18	69.37	71.48	73.64	75.68	77.72	71.78	68.34	73.43
KGRAT-NET	72.21	**73.43**	**74.24**	71.2	70.99	72.41	74.96	76.48	**78.08**	72.8	69.89	74.44
PSYLING	71.95	61.38	63.01	60.16	60.98	63.5	n.a	n.a	n.a	n.a	n.a	n.a
GAtnN	70.1	69.2	66.5	64.8	69.0	67.9	n.a	n.a	n.a	n.a	n.a	n.a
CGTN	n.a	n.a	n.a	n.a	n.a	n.a	72.17	76.21	78.78	77.12	70.87	75.03
DeepPerson	n.a	n.a	n.a	n.a	n.a	n.a	58.3	61.0	60.3	59.6	62.4	60.3
Proposed	**73.1**	70.7	69.9	**75.7**	**73.6**	**72.6**	**75.7**	**78.8**	72.5	**75.0**	**76.7**	**75.2**

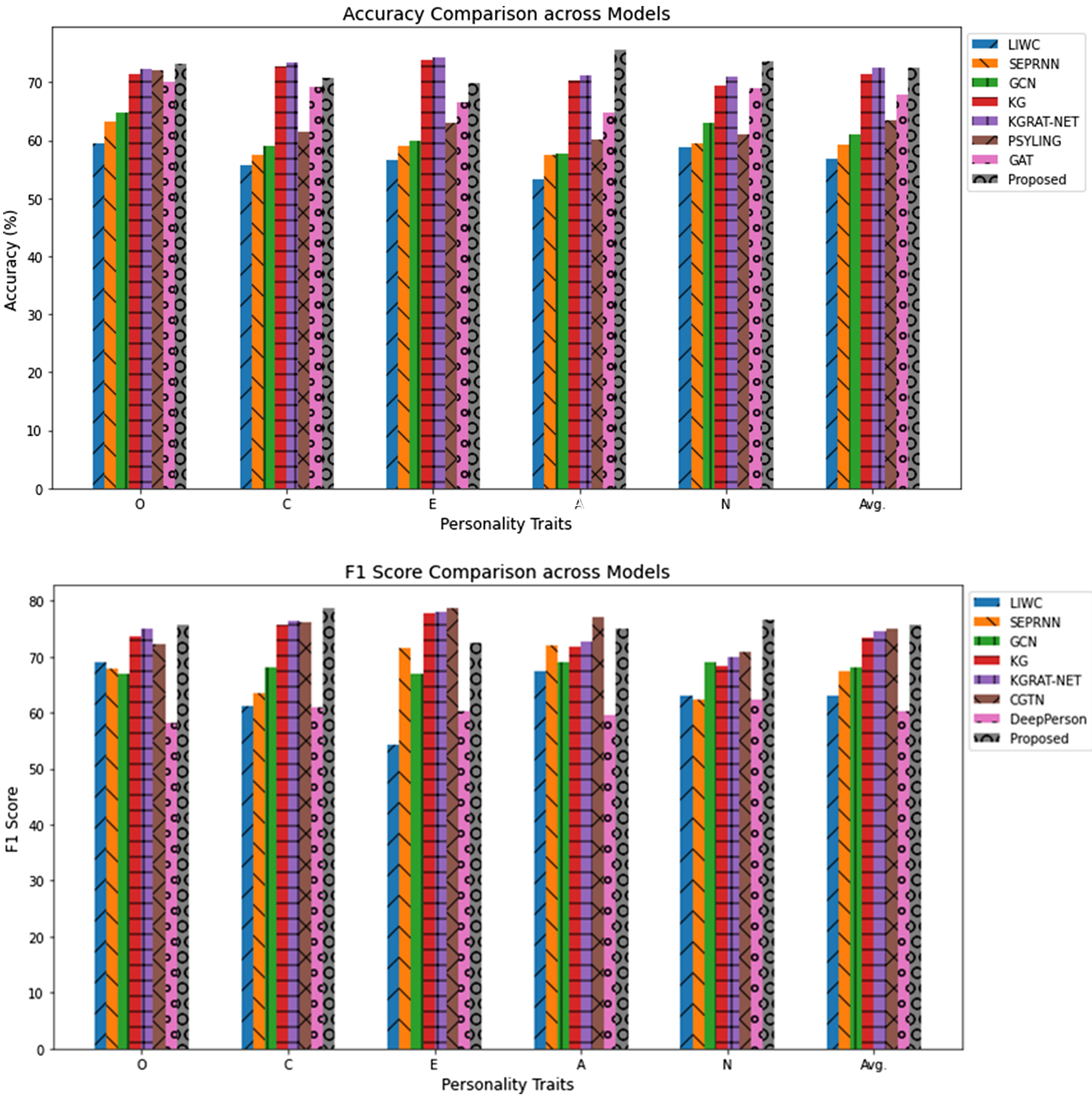
Note: where the baseline models have not used a metric (Accuracy/$F1$ Score).

Table 4
Ablation study with different interaction flows

Model configuration	Accuracy					
	Opn	Con	Ext	Agr	Neu	Accuracy
PersonaG (Full Model)	**73.1**	**70.7**	**69.9**	**75.7**	**73.6**	**72.6**
"d <=> w <=> d"	71.2	69.0	68.4	73.5	72.3	70.4
"d <=> s <=> d"	70.4	68.1	67.5	72.0	71.8	69.9
"d <=> c <=> d"	71.0	69.3	68.7	73.2	72.7	71.3
"d <=> m <=> d"	70.6	68.5	68.2	72.6	71.9	70.8

The superior performance of PersonaG highlights the benefits of combining linguistic knowledge from psychological lexicons with semantic information from knowledge bases like WordNet. This approach enables the model to capture the multifaceted nature of personality expressions more effectively than methods that rely solely on either linguistic features or semantic representations. To sum it up, PersonaG model represents a significant advancement in the field of APRT, offering a robust framework for extracting meaningful insights from textual data.

Figure 2
Comparison with state of the art



6. Conclusion

Our research contributes to advanced APRT systems, offering a robust framework for personality profiling from textual data. Moreover, the interpretable nature of the quinpartite graph structure and the learned node representations will offer valuable insight into the underlying relationships between language and personality traits. This is of extreme importance in applications where understanding the reasoning behind personality predictions is crucial, such as in personalized user experiences, targeted marketing strategies, and psychological research. The PersonaG model holds promise for applications in various domains, including social sciences, marketing, and human-computer interaction, where understanding and adapting to individual personalities are crucial for tailored communication and user engagement. Additionally, the model's capability for knowledge graph generation opens up opportunities for exploring personality recognition in diverse textual data sources, such as social media posts, customer reviews, and personal narratives.

6.1. Limitations and future work

PersonaG represents a significant advancement in the field, but there remain challenges and opportunities for future research. Incorporating multimodal data, such as audio and visual cues, could further enhance the model's ability to capture the multifaceted nature of personality expression. Furthermore, APRT is a data-driven task, and labeling is very costly; we can also use a generative model based on PersonaG to augment the existing dataset, which is fed back to the model in an iterative manner.

Additionally, exploring personalized and context-aware personality recognition systems tailored to specific domains or user groups could yield valuable insights and applications. To conclude, our research integrates psycholinguistic knowledge and semantic relationships in a dynamic graph-based framework for accurate and interpretable personality recognition from textual data.

Acknowledgement

We extend our sincere gratitude to Dr. M.A. Wani and the Department of Computer Science, University of Kashmir for providing access to the NVIDIA DGX A100 server, which played a pivotal role in the computational tasks required for this research.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://psycnet.apa.org/doi/10.1037/0022-3514.77.6.1296, reference number [17].

Author Contribution Statement

Mohmad Azhar Teli: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Manzoor Ahmad Chachoo:** Conceptualization, Resources, Supervision, Project administration.

References

[1] Bergner, R. M. (2020). What is personality? Two myths and a definition. *New Ideas in Psychology*, *57*, 100759. https://doi.org/10.1016/j.newideapsych.2019.100759

[2] Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. *IEEE Transactions on Affective Computing*, *5*(3), 273–291. https://doi.org/10.1109/TAFFC.2014.2330816

[3] Leese, M. (2024). Preferences, personality, and rational choice in Aristotle and Theophrastus. *ΠΗΓΗ/FONS*, (7), 179–196.

[4] Cutler, A., & Condon, D. M. (2023). Deep lexical hypothesis: Identifying personality structure in natural language. *Journal of Personality and Social Psychology*, *125*(1), 173–197. https://psycnet.apa.org/doi/10.1037/pspp0000443

[5] Hassanein, M., Hussein, W., Rady, S., & Gharib, T. F. (2018). Predicting personality traits from social media using text semantics. In *2018 13th International Conference on Computer Engineering and Systems*, 184–189. https://doi.org/10.1109/ICCES.2018.8639408

[6] Saha, P., Sengupta, A., & Gupta, P. (2024). Influence of personality traits on generation Z consumers' click-through intentions towards personalized advertisements: A mixed-methods study. *Heliyon*, *10*(15), e34559. https://doi.org/10.1016/j.heliyon.2024.e34559

[7] Filippi, S. (2020). PERSEL, a ready-to-use PERsonality-based user selection tool to maximize user experience redesign effectiveness. *Multimodal Technologies and Interaction*, *4*(2), 13. https://doi.org/10.3390/mti4020013

[8] Alqahtani, F., Meier, S., & Orji, R. (2022). Personality-based approach for tailoring persuasive mental health applications. *User Modeling and User-Adapted Interaction*, *32*(3), 253–295. https://doi.org/10.1007/s11257-021-09289-5

[9] Teli, M. A., & Chachoo, M. A. (2022). Lingual markers for automating personality profiling: Background and road ahead. *Journal of Computational Social Science*, *5*(2), 1663–1707. https://doi.org/10.1007/s42001-022-00184-6

[10] Fang, Q., Giachanou, A., Bagheri, A., Boeschoten, L., van Kesteren, E. J., Kamalabad, M. S., & Oberski, D. (2023). On text-based personality computing: Challenges and future directions. In *Findings of the Association for Computational Linguistics: ACL 2023*, 10861–10879. https://doi.org/10.18653/v1/2023.findings-acl.691

[11] Biel, J. I., & Gatica-Perez, D. (2012). The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Transactions on Multimedia*, *15*(1), 41–55. https://doi.org/10.1109/TMM.2012.2225032

[12] Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70*(6), 543–556. https://doi.org/10.1037/a0039210

[13] Gjurković, M., Karan, M., Vukojević, I., Bošnjak, M., & Šnajder, J. (2021). PANDORA talks: Personality and demographics on reddit. In *9th International Workshop on Natural Language Processing for Social Media*, 138–152. https://doi.org/10.18653/v1/2021.socialnlp-1.12

[14] Kong, S., & Sokolova, M. (2024). Explainable muti-label classification of MBTI types. *arXiv Preprint:2405.02349*.
[15] Verhoeven, B., Daelemans, W., & Plank, B. (2016). Twisty: A multilingual twitter stylometry corpus for gender and personality profiling. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation,* 1632–1637.
[16] Rangel Pardo, F. M., Celli, F., Rosso, P., Potthast, M., Stein, B., & Daelemans, W. (2015). Overview of the 3rd author profiling task at PAN 2015. In *CLEF 2015 Evaluation Labs and Workshop Working Notes Papers, 1391,* 1–8.
[17] Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296–1312. https://psycnet.apa.org/doi/10.1037/0022-3514.77.6.1296
[18] Arambašic, L., Bicanic, M., & Rajic, F. (2021). Essays are a fickle thing. In J. Jukic & J. Šnajder (Eds.), *Text analysis and retrieval 2021 course project reports* (pp. 1–5). University of Zagreb.
[19] Briggs, K. C. (1976). *Myers-Briggs type indicator*. USA: Consulting Psychologists Press.
[20] Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34. https://psycnet.apa.org/doi/10.1037/0003-066X.48.1.26
[21] Celli, F., & Lepri, B. (2018). Is big five better than MBTI? A personality computing challenge using Twitter data. In *Proceedings of the Fifth Italian Conference on Computational Linguistics,* 93–98. https://doi.org/10.4000/books.aaccademia.3147
[22] Furnham, A. (2022). The big five facets and the MBTI: The relationship between the 30 NEO-PI (R) facets and the four Myers-Briggs type indicator (MBTI) scores. *Psychology*, *13*(10), 1504–1516. https://doi.org/10.4236/psych.2022.1310095
[23] Yang, T., Yang, F., Ouyang, H., & Quan, X. (2021). Psycholinguistic tripartite graph network for personality detection. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 1*, 4229–4239. https://doi.org/10.18653/v1/2021.acl-long.326
[24] Yang, T., Deng, J., Quan, X., & Wang, Q. (2023). Orders are unwanted: Dynamic deep graph convolutional network for personality detection. In *Proceedings of the AAAI Conference on Artificial Intelligence, 37*(11), 13896–13904. https://doi.org/10.1609/aaai.v37i11.26627
[25] Tighe, E. P., Ureta, J. C., Pollo, B. A. L., Cheng, C. K., & de Dios Bulos, R. (2016). Personality trait classification of essays with the application of feature reduction. In *Proceedings of the 4th Workshop on Sentiment Analysis Where AI Meets Psychology of IJCAI, 1619*, 22–28.
[26] Xue, X., Feng, J., & Sun, X. (2021). Semantic-enhanced sequential modeling for personality trait recognition from texts. *Applied Intelligence*, *51*, 7705–7717. https://doi.org/10.1007/s10489-021-02277-7
[27] Ramezani, M., Feizi-Derakhshi, M. R., Balafar, M. A., Asgari-Chenaghlu, M., Feizi-Derakhshi, A. R., Nikzad-Khasmakhi, N., . . . , & Akan, T. (2022). Automatic personality prediction: An enhanced method using ensemble modeling. *Neural Computing and Applications*, *34*(21), 18369–18389. https://doi.org/10.1007/s00521-022-07444-6
[28] Wang, Z., Wu, C. H., Li, Q. B., Yan, B., & Zheng, K. F. (2020). Encoding text information with graph convolutional networks for personality recognition. *Applied Sciences*, *10*(12), 4081. https://doi.org/10.3390/app10124081
[29] Ramezani, M., Feizi-Derakhshi, M. R., & Balafar, M. A. (2022). Knowledge graph-enabled text-based automatic personality prediction. *Computational Intelligence and Neuroscience*, *2022*(1), 3732351. https://doi.org/10.1155/2022/3732351
[30] Ramezani, M., Feizi-Derakhshi, M. R., & Balafar, M. A. (2022). Text-based automatic personality prediction using KGrAt-Net: A knowledge graph attention network classifier. *Scientific Reports*, *12*(1), 21453. https://doi.org/10.1038/s41598-022-25955-z
[31] Kerz, E., Qiao, Y., Zanwar, S., & Wiechmann, D. (2022). Pushing on personality detection from verbal behavior: A transformer meets text contours of psycholinguistic features. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis,* 182–194. https://doi.org/10.18653/v1/2022.wassa-1.17
[32] Roy, S. S., Mercer, R. E., & Kundu, S. (2023). Personality trait detection using an hierarchy of tree-transformers and graph attention network. In *36th Canadian Conference on Artificial Intelligence*, 2023L8. https://doi.org/10.21428/594757db.62e766bc
[33] Zhu, Y., Hu, L., Ge, X., Peng, W., & Wu, B. (2022). Contrastive graph transformer network for personality detection. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence,* 4559–4565. https://doi.org/10.24963/ijcai.2022/633
[34] Yang, K., Lau, R. Y., & Abbasi, A. (2023). Getting personal: A deep learning artifact for text-based measurement of personality. *Information Systems Research*, *34*(1), 194–222. https://doi.org/10.1287/isre.2022.1111
[35] Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. *University of Texas at Austin*, *10*, 1–47.
[36] Miller, G. A., & Fellbaum, C. (2007). WordNet then and now. *Language Resources and Evaluation*, *41*, 209–214. https://doi.org/10.1007/s10579-007-9044-6
[37] Kenton, J. D. M. W. C., & Toutanova, L. K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1*, 4171–4186. https://doi.org/10.18653/v1/N19-1423
[38] Asadi, B., & Jiang, H. (2020). On approximation capabilities of ReLU activation and softmax output layer in neural networks. *arXiv Preprint: 2002.04060*.
[39] Tucaković, L., & Nedeljković, B. (2023). From the BFI-44 to BFI-20: Psychometric properties of the short form of the Big Five Inventory. *Psychological Reports*, *128*(2), 1230–1247. https://doi.org/10.1177/00332941231161754
[40] Imambi, S., Prakash, K. B., & Kanagachidambaresan, G. R. (2021). PyTorch. In I. Chlamtac (Ed.), *EAI/Springer innovations in communication and computing* (pp. 87–104). Springer. https://doi.org/10.1007/978-3-030-57077-4_10
[41] Openja, M., Majidi, F., Khomh, F., Chembakottu, B., & Li, H. (2022). Studying the practices of deploying machine learning projects on docker. In *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering*, 190–200. https://doi.org/10.1145/3530019.3530039
[42] Choquette, J., & Gandhi, W. (2020). NVIDIA A100 GPU: Performance & innovation for GPU computing. In *2020 IEEE Hot Chips 32 Symposium*, 1–43. https://doi.ieeecomputersociety.org/10.1109/HCS49909.2020.9220622

How to Cite: Teli, M. A., & Chachoo, M. A. (2025). PersonaG: A Quinpartite Graph Convolutional Network for Interpretable Personality Recognition from Text. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA52024561

Appendices

Table 1
LIWC 2015 categories

Category	Description
Linguistic Processes	
Word count	Total number of words in the text
Analytic	Formal, logical, and structured language
Clout	Confidence, power, and influence expressed in the writing
Authentic	Emotional expressiveness and authenticity
Tone	Overall emotional tone of the text (positive/negative)
Words/sentence	Average number of words per sentence
Words>6 letters	Percentage of words with more than six letters
Dictionary words	Percentage of words in the text found in the dictionary
Total function words	Percentage of function words in the text (e.g., 'the', 'to')
Psychological Processes	
Affective Processes	
Positive Emotion	Words indicating positive emotion (e.g., happy, love)
Negative Emotion	Words indicating negative emotion (e.g., sad, angry)
Anxiety	Words related to anxiety or fear
Anger	Words related to anger or frustration
Sadness	Words related to sadness or grief
Fear	Words related to fear or apprehension
Cognitive Processes	
Insight	Words that show thinking, understanding, or awareness
Causation	Words related to cause and effect
Discrepancy	Words indicating a difference or mismatch from expectation
Tentative	Words indicating uncertainty, hedging, or tentativeness
Social Processes	
Social	Words related to social interaction (e.g., people, friends)
Family	Words related to family life and relationships
Friend	Words related to friendships
Humans	Words related to humans in general
Perceptual Processes	
See	Words related to sight or vision (e.g., see, look, watch)
Hear	Words related to sound or hearing (e.g., listen, hear)
Feel	Words related to physical touch or feelings
Sensory	Words related to sensory experiences (e.g., taste, feel)
Personal Concerns	
Work	Words related to work, career, and profession
Achievement	Words related to personal success or achievement
Leisure	Words related to recreation, leisure, and entertainment
Money	Words related to financial concerns or money
Religion	Words related to religion and spirituality
Death	Words related to death, dying, and mortality
Home	Words related to home, family, or domestic life
Health	Words related to physical health, illness, or wellness
Food	Words related to food, eating, and nourishment
Other Dimensions	
Relativity	Words related to time, space, or perspective
Motion	Words indicating motion or movement
Space	Words related to physical space or location
Time	Words related to the concept of time (e.g., past, present, future)
Verbs	Words that are verbs in the text
Adjectives	Words that are adjectives in the text
Pronouns	Words that are pronouns (e.g., I, you, they, we)
Negations	Words related to negations (e.g., no, not, never)
Certainty	Words indicating certainty or decisiveness (e.g., always, must)
Tentative	Words indicating uncertainty (e.g., might, could, maybe)
Inhibition	Words indicating control, restriction, or self-discipline
Perceptual	Words related to sensory experiences or perceptions
Self-Reflection	Words related to self-reflection, introspection, or self-analysis
Self-Esteem	Words related to self-worth or self-confidence
Ego	Words related to ego, self-centeredness, or narcissism
Religion	Words related to religious beliefs, practices, or terms
Morality	Words reflecting moral or ethical judgment

Table 2
Primary WORDNET Synsets used

Category	Definition
emotion.n.01 (emotion)	Captures words related to emotional states.
feeling.n.01 (feeling)	Refers to subjective experiences and personal emotional reactions.
mood.n.01 (mood)	Represents a person's emotional state over a longer period, such as a general disposition.
social_behavior.n.01 (social behavior)	Describes patterns of interaction with others, including cooperation and communication.
cooperation.n.01 (cooperation)	Words associated with teamwork, collaboration, and mutual effort.
extraversion.n.01 (extraversion)	Reflects sociability, talkativeness, and outward energy in social situations.
cognition.n.01 (cognition)	Captures mental processes such as perception, thinking, and understanding.
reasoning.n.01 (reasoning)	Relates to the logical and rational process of thinking, deduction, and argumentation.
intelligence.n.01 (intelligence)	Describes general cognitive ability, including learning, problem-solving, and comprehension.
personality.n.01 (personality)	Encompasses overall personality traits and characteristics that define an individual.
temperament.n.01 (temperament)	Describes innate emotional and behavioral tendencies, often seen as a foundation for personality.
self_esteem.n.01 (self-esteem)	Reflects an individual's sense of self-worth, confidence, and self-regard.

Table 3
Secondary WORDNET Synsets used

Primary Category	Secondary Synset 1 (Related)	Secondary Synset 2 (Related)	Secondary Synset 3 (Related)	Secondary Synset 4 (Related)	Secondary Synset 5 (Related)	Secondary Synset 6 (Related)
Emotion (emotion.n.01)	joy.n.01 (joy)	anger.n.01 (anger)	fear.n.01 (fear)	sadness.n.01 (sadness)	trust.n.01 (trust)	love.n.01 (love)
Feeling (feeling.n.01)	warmth.n.01 (warmth)	coldness.n.01 (coldness)	affection.n.01 (affection)	disgust.n.01 (disgust)	envy.n.01 (envy)	happiness.n.01 (happiness)
Mood (mood.n.01)	melancholy.n.01 (melancholy)	optimism.n.01 (optimism)	pessimism.n.01 (pessimism)	joyfulness.n.01 (joyfulness)	despair.n.01 (despair)	excitement.n.01 (excitement)
Social Behavior (social_behavior.n.01)	cooperation.n.01 (cooperation)	aggression.n.01 (aggression)	conformity.n.01 (conformity)	altruism.n.01 (altruism)	shyness.n.01 (shyness)	assertiveness.n.01 (assertiveness)
Cooperation (cooperation.n.01)	collaboration.n.01 (collaboration)	negotiation.n.01 (negotiation)	competition.n.01 (competition)	teamwork.n.01 (teamwork)	collective.n.01 (collective)	mutual_aid.n.01 (mutual aid)
Extraversion (extraversion.n.01)	sociability.n.01 (sociability)	assertiveness.n.01 (assertiveness)	talkativeness.n.01 (talkativeness)	leadership.n.01 (leadership)	enthusiasm.n.01 (enthusiasm)	openness.n.01 (openness)
Cognition (cognition.n.01)	reasoning.n.01 (reasoning)	thinking.n.01 (thinking)	knowledge.n.01 (knowledge)	memory.n.01 (memory)	perception.n.01 (perception)	concentration.n.01 (concentration)
Reasoning (reasoning.n.01)	deduction.n.01 (deduction)	intuition.n.01 (intuition)	logic.n.01 (logic)	analysis.n.01 (analysis)	judgment.n.01 (judgment)	argumentation.n.01 (argumentation)
Intelligence (intelligence.n.01)	wisdom.n.01 (wisdom)	insight.n.01 (insight)	cleverness.n.01 (cleverness)	brilliance.n.01 (brilliance)	acumen.n.01 (acumen)	problem_solving.n.01 (problem-solving)
Personality (personality.n.01)	temperament.n.01 (temperament)	traits.n.01 (traits)	disposition.n.01 (disposition)	character.n.01 (character)	individuality.n.01 (individuality)	identity.n.01 (identity)
Temperament (temperament.n.01)	nature.n.01 (nature)	constitution.n.01 (constitution)	personality_type.n.01 (personality type)	disposition.n.01 (disposition)	moodiness.n.01 (moodiness)	volatility.n.01 (volatility)
Self-Esteem (self_esteem.n.01)	confidence.n.01 (confidence)	self-worth.n.01 (self-worth)	self-respect.n.01 (self-respect)	self-assurance.n.01 (self-assurance)	pride.n.01 (pride)	vanity.n.01 (vanity)