SN Computer Science (2024) 5:712
https://doi.org/10.1007/s42979-024-03055-1

ORIGINAL RESEARCH

Behavior Based Group Recommendation from Social Media Dataset by Using Deep Learning and Topic Modeling

Md. Saddam Hossain Mukta[1] · Jubaer Ahmed[2] · Mohaimenul Azam Khan Raiaan[2] · Nur Mohammad Fahad[2] · Muhammad Nazrul Islam[3] · Nafiz Imtiaz[3] · Md. Adnanul Islam[4] · Mohammed Eunus Ali[5] · Sami Azam[6]

Received: 6 November 2023 / Accepted: 13 June 2024
© The Author(s) 2024

Abstract

In this digital era, users frequently share their thoughts, preferences, and ideas through social media, which reflect their Basic Human Values. Basic Human Values (aka values) are the fundamental aspects of human behavior, which define what we consider important, and worth having and pursue them. Existing studies identify the values of individuals from different social network usages such as Facebook and Reddit. However, discovering the similarity (or diversity) of value priorities among the members in a group is important since we can reveal many interesting insights such as finding a set of target customers, identifying the chain of misdeed groups, searching for similar acquaintances in workplaces, etc. In this paper, a graph dataset is compiled using the strongest correlation among the features and then we apply a graph clustering technique to identify a suitable hedonist group (i.e., one dimension of values) for users' recommendations. Then, we also propose a behavior based (i.e., value) group recommendation technique by analyzing users' contextual and psychological attributes. Finally, we validate those group members in real life by introducing two hypotheses. In particular, we analyze the tweets of a total of 1140 users collected from Twitter. We obtain a substantial *intra-cluster correlation coefficient (ICC)* and *silhouette clustering coefficient (SCC)* scores of 65% and 76%, respectively, among the members in our discovered group.

Keywords Psychological attributes · Values · Twitter · Graph neural network (GNN) · Social media · Group recommendation · Topic modeling

Introduction

Basic Human values (*values* in short) [1, 2] are important attributes of human behavior that distinguish individuals from one another. These attributes play a crucial role in our decision-making process and influence our personal and social life. In this digital era, people express their thoughts, ideas, and opinions with their family, friends, and acquaintances through social media platforms like Facebook, Twitter [3, 4], and Reddit [5]. In this paper, we identify group of users who have similar value dimensions from their social media interactions and validate how the similarity in value dimensions can lead to perform identical actions in real life.

Social media content has been widely used to discover diverse patterns of human behavior [6–9] such as values [10, 11], personality [12], and self-efficacy [13]. Prior studies [10, 14] demonstrate that it is possible to recommend based on individual user's behavior (i.e., values, personality) by observing their social media interactions. Several studies [15, 16] also identify group of users based on certain aspect/feature to

✉ Md. Saddam Hossain Mukta
saddam.mukta@lut.fi

1 LUT School of Engineering Sciences, Lappeenranta-Lahti University of Technology, Lappeenranta, Finland

2 Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh

3 Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh

4 Department of HCC Action Lab, Monash University, Melbourne, Australia

5 Department of Computer Science and Engineering Bangladesh, University of Engineering and Technology, Dhaka, Bangladesh

6 Faculty of Science and Technology, Charles Darwin University, Darwin, NT, Australia

Published online: 16 July 2024

find target audience for recommendation. In majority of the group recommendation based studies [17–19], authors find groups based on structural properties, social ties, interpersonal relationship, information propagation, etc. The existing studies largely lack the aspect of identifying groups based on human psychological attributes (e.g., value dimensions). For example, Bisgin et al. [20] analyze users' characteristics that construct a group based on social ties. Our study largely identifies the research gap and determines the group of users by using their social media interactions who possess similar behavioral traits, i.e., value dimensions.

In our study, we randomly select 1140 Twitter users and compute values by using IBM personality insight API.[1],[2] We mainly perform two steps. First, we discover a group of users who share similar values (i.e., hedonism dimension). Hedonism value refers to pleasure and sense of gratification of an individual [21]. Second, we validate the group members who have similar preferences in real life. In the first step, we identify group of users by using two approaches, (1) *graph-based* group identification from *hedonism value* scores (GHV) by using graph neural network (GNN) [22] and (2) *contextual-psychological* based *hedonism value* scores (CPHV). We initially identify group of users, $G(h)$, based on their interrelated value scores. These users have similar value scores within a threshold of δ_d and $|h_i - h_j| \leq \delta_d$ where h_i and h_j indicate value scores of two different users. Based on the similarity scores, we construct graphs and identify groups by using GNN. In our second approach of group identification, we exploit transformer and topic based techniques to discover contextual meaning of the content. Thus, we apply both latent semantic analysis (LSA) [23] and bidirectional encoder representations from transformers (BERT) [24] to find a set of similar users who write similar topics. In addition with the contextual patterns, we adopt a psychological lexicon based approach namely, linguistic inquiry and word count (LIWC) [25] over users' writing content. We discover a more compact set of user group who have similar value dimensions. Similarly, in our second step (validation), we propose two hypotheses to validate whether these extracted group members have similarities in their real life preferences. Finally, we compare our proposed approach with popular recent group recommendation models: COM [26], HBGG [27], GGC [28], and HGGC [28]. Our approach shows better performance than that of the state-of-the-art approaches. Identifying a group of similar users from social media interaction can have several real-life applications. For instance, finding a target customer set for an existing or upcoming product. Legal agencies and government can find chain of misdeed groups, privacy-based social network detection [29], a user can find similar acquaintances in workplaces, observing the social dynamics (e.g., group behavior), parental approval in friends searching, and accepting friend requests [30, 31].

In a nutshell, our contributions of this study are as follows:

- We build a new dataset from Twitter and construct a graph dataset for grouping users based on their value (i.e., hedonism) scores.
- We apply a novel fusion model incorporating graph neural networks (GNNs) and spectral clustering for effective group identification, i.e., GHV.
- We further apply another method of group identification by discovering users' identical contextual and psychological patterns derived from tweets, i.e., CPHV.
- We also propose an evaluation method to validate the values of the group members in real life.
- We conduct a comparison with the existing models to evaluate the efficacy of the proposed approaches, i.e., GHV and CPHV.

We organize the rest of the paper as follows. We discuss previous works related to group recommendation in "Related work". Then, we present data collection and methodology in "Data collection" and "Methodology", respectively. Later, we describe about the data preprocessing and value computation process by IBM Personality Insight API in "Pre-processing and computation of users' value scores". Then, we present our group identification techniques in "Group identification". Next, we validate our users by using hypotheses in "Validation". We compare our method with the state-of-the-art techniques in "Comparison with state-of-the-art group recommendation models" and discuss our findings in "Discussion". In "Conclusion", we conclude the paper.

Related Work

In the current literature, several studies address different psychological attributes, i.e., values [32], personality [33], mood [34], emotion [35, 36], and sentiment [37, 38] from users' social media interactions.

Chen et al. [10] conduct a study to predict individual's values from their word usage pattern using Reddit's content. This study uses Linguistic Inquiry and Word Count (LIWC)[3] to extract features from user's posts as independent variables and their values as dependent variables. On the other hand, Bisgin et al. [20] analyze users' tweets to investigate how

[1] https://cloud.ibm.com/apidocs/personality-insights.

[2] Note that IBM personality-insight system infers user intrinsic personality characteristics, including Big5, Needs, and *Values* based on social media content, email and text messages.

[3] https://liwc.wpengine.com/ [39].

group of similar users exists in social media based on people's interest by analyzing their posts, and comments. They use clustering algorithm to find out users' interests in a group from social media. Similarly, Yen et al. [40] explore student groups in student-student and student-teacher relationships in an online discussion by using text-mining approaches. They use *K-means* clustering and find a significant amount of students group who share the same interests.

Zhang et al. [41] use a novel structure, attribute, and homophily preserved (SAHP) algorithm, which uses three information within a connection space, such as network structure, node attributes and homophily for learning a network representation thus training a vector representation for a social network. In another study, Solomon et al. [42] perform a study to find out the behavioral similarities of users by analyzing the personality and values of twitter users using the Big5 personality model. In this work, link prediction is performed using three broad genres for community detection, such as: (1) similarity-based approach which is done by calculating the cosine similarity between nodes based on their psychological and sociological attributes, (2) path-based approach that uses the node2vec model and (3) learning-based approach, by creating a node2vec link prediction system which uses the path-based approach.

The study in [43] explores group in social networks where they find out the moral similarities which connect the communities. Besides, Yohsuke et al. [44] propose a model to investigate how group influences the social network. Hammou et al. [45] propose a system to handle a huge scale of data for distributed group recommendations based on Apache Spark. They do not apply any group recommendation technique based on users' soft attributes such as values or personality. Jeong et al. [28] propose a group recommendation model where they consider group cohesion as cluster forming measure.

In the light of above discussion, Solomon et al. [42] show that friends can only form network of similar users based on their Big5 personality and values, while Mukta et al. [12] show that users with close Big5 personality scores likely to fall under the same homophily. Both of these studies report that users' group cannot be identified from an open network by using a clustering technique. Previous studies [12, 42] need self reported survey tests in both group formation and validation steps. In our study, we reduce the effort, yet we devise a potential technique with the help of users' digital footprint extracted from social media interactions.

Several studies [46, 48] introduced a graph approach for group recommendation. Wang et al. [46] introduce the deep adaptive collaborative graph neural network for social recommendation framework to address certain limitations observed in GCN-based recommender systems. Li et al. [48] attempted to resolve the challenges of occasional group recommendation by introducing the Self-Supervised Group Graph Collaborative Filtering (SGGCF) model. This model addresses the complex interactions between users, items, and groups while efficiently handling data sparsity issues. Zhou et al. [47] introduce a multi-graph neural model that incorporates meta-learning and multi-teacher distillation techniques to address challenges in group recommendation.

In the light of above discussion, our approach first proposes a fusion model merging GNNs and spectral clustering for a cognitive approach (i.e., value) based group recommendation. Our study also considers users' contextual and psychological attributes for personalized suggestions and validates our approach with real-world testing, enhancing credibility. Our study stands out by combining diverse techniques for more effective group recommendations, addressing prior graph-based approaches' limitations. Table 1 provides a comparative analysis of our approach with several state-of-the-art group recommendation methods.

Data Collection

Our data collection process consists of two steps. First, we collect tweets from a set of selected users for identifying groups based on values dimensions. Then, we also collect data for validating the identified group in previous step for a specific value dimension, i.e., *hedonism*. Value dimensions proposed by Schwartz can be presented into five higher-level value dimensions (see Fig. 5). We observe that users are likely to publish posts in social media which are closely related to hedonism [49]. Therefore, we only demonstrate group identification and validation for hedonism value dimension due to the availability of data on social media platforms. For example, traveling to a destination, purchasing an expensive gadget, owning a luxurious car, visiting a lavish restaurant, etc. [50]. We find several studies that show that users share their such posts. However, the same method can be applicable to identify and validate other value dimensions. We only focus on English content in our analysis. Thus, we extract tweets from the United States and the United Kingdom to confirm users' English language proficiency. To collect tweets we used *tweepy*[4] Python implementation package. We describe the two steps as follows.

Dataset for Group Identification

For group identification task, as reported in Fig. 1, we selected 1140 active unverified Twitter users and collected a total of 3,275,417 tweets posted by these users. Table 2 presents the statistics of our group identification dataset. We also refer this dataset as our training set from which

[4] https://www.tweepy.org/.

Table 1 Comparative analysis of reviewed papers

Study	Contributions	Psychological attribute	Content-based graph (GCN)	Validation approach	Use of deep learning
Yen et al. [40]	Identifying student groups	No	No	No validation of student groups	No
Zhang et al. [41]	Learning network representation for recommendation	No	No	No validation of network representation	No
Solomon et al. [42]	Link prediction for personality traits	Big5 Personality Traits	No	Link Prediction	Limited exploration of user values
Hammou et al. [45]	Distributed group recommendation	No	No	No validation of group recommendations	No
Jeong et al. [28]	Group recommendation based on cohesion	No	No	No validation of group cohesion	No
Wang et al. [46]	Adaptive graph neural network for social recommendation	No	GCN	No validation of network representation	Yes
Zhou et al. [47]	Multi-graph neural model for group recommendation	No	GCN	No validation of group recommendations	Yes
Our approach	Fusion of GNNs and spectral clustering for value based group recommendation	Cognitive attributes, i.e., Schwartz value dimension	GCN	Validation with real data	Yes

Table 2 Statistics for the dataset developed for group identification

Description	# of items
Number of users	1140
Number of Tweets	3,275,417
Maximum number of Tweets per user	3250
Minimum number of Tweets per user	29
Number of profession	31
Number of countries	0

Table 3 Statistics for group validation dataset consisting of two hedonism characteristics (i.e., movie and technology enthusiasts)

Description	Movies enthusiast:(#)	Tech. enthusiast:(#)
Number of users	150	150
Number of Tweets	449,092	391,416
Max. # of Tweets per user	3200	3219
Min. # of Tweets per user	112	84
Number of professions	19	13
Number of countries	2	3

we extract users' different linguistic and psychological patterns. Since we collect users' pleasure seeking behavior, i.e., hedonism, we observe that we can easily extract digital footprints of these two categories from Twitter. We could also collect users' food related data from their *check-ins* of different restaurants. Collecting those data from user's check-ins poses dependency on third party apps, i.e., *Swarm*.[5]

Dataset for Group Validation

To validate our findings by using our group identification dataset, we collect another set of users who frequently post on "movie" and "gadget" topics in Twitter. We collect a total of 840,508 tweets from 300 Twitter users (for each of the keywords, we collect 150 users).

1. Gadget enthusiasts: iphone, gadgets, apple, samsung, oppo, technology, ipod, artificial intelligence, smartphone, smartwatch, ai google, android, and gaming
2. Movie enthusiasts: movie, comedy, action, dc, marvel, superhero, show, watching, enjoying, avengers, funny, and movienight

The number of tweets in the first set consists of 391,416 from 150 users, and the number of tweets in the second set

[5] https://www.swarmapp.com/.

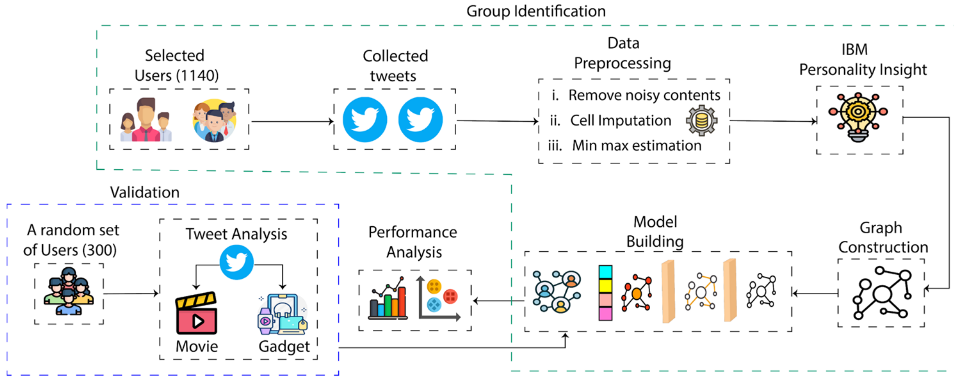


Fig. 1 A complete pipeline of group identification and validation process

Table 4 Values distributions among the users

Values	Min	Max	Mean	Std. dev.
Self-transcendence	0.0106	0.9219	0.2733	0.1779
Self-enhancement	0.0112	0.9038	0.2447	0.1629
Conservation	0.0181	0.8804	0.3978	0.1860
Hedonism	0.0961	0.9989	0.6529	0.0513
Openness	0.0410	0.9978	0.4810	0.188

is 449,092 from the same number of users. In Table 3, we report the distribution of this dataset.

Methodology

Figure 1 presents different steps of our methodology. We briefly explain all the steps as follows:

Pre-processing and Computation of Users' Value Scores

To remove noisy conent from tweets, we preprocess the tweet text by filtering links, non-Ascii characters, words in a different language other than English, stops words, words with numbers, and mentions. For raw data preprocessing, we discard the username and mentions. We keep the retweets of users because the act of retweeting could indicate the behavioral pattern of users. We remove hashtags (eg, "#weStand,") and converted them into text (eg, "we stand").

To understand different patterns of user groups, we use IBM Personality Insight API and compute users' value scores from their tweets. Since Arnoux et al. [51] show that IBM Personality Insight API performs better that any other techniques and different prior studies [52] also used the API and achieved reasonable performance. For every value, the service provides a score (0–1) for each value dimension. We prepare the data set with each user having the five

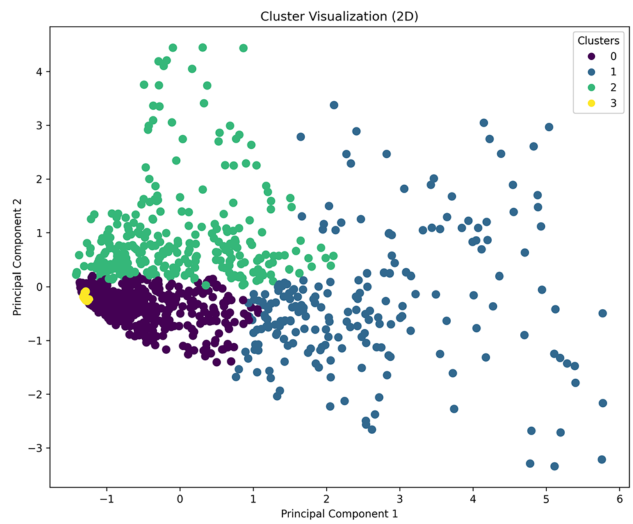


Fig. 2 Distribution of users' values scores of different dimensions of 4 cluster

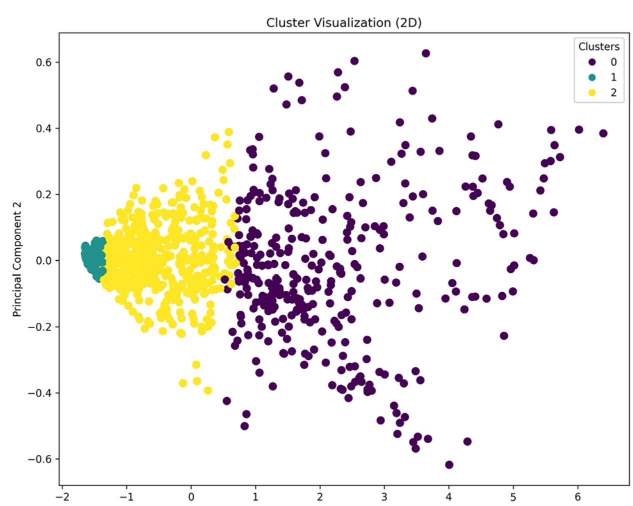


Fig. 3 Distribution of users' values scores of different dimensions of 3 cluster

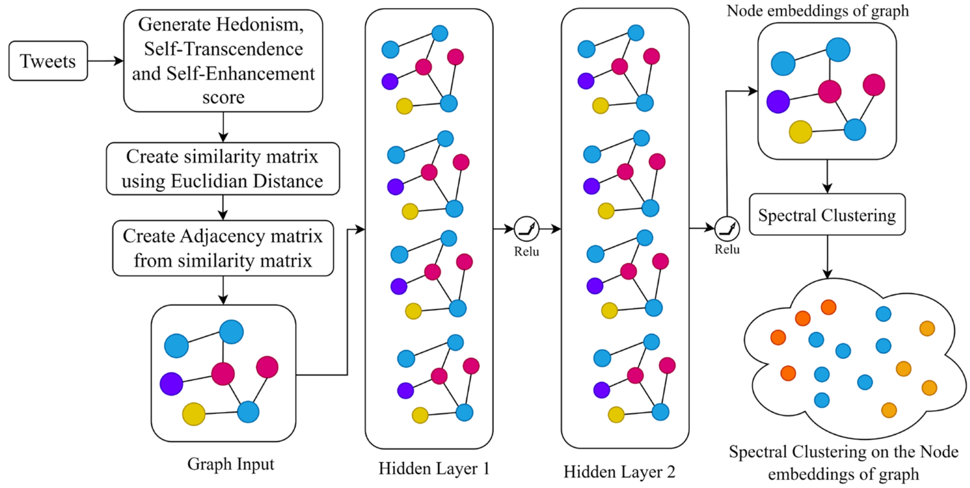


Fig. 4 Graph architecture of group recommendation

values dimensions: *conservation*, *openness to change*, *self-enhancement*, *self-transcendence* and *hedonism*.

In Table 4, we present the distributions of all users value scores that we compute through IBM personality insight API. The table shows users' minimum, maximum, mean and standard deviation of scores of five different value dimensions. We find that hedonism has the highest mean score (69.59%) and for other value dimensions have an average mean score of 34.92%. Users have an average standard deviation of 17.85%. Figures 2 and 3 show distribution of different value scores for 1140 users.

Group Identification

For group recommendation, we follow two different approaches.

- *Group identification based on hedonism scores (GHV)* In the first approach, we clustered the self-transcendence with hedonism and alternatively self-enhancement with hedonism to derive users' groups. We justify later why we choose the above combinations of values with hedonism.
- *Group identification from contextual and psychological attributes (CPHV)* In our second approach, we consider users' contextual and psychological attributes to find an accurate group of users who have similar values.

We discuss these two approaches in the following subsections:

Graph Based Group Identification from Hedonism Scores (GHV)

We investigate the variant effects on hedonism score from other value dimensions (i.e., self-enhancement and self-transcendence) in user group formation by using the fusion of *GNN* and *spectral clustering* techniques. Following the study [53], we select a combination of dimension: hedonism, self-transcendence and self-enhancement (HSTSE) to identify group using hedonism score. We also give a detail explanation for combining these dimensions to compute hedonism score in "Group identification from contextual and psychological attributes (CPHV)".

In our study, we construct a graph using a dataset of tweets, focusing on the above three distinct features, i.e., HSTSE. For constructing the graph, connections between nodes are established based on their similarity values. Specifically, nodes are linked if their similarity value exceeds 0.7, motivated from the study [54]. The computation of similarity involves the following procedure:

$$similarityMatrix = \exp\left(-1 \cdot \|X[:, \mathbf{None}] - X\|_2\right)$$

where X represents the normalized data values. The $\|\cdot\|_2$ denotes the Euclidean norm, and $[:, \text{None}]$ is used to broadcast the subtraction operation across the array dimensions. The result of the subtraction is exponentiated using the exponential function exp.

These features exhibit a robust correlation, making them effective for group identification. To optimize this process, we propose a fusion of GNNs [55] and spectral clustering [56]. GNNs are like smart learners that understand how things connect in a graph. They learn subtle patterns in the graph. We use what they learn to help us organizing the data using spectral clustering. This helps us to deal with the problems that come with lots of data. By making the data simpler, we can see the

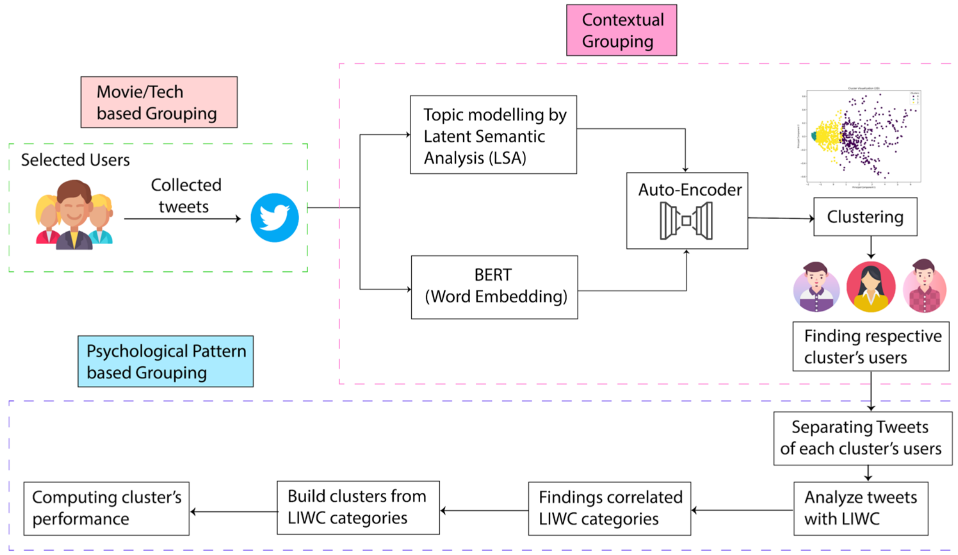


Fig. 5 Contextual-psychological based group recommendation technique

important parts of the graph better and understand it more easily. By integrating GNNs into our approach, we enhance the node representations utilized during spectral clustering. This improvement helps us divide things in the graph into smaller groups. It uses the patterns in the graph to do this better. Because of this teamwork, we can group things more accurately, and that gives us better results. The workflow of this clustering process is depicted in Fig. 4.

GNNs have gained attention for their ability to work with graph-structured data. A GNN processes a graph as a series of layers, each of which refines node representations based on information from neighboring nodes. At the core of a GNN is the graph convolution operation, which mathematically updates the representation of a node $h_i^{(l+1)}$ in layer $l+1$ as follows:

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in N(i)} \frac{1}{c_i}\mathbf{W}^{(l)}h_j^{(l)}\right) + \mathbf{W}^{(l)}h_i^{(l)} + \mathbf{b}^{(l)}$$

where $h_i^{(l+1)}$ is the updated representation of node i at layer $l+1$. $N(i)$ represents the neighborhood of node i. σ is an activation function. c_i is the normalization factor for node i. $\mathbf{W}^{(l)}$ is the weight matrix at layer l. $\mathbf{b}^{(l)}$ is the bias term at layer l.

Then we apply spectral clustering for grouping nodes in a graph by leveraging the graph's Laplacian matrix. Given an affinity matrix $\mathbf{A}$, which encodes pairwise relationships between nodes, the normalized graph Laplacian matrix is defined as:

$$\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$$

where $\mathbf{D}$ is the degree matrix, a diagonal matrix with node degrees on the diagonal. Spectral clustering involves computing the eigenvalues and eigenvectors of the normalized Laplacian matrix. The eigenvectors corresponding to the k smallest eigenvalues form a low-dimensional embedding of the graph. Subsequently, traditional clustering algorithms like K-means can be applied to this embedding for clustering.

This process essentially extends the same methodology used for clustering to a single user. The difference is that we are computing embedding and similarity scores for a single user's tweets instead of the entire dataset. To generate node embedding, begin by extracting the characteristics (hedonism, self-transcendence, self-enhancement) from the tweets of the individual user. Lastly, we conduct spectral clustering utilizing the embeddings of both the original nodes and the newly introduced node for the individual user. This process allows us to assign the single user's tweets to the existing clusters based on their similarity to the cluster centroids.

Group Identification from Contextual and Psychological Attributes (CPHV)

In this group identification technique, we again discover a group of hedonist users. We consider users' contextual patterns and psychological attributes from their writing content. We follow three different steps: (1) feature filtering for correlated values scores, (2) conducting contextual analysis of users' writing content and (3) analyzing users' psychological attributes. Figure 5 presents the detailed steps of computing group identification based on contextual and psychological lexicon based approaches. To the best of our knowledge, our study first proposes such a technique where a sequence of contextual and psychological patterns have been adopted to discover a group of users based on their values dimensions.

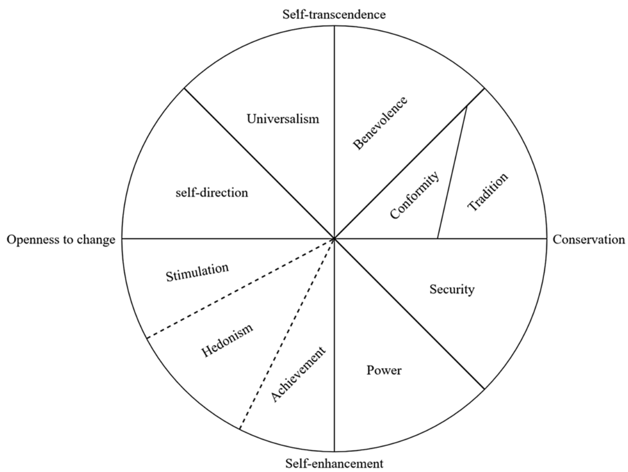


Fig. 6 Structure of Schwartz value dimension

Feature filtering from correlated value dimensions: We determine an initial set of users by filtering correlated value scores of 1140 users. Following the socio-psychological study of Ma et al. [57] *self-enhancement* and *hedonism* (i.e., happiness, joy and cheer) values are aligned with the same direction, while *self transcendence* (i.e., helping others) lies to the opposite orientation. Schwartz [58] in his seminal work also presents that *self-enhancement* and *hedonism* value dimensions both are positively correlated with each other.

Schwartz also finds that some value dimensions show conflicting properties to each other. He further finds reason that simultaneous pursuit of values may reveal conflicting value dimensions. Figure 6 shows a circle of value structure where diagonal value dimensions show negative correlation (i.e., hedonism and self-transcendence). Motivated by the above findings, we initially filter a set of users for discovering further clusters of hedonist users by setting ranges of values scores (according to Table 5). To this end, we first find the users who have high hedonism and self-enhancement scores. At the same time, these individuals have low self-transcendence scores. Among our selected 1140 random users, we find our initial set of users whose value scores are selected from the three value dimensions (*self-enhancement, hedonism* and *self-transcendence*). Then we filter out our extracted users and obtain a total of around 75 users who follow the constraints that we discover from our socio-psychological studies [58, 59]

Table 5 presents different ranges of value dimensions heuristically for selecting initial users. From our dataset, we discard categories of 1, 4, 5 and 6 due to poor number of user selection (according to Table 5). We mainly focus on the categories of 2 and 3 for moderate amount of user selection (around 75 in numbers).

Contextual clustering among the hedonistic users: We further build compact clusters of users from the previously

Table 5 Users selection criteria for different ranges of hedonism, self-enhancement and self-transcendence dimensions

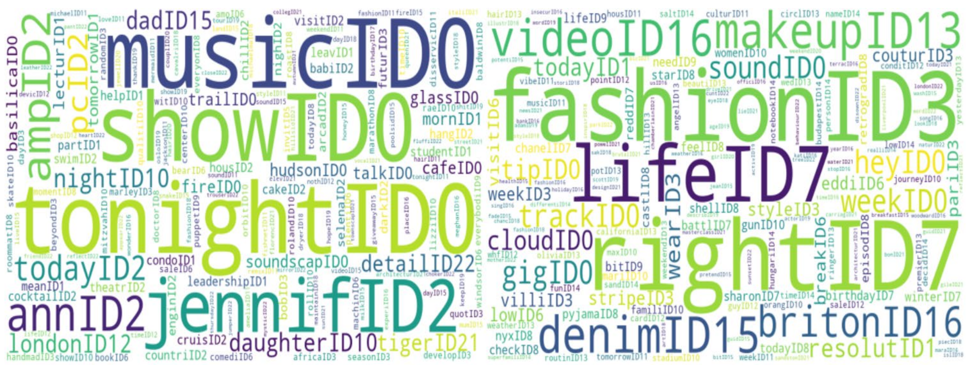
Cat.	Hedonism	Self-enhan.	Self-tran.	# of selected users
1	>0.75	>0.6	<0.5	53
2	**>0.70**	**> 0.55**	**<0.5**	**73**
3	**>0.6**	**<0.55**	**<0.5**	**75**
4	>0.75	>0.7	<0.5	29
5	>0.75	>0.7	<0.6	41
6	>0.75	>0.75	<0.7	39

The best-selected values are in bold

selected 75 users. For identifying important context from these users' tweets, we use topic modeling by using *latent semantic analysis* (LSA) [60]. Over these topics, we apply *bidirectional encoder representations from transformers* (BERT) [61] to embed topics into a vector space where the vectors capture the contextual meaning of the sentences. Figure 7 presents important topics who have high hedonism and self-enhancement scores and weak self-transcendence score by using our encoder based technique. We observe major shared topics by tweets are: *music, tonight, cafe, video, fashion,* etc. which likely reveal users' happiness and enjoyment in life.

Psychological lexicon based approach: In this section, we compress our users' cluster than previous step by using contextual topics. Towards this direction, we conduct a psychological based analysis over the tweets of these users of groups. Thus, we analyze these users' tweets by using Linguistic Inquiry and Word Count (LIWC) by using LIWC 2015. LIWC2015 identifies approximately 90 different features from texts into 7 different categories, where each category contains hundreds of words [58]. These categories include summary language variables (analytical thinking, clout, authenticity, and emotional tone), general descriptor categories (words per sentence, percent of target words captured by the dictionary, etc.), standard linguistic dimensions (articles, auxiliary verbs, etc.), word categories tapping psychological constructs (affect, cognition, etc.), personal concern categories (work, home, leisure activities, etc.), informal language markers (assents, fillers, swear words, netspeak) and punctuation categories (periods, commas, etc).

Later, we compute correlation between users' LIWC categories and their corresponding hedonism scores. We only select the relevant features which are significant where p value 0.05. We find that different numbers of LIWC categories are correlated. Later, we compute clustering technique over these data by using a density based clustering [62], because we do not want to mention a predefined number of clusters. Based on the density, we make



Word clouds from the tweets of hedonist user group

Word clouds from the tweets of another hedonist user group

Fig. 7 Word clouds extracted from contextual analysis (i.e., LSA and BERT Word embedding) over users' tweets

Table 6 LIWC based cluster description

Cat.	# of clusters	Cluster name	Size of the cluster (users)	SCC	Selected cluster	# of selected LIWC categories
2	3	Cluster 1	27	0.51	Cluster 2	11
		Cluster 2	26	0.553		12
		Cluster 3	20	0.486		4
	4	Cluster 1	19	0.699	Cluster 1	10
		Cluster 2	21	0.627		23
		Cluster 3	15	0.578		9
		Cluster 4	18	0.443		21
3	**3**	**Cluster 1**	**28**	**0.69**	**Cluster 1**	**11**
		Cluster 2	23	0.518		17
		Cluster 3	24	0.565		2
	4	Cluster 1	20	0.56	Cluster 2	16
		Cluster 2	16	0.757		12
		Cluster 3	18	0.58		12
		Cluster 4	21	0.70		4

The best-selected values are in bold

Table 7 Selected LIWC categories for different cluster categories

Cluster category	Selected cluster	LIWC categories
2	2	Unique, funct, i, you, past, social family, pronoun, ppron, swear, present, verb
	1	Percept, funct, body, pronoun, ppron, verb cogmech, social, ipron, present
3	1	**Posemo, home, insight, achieve, family, swear health, friend, leisure, present, money**
	2	Dic, funct, past, you, family, swear cogmech, adverb, ppron, health, future, pronoun

The best-selected values are in bold

separate clusters by using its silhouette coefficient [63]. Table 6 presents cluster category (i.e., configurations of values), number of clusters, cluster name, SCC values, selected clusters and corresponding # of correlated LIWC categories. From the above configurations (for each category two cluster types: 3 and 4), we select the best cluster which is the more compact version of our previous clusters

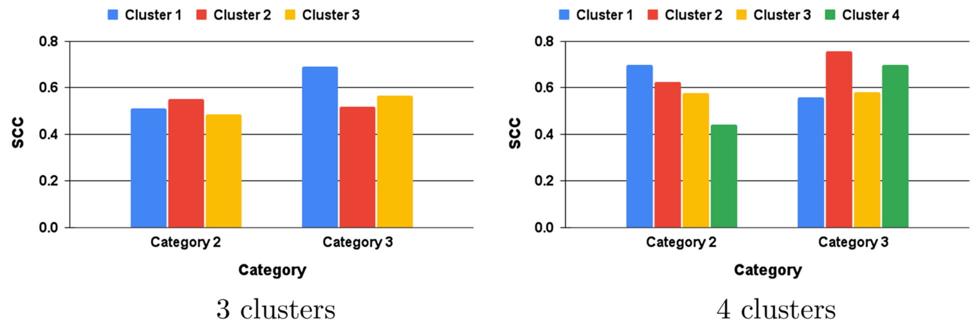


Fig. 8 SCC scores for each generated cluster in each user selection category

Table 8 Sample tweets of our discovered group: (i) movie lovers, and (ii) technological enthusiasts

Tweets of movie lovers	Tweets of tech. enthusiasts
Meryl Streep, one of the all-time greats. What's her best performance to date?	Let's work on 32 films
Symas Alexa Google Assistant has pretty good Scottish accents. I can change that	I remember watching it a hundred times a day. I still get chills. My testing framework is currently on top of /r/Python!
So, what's the best Batmobile chase sequence, huh? #TheBatman	This one has the coolest game mechanics ever
Obviously, I care enough about classic Disney films to the point where I have issues	It's a great honor to be one of the top #ArtificialIntelligence influencers in 2020. #AI #MachineLearning
I've actually never seen Scary Movie 3, but I've seen it still	Our biggest fans of the week: pbyrond, PloybCG, iphones4sale145. Thank you!

where users are similar to each other in terms of hedonism score.

Table 7 presents correlated LIWC categories for the selected clusters for category 2 and 3 for their number of clusters of 3 and 4. However, we finally selected for our final recommendation as category 3 and cluster 1 (# of users-28) which is marked as bold in Table 6. From our LIWC features, we find that these categories are highly intuitive [64–66] with a user's happiness such as positive emotion, friends, leisure times, money, achievement, good health and so on. The summary of the results in Table 6, i.e., SCC score of each cluster per category can be better visualized in Fig. 8 which shows that the average SCC score for category 3 is higher than that of category 2. Thus, category 3 (in Table 5) is our preferred configuration for choosing users.

Validation

To justify our findings, we address the following two hypotheses. The reason for selecting these two hypotheses is that individuals usually post about their favorite movies and gadgets frequently in Twitter.

- H1: People with high hedonism scores are likely to be highly interested in using gadgets, i.e., tabs, smartphones, etc [67].
- H2: People with high hedonism scores have strong interest in watching movies [67].

Table 9 ICC for users with different interests

Domain	Hedonistic	Movie lovers	Tech. enthusiasts
ICC	0.440	0.653	0.615

We validate these two hypotheses by using the group validation datasets which are discussed as follows.

Lexical Analysis Over the Tweets of Tech Enthusiasts (H1)

To verify H1 hypothesis, we analyze the tweets and users that we collect by using gadget keywords (see "Dataset for group validation"). We find that these 150 users prefer topics like *innovations, gadgets, science,* and *technology*. Table 8 presents several tweets, which reveal users' propensity related to technological issue. Some users become happy by watching gaming devices. For example, a sample tweet supporting this observation is *"This one coolest game mechanics ever ( )"*. Some users like artificial intelligence and machine learning related content. A related tweet supporting this claim is *"Great Honor Top #ArtificialIntelligence Influencers 2020 #AI #MachineLearning"*. Sometimes users share their excitement while experimenting with a programming framework. A sample tweet related users' excitement is *"I think great. And like right Framework (.NET)"*. These users tend to be fascinated by observing the gadgets, high-tech products, their versions, etc.

Table 10 SCC scores for different numbers of clusters

# of clusters	SCC score (hedonists)	SCC score (tech enthu.)	SCC score (movie lovers)
2	0.6415	0.5315	0.5484
3	0.6007	0.5727	0.5506
4	0.5982	0.5413	0.5467
5	0.5881	0.5720	0.508

Table 11 SCC and ICC scores for different models

Models	SCC (20)	SCC (12)	ICC (20)	ICC (12)
COM [69]	0.38	0.47	0.33	0.36
HBGG [27]	0.30	0.34	0.24	0.27
GGC [28]	0.43	0.49	0.35	0.37
HGGC [28]	0.60	0.63	0.51	0.56
GHV	0.64	0.64	0.44	0.44
CPHV	0.70	0.76	0.58	0.60

Lexical Analysis Over the Tweets of Movie Lovers (H2)

We address our H2 hypothesis by using our another dataset of 150 Twitter (see "Dataset for group validation"). Table 8 also shows several sample tweets which express users' inclination related to movies. Some users prefer to watch movies repetitively with a strong gratification. A representative tweet is *"I remember watching hundred times day. Still get chills"*. We also observe that some Twitter users also convey their self-contentment watching a movie and want to reflect the movie events in their lives. A sample tweet that supports our claim is *"The last super hero movie I saw full Endgame theaters. I 'm sweets, need go gym"*. We also find that some users express complete satisfaction after watching a movie. A sample tweet is *"YES ! Loved movie. One fave Oscar runs !"*.

Tables 9 and 10 show (see columns 3 and 4) that both tech enthusiasts and movie lovers have near similar hedonism scores for different number of clusters.

Implication of Users' Values and Their Preferences

In this subsection, we make a connection between a group of users' value dimension and their real life preferences. We use *IBM personality insight API* to get the raw scores for the group validation dataset. From the collected ratings, we find that most of the users scored high in hedonism with an average of 0.71 for movie lovers and 0.70 for tech enthusiasts. To precisely, Eq. (1) presents the validation process of identifying group of similar hedonist users from their tweets and their preferences:

$$hed(IBMapi(movie\_lovers(t) \vee tech\_enthu(t))) \implies grp(IBMapi(t)_{hed}) \quad (1)$$

In Eq. (1), *grp* refers to a function for building *groups*, *t* indicates *tweets*, and *hed* represents function for computing hedonism score. Left side of the equation presents the group formation process of similar hedonistic users who have approximate value scores of 0.70 by using IBM personality insight API derived from their tweets. Right side of Eq. (1) describes that hedonism scores derived from the tweets of movie lovers or tech enthusiasts also produce similar score, i.e., 0.70, by using IBM personality insight API derived from tweets. Then we compute *Z test* [68] between the *hedonism* scores of these two samples where variance (σ^2) is 0.03. We obtain a *Z test* value of 1.35[6] which also represents that hedonism scores of these two users' samples are similar. Thus, it is likely that the users who have high hedonism scores usually have preferences of watching movies or inclination of using tech products.

Comparison with State-of-the-Art Group Recommendation Models

Next, we show comparative analysis of our proposed approaches with several popular state-of-the-art approaches. We refer to hedonism and contextual-psychological based group identification approaches as **GHV** and **CPHV**, respectively. In particular, we consider topic-based clustering models, since our approaches also rely on topic modeling. We consider four such recent models—COM [69], HBGG [27], GGC [28], and HGGC [28].

Experimental Settings and Evaluation Metrics

Both COM and HBGG use statistical topic modeling concepts for group recommendation (i.e., clustering). GGC is an LDA topic model based clustering approach. Whereas, HGGC is a hybrid approach, which integrates content information of each member item to GGC based on probabilistic matrix factorization. For all these models, hyperparameters are fixed as $\alpha = \beta = \eta = \rho = 0.01$. Here, α (document-topic density), β (topic-word density), η, and ρ are dirichlet constants for calculating probability distributions in LDA. For preference score calculation in HGGC, value of λ (parameter for Bernoulli distribution) is set as 0.5, as

[6] https://bit.ly/2Kz0iPP.

suggested in [28]. Besides, initially, we set the number of topics as 20 for all the models. Afterwards, we tune this number based on the topics relevant to our approaches to perform more robust evaluation of our proposed approaches.

Results

We evaluate the performances of the models using two popular evaluation metrics for group recommendation—SCC and ICC. The silhouette value in SCC measures an object's suitability within its own cluster (cohesion) compared to other clusters (separation). Table 11 reflects the performance of our approaches and four state-of-the-art models using our dataset in terms of both SCC and ICC scores. To get SCC and ICC scores, we first generate the clusters for each of these four models by calculating and interpreting recommendation scores, given a group g (i.e., hedonistic users) and an item i [28]. Note that we evaluate the model performances based on two values of the number of topics—initially, set as 20 [28], and afterward, set as 12 (based on the best performance scenario of our proposed approaches). In our CPHV, the number of topics has been identified either through calculating their correlations with users' hedonism scores (12 topics) or based on the topics relevant to other models (20 topics) for calculating corresponding SCC and ICC values. Note that our GHV model is independent of number of topics. SCC of a particular cluster object is measured between its own cluster and all the other clusters where the object is not belonged [70]. For example, if a Dataset D, is partitioned into C clusters, and for each of object $o \in D$, then $a(o)$ indicates the average distance between its own cluster and $b(o)$ denotes the minimum average distance of all other clusters where o is not presented, and these calculations are expressed using Equations (2) and (3), respectively.

$$a(o) = \frac{\sum_{o' \in C_{i,0} \neq o'} \mathrm{dist}(o, o')}{|C_i| - 1} \quad (2)$$

$$b(o) = \min_{C_j : 1 \leq j \leq k, j \neq i} \frac{\sum_{o' \in C_j} \mathrm{dist}(o, o')}{|c_j|} \quad (3)$$

Based on this calculation, SCC of that object $s(o)$, can be determined using the following equation.

$$s(o) = \frac{b(o) - a(o)}{\max\{a(o), b(o)\}} \quad (4)$$

The output value is ranged between − 1 and +1, where smaller value reflects the compactness of the clusters and larger value denotes the separation of individual clusters.

ICC is another statistical measurement that evaluates the degree of similarity inside the same cluster and estimates the data variance [71].

$$\rho = \sigma\_b^2 / \left(\sigma\_b^2 + \sigma\_w^2\right) \quad (5)$$

In Eq. (5), ρ refers to the ICC, σ_b^2 indicates the average values of outcome variability between the clusters; on the other hand, outcome variability within each cluster is denoted by σ_w^2. The output value ranges between 0 and 1. Score approaching 1 signifies stronger similarity between clusters, whereas values tend to 0, suggesting weaker similarity.

Table 11 reflects the comparison among different approaches. For SCC and ICC scores, the proposed CPHV outperforms all other approaches by a big margin e.g., up to 21% and 16% of improvements. GHV obtained the best performance in hedonism based approaches. However, HGGC performs the best among our employed state-of-the-art models, followed by HGGC, COM, and HBGG, respectively. The better SCC and ICC scores indicate higher quality of clusters and stronger similarity in the cluster's data. These outcomes validate the effectiveness of the proposed approaches substantially.

Discussion

Although recommendation systems have achieved success in predicting users' preferences, group recommendation system can leverage in providing powerful suggestion to several members at a time. Our work introduces a novel technique to provide group recommendation from social media by using an important human psychological factor, i.e., values, by using GNN, contextual and psychological attributes of social media interactions.

Analyzing graph-based approach We leverage the combined strengths of GNNs and spectral clustering to significantly enhance the efficiency and effectiveness of the system. By integrating GNNs, we tap into their capability to uncover intricate relationships of individual nodes. This iterative learning process of GNNs allows us to capture subtle dependencies among the features present in our tweet dataset, resulting in more informative node representations. Subsequently, we apply spectral clustering to transform the graph into a more interpretable embedding, facilitating a clearer understanding of the underlying structure. This fusion of GNNs and spectral clustering ensures that the clustering process benefits from the learned patterns, leading to more accurate and meaningful groupings.

Analysis over hedonistic users during group identification Prilyantinasari et al. [72] describe that hedonistic people likely to share vacation photos of different places with higher esteem and luxury. Knox et al. [73] also describe that hedonistic people likely to travel different places for seeking pleasure. We find top 5 latent topics that are obtained from the tweets of 169 hedonistic users out of initial 1140 Twitter users. For example, we find Topic 1 (i.e., *Capricorn, Chicago, beach, Miami,* etc.) which is likely to describe exciting places for traveling. Chaney et al. [74] describe that people with high hedonism scores prone to spend more time outside of home, enjoy city crowd, and become happy to be center of attraction. Knox et al. [73] also discover that hedonistic people become happy during food consumption and drinking, relaxing at pool and visiting beaches. Our graph based approach, *GHV* supports these findings and outlines similar observation from hedonistic users derived from social media content. Topics 2 and 3 describe *happy, night, York, party, vegas* etc. topics that are related to night clubs and enjoying lives. Topic 4 describes *food, restaurant, yummy, delicious* for exploring food related topics and topic 5 likely to describe love and affair related issues. These topics tend to show users' pattern of enjoying lives.

In our context based approach, CPHV, we employ a sequence of cognitive attributes based (i.e., value) techniques to extract most suitable group by filtering contextual topics. Figure 7 shows the word clouds derived from the tweets by using the fusion of LSA and BERT of the selected hedonist users in the second step. These words largely reflect the nature of hedonist users such as they perhaps like fashionable products (i.e., Denim), they enjoy music and movies, they are prone to travel to different places, etc. In the last step, we discover better insights about the users' affective behavior by using LIWC. Our correlated LIWC categories most likely represent necessary components of a user to be happy in life. For example (see Table 7), we need to be positive in every step of our lives, we may need to be concerned about our home, family and friends, we need to earn money and achieve success in life, etc. In this way, we find the actual cluster with users' hedonistic nature by our study.

Analysis over validation dataset We observe that contents tweeted by the tech enthusiast users reflect their strong interest. We find relevant words such as *new, data, first, robot, phone, artificial intelligence, watch, thank*, etc. that can be connected with the users who have keen interest in technology. We observe in Table 9 that cluster of technologically enthusiast users show a good (0.615) scores of ICC correlation coefficient [75]. Table 8 shows (see second column) that technologically enthusiast users post tweets related to exciting features of games, outstanding features of a software, performance of a framework, etc. We also notice that contents tweeted by the movie lovers reveal their preference through social media interactions. We find that *movie, film, good, time, fan, show, love,* etc. frequent words from users' tweets who have emotional attachment on watching movies and write critical analyses on the storyline. We observe in Table 9 that cluster of movie enthusiast users also shows a good (0.653) ICC correlation coefficient [75]. Scarpi et al. [76] show that Internet and media help to propagate peer influence among the hedonistic users because they are fun loving, and have intention to fulfill personal desires without thinking long and rational. Table 8 (see first column) shows that sample tweets by the hedonistic users which reveal their opinion, excitement, comments, and suggestion regarding movies.

Correlations between hedonistic users' preferences While validating our groups where people with high hedonism score are clustered, SCC is applied to determine the validation of consistency on different number of clusters of two user domains. We see from Table 10, that both technologically enthusiasts and movie lovers have near similar score for different number of clusters. For example, SCC score of movie lovers show good (0.53–0.57) similarities within their own cluster. We also notice that the SCC score of movie lovers show fair (0.50–0.55) similarities within their own cluster. In these clusters data points are tightly bounded among the group members and explicitly separated from the members of other groups. Thus, members with similar values show identical nature in their respective domain.

Since values influence our real world preferences, it is possible to connect values with with a group of similar people. A similar study [77] also show that people exhibit on similar TV program who possess similar personality traits. Some studies [78–80] show that profile similarity, interaction activity among group members, and context play a major role. A few socio-psychological studies [81–83] also support to our claim. It is also possible to extend our study to recommend users based on other value dimensions by applying our technique using relevant hypotheses.

Our research is not without any limitations. We may face difficulty when adequate digital footprints might not be available to compute groups of users based on their values. For example, self-enhancement users likely generate less content in social media, because they consider usage of social media is waste of time [84]. Sometimes, values might be change over time [85], in those cases determining value dimensions might be challenging. However, based on the above discussion, Table 12 presents a few research avenues which might be intriguing for future researchers.

Conclusion

In this paper, we have presented group identification and validation techniques for hedonistic users in an open social network, i.e., Twitter. We have presented two novel group identification

Table 12 Future research avenues of this study

Research agenda	Challenge/research question	Research directions
Extension to other social media platforms	Can the proposed approach be adapted to analyze user values on platforms other than Twitter?	Investigate the feasibility and effectiveness of extending the methodology to platforms like Facebook, Instagram, or LinkedIn
Enhancement of psychological analysis techniques	How can the accuracy and depth of psychological analysis techniques, such as LSA and BERT, be improved for better understanding user preferences?	Explore advanced natural language processing (NLP) models, sentiment analysis techniques, and psychological lexicons to enhance the depth and accuracy of psychological analysis
Dynamic group identification	Can the proposed approach adapt to the dynamic nature of user preferences and interactions on social media?	Develop dynamic clustering algorithms that can continuously update user groupings based on evolving preferences and interactions over time
Integration of external data sources	How can external data sources, such as demographic data or user activity logs, be integrated to enrich the analysis of user values and preferences?	Investigate methods for integrating external data sources into the analysis pipeline to provide additional context and insights into user behavior
Content type	Do pattern of image sharing, brain's EEG signal, audio mel scale, etc. can identify users' group?	Using latest computer vision and ML techniques such as CNN, LSTM, and GRU may address the above issue

approaches. A graph dataset is initially constructed by selecting the most robust correlations among the features. Furthermore, a graph clustering technique is utilized to discover the most optimal group recommendations. Our contextual and psychological based approach is a novel technique to find group based on users' cognitive attributes derived from their social media interactions. We have demonstrated how both technologically enthusiasts and movie lovers show strong hedonistic characteristics in social media. To do this, we have proposed two hypotheses and systematically proved those two hypotheses with users' digital footprint. Our experimental results have shown high correlation coefficient for finding association between characteristics of the hedonistic users and their posted tweets. We have also compared our method with recent group recommendation techniques and our approach has outperformed the previous approaches. Our technique has opened up a new mechanism to recommend a group of users based on their value dimensions.

Author Contributions Conceptualization, M.S.H.M., S.A., M.N.I., M.A.I. and M.E.A.; methodology, M.S.H.M., J.A., M.A.K.R, M.N.I., M.A.I. and N.M.F; validation, M.S.H.M., S.A., M.N.I., M.A.I. and M.E.A.; formal analysis, M.S.H.M., J.A., N.M.F and M.A.K.R; investigation, M.S.H.M., J.A. and M.A.K.R; writing-original draft preparation, M.S.H.M., J.A., N.M.F and M.A.K.R; writing-review and editing, M.S.H.M. and S.A.; supervision, M.S.H.M., S.A., M.N.I. and M.E.A.; project administration, M.S.H.M.; All authors have read and agreed to the published version of the manuscript.

Funding Open Access funding provided by LUT University (previously Lappeenranta University of Technology (LUT)).

Data Availability Statement Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Research involving human Not applicable.

Informed consent Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Atif M, Shafiq M, Farooq M, Ayub G, Hussain M, Waqas M. Evolution of basic human values orientations: an application of monitoring changes in cluster solutions. PLoS One. 2022;17(9):0274600.
2. Schwartz SH (2005) Basic human values: their content and structure across countries. Valores e comportamento nas organizações (nd)
3. Acharya M, Sharma S. Semantic analysis of twitter posts. In: 2018 International conference on computational techniques, electronics and mechanical systems (CTEMS). 2018. p. 104–109.
4. Mohbey KK. Multi-class approach for user behavior prediction using deep learning framework on twitter election dataset. J Data Inf Manag. 2020;2(1):1–14.
5. Sansonetti G, Gasparetti F, Micarelli A, Cena F, Gena C. Enhancing cultural recommendations through social and linked open data. User Model User Adapt Interact. 2019;29(1):121–59.
6. Nawshin S, Mukta MSH, Ali ME, Islam AN. Modeling weather-aware prediction of user activities and future visits. IEEE Access. 2020;8:105127–38.
7. Javornik A, Marder B, Barhorst JB, McLean G, Rogers Y, Marshall P, Warlop L. What lies behind the filter? Uncovering the motivations for using augmented reality (ar) face filters on social media and their effect on well-being. Comput Hum Behav. 2022;128: 107126.
8. Khan EM, Mukta MSH, Ali ME, Mahmud J. Predicting users' movie preference and rating behavior from personality and values. ACM Trans Interact Intell Syst (TiiS). 2020;10(3):1–25.
9. Suman C, Reddy SM, Saha S, Bhattacharyya P. Why pay more? A simple and efficient named entity recognition system for tweets. Expert Syst Appl. 2020;167: 114101.
10. Chen J, Hsieh G, Mahmud JU, Nichols J. Understanding individuals' personal values from social media word use. In: Proceedings of the 17th ACM conference on computer supported cooperative work and social computing. 2014. p. 405–414
11. Mukta MSH, Ali ME, Mahmud J. User generated vs. supported contents: Which one can better predict basic human values? In: International conference on social informatics. Springer; 2016. p. 454–470.
12. Mukta MSH, Ali ME, Mahmud J. Identifying and validating personality traits-based homophilies for an egocentric network. Soc Netw Anal Min. 2016;6(1):74.
13. Mukta MSH, Ahmad J, Zaman A, Islam S. Attention and meta-heuristic based general self-efficacy prediction model from multimodal social media dataset. IEEE Access. 2024;12:36853–73. https://doi.org/10.1109/ACCESS.2024.3373558.
14. Sakib AS, Mukta MSH, Huda FR, Islam AN, Islam T, Ali ME. Identifying insomnia from social media posts: psycholinguistic analyses of user tweets. J Med Internet Res. 2021;23(12):27613.
15. Huang Z, Xu X, Zhu H, Zhou M. An efficient group recommendation model with multiattention-based neural networks. IEEE Trans Neural Netw Learn Syst. 2020;31(11):4461–74.
16. Liang R, Zhang Q, Wang J, Lu J. A hierarchical attention network for cross-domain group recommendation. IEEE Trans Neural Netw Learn Syst. 2022;35:3859–73.
17. Chen J, Chen W, Huang J, Fang J, Li Z, Liu A, Zhao L. Co-purchaser recommendation for online group buying. Data Sci Eng. 2020;5(3):280–92.
18. Raiaan MAK, Fatema K, Khan IU, Azam S, Rashid MR, Mukta MSH, Jonkman M, De Boer F. A lightweight robust deep learning model gained high accuracy in classifying a wide range of diabetic retinopathy images. IEEE Access. 2023;11:42361–88.
19. Shevtsov A, Oikonomidou M, Antonakaki D, Pratikakis P, Ioannidis S. What tweets and youtube comments have in common? Sentiment and graph analysis on data related to us elections 2020. PLoS One. 2023;18(1):0270542.
20. Bisgin H, Agarwal N, Xu X. Investigating homophily in online social networks. In: 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, vol. 1. IEEE; 2010. p. 533–36.
21. Sagiv L, Schwartz SH. Personal values across cultures. Ann Rev Psychol. 2022;73:517–46.
22. Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY. A comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst. 2020;32(1):4–24.
23. Landauer, T.: Latent semantic analysis: theory, method and application. In: Computer support for collaborative learning. Routledge; 2023. p. 742–43.
24. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018.
25. Koutsoumpis A, Oostrom JK, Holtrop D, Van Breda W, Ghassemi S, Vries RE. The kernel of truth in text-based personality assessment: a meta-analysis of the relations between the big five and the linguistic inquiry and word count (liwc). Psychol Bull. 2022;148(11–12):843.
26. Huang Z, Liu Y, Zhan C, Lin C, Cai W, Chen Y. A novel group recommendation model with two-stage deep learning. IEEE Trans Syst Man Cybern Syst. 2021;52(9):5853–64.
27. Lu Z, Li H, Mamoulis N, Cheung DW. Hbgg: a hierarchical Bayesian geographical model for group recommendation. In: Proceedings of the 2017 SIAM International Conference on Data Mining. SIAM; 2017. p. 372–80.
28. Jeong HJ, Kim MH. Hggc: a hybrid group recommendation model considering group cohesion. Expert Syst Appl. 2019;136:73–82.
29. Acharya M, Mohbey KK. Differential privacy-based social network detection over spatio-temporal proximity for secure poi recommendation. SN Comput Sci. 2023;4(3):252.
30. Khanam KZ, Srivastava G, Mago V. The homophily principle in social network analysis: a survey. Multimed Tools Appl. 2023;82(6):8811–54.
31. Halberstam Y, Knight B. Homophily, group size, and the diffusion of political information in social networks: evidence from twitter. J Public Econ. 2016;143:73–88.
32. Cingel DP, Carter MC, Krause H-V. Social media and self-esteem. Curr Opin Psychol. 2022;45: 101304.
33. Golbeck J, Robles C, Turner K. Predicting personality with social media. In: CHI '11 extended abstracts on human factors in computing systems. New York, NY, USA: ACM; 2011.
34. Alqahtani G, Alothaim A. Predicting emotions in online social networks: challenges and opportunities. Multimed Tools Appl. 2022;81(7):9567–605.
35. Liu D, Fu Q, Wan C, Liu X, Jiang T, Liao G, Qiu X, Liu R. Suicidal ideation cause extraction from social texts. IEEE Access. 2020;8:169333–51.
36. Mohammad SM, Kiritchenko S. Using hashtags to capture fine emotion categories from tweets. Comput Intell. 2015;31(2):301–26.
37. Liu B. Sentiment analysis: mining opinions, sentiments, and emotions. Cambridge: Cambridge University Press; 2015.
38. Zhang L, Ghosh R, Dekhil M, Hsu M, Liu B. Combining lexicon-based and learning-based methods for twitter sentiment analysis. HP Laboratories, Technical Report HPL-2011, vol. 89. 2011.
39. Pennebaker JW, Francis ME, Booth RJ. Linguistic inquiry and word count: Liwc 2001, vol. 71. Mahway: Lawrence Erlbaum Associates. 2001.
40. Shih Y-A, Chang B, Chin JY. Data-driven student homophily pattern analysis of online discussion in a social network learning environment. J Comput Educ. 2020;7:1–22.
41. Zhang L, Li X, Shen J, Wang X. Structure, attribute and homophily preserved social network embedding. In: International conference on neural information processing. Springer; 2018. p. 118–30.

42. Solomon RS, Srinivas P, Das A, Gamback B, Chakraborty T. Understanding the psycho-sociological facets of homophily in social network communities. IEEE Comput Intell Mag. 2019;14(2):28–40.
43. Dehghani M, Johnson K, Hoover J, Sagi E, Garten J, Parmar NJ, Vaisey S, Iliev R, Graham J. Purity homophily in social networks. J Exp Psychol Gen. 2016;145(3):366.
44. Murase Y, Jo H-H, Török J, Kertész J, Kaski K. Structural transition in social networks: the role of homophily. Sci Rep. 2019;9(1):1–8.
45. Hammou BA, Lahcen AA, Mouline S. A distributed group recommendation system based on extreme gradient boosting and big data technologies. Appl Intell. 2019;49(12):4128–49.
46. Wang L, Zhou W, Liu L, Yang Z, Wen J. Deep adaptive collaborative graph neural network for social recommendation. Expert Syst Appl. 2023;229: 120410.
47. Zhou W, Huang Z, Wang C, Chen Y. A multi-graph neural group recommendation model with meta-learning and multi-teacher distillation. Knowl Based Syst. 2023;276: 110731.
48. Li K, Wang C-D, Lai J-H, Yuan H. Self-supervised group graph collaborative filtering for group recommendation. In: Proceedings of the sixteenth ACM international conference on web search and data mining. New York, NY, USA: ACM; 2023.
49. Daudi A. Social media hedonism and the case of 'fitspiration': a nietzschean critique. Sport Ethics Philos. 2022;17:1–16.
50. Intason M. The dilemma between cultural rituals and hedonism for tourism in a cultural festival: a case study of the Lanna Songkran festival in Chiang Mai, Thailand. Int J Event Festiv Manag. 2023;15:121–49.
51. Arnoux P-H, Xu A, Boyette N, Mahmud J, Akkiraju R, Sinha V. 25 tweets to know you: a new model to predict personality with social media. arXiv preprint arXiv:1704.05513. 2017.
52. Mukta MSH, Khan EM, Ali ME, Mahmud J. Predicting movie genre preferences from personality and values of social media users. In: Proceedings of the international AAAI conference on web and social media, vol. 11, no. 1; 2017. p. 624–7. https://ojs.aaai.org/index.php/ICWSM/article/view/14910.
53. Veenhoven R. Hedonism and happiness. J Happiness Stud. 2003;4(4):437–57.
54. Mukta MSH, Sakib AS, Islam MA, Ali ME, Ahmed M, Rifat MA. Friends' influence driven users' value change prediction from social media usage. SBP-BRiMS (2021)
55. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The graph neural network model. IEEE Trans Neural Netw. 2008;20(1):61–80.
56. Bianchi FM, Grattarola D, Alippi C. Spectral clustering with graph neural networks for graph pooling. In: International conference on machine learning. PMLR; 2020. p. 874–83.
57. Ma YJ, Lee H-H. Understanding consumption behaviours for fair trade non-food products: focusing on self-transcendence and openness to change values. Int J Consum Stud. 2012;36(6):622–34.
58. Schwartz, S.H.: Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In: Advances in experimental social psychology, vol. 25. Elsevier; 1992. p. 1–65.
59. Raiaan MAK, Mukta MSH, Fatema K, Fahad NM, Sakib S, Mim MMJ, et al. A review on large language models: Architectures, applications, taxonomies, open issues and challenges. IEEE Access. 2024;12:26839–74. https://doi.org/10.1109/ACCESS.2024.3365742.
60. Zhang Y, Ge C, Hong S, Tian R, Dong C, Liu J. Delesmell: code smell detection based on deep learning and latent semantic analysis. Knowl Based Syst. 2022;255: 109737.
61. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018.
62. Lyu S, Ren X, Du Y, Zhao N. Detecting depression of Chinese microblog users via text analysis: combining linguistic inquiry word count (liwc) with culture and suicide related lexicons. Front Psychiatry. 2023;14:1121583.
63. Dinh D-T, Fujinami T, Huynh V-N. Estimating the optimal number of clusters in categorical data clustering by silhouette coefficient. In: International symposium on knowledge and systems sciences. Springer; 2019. p. 1–17.
64. Crisp R. Hedonism reconsidered. Philos Phenomenol Res. 2006;73(3):619–45.
65. Schwartz SH, Boehnke K. Evaluating the structure of human values with confirmatory factor analysis. J Res Personal. 2004;38(3):230–55.
66. Schwartz SH. An overview of the schwartz theory of basic values. Online Read Psychol Cult. 2012;2(1):11 arXiv:2307.0919.
67. Ksendzova M, Iyer R, Hill G, Wojcik SP, Howell RT. The portrait of a hedonist: the personality and ethics behind the value and maladaptive pursuit of pleasure. Personal Individ Differ. 2015;79:68–74.
68. Sprinthall RC, Fisk ST. Basic statistical analysis. Englewood Cliffs: Prentice Hall; 1990.
69. Yuan Q, Cong G, Lin C-Y. Com: a generative model for group recommendation. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '14. New York: Association for Computing Machinery; 2014. p. 163–72.
70. Han J, Kamber M, Pei J. 10-cluster analysis: basic concepts and methods. In: Han J, Kamber M, Pei J, editors. Data mining. The Morgan Kaufmann series in data management systems. 3rd ed. Boston: Morgan Kaufmann; 2012. p. 443–95.
71. Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. BMJ. 1998;316(7142):1455–60.
72. Prilyantinasari P, Mulyana A. The effect of instagram exposure of hedonic lifestyle on dissonance rates for digital native. Int J English Lit Soc Sci (IJELS). 2020;5(2):396–402.
73. Knox D, Hannam K, Margry PJ, Olsen DH, Salazar NB. Is tourist a secular pilgrim or a hedonist in search of pleasure? Tour Recreat Res. 2014;39(2):235–67.
74. Chaney D. Lifestyle atau lifestyle: Sebuah pengantar komprehensif. Terj. Nuraeni. Yogyakarta: Jalasutra; 2009.
75. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6(4):284.
76. Scarpi D. Work and fun on the internet: the effects of utilitarianism and hedonism online. J Interact Mark. 2012;26(1):53–67.
77. Asabere NY, Acakpovi A. Roppsa: Tv program recommendation based on personality and social awareness. Math Probl Eng. 2020;2020:1971286.
78. Gartrell M, Xing X, Lv Q, Beach A, Han R, Mishra S, Seada K. Enhancing group recommendation by incorporating social relationship interactions. In: Proceedings of the 16th ACM international conference on supporting group work. 2010. p. 97–106.
79. Namisango F, Kang K, Rehman J. Examining the relationship between sociomaterial practices enacted in the organizational use of social media and the emerging role of organizational generativity. Int J Inf Manag. 2023;71: 102643.
80. McKenzie J. The economics of movies (revisited): a survey of recent literature. J Econ Surv. 2023;37(2):480–525.
81. Bruning PF, Alge BJ, Jackson CL. Explaining the relational mechanisms and outcomes of multi-modal leader–member-exchange differentiation. Can J Adm Sci (Revue Canadienne des Sciences de l'Administration). 2022;39(1):5–20.
82. Oluwajana D, Adeshola I, Clement S. Does the use of a web-based collaborative platform reduce cognitive load and influence project-based student engagement? Curr Psychol. 2023;42(10):8265–78.
83. Hu H, Yu G, Xiong X, Guo L, Huang J. Cultural diversity and innovation: an empirical study from dialect. Technol Soc. 2022;69: 101939.

84. Park JY, Im I, Sung C-S. Is social networking a waste of time? the impact of social network and knowledge characteristics on job performance. Knowl Manag Res Pract. 2017;15:560–71.
85. Mukta MSH, Ali ME, Mahmud J. Temporal modeling of basic human values from social network usage. J Assoc Inf Sci Technol. 2019;70(2):151–63.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.