Heliyon 8 (2022) e09664

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Predicting verbal reasoning from virtual community membership in a sample of Russian young adults

Pavel Kiselev [a,*], Valeriya Matsuta [b], Artem Feshchenko [b], Irina Bogdanovskaya [c], Boris Kiselev [d]

[a] Career Consultants Association, Russian Federation
[b] National Research Tomsk State University, Russian Federation
[c] Herzen State Pedagogical University of Russia, Russian Federation
[d] National Research Nuclear University MEPhI, Russian Federation

HIGHLIGHTS

- Investigating if verbal reasoning can be predicted by virtual community membership.
- Data from 3646 Russian young adults' social networking accounts were collected.
- Binary classification machine-learning models were used for analysis.
- Results showed reasonably good performance for verbal-reasoning prediction.
- Influence of community genres for predictions based on sex were also examined.

ARTICLE INFO

Keywords:
Verbal reasoning
Social networking site
Virtual community
Machine learning

ABSTRACT

Predicting personality traits from social networking site profiles can help to assess individual differences in verbal reasoning without using long questionnaires. Inspired by earlier studies, which investigated whether abstract-thinking ability are predictable by social networking sites data, we used supervised machine learning to predict verbal-reasoning ability based on a proposed set of features extracted from virtual community membership. A large sample (N = 3,646) of Russian young adults aged 18–22 years approved access to the data from their social networking accounts and completed an online test on verbal reasoning. We experimented with binary classification machine-learning models for verbal-reasoning prediction. Prediction performance was tested on isolated control subsamples for men and women. The results of prediction on AUC-ROC metrics for control subsamples over 0.7 indicated reasonably good performance on predicting verbal-reasoning level. We also investigated the contribution of virtual community's genres to verbal reasoning level prediction for male and female participants. Theoretical interpretations of results stemming from both Vygotsky's sociocultural theory and behavioural genomics are discussed, including the implication that virtual communities make up a non-shared environment that can cause variance in verbal reasoning. We intend to conduct studies to explore the implications of the results further.

1. Introduction

Circa 2005, Internet social networking sites, or SNS[1], have become one of the most important channels for human communication and socialisation (Brailovskaia and Bierhoff, 2016). Boyd and Ellison (2007) defined SNS as web-based services that allow individuals to "(1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.'' (p. 211). Ellison et al. (2007) defined SNS as an online application that allows individuals to present themselves, articulate their offline social networks, establish or maintain connections with others, and join virtual groups based on common interests. As with the advent of television, SNS transformed society (Shapiro and Margolin, 2014). Since their introduction, spending time on SNS has become part of young adults' daily routines (Casale and Fioravanti, 2018).[1]

* Corresponding author.
E-mail address: forestfield@yandex.ru (P. Kiselev).
[1] Social Networking Services.

https://doi.org/10.1016/j.heliyon.2022.e09664
Received 1 November 2021; Received in revised form 2 February 2022; Accepted 31 May 2022

Human behaviour is manifested by one's actions, and these actions on SNS are largely recorded, creating digital footprints (Gencoglu et al., 2015). The rapid spread of social media and smartphones enables us to collect and process data about human behaviour on a previously unimaginable scale (Salganik, 2019). SNS data provides ecologically valid measures of people's real-world behaviour, as opposed to data collected during experimental sessions; thus, they are less susceptible to bias (Azucar et al., 2018; Meshi et al., 2015).

Additionally, personality is strongly related to human behaviour on the Internet in general (Landers and Lounsbury, 2006) and SNS in particular (Amichai-Hamburger and Vinitzky, 2010). To effectively predict personality, it is essential to extract helpful features from SNS digital footprints. Moreover, judgements of people's personalities based on supervised machine learning with features extracted from digital footprints are more accurate and valid than judgements made by their friends, family, spouses, or colleagues (Youyou et al., 2015). For example, Segalin et al. (2017) reasoned that SNS store a tremendous amount of information and machine-learning models could use this information to optimise the accuracy of judgements and examine how humans are often affected by various motivational biases. Yarkoni and Westfall (2017) argued that machine-learning concepts and methods allow us to predict human behaviour with appreciable accuracy.

Much of the research on online content sharing has focused on prediction of the Big Five model of personality traits, represented by the acronym OCEAN: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. In contemporary literature, the Big Five model is the most widespread and validated method (McCrae and Costa, 1997), as these five fundamental traits are repeatedly obtained in factor analyses of personality questionnaires (Goldberg, 1990). Other studies have examined curiosity (Menk and Sebastiá, 2016), anxiety (Gruda and Hasan, 2019), and the Dark Triad of personality types (Garcia and Sikström, 2014).

SNS digital footprints include all possible SNS data; however, researchers often use only parts of it. Kosinski et al. (2013) predicted the Big Five personality traits of Facebook users by analysing the behaviour of 'liking' other users' posts and the content of those posts. Big Five personality traits have also been predicted using text mining (Golbeck, 2016; Wald et al., 2012) and picture mining (Celli et al., 2014; Liu et al., 2016).

Kosinski et al. (2013) demonstrated that abstract thinking measured using Raven's Standard Progressive Matrices could also be predicted by data on the 'liking' of other users' posts. These results were repeated by Wei and Stillwell (2017) through the analysis of Facebook user's avatar. Mori and Haruno (2020) also obtained similar results for Japanese adults by analysing the content of Twitter posts.

Thus, although these studies examined online behaviour and abstract thinking, to the best of our knowledge, there are no studies concerning the prediction of verbal abilities that use digital SNS footprints. This study explored the prediction of verbal-reasoning abilities using features extracted from virtual community membership on SNS to contribute to the literature.

1.1. Verbal reasoning

Verbal skills have been identified as indicators of cognitive functioning since the earliest modern theories of intelligence (Conte et al., 2020). The Cattell and Horn Fluid-Crystalized (Gf–Gc) theory is probably the best known and most widely used theory of intelligence (Stankov et al., 1995; Kaya et al., 2015).

Gf comprises reasoning as well as memory and perceptual speed (Beauducel et al., 2001). Fluid intelligence is often measured with figural tests, whereas crystalized intelligence is often assessed with verbal tests (Beauducel et al., 2001). Hence, previous studies have dealt with the relationship of Gf to digital SNS footprints.

Horn and Noll (1997) conceptualised Gc as 'acculturation knowledge', expressing the importance of the knowledge domain for the conceptualisation of Gc. Most researchers agree that Gc is influenced by education and cultural exposure (Brody, 1992; Moutafi et al., 2004). It suggests that Gc is no less likely to be associated with online behaviour on SNS than on Gf.

1.2. Virtual communities

Virtual communities, sometimes called online communities (Shen and Khalifa, 2013), are groups of people sharing interests or goals, for whom electronic communication is their primary form of interaction (Dennis et al., 1998). Although virtual communities appeared long before SNS, through bulletin boards or online forums, they achieved their high point of connections with the global proliferation of SNS. For example, the virtual communities on VKontakte, a Russian SNS, engage millions of users.

From the viewpoint of social neuroscience, Weaverdyck and Parkinson (2018) suggested that the ability to navigate large, complexly bonded social groups on SNS has been shaped evolutionary and has a neural representation. Similarly, a previous study by Dunbar et al. (2015) confirmed that virtual communities on SNS have similar structural characteristics as offline face-to-face networks. Notwithstanding the evolutionary basis, virtual communities overcome the limitations of accommodating face-to-face interactions in offline communities, such as synchronicity, physical proximity, and spatial cohesiveness (Abfalter et al., 2012).

Virtual communities can have a significant influence on individuals' attitudes and behaviour, particularly for young people (Sirola et al., 2019) as virtual community identification guides its members' feelings, beliefs, and behaviour (Kim et al., 2012). This identification significantly relates to trust in community members and collective efficacy. People in virtual communities tend to be relatively homogeneous in their interests and less in age, social class, ethnicity, life-cycle stage, and other aspects of their social backgrounds (Wellman and Gulia, 1999).

1.3. Relationship between virtual communities' membership and general abilities

A methodological challenge in research relates to understanding the relationship between virtual communities' participation and ability level. Virtual communities can be understood through the lens of social theory of learning (Lave and Wenger, 1991), which argues that social participation is at the centre of the learning process, and community is a social configuration defined by action, occurring through discourse. Communities of practice are considered open communities, where users regularly share common interests, create content, and negotiate knowledge (Wenger et al., 2002). Individual membership within communities of practice influences individuals' knowledge and cognitive changes (Billett, 1998). There is no exact border between a community of practice and a community of interest. Virtual communities can correspond to all these categories (Reyes, 2018).

Young adults' collaborations in social environments may be understood as peer tutoring for education that benefits both tutors and tutees (Lieberman, 2012). Social interactions involve brain areas and mechanisms that assist and support learning by strengthening learning experiences, consequently making them more memorable (Laiti and Frangou, 2019).

Further, Sarmiento and Shumar (2010) suggested using positioning theory as a framework for research on the construction of virtual mathematical communities. The positional theory provides insight into the discursive construction of knowledge and virtual community participants' identity self-construction in activities that are constituted by and performed through social interaction.

In the context of individual differences in general abilities, Alloway and Alloway (2012) experimentally examined the positive impact of Facebook engagement on young adults' working memory. Quiroga et al. (2015) reported a strong correlation between high-order latent factors

capturing the variance common to a heterogeneous set of commercial video games and general intelligence. Additionally, knowledge construction in virtual communities through knowledge sharing positively correlates with high levels of self-esteem, need for social interaction, and public individuation (Lee and Jang, 2010).

1.4. This study

Based on the previously discussed literature highlighting the relationship between virtual communities participation and general abilities level, we proposed to investigate two main research questions:

Research Question 1: How well can the verbal reasoning ability level be predicted from virtual community membership?

Research Question 2: What virtual community genres contribute to verbal reasoning level prediction?

The current study extends the literature by demonstrating the psychological mechanisms behind the machine learning algorithm. This theoretical perspective can provide some explanation for the findings of earlier studies in the prediction of abstract thinking by digital footprint in SNS.

This study focused on Russian young adults, 92% of whom use various types of SNS (Poushter et al., 2018). Although Facebook is the most popular SNS worldwide (Alexa, 2019), this study focused on VKontakte, as this SNS is the most popular in Russia (Baran and Stock, 2015). Like other SNS, VKontakte enables users to create visible profiles. Compared to Facebook, wall posts on VKontakte contain a photo or video with little text information, which makes text mining using a 'bag of words' with term frequency-inverse document frequency metrics for prediction and computational instruments, such as the Linguistic Inquiry and Word Count used in psychology (Pennebaker et al., 2007), inefficient.

One function of VKontakte is that it enables users to create and maintain official virtual communities. Users can create both groups and 'publics'. According to VKontakte, groups are more intended for users' associations by interests and discussions, while publics are intended for news publications from famous people or companies. However, in practice, groups and publics are not clearly distinguished. Therefore, we will refer to both groups and publics as 'groups' and examine them as examples of virtual communities.

2. Method

2.1. Sample

This study was conducted with a cohort of 4,044 Russian young adults, aged 18–22 years. The research survey was presented in Russian. Participants were recruited through Internet advertising of an online battery of 17 tests for career guidance. This battery was designed for career guidance purposes. All participants confirmed they were aged 18–22 years and provided informed consent. Thus, participants were informed about the aim of the data collection and that they retained the right to withdraw from the study at any time (no one withdrew). Additionally, participants approved access to their VKontakte account through the VKontakte API.

Ethical approval was obtained from the Ethical Committee at Career Consultants Association. The study complied with all regulations and confirmation of Russian Federation.

Completing the whole battery of tests took participants about 50 min. For this study, we used the results of only one of the tests related to verbal reasoning. The test took approximately five minutes to complete.

VKontakte users tend to follow many groups and, on average, participants followed 123 groups. A total of 398 respondents (9.8%) reported membership in less than 10 groups, which may have meant the SNS VKontakte account provided to us was not one they had used permanently, or they hid their main account and used a fictitious account for online questionnaires. Therefore, we chose to remove those 398 respondents from our sample. Among the remaining 3,646 respondents, there were 2,241 women (61.5%) and 1,405 men (38.5%).

2.2. Measures

To measure verbal reasoning, items similar to those developed by Amthauer et al. (2001) for test no. 3, 'Analogies', on verbal reasoning have been developed. For each item, a word pair is provided, along with the first word of a second pair (e.g., noun: decline = verb). Five response options are given, one of which best completes the pairing (Change, Form, Use, Conjugate, Write; Solution: Conjugate). The test has no time limit.

The development was accomplished with two experts who worked with the measurement of intelligence in young adults. Each of these experts had more than 10 years of experience in the fields of academia and counselling. The process resulted in a final cohort of 20 questionnaire items. A group of experts comprising two psychologists (with doctoral degrees) and career-guidance counsellors evaluated the content validity of the items. Based on their comments, the final items were formulated. The next stage was pilot work, after which six items were excluded. The instrument was then tested among a sample of 376 Russian young adults aged 18–22 years. The results of confirmatory factor analyses revealed acceptable model fits where $\chi 2/df = 1.42$, CFI = .97, SRMR = .039, and RMESA = .033 (95% C.I.; .017, .047). Appendix A lists the final test items.

2.3. Statistical analyses

Statistical analyses were performed using SPSS v26.0 (IBM Corp.: Armonk, NY, USA). A Lilliefors test confirmed that the shape of the acquired data was not normally distributed ($p < .05$). Therefore, to avoid prediction by sex or because of the presence of abnormally distributed data, we used a non-parametric Mann-Whitney U test. The highly significant result ($p < .001$) showed that women tended to have significantly higher verbal reasoning ability than men. Hence, further statistical analysis, feature extraction, and the construction of machine-learning models were conducted separately for male and female samples.

To define high-level verbal reasoning, we also calculated the 75th percentile. Participants with results above the 75th percentile were considered to possess high levels of verbal reasoning. To show the appropriateness of test data quality, we calculated Cronbach's alpha for the verbal-reasoning scale that was used in the career guidance tests. We also calculated correlations for all test items and total scores.

2.4. Subsamples

To minimise the machine learning model's overfitting, the participant sample was randomly split into a development subsample (90%) and a control subsample (10%). Men and women were selected separately to keep sex distribution generalisability in the control subsample, as participants below and above the 75th percentile (Table 1). We did not use data for verbal-reasoning from the control subsample for feature extraction, machine learning model fitting, or parameter selection.

2.5. Feature extraction

Feature extraction is a critical step in the development of any machine-learning model (Bayat et al., 2014; Flach, 2012). The aim of feature extraction in our study is to extract valuable information from virtual community memberships to predict verbal reasoning. Feature extraction was conducted for male and female samples separately.

For feature extraction in the male sample, we used virtual communities with a high number of members. Each community had at least 100,000 members, and at least 35 participants were from the male sample. We assessed the strength of the relationships between virtual

Table 1. Subsample size for sex and verbal reasoning.

Percentile	Development		Control		Total	
	Men n (% all develop. men)	Women n (% all develop. women)	Men n (% all control men)	Women n (% all control women)	All Men develop. + control (% total men)	All Women develop. + control (% total women)
≤75th	973 (76.9%)	1,448 (71.8%)	108 (77.1%)	161 (71.9%)	1,081 (76.9%)	1,609 (71.8%)
>75th	292 (23.1%)	569 (28.2%)	32 (22.9%)	63 (28.1%)	324 (23.1%)	632 (28.2%)
Total	1,265	2017	140	224	1,405	2,241

communities as predictive features for males by calculating the difference between mean verbal-reasoning scores of males from the development subsample and the mean verbal-reasoning scores for males in our sample. In accordance with this approach, we calculated a score for all virtual communities with 100,000 members and 35 male participants. Restricting the participant number ensures the robustness of machine-learning models.

Although every group selected for feature construction had at least 35 participants, using membership in concrete groups as a binary feature (i.e., participant is member of a group or not) is inefficient for machine-learning model construction. Multiple number of features will make models overfit rare training data and misguide the prediction, and features that relate to only a few participants have no generalizability (Zhong et al., 2013). Thus, to perform dimension reduction, we used two aggregated positive and negative indices: the sum of membership in groups with, positives scores and the sum of membership in groups with negative scores. Concerning psychometrics, we can compare membership in concrete groups with questionnaire items and scale indices. Virtual community cut-off scores for positive and negative indices were parameters for machine-learning models.

Aside from positive and negative indices, to better understand association between virtual communities' membership and verbal-reasoning level, we manually selected groups with three special genres. These genres were selected as the most common for groups with high scores, for both males and females. The first genre was science and technology, and it included discussions on actual science challenges in fields from astrophysics to neurobiology and on modern technological achievements such as the SpaceX project or MIT's dog-like robots. The second genre was abstract humour and memes. These group discussions concentrated on pictures and videos similar to *Monty Python's Flying Circus*. The third genre was art and aesthetics, in which discussion subjects were pictures or stories that had artistic value.

We identified group genres through a qualitative analysis of posts. We reduced the requirement for the number of participants for these groups from 35 to 10. Groups belonging to each of the three genres should have a positive index of at least 0.9. Many groups satisfying this condition were not classified as belonging to any of the three genres. We achieved a balance for the number of participants for each genre. Therefore, for male participants, we selected 10 groups on abstract humour and memes, 10 groups on science and technology, and 10 groups on art and aesthetics. Appendix B lists items classified by group for the male subsample. Thus, we calculated three additional indices for all males as the sum of group membership for the corresponding genre.

Feature extraction for the female sample was performed using a similar method. For all groups with 100,000 members and at least 35 females, we calculated scores as the difference of mean verbal-reasoning scores of female members from the development subsample and mean values from Table 1 for female participants. We calculated positive and negative indices, and virtual community cut-off scores for positive and negative indices were machine-learning model parameters. We also selected groups with the three genres listed in Appendix C and calculated three additional indices. Therefore, we extracted five features for verbal-reasoning prediction in the male and female subsamples: positive index, negative index, and positive indices by genre (science and technology, abstract humour and memes, and art and aesthetics).

2.6. Machine-learning modelling

Considering that binary classification problems are well-known, we transformed verbal-reasoning prediction into a binary classification task. Given the five features, the classifier needed to determine whether participants had verbal-reasoning abilities above 75th percentile (Class '1' or positive class) or not (Class '0' or negative class). All participants received one class label, and there were no excluded participants.

Binary classification tasks have been investigated with several machine-learning methods. In this study, we chose decision tree and CatBoost classifiers. CatBoost classifiers are the most complicated data-mining technique and outperform leading packages such as XGBoost and LightGBM (Prokhorenkova et al., 2018). CatBoost classifiers are an example of stochastic gradient boosting, using a decision tree approach. Stochastic gradient boosting combines decision tree classifiers into an ensemble in an iterative way.

Proposed by Quinlan (1986), decision tree classifiers are one of the most well-known (Stein et al., 2005) and traditional machine-learning techniques. Decision tree classifiers construct a tree-like structure. The tree comprises nodes and leaves, and each node can have a child node. If a node has no child node, it is called a leaf or terminal node and has a probability of being a 'positive class' (Fehrman et al., 2017). A terminal node represents an available conclusion based on the information that led to it once no further information is needed to make the determination. In our study, it is probable to have high-level verbal-reasoning abilities. The downside to using decision tree classifiers is their susceptibility to overfitting (Uddin and Lee, 2017).

The classes were not balanced because the positive class (participants with high-level verbal-reasoning abilities) constituted around a quarter of all participants, and three-quarters of participants fell into the negative class (low or medium level of verbal-reasoning ability). Machine learning models were evaluated using descriptive statistics—accuracy, precision, recall, F-measure, AUC-ROC—as metrics, as opposed to inferential statistics such as p-values. For binary classification problems with an imbalance in classes, most scholars agree to recommend F-measure and AUC-ROC (Flach, 2012; Procaci et al., 2019). The advantage of an AUC-ROC measure is the quality interpretability concerning 'excellent', 'good', 'fair', and 'poor'.

For both classifiers, model parameters included cut-offs for positive and negative indices, as described below. For CatBoost classifiers, parameters also included maximum tree depth, number of iterations of stochastic gradient boosting, and learning rate. For decision tree classifiers, parameters also included maximum tree depth and learning rate.

Parameter selection was performed by 5-fold cross-validation ($k = 5$), as K-fold cross-validation is the most common method in machine learning (Seni and Elder, 2010). Development subsamples for male or female sex were partitioned into five subsets, with similar sizes and percentages for positive classes. The union of four subsets was then used as the training set, while the remaining subset was used as the test set, which was repeated five times so that every subset was used as the test set once. For both classifiers, parameter selection was performed on development subsamples. We optimised decision tree and CatBoost classifiers using the AUC-ROC metric on the development subsample.

Machine-learning modelling was performed with Python 3.7, using the scikit-learn package (version 0.22.1), and realisation of decision tree classifiers and parameter choosing with cross-validation based on

GridSearchCV, CatBoost package (version 0.21) for CatBoost classifiers. There were no features selected as categorical for CatBoost classifiers. The random state parameter was fixed at '0' to ensure reproducible results for all classifiers.

2.7. Feature analysis

For the second research question, a path analysis was used. Positive and negative index features are effective for dimension reduction and machine-learning purposes. Concurrently, our study aimed to gain insight into the contribution of virtual communities' genres to verbal reasoning level prediction. Examining such issues might be informed by discovering relationships between a positive index and genre features. At a finer level of analysis, it would be possible to assess how genres, directly and indirectly, influence the positive index. In particular, we used a classical path model approach. For path analysis, IBM SPSS AMOS v. 26 was utilised. We tested relationships between genre features and the positive index. Separate control subsamples were used for male and female participants. Analysis was conducted using maximum-likelihood estimates.

3. Results

3.1. Descriptive statistics

Mean, standard deviation, and 75th percentile are shown in Table 2 for males and females.

3.2. Measurement validation

Cronbach's alpha was 0.789 for men and 0.776 for women, which was greater than the accepted level of 0.7 recommended by Nunnally (1978).

Table 3 reports item-total correlations and Cronbach's alpha for each item in the verbal-reasoning scale. The item-total correlations for all items on the verbal-reasoning test for males and females exceeded 0.2 and were considered acceptable (Streiner et al., 2015). The correlations indicated that all items measured the same construct.

When any item was removed, Cronbach's alpha decreased equally for both men and women. Therefore, the 14 items for this study indicated good reliability. Table 3 shows the reliability analysis of the verbal-reasoning scale and indicates a good homogeneity among the sample items.

3.3. Research Question 1: machine learning results

Using cross-validation of development subsamples, optimal parameters for decision trees and CatBoost classifiers (Table 4) were found to maximise AUC-ROC metric values.

Predictions by classifiers with these parameters were calculated. Performance results for AUC-ROC (Table 5) and F-1 metrics (Table 6) on development and control subsamples were reported. For development subsamples, we also calculated standard deviations based on cross-validation data.

Although CatBoost classifiers showed better performance for both metrics used on development subsamples, the difference did not exceed the standard deviation. Further, CatBoost classifiers showed better results than Decision tree classifiers for females than males in the control subsample. The results of CatBoost classifiers on AUC-ROC metrics for both male and female control subsamples over 0.7 were fair, indicating reasonably good performance on predicting verbal-reasoning levels (Carter et al., 2016). According to Rice and Harris (2005), AUC-ROC values of 0.71 and greater correspond to Cohen's d-values of 0.80, which are regarded as large effects. This result is consistent with a meta-analysis of intelligence prediction by SNS digital footprint (Settanni et al., 2018) that found an association between digital traces and intelligence. Notably, our AUC-ROC results were lower than the value over 0.9 reported by scholars predicting Big Five personality traits (Markovikj et al., 2013).

Figure 1 for the male control subsample and Figure 2 for the female control subsample show the results of CatBoost classifier predictions as a confusion matrix. The rows indicate the actual verbal-reasoning level (high or not) from each data segment, and the columns indicate the predicted level.

The confusion matrix showed that 64% of female participants and 71% of male participants from the control subsample were correctly classified by verbal-reasoning level.

Table 2. Descriptive statistics for verbal-reasoning ability.

Sex	Mean	n	SD	Min	Max	75th percentile
Men	7.62	1,405	3.38	0	14	10
Women	8.15	2,241	3.24	0	14	10

Note. SD, standard deviation; Min, minimum; Max, maximum.

3.4. Research Question 2: path model results

Path model results are shown in Figure 3 for the male development subsample and in Figure 4 for the female development subsample.

Direct effect relationships between machine-learning model features. Note. Male development subsample. Path entries are standardised coefficients. All path weights between genre features and the positive index are significant ($p < .001$).

Direct effect relationships between machine-learning model features. Note. Male development subsample. Path entries are standardised coefficients. All path weights between genre features and the positive index are significant ($p < .001$).

The three genre indices explained 36% of the variance in the positive index for the female development subsample and 59% of the variance for the male development subsample.

4. Discussion

In this study, we aimed to elucidate association between virtual communities' membership and verbal reasoning ability among young adults. Guided by earlier research on predicting abstract thinking using SNS digital footprints, we used machine-learning models to predict levels of verbal reasoning. Indeed, the current results showed that features extracted from membership in virtual communities could be an excellent predictor of high-level verbal reasoning in Russian young adults. The results highlighted the role of virtual communities for feature extraction in individual difference prediction.

At the theoretical level, our results can be interpreted as being in line with social constructivism. From a social constructivist perspective, each internal cognitive change is expressed by the causal effect of social interaction (Vygotsky, 1980). Thus, social processes promote changes in verbal reasoning through the process of interacting socially using SNS.

Social constructivism is the umbrella framework for well-established theoretical models that underline the importance of social contexts in understanding individual differences, such as social cognitive theory (Bandura, 1999), social theory of learning (Lave and Wenger, 1991), cognitive mediation networks theory (De Souza et al., 2010), and positioning theory (Sarmiento & Shumar, 2010). Thus, our findings are consistent with the existing literature concerning the relationship between ability level and virtual communities' participation, as is considered in the theoretical section of this paper.

Path analysis results confirmed an important role of the abstract humour and memes genre. Additionally, our results indicated the value of the art and aesthetics genre. For men, we should underline the indirect effects of art and aesthetics' contribution to the positive index through abstract humour and memes.

Table 3. Mean, standard deviation, corrected item-total correlation, and Cronbach's alpha of the verbal-reasoning test.

Item	Mean		SD		Item-total correlation		Cronbach's alpha if item deleted	
	Men	Women	Men	Women	Men	Women	Men	Women
Q_1	0.63	0.70	0.484	0.459	0.334	0.342	0.782	0.767
Q_2	0.50	0.50	0.500	0.500	0.449	0.466	0.772	0.755
Q_3	0.70	0.79	0.459	0.408	0.295	0.236	0.785	0.775
Q_4	0.73	0.78	0.446	0.416	0.510	0.476	0.768	0.755
Q_5	0.51	0.54	0.500	0.499	0.346	0.395	0.782	0.762
Q_6	0.68	0.71	0.468	0.456	0.395	0.452	0.777	0.757
Q_7	0.28	0.34	0.447	0.475	0.307	0.303	0.784	0.771
Q_8	0.53	0.53	0.500	0.499	0.572	0.522	0.761	0.749
Q_9	0.61	0.65	0.489	0.476	0.404	0.360	0.776	0.765
Q_10	0.58	0.65	0.493	0.477	0.473	0.458	0.770	0.756
Q_11	0.67	0.70	0.471	0.457	0.496	0.443	0.768	0.758
Q_12	0.30	0.32	0.458	0.466	0.438	0.403	0.773	0.761
Q_13	0.17	0.18	0.380	0.387	0.324	0.255	0.782	0.773
Q_14	0.74	0.77	0.438	0.423	0.319	0.318	0.783	0.769

Note. SD, standard deviation.

Table 4. Best Parameters of Classifiers for Development of Subsample of Russian young adults by Five-Fold Cross-Validation in Measuring Verbal-Reasoning Ability.

Classifier		Sex	
		Men	Women
CatBoost	Depth	5	4
	Learning rate	0.17	0.1
	Number of estimators	25	100
	Scale positive weight	4	3
	Positive index cut-off	0.5	0.3
	Negative index cut-off	–0.7	–0.7
Decision tree	Max depth	5	5
	Class weight	4	3
	Positive index cut-off	0.5	0.3
	Negative index cut-off	–0.7	–0.7

Table 5. Classifier AUC-ROC Metrics on Development (average for five runs) and Control Subsamples.

Classifier	Development (mean ± SD)		Control	
	Men	Women	Men	Women
CatBoost	0.72 ± 0.05	0.72 ± 0.03	0.72	0.74
Decision tree	0.68 ± 0.04	0.69 ± 0.04	0.72	0.69

Note. SD, standard deviation.

Table 6. Classifier F1 Metrics on Development (average for five runs) and Control Subsamples.

Classifier	Development (mean ± SD)		Control	
	Men	Women	Men	Women
CatBoost	0.48 ± 0.07	0.52 ± 0.03	0.51	0.55
Decision tree	0.45 ± 0.06	0.51 ± 0.01	0.51	0.51

Note. SD, standard deviation.

All groups that we categorised by the three genres had positive scores for another sex or had a negligible number of participants of another sex (less than 10). However, it appears that one group can have a large positive score for one sex and a large negative score for the other, which

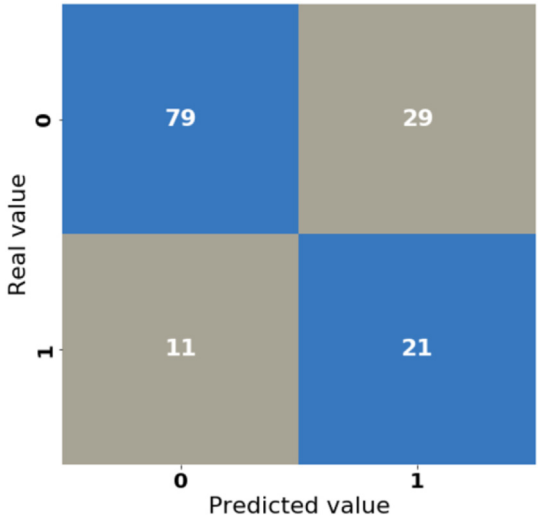

Figure 1. Confusion matrix for the male control subsample.

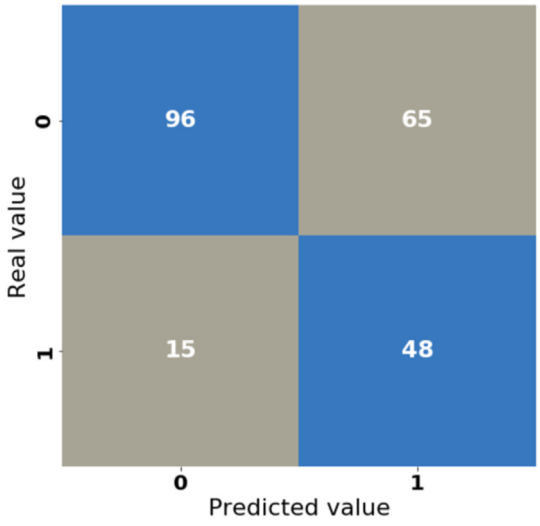

Figure 2. Confusion matrix for the female control subsample.

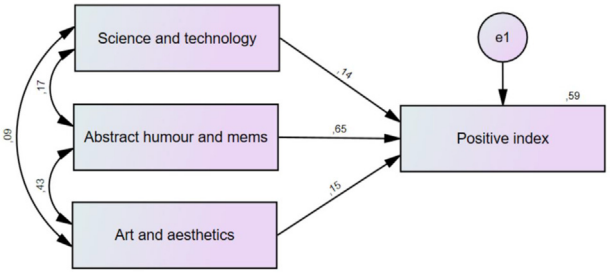

Figure 3. Positive index via genre indexes for male participants.

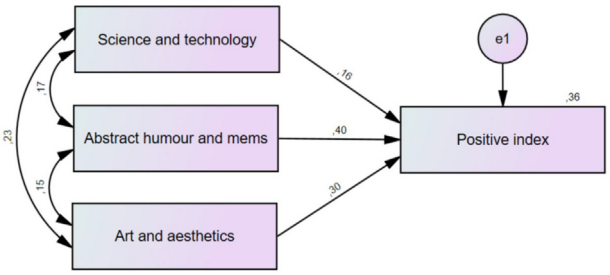

Figure 4. Positive index via genre indexes for female participants.

means that a virtual community's association on verbal reasoning is not absolute but may depend on sex.

According to path analysis, genre features explained most of the positive index variance for males and less than half of the variance for females. We can assume that there are unaccounted for genres, which are important for women and contribute to verbal-reasoning level. We believe that a finite number of group genres exist, just like in literature or cinema. For example, for the female development subsample, we can speculate there were more practically oriented high scoring groups. First, we identified groups with English learning topics, such as 'English yo', 'English | Английский язык', 'Proper English'. Further, there were groups on different practical subjects, such as 'Конференции, СеМинары, Гранты, Бизнес Идеи' (conferences, seminars, grants, business ideas) and 'Vandrouki | Путешествия почти бесплатно (RU)' (travel for almost free). However, some gamer groups, such as 'Dota 2 RuHub' and 'Heroes of the Storm', had high positive scores for males and low negative scores for females in the development subsample. 'ИгроМан' group (in English, 'a man fond of games') also had a high positive score for the male development subsample. That group is concentrated on intellectual humour regarding games and could also be categorised as part of the abstract humour and memes genre.

It seems that groups can be selected by genre without scoring, using only qualitative evaluation of group discussions. In fact, they were similar at first sight groups can have reverse scoring. For example, both the female and male development subsamples had negative scores for the groups 'Наука и Образование' ('science and education'), 'Факты Истории • Доисторические Цивилизации' ('historical facts, prehistoric civilisations'), and 'art'. Hence, groups can be selected for prediction only using quantitative scoring, based on common psychometric tests or other quantitative assessments.

Additionally, the positive index was the main feature for both males and females. The negative index had almost the same value for predicting verbal reasoning. In this study, we did not investigate the nature of groups with low negative scores. Nevertheless, we can suppose that groups that are obviously connected with unlawful or aggressive discourse have a negative bias towards verbal-reasoning results, such as 'BLATATA' (which comes from the Russian word 'блатной', meaning organised criminal) and 'Околофутбола 2 | Хулиганы | A.C.A.B.' ('near football, hooligans'). The same situation was found with groups focused on sexual discourse, including 'Пошлые и интиМные истории' ('vulgar and intimate stories') and '69 ПОШЛЫХ' ('69 vulgar people'). Regarding these issues, it may be interesting in the future to conduct research with the groups with low negative scores divided by genres, rather than considering them as one homogeneous construct, as we did in this study.

This study was in line with a social constructivism approach; however, an alternative explanation for our results may be that young adults with higher levels of verbal reasoning tend to get together. One could interpret the results reported in this study as individual differences in verbal reasoning can lead to one's selection of virtual communities. Similarly, Meldrum et al. (2019) suggested that adolescents become friends with others with whom they share similar intellectual abilities, as opposed to there being peer effects on intelligence. To explain individual differences in verbal reasoning, some researchers considered the role of additive genetic factors and shared environment, such as parenting strategies. For example, Haworth et al. (2010) demonstrated that heritability explains 66% of the variance in general cognitive abilities in young adulthood.

In line with comprehensive research on behavioural genomics, we can assume that both explanations complement each other. Schwartz et al. (2019), in a study on the role of gene-environment interplay in antisocial delinquency, revealed that groups of youth with similar traits and social influence need not be opposed to one another, but can complement each other and operate together; this observation was used to explain peer-delinquency homophily. To illustrate this in the context of intelligence, mathematically gifted adolescents have been shown to participate in dedicated math virtual communities, and such participation may further develop their mathematical abilities (Kovas et al., 2016). Thus, virtual communities make up a non-shared environment that can cause variance in verbal reasoning, which corresponds to a social constructivism approach. Further research applying a longitudinal design would allow researchers to capture a more precise picture of the interplay between social learning and heritability.

5. Limitations

Along with no clear verbal-reasoning variance and the possibility of virtual community participation causality, our study has some other limitations. First, this study focused only on virtual community subscription. We did not consider activity in virtual communities, such as likes, comments, and so on. Such methods consider all information about observers (or 'lurkers'), who represent passive involvement in virtual communities. According to the '90-9-1' principle, in a typical virtual community, 90% of the users only read content; 9% edit, comment, and repost content; and only 1% actively create and share new content (Chen et al., 2019). However, future research should explore extracting features from activity in virtual communities. Second, we focused on virtual communities that formally organised as groups or publics and had more than 100,000 members. However, groups and publics with fewer members, as well as informal virtual communities, should also be considered in future research. Third, we manually selected groups by genres using qualitative analysis of group content; however, it is clear, there is no explicit group genre, and groups focused on abstract humour and memes can have aesthetic discussions and *vice versa*. Furthermore, we could not exactly measure the abstractness of humour as a quantitative value, which was a limitation in feature extraction for our study. Fourth, this study only included young adults from Russia, which limits the generalisability of the findings to other age groups and/or different countries. Further, although the current study was conducted based on data from VKontakte, the research approach and method could be applied to other SNSs.

6. Conclusion

In summary, from a practical viewpoint the current study adds to the growing list of studies that predict human behaviour through the use of

SNS digital footprints. The ability to use digital SNS footprints for prediction may represent a rapid, cost-effective alternative to surveys and is a method for reaching larger populations. Therefore, it could be beneficial for academic, health-related, and commercial purposes (Azucar et al., 2018). Verbal reasoning is related to a broad spectrum of human activities and behaviours, including academic performance (Kotzé and Massyn, 2019), leadership (Mumford et al., 2000) and job performance (Lang et al., 2010). Because many individuals from all different lifestyles regularly use SNS, knowledge regarding the abilities of individuals could allow us to make predictions about each of these spheres.

From a theoretical perspective, our study shed light on the psychological mechanisms behind prediction with the machine-learning algorithm and SNS data. This study brought us a step closer to an awareness of how verbal reasoning is associated within a social context, which is the main contribution of our study. Our results should be considered as an early attempt to model the relationship between verbal reasoning level and activities on SNS. Although non-shared environmental association with ability level has been well established (Bishop et al., 2003; Nisbett et al., 2012), only a few studies have examined the mechanisms that underlie such effects.

Declarations

Author contribution statement

Pavel Kiselev:Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Valeriya Matsuta; Artem Feshchenko: Conceived and designed the experiments; Analyzed and interpreted the data.

Irina Bogdanovskaya:Conceived and designed the experiments; Performed the experiments.

Boris Kiselev: Analyzed and interpreted the data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

The authors do not have permission to share data.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e09664.

References

Alexa, 2019. Top Sites. Retrieved from. https://www.alexa.com/topsites.
Abfalter, D., Zaglia, M.E., Mueller, J., 2012. Sense of virtual community: a follow up on its measurement. Comput. Hum. Behav. 28, 400–404.
Alloway, T.P., Alloway, R.G., 2012. The impact of engagement with social networking sites (SNSs) on cognitive skills. Comput. Hum. Behav. 28, 1748–1754.
Amichai-Hamburger, Y., Vinitzky, G., 2010. Social network use and personality. Comput. Hum. Behav. 26, 1289–1295.
Amthauer, R., Brocke, B., Liepmann, D., Beauducel, A., 2001. Intelligenz-Struktur-Test 2000 R, Vol.2. Hogrefe, Verlag für Psychologie.
Azucar, D., Marengo, D., Settanni, M., 2018. Predicting the Big 5 personality traits from digital footprints on social media: a meta-analysis. Pers. Indiv. Differ. 124, 150–159.
Bandura, A., 1999. Social cognitive theory of personality. In: Pervin, L.A., John, O.P. (Eds.), Handbook of Personality: Theory and Research. Guilford Press, pp. 154–196.
Baran, K.S., Stock, W.G., 2015. Facebook has been smacked down. The Russian special way of SNSs: vkontakte as a case study. In: Proceedings of the 2nd European Conference on Social Media (ECSM 2015), 9–10, pp. 574–582.
Bayat, A., Pomplun, M., Tran, D.A., 2014. A study on human activity recognition using accelerometer data from smartphones. Procedia Comput. Sci. 34, 450–457.
Beauducel, A., Brocke, B., Liepmann, D., 2001. Perspectives on fluid and crystallized intelligence: facets for verbal, numerical, and figural intelligence. Pers. Indiv. Differ. 30, 977–994.
Billett, S., 1998. Ontogeny and participation in communities of practice: a socio-cognitive view of adult development. Stud. Educ. Adults 30, 21–34.
Bishop, E.G., Cherny, S.S., Corley, R., Plomin, R., DeFries, J.C., Hewitt, J.K., 2003. Development genetic analysis of general cognitive ability from 1 to 12 years in a sample of adoptees, biological siblings, and twins. Intelligence 31, 31–49.
Boyd, D.M., Ellison, N.B., 2007. Social network sites: definition, history, and scholarship. J. Comput.-Mediat. Commun. 13, 210–230.
Brailovskaia, J., Bierhoff, H.W., 2016. Cross-cultural narcissism on Facebook: relationship between self-presentation, social interaction and the open and covert narcissism on a social networking site in Germany and Russia. Comput. Hum. Behav. 55 (PART A), 251–257.
Brody, N., 1992. Intelligence, second ed. Academic Press.
Casale, S., Fioravanti, G., 2018. Why narcissists are at risk for developing Facebook addiction: the need to be admired and the need to belong. Addict. Behav. 76, 312–318.
Carter, J.V., Pan, J., Rai, S.N., Galandiuk, S., 2016. ROC-ing along: evaluation and interpretation of receiver operating characteristic curves. Surgery 159, 1638–1645.
Celli, F., Bruni, E., Lepri, B., 2014. Automatic personality and interaction style recognition from Facebook profile pictures. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 1101–1104. ACM.
Chen, X., Tao, D., Zhou, Z., 2019. Factors affecting reposting behaviour using a mobile phone-based user-generated-content online community application among Chinese young adults. Behav. Inf. Technol. 38, 120–131.
Conte, F., Costantini, G., Rinaldi, L., Gerosa, T., Girelli, L., 2020. Intellect is not that expensive: differential association of cultural and socio-economic factors with crystallized intelligence in a sample of Italian adolescents. Intelligence 81, 101466.
De Souza, B.C., De Lima e Silva, L.X., Roazzi, A., 2010. MMORPGS and cognitive performance: a study with 1280 Brazilian high school students. Comput. Hum. Behav. 26, 1564–1573.
Dennis, A.R., Pootheri, S.K., Natarajan, V.L., 1998. Lessons from the early adopters of web groupware. Manag. Inf. Syst. 14, 65–86.
Dunbar, R.I., Arnaboldi, V., Conti, M., Passarella, A., 2015. The structure of online social networks mirrors those in the offline world. Soc. Netw. 43, 39–47.
Ellison, N.B., Steinfield, C., Lampe, C., 2007. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. J. Comput.-Mediat. Commun. 12, 1143–1168.
Fehrman, E., Muhammad, A.K., Mirkes, E.M., Egan, V., Gorban, A.N., 2017. The five factor model of personality and evaluation of drug consumption risk. In: Palumbo, F., Montanari, A., Vichi, M. (Eds.), Data Science: Studies in Classification, Data Analysis, and Knowledge Organization. Springer, pp. 231–242.
Flach, P., 2012. Machine Learning: the Art and Science of Algorithms that Make Sense of Data. Cambridge University Press.
Garcia, D., Sikström, S., 2014. The dark side of Facebook: semantic representations of status updates predict the Dark Triad of personality. Pers. Indiv. Differ. 67, 92–96.
Gencoglu, O., Similä, H., Honko, H., Isomursu, M., 2015. Collecting a citizen's digital footprint for health data mining. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 7626–7629.
Golbeck, J., 2016. Predicting personality from social media text. AIS Transact. Replicat. Res. 2, 2.
Goldberg, L.R., 1990. An alternative "description of personality": the big-five factor structure. J. Pers. Soc. Psychol. 59, 1216.
Gruda, D., Hasan, S., 2019. Feeling anxious? Perceiving anxiety in tweets using machine learning. Comput. Hum. Behav. 98, 245–255.
Haworth, C.M., Wright, M.J., Luciano, M., Martin, N.G., de Geus, E.J., van Beijsterveldt, C.E., Bartels, M., Posthuma, D., Davis, O.S.P., Kovas, Y., Corley, R.P., Defries, J.C., Hewitt, J.K., Olson, R.K., Rhea, S.-A., Wadsworth, S.J., Iacono, W.G., McGue, M., et al., 2010. The heritability of general cognitive ability increases linearly from childhood to young adulthood. Mol. Psychiatr. 15, 1112–1120.
Horn, J.L., Noll, J., 1997. Human cognitive capabilities: Gf±Gc theory. In: Flanagan, D.P., Genshaft, J.L., Harrison, P.L. (Eds.), Contemporary Intellectual Assessment. Theories, Tests, and Issues. The Guilford Press.
Kaya, F., Juntune, J., Stough, L., 2015. Intelligence and its relationship to achievement. Elem. Educ. Online 14, 1060–1078.
Kim, C., Lee, S.G., Kang, M., 2012. I became an attractive person in the virtual world: users' identification with virtual communities and avatars. Comput. Hum. Behav. 28, 1663–1669.
Kosinski, M., Stillwell, D., Graepel, T., 2013. Private traits and attributes are predictable from digital records of human behavior. Proc. Natl. Acad. Sci. U.S.A. 110, 5802–5805.
Kotzé, M., Massyn, L., 2019. Predictors of academic performance in an adult education degree at a Business School in South Africa. Innovat. Educ. Teach. Int. 56, 628–638.
Kovas, Y., Tikhomirova, T., Selita, F., Tosto, M.G., Malykh, S., 2016. How genetics can help education. In: Kovas, Y., Malykh, S., Gaysina, D. (Eds.), Behavioural Genetics for Education. Palgrave Macmillan, pp. 1–23.
Laiti, O.K., Frangou, S.M., 2019. Social aspects of learning: Sámi people in the Circumpolar North. Int. J. Multicult. 21, 5–21.
Landers, R.N., Lounsbury, J.W., 2006. An investigation of Big Five and narrow personality traits in relation to Internet usage. Comput. Hum. Behav. 22, 283–293.
Lang, J.W., Kersting, M., Hülsheger, U.R., Lang, J., 2010. General mental ability, narrower cognitive abilities, and job performance: the perspective of the nested-factors model of cognitive abilities. Person. Psychol. 63, 595–640.

Lave, J., Wenger, E., 1991. Situated Learning: Legitimate Peripheral Participation. Cambridge University Press.
Lee, E.J., Jang, J.W., 2010. Profiling good Samaritans in online knowledge forums: effects of affiliative tendency, self-esteem, and public individuation on knowledge sharing. Comput. Hum. Behav. 26, 1336–1344.
Lieberman, M.D., 2012. Education and the social brain. Trends Neurosci. Educ. 1, 3–9.
Liu, L., Preotiuc-Pietro, D., Samani, Z.R., Moghaddam, M.E., Ungar, L., 2016. Analyzing personality through social media profile picture choice. In: Tenth International AAAI Conference on Web and Social Media. Retrieved from: aai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13102/12741.
Markovikj, D., Gievska, S., Kosinski, M., Stillwell, D.J., 2013. Mining Facebook data for predictive personality modeling. In: Seventh International AAAI Conference on Weblogs and Social Media. Retrieved from. https://www.gsb.stanford.edu/sites/gsb/files/conf-presentations/miningfacebook.pdf.
McCrae, R.R., Costa Jr., P.T., 1997. Personality trait structure as a human universal. Am. Psychol. 52, 509.
Meldrum, R.C., Young, J.T., Kavish, N., Boutwell, B.B., 2019. Could peers influence intelligence during adolescence? An exploratory study. Intelligence 72, 28–34.
Menk, A., Sebastiá, L., 2016. Predicting the human curiosity from users' profiles on Facebook. In: Proceedings of the 4th Spanish Conference on Information Retrieval. Association for Computing Machinery, pp. 1–8.
Meshi, D., Tamir, D.I., Heekeren, H.R., 2015. The emerging neuroscience of social media. Trends Cognit. Sci. 19, 771–782.
Mori, K., Haruno, M., 2020. Differential ability of network and natural language information on social media to predict interpersonal and mental health traits. J. Pers. 89, 228–243.
Moutafi, J., Furnham, A., Paltiel, L., 2004. Why is conscientiousness negatively correlated with intelligence? Pers. Indiv. Differ. 37, 1013–1022.
Mumford, M.D., Zaccaro, S.J., Johnson, J.F., Diana, M., Gilbert, J.A., Threlfall, K.V., 2000. Patterns of leader characteristics: implications for performance and development. Leader. Q. 11, 115–133.
Nisbett, R.E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D.F., Turkheimer, E., 2012. Intelligence: new findings and theoretical developments. Am. Psychol. 67, 130–159.
Nunnally, J.C., 1978. Psychometric Theory. McGraw-Hill.
Pennebaker, J.W., Booth, R.J., Francis, M.E., 2007. Linguistic Inquiry and Word Count: LIWC [Computer Software], 135. LIWC. net.
Poushter, J., Bishop, C., Chwe, H., 2018. Social media Use Continues to Rise in Developing Countries but Plateaus across Developed Ones, 22. Pew Research Center. Retrieved from. https://www.pewresearch.org/global/2018/06/19/social-media-use-continues-to-rise-in-developing-countries-but-plateaus-across-developed-ones/.
Procaci, T.B., Siqueira, S.W.M., Nunes, B.P., Nurmikko-Fuller, T., 2019. Experts and likely to be closed discussions in question and answer communities: an analytical overview. Comput. Hum. Behav. 92, 519–535.
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. In: Bengio, S., M Wallach, H., Larochelle, H., Lorraine, K., Cesa-Bianchi, N. (Eds.), NIPS' 18: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Curran Associates, pp. 6638–6649.
Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81–106.
Quiroga, M.Á., Escorial, S., Román, F.J., Morillo, D., Jarabo, A., Privado, J., Hernándeza, M., Gallegoa, B., Colom, R., 2015. Can we reliably measure the general factor of intelligence (g) through commercial video games? Yes, we can. Intelligence 53, 1–7.
Reyes, A., 2018. Virtual communities: interaction, identity and authority in digital communication. Text Talk 39, 99–120.
Rice, M.E., Harris, G.T., 2005. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. Law Hum. Behav. 29, 615–620.
Salganik, M., 2019. Bit by Bit: Social Research in the Digital Age. Princeton University Press.
Sarmiento, J.W., Shumar, W., 2010. Boundaries and roles: positioning and social location in the Virtual Math Teams (VMT) online community. Comput. Hum. Behav. 26, 524–532.
Schwartz, J.A., Solomon, S.J., Valgardson, B.A., 2019. Socialization, selection, or both? The role of gene–environment interplay in the association between exposure to antisocial peers and delinquency. J. Quant. Criminol. 35, 1–26.
Shapiro, L.A.S., Margolin, G., 2014. Growing up wired: social networking sites and adolescent psychosocial development. Clin. Child Fam. Psychol. Rev. 17, 1–18.
Segalin, C., Celli, F., Polonio, L., Kosinski, M., Stillwell, D., Sebe, N., Cristani, M., Lepri, B., 2017. What your Facebook profile picture reveals about your personality. In: Proceedings of the 25th ACM International Conference on Multimedia. ACM, pp. 460–468.
Seni, G., Elder, J.F., 2010. Ensemble methods in data mining: improving accuracy through combining predictions. In: Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan and Claypool Publishers.
Settanni, M., Azucar, D., Marengo, D., 2018. Predicting individual characteristics from digital traces on social media: a meta-analysis. Cyberpsychol., Behav. Soc. Netw. 21, 217–228.
Shen, K.N., Khalifa, M., 2013. Effects of technical and social design on virtual community identification: a comparison approach. Behav. Inf. Technol. 32, 986–997.
Sirola, A., Kaakinen, M., Savolainen, I., Oksanen, A., 2019. Loneliness and online gambling-community participation of young social media users. Comput. Hum. Behav. 95, 136–145.
Stankov, L., Boyle, G.J., Cattell, R.B., 1995. Models and Paradigms in Personality and Intelligence Research. Springer.
Stein, G., Chen, B., Wu, A.S., Hua, K.A., 2005. Decision tree classifier for network intrusion detection with GA-based feature selection. In: ACM-SE 43: Proceedings of the 43rd Annual Southeast Regional Conference 2, pp. 136–141.
Streiner, D.L., Norman, G.R., Cairney, J., 2015. Health Measurement Scales: A Practical Guide to Their Development and Use, fifth ed. Oxford University Press.
Uddin, M.F., Lee, J., 2017. Proposing stochastic probability-based math model and algorithms utilizing social networking and academic data for good fit students prediction. Soc. Netw. Anal. Min. 7, 29.
Vygotsky, L.S., 1980. Mind in Society: the Development of Higher Psychological Processes. Harvard University Press.
Wald, R., Khoshgoftaar, T., Sumner, C., 2012. Machine prediction of personality from Facebook profiles. In: 2012 IEEE 13th International Conference on Information Reuse & Integration (IRI). IEEE, pp. 109–115.
Weaverdyck, M.E., Parkinson, C., 2018. The neural representation of social networks. Curr. Opin. Psychol. 24, 58–66.
Wei, X., Stillwell, D., 2017. How smart does your profile image look? Estimating intelligence from social network profile images. In: WSDM '17: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining 33–40. Association for Computing Machinery.
Wellman, B., Gulia, M., 1999. Net surfers don't ride alone: virtual communities as communities. In: Smith, M.A., Kollock, P. (Eds.), Communities in Cyberspace. Routledge, pp. 167–194.
Wenger, E., McDermott, R.A., Snyder, W., 2002. Cultivating Communities of Practice: A Guide to Managing Knowledge. Harvard Business Press.
Yarkoni, T., Westfall, J., 2017. Choosing prediction over explanation in psychology: lessons from machine learning. Perspect. Psychol. Sci. 12, 1100–1122.
Youyou, W., Kosinski, M., Stillwell, D., 2015. Computer-based personality judgments are more accurate than those made by humans. Proc. Natl. Acad. Sci. U.S.A. 112, 1036–1040.
Zhong, E., Tan, B., Mo, K., Yang, Q., 2013. User demographics prediction based on mobile data. Pervasive Mob. Comput. 9, 823–837.