

# Semantic Interpretation of Social Network Communities (AAAI Student Poster Additional)

Aishwarya N Reganti, Tushar Maheshwari, Upendra Kumar, Tanmoy Chakraborty\*, Amitava Das

Indian Institute of Information Technology (IIIT), Sri City, AP, India, 517588

\*University of Maryland, College Park, USA, 20742

{aishwarya.r14, tushar.m14, upendra.k14, amitava.das}@iiits.in

\*tanchak@umiacs.umd.edu

## Introduction

The Big 5 Personality traits (Goldberg 1990), aka the five factor model (FFM) or the OCEAN model (for the initials of the five traits), as given in Figure 1, is a widely used *Personality model*.<sup>1</sup> The five factors are:

- **Openness (O):** A personality trait possessed by individuals who are imaginative, insightful and have wide interests;
- **Conscientiousness (C):** Refers to those who are organised, thorough, and planned;
- **Extroversion (E):** Refers to the Personality of those who are talkative, energetic, and assertive;
- **Agreeableness (A):** Individuals with this Personality trait are sympathetic, kind, and affectionate; and
- **Neuroticism (N):** Individuals who are mostly tense, moody, and anxious.

To define *societal sentiment*, we use the well-established “Schwartz Theory of Basic Human Values” (Schwartz 1992) as given in Figure 2, which defines ten basic and distinct Values:<sup>2</sup>

- **Achievement (AC):** The value here comes from setting goals and then achieving them;
- **Benevolence (BE):** Those who tend towards being benevolent are very philanthropic, they seek to help others and provide general welfare;
- **Conformity (CO):** This category of people obey clear rules and structures;
- **Hedonism (HE):** Hedonists are those who simply enjoy themselves;
- **Power (PO):** The ability to control others is important to people who possess this value and power will be actively sought by dominating others and control over resources;

Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup>Image from Greg Ver Steeg, 2014, [www.apparenthorizons.com](http://www.apparenthorizons.com).

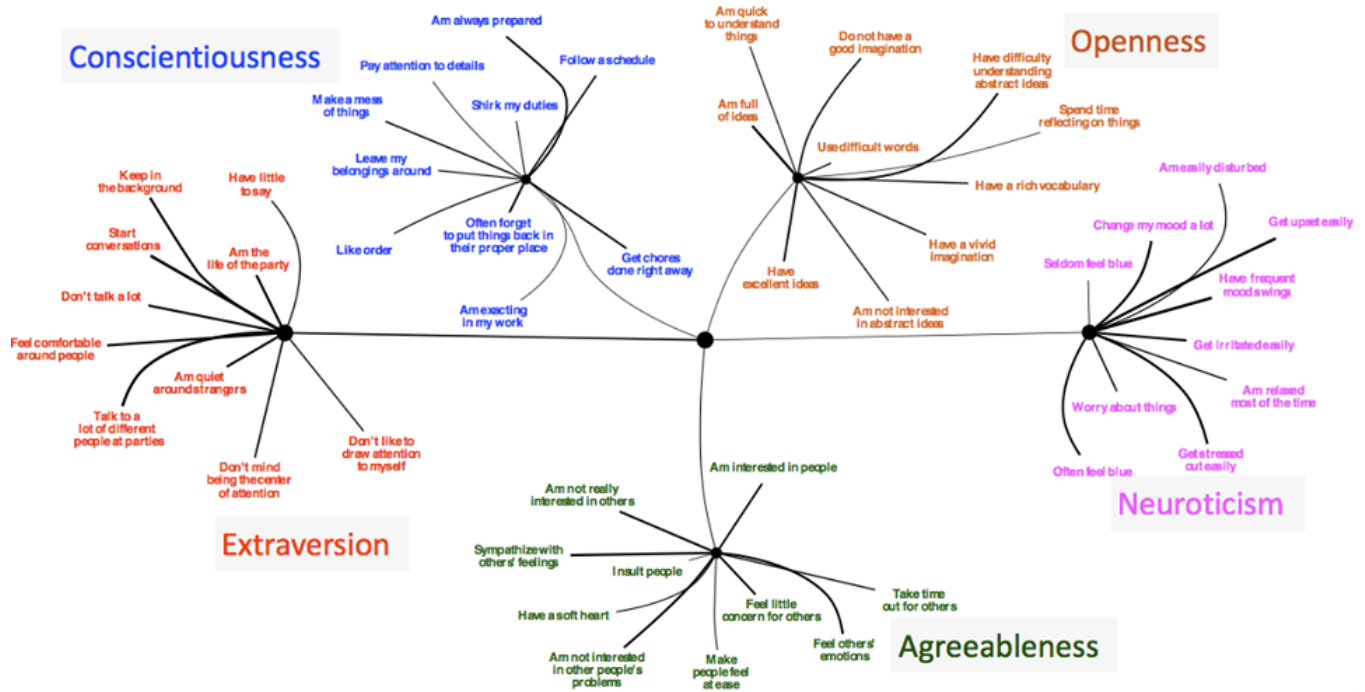
<sup>2</sup>Image from Common Cause Foundation, 2011, <http://valuesandframes.org>

- **Security (SE):** Those who seek security value, health and safety to a greater extent than other people (perhaps because of childhood woes);
- **Self-direction (SD):** Individuals who are self-directed, enjoy being independent and are outside the control of others;
- **Stimulation (ST):** Is closely related to hedonism, nevertheless the goals are slightly different. In this case, pleasure is acquired specifically from excitement and thrill;
- **Tradition (TR):** A traditionalist respects practices of the past, doing things blindly because they are customary; and
- **Universalism (UN):** Individuals who seek social justice and tolerance for all.

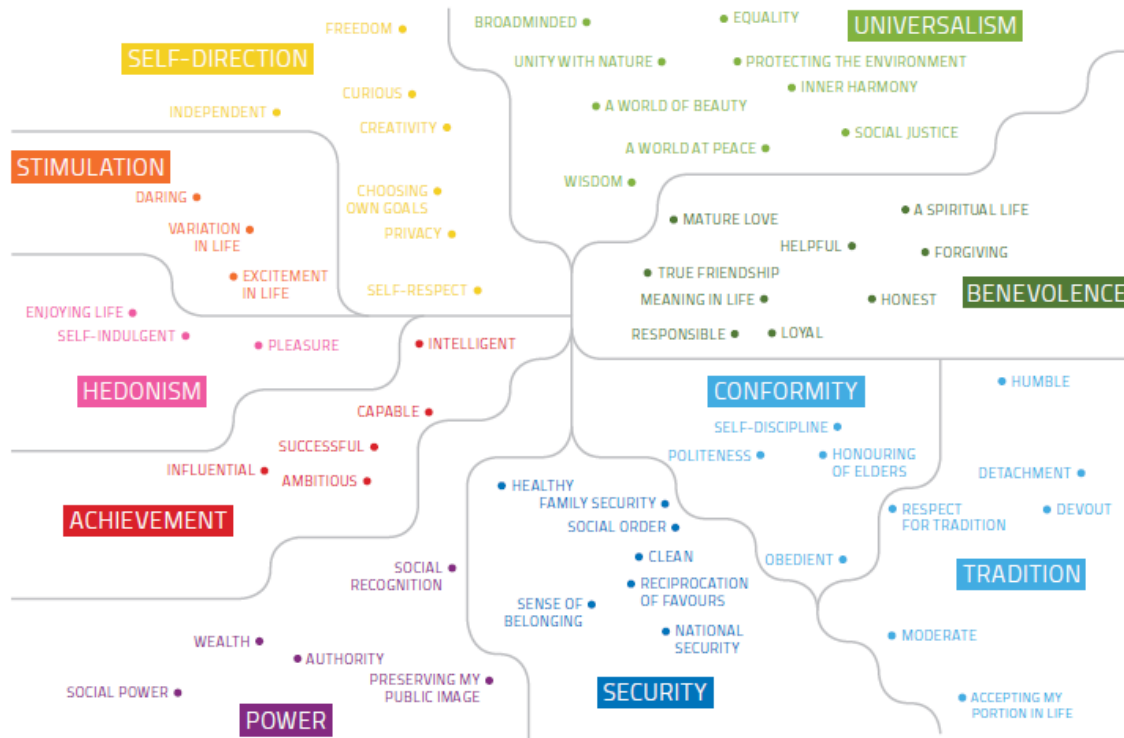
In addition to identifying the ten basic values, Schwartz’ theory also explains how the values are interconnected and influence each other, since the pursuit of any of the values results in either an accordance with another (e.g., Conformity and Security) or a conflict with at least one other Value (e.g., Benevolence and Power). Similar values can be grouped into smaller super-classes.

- **Self-Enhancement (SC1):** (Achievement, Power, Hedonism)  
Refers to individuals who yearn for power and urge to enhance themselves.
- **Openness to Change (SC2):** (Stimulation, Self-direction, Hedonism)  
This super-class is composed of all those individuals who urge to succeed, prosper and are open to changes around them and mould themselves accordingly.
- **Self-Transcendence (SC3):** (Universalism, Benevolence)  
Individuals possessing this super-class of Values are universal and generous, they believe in helping and supporting all forms of life.
- **Conservation (SC4):** (Security, Tradition, Conformity)  
This value is possessed by people who confine to rules set by the society and prefer to stay secure rather than take risks.

We believe that these kinds of models may become extremely useful in the future for various purposes such as



**Figure 1:** Big Five personality traits, also known as the five factor model (FFM) or OCEAN model. Each of the Big Five factors is quite broad and consists of a range of more specific traits.



**Figure 2:** Schwartz' Values model consists of ten human value types. The figure shows sets of more specific values that express each basic value.

Internet advertising (specifically social media advertising), computational psychology, recommendation systems, and sociological analysis over social media (e.g., East vs West cultural analysis).

### Corpus Acquisition

To start with, we ask a very fundamental question: *Is social media a good proxy of the original society or not?* (Back et al. 2010) and (Golbeck, Robles, and Turner 2011) provide empirical answers to this question. Their results respectively indicate that, in general, people do not use virtual desired/bluffed social media profiles to promote an idealised-virtual-identity and that a user's Personality can be predicted from his/her social media profile. This does not mean that there are no outliers, but we grounded our corpus collection on the assumption that is true for a major portion of the population.

Two of the most popular social media platforms, Twitter and Facebook, were chosen as sources for the corpora to validate this assumption. In addition, an essay corpus was collected. These three diverse corpora were then used for training and testing Schwartz values analysis methods.

### Self-Assessment through Portrait Values Questionnaire

A standard method of psychological data collection is through self-assessment tests, popularly known as psychometric tests. In our experiments, self-assessments were obtained using male/female versions of PVQ, the Portrait Values Questionnaire (Schwartz et al. 2001). The participants were asked to answer each question on a 1–6 Likert rating scale.<sup>3</sup> A rating of 1 means “*not like me at all*” and 6 means “*very much like me*”. An example question is “*He likes to take risks. He is always looking for adventures.*” where the user should answer while putting himself in the shoes of “He” in the question. A few exemplary items as well as the instructions and format of the written form of the PVQ are presented in Table 1. The standard practice is to ask a fixed number of questions per psychological dimension. Therefore, there are five questions for each of the ten Values classes, resulting in a 50 item PVQ questionnaire. Once all the questions in the PVQ have been answered, for each user and for each Values class, a score is generated by averaging all the scores (i.e., user responses) corresponding to the questions in that class, as described by (Schwartz 2012). Further, the rescaling strategy proposed by (Schwartz 2012) was used to eliminate randomness from each response given by a user as follows: For each user, the mean response score was first calculated considering all the responses s/he provided, and then the mean score from each response was subtracted. See (Schwartz 2012) for more details on PVQ and the score computation mechanism.

The ranges of scores obtained from the previous rescaling method may vary across different Values classes. For instance, the ranges of the rescaled scores for the Twitter

<sup>3</sup><http://www.simplypsychology.org/likert-scale.html>

corpus are as follows: Achievement [−4.12, 3.36], Benevolence [−1.56, 3.39], Conformity [−3.35, 3.01], Hedonism [−5.18, 4.35], Power [−6.0, 2.27], Security [−2.60, 2.40], Self-Direction [−1.61, 3.40], Stimulation [−5.0, 2.63], Tradition [−4.49, 3.35], and Universalism [−3.33, 3.30].<sup>4</sup> Hence the standard normalisation formula<sup>1</sup> was applied to move the ranges of the different Values classes to the [−1, 1] interval:

$$x_{scaled} = \frac{2 * (x - x_{min})}{x_{max} - x_{min}} - 1 \quad (1)$$

A ‘Yes’ or ‘No’ binary value was assigned to each Values class: if the score was less than 0, the class was considered to be negative, indicating absence of that Values trait for the particular user; while scores  $\geq 0$  were considered to be positive, indicating the presence of that trait for the user. We will use the real scores ranging [−1, 1] for the regression experiments.

Reports of psychological analysis always depend on how the target population is chosen. Therefore while we are hypothesising that a few people are more Power oriented, an open question that remains unanswered is whom they are more Power oriented than. For example, if we (hypothetically) choose parliamentarians / politicians as participants in an experiment, then the entire examined population will likely turn out to be Power oriented. Therefore, it makes sense to normalise the obtained data into two groups [−1, 0) and [0, 1] and proclaim that people with [0, 1] range scores are relatively more Power (or any other Value) oriented than the people having score ranging [−1, 0). The same normalisation mechanism was applied to all the corpora, but also after normalisation the different Values distributions were imbalanced (with the Facebook data being the most imbalanced). One possible reason behind such imbalanced distributions is that the portion of the real population using social media is slightly biased towards some Values types due to several societal reasons such as educational/family background, age group, occupation, etc. Another reason could be that the divisions between different value types simply never are balanced in any population. However, analysing such societal traits is a separate research direction altogether and out of the scope of the current study.

The PVQ questionnaire setting described above was used to separately collect textual user data separately for the Essay, Facebook, and Twitter corpora, as discussed in the rest of this section.

### Twitter Values Corpus

In the first quarter of 2016, the micro blogging service Twitter averaged 310 million monthly active users,<sup>5</sup> with around 6,000 tweets being posted every second. Therefore, Twitter came as the second natural choice for the data collection. The data collection was crowd-sourced using Ama-

<sup>4</sup>The distribution of a particular value type over a corpus was analysed using the Bienaymé-Chebyshev Inequality (Bienaymé 1853; Tchébichef 1867), showing that, for example, most of the Achievement instances (89%) were in the range [−2.96, 2.84].

<sup>5</sup>[statista.com/statistics/282087/number-of-monthly-active-twitter-users](http://statista.com/statistics/282087/number-of-monthly-active-twitter-users)

**Table 1:** An example of the instructions and format of the Portrait Values Questionnaire (PVQ). For each statement, the respondents should answer the question "How much like you is this person?" by checking one of the six boxes.

Here we briefly describe some people. Please read each description and think about how much each person is or is not like you. Tick the box to the right that shows how much the person in the description is like you.

	HOW MUCH LIKE YOU IS THIS PERSON?					
	Very much like me	Like me	Some- what like me	A little like me	Not like me	Not like me at all
1. Thinking up new ideas and being creative is important to her. She likes to do things in her original way. <b>SD</b>	6	5	4	3	2	1
2. It is important to her to be rich. She wants to have a lot of money and expensive things. <b>PO</b>	6	5	4	3	2	1
3. She thinks it is important that every person in the world be treated equally. She believes everyone should have equal opportunities in life. <b>UN</b>	6	5	4	3	2	1
4. Its important to her to show her abilities. She wants people to admire what she does. <b>AC</b>	6	5	4	3	2	1
5. It is important to her to live in secure surroundings. She avoids anything that might endanger her safety. <b>SE</b>	6	5	4	3	2	1

zon Mechanical Turk as a service, while ensuring that the participants came from various cultures and ethnic backgrounds: the participants were equally distributed, and consisted of Americans (Caucasian, Latino, African-American), Indians (East, West, North, South), and a few East-Asians (Singaporeans, Malaysian, Japanese, Chinese). The selected Asians were checked to be mostly English speaking.

The participants were requested to answer the PVQ questionnaire and to provide their Twitter IDs, so that their tweets could be crawled. However, several challenges have to be addressed when working with Twitter, and a number of iterations, human interventions and personal communications were needed to resolve all the issues. For example, several users had protected Twitter accounts, so that their tweets were not accessible when using the Twitter API. In addition, many users had to be discarded since they had published less than 100 tweets, making them uninteresting for statistical analysis. In addition, some extreme cases when users mentioned someone else's (some celebrity's) Twitter account, had to be discarded. The open source free Twitter API: Twitter4J<sup>6</sup> also has a limit of accessing only the current 3,200 tweets from any user. To resolve this issue, an open source Java application (Henrique 2015) was used. At the end of the data collection process, data from 367 unique users had been gathered. The highest number of tweets for one user was 15K, while the lowest number of tweets for a user was a mere 100; the average number of messages per user in the Twitter corpus was found to be 1,608.

## References

Back, M. D.; Stopfer, J. M.; Vazire, S.; Gaddis, S.; Schmukle, S.; Egloff, B.; and Gosling, S. D. 2010. Facebook profiles reflect actual personality, not self-idealization. *Psychological Science* 21:372–374.

Bienaymé, I.-J. 1853. *Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la*

*méthode des moindres carrés*. Imprimerie de Mallet-Bachelier.

Golbeck, J.; Robles, C.; and Turner, K. 2011. Predicting personality with social media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, 253–262. New York, NY, USA: ACM.

Goldberg, L. R. 1990. An alternative "description of personality": the big-five factor structure. *Journal of personality and social psychology* 59(6):1216.

Henrique, J. 2015. <https://github.com/Jefferson-Henrique/GetOldTweets-java>.

Schwartz, S. H.; Melech, G.; Lehmann, A.; Burgess, S.; Harris, M.; and Owens, V. 2001. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of cross-cultural psychology* 32(5):519–542.

Schwartz, S. H. 1992. Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In Zanna, M., ed., *Advances in Experimental Social Psychology*, volume 25. New York: Academic Press. 1–65.

Schwartz, S. H. 2012. An overview of the schwartz theory of basic values. *Online Readings in Psychology and Culture* 2(1):11.

Tchébichef, P. L. 1867. Des valeurs moyennes (translated into French by N.V. Khanykov). *Journal de Mathématiques Pures et Appliquées* 12(2):177–184.

<sup>6</sup><http://twitter4j.org/>