Semantic Interpretation of Social Network Communities

Tushar Maheshwari, Aishwarya N. Reganti,
Upendra Kumar, Tanmoy Chakraborty,* Amitava Das
Indian Institute of Information Technology, Sri City, AP, India, 517588
*University of Maryland, College Park, USA, 20742
tushar.m14,aishwarya.r14, upendra.k14, amitava.das}@iiits.in
*tanchak@umiacs.umd.edu

Abstract

A community in a social network is considered to be a group of nodes densely connected internally and sparsely connected externally. Although previous work intensely studied network topology within a community, its semantic interpretation is hardly understood. In this paper, we attempt to understand whether individuals in a community possess similar Personalities, Values and Ethical background. Finally, we show that Personality and Values models could be used as features to discover more accurate community structure compared to the one obtained from only network information.

Personality & Values Model

In this paper we examine Personality and Values models to understand semantic homogeneity of network communities. The Big 5 Personality model [**Openness (O)**, **Conscientiousness (C)**, **Extroversion (E)**, **Agreeableness (A)**, **Neuroticism(N)**] is being used to understand the characteristics or blend of characteristics at individual level, whereas the Schwartz Values model [**Achievement (AC)**, **Benevolence (BE)**, **Conformity (CO)**, **Hedonism (HE)**, **Power (PO)**, **Security (SE)**, **Self-direction (SD)**, **Stimulation (ST)**, **Tradition (TR)**, **Universalism (UN)**] is being used to understand and analyze inter-personal dynamics of societal sentiment (see SI Text for further details).[1]

The Personality labeled gold corpus (10K Facebook status updates of 250 users and their Facebook network properties), released in WCPR'13 workshop, was used to build the Personality model. For the Values model we crowd-sourced a Twitter corpus using the Amazon Mechanical Turk. Self-assessments were obtained using the Portrait Values Questionnaire (PVQ). At the end of the data collection process, data from 367 unique users had been gathered, having 1,608 average tweets per user (see SI Text for details).

For the automatic categorization of Personalities and Values, several psycholinguistic features were tested including Linguistic features (LIWC[2], Harvard General Inquirer, MRC psycholinguistic feature, and Sensicon[3]), network properties (Network size, betweenness centrality, density and transitivity), and Speech-Act classes.

Table 1: Performance of Personality and Values Models.

Features	Model	F-Score (SVM)	F-Score (LR)	F-Score(RF)
Lexicon	Personality	0.78	0.62	0.65
	Values	0.74	0.59	0.62
+Non-Linguistic	Personality	0.79	0.66	0.68
	Values	0.76	0.61	0.65
+Speech-Act	Personality	**0.80**	0.70	0.71
	Value	**0.81**	0.63	0.67

Support Vector Machine (SVM), Logistic Regression (LR), and Random Forest (RF) were tested for the classification tasks. The best performing Personality classification system (Verhoeven, Daelemans, and De Smedt 2013) at WCPR'13 achieved an average F-Score of 0.73. In Table 1, we see that our SVM-based model outperforms their system by achieving an average F-Score of 0.80. The Values model achieves an accuracy of 0.81.

Semantic Interpretation of Communities

The Twitter network, released by SNAP (Leskovec and Krevl, 2014) (nodes: 81,306, edges: 1,768,149) has been used to study community structure. We considered 1,562 ground-truth communities (after discarding communities having size less than 5 and with tweets less than 100).

In order to analyse whether people within the same community tend to be homogeneous with respect to their Personality and background Values/Ethics, we measure Shannon's Entropy (measure of the uncertainty) for each dimension separately. Higher entropy scores suggest lower similarity. Cross-relational entropy scores both for Personality and Values models are reported in Table 2a and Table 2b respectively. In these tables, rows represent communities having less entropy scores for the corresponding psychological dimension. For example in Table 2a, the first row AC (Achievement) represents all the communities having less entropy scores for achievement. Columns represent the fuzzy orientations of community members in rest of the dimensions. Resulting entropy scores for different Personality and Values differ greatly across communities. Therefore we normalize entropy scores using: $x_{scaled} = \frac{(x-x_{min})}{x_{max}-x_{min}}$ which keeps the range between (0,1). Entropy scores are further normalized based on the community size since the

[1]https://aishwarya-nr.github.io/AAAI.pdf

[2]http://www.liwc.net/

[3]https://hlt-nlp.fbk.eu/technologies/sensicon

Table 2: Cross-relational entropy scores for (a) Personality and (b) Values Models within a community.

Class	AC	BE	CO	HE	PO	SE	SD	ST	TR	UN
AC	–	0.01	0.13	0.09	**0.76**	**0.78**	0.08	0.78	**0.79**	0.17
BE	0.01	–	0.13	0.18	**0.79**	0.18	0.18	0.18	**0.83**	0.27
CO	0.00	0.01	–	0.09	**0.76**	0.09	**0.89**	0.89	**0.76**	0.14
HE	0.01	0.00	0.15	–	**0.73**	0.00	0.01	0.00	**0.78**	0.09
PO	0.01	0.00	0.13	0.06	–	0.07	0.06	0.07	**0.79**	0.15
SE	0.00	0.00	0.15	0.00	**0.73**	–	0.02	0.02	**0.77**	0.09
SD	0.00	0.00	0.16	0.00	**0.72**	0.01	–	0.00	**0.77**	0.10
ST	0.01	0.00	0.14	0.01	**0.73**	0.01	0.00	–	**0.79**	0.09
TR	0.00	0.00	0.10	0.05	**0.75**	0.05	0.05	0.06	–	0.13
UN	0.01	0.00	0.12	0.00	**0.73**	0.00	0.01	0.01	**0.75**	–

(a)

Class	O	C	E	A	N
O	–	**0.76**	**0.74**	0.31	0.32
C	0.00	–	**0.56**	0.21	0.20
E	0.01	**0.59**	–	0.21	0.19
A	0.00	**0.57**	**0.55**	–	0.18
N	0.20	**0.59**	**0.54**	0.19	–

(b)

Table 3: Cross-relational entropy scores among (a) Personality vs. Values and (b) Values vs. Personality Models within a community

Class	AC	BE	C0	HE	PO	SE	SD	ST	TR	UN
O	0.01	0.00	0.15	0.02	**0.70**	0.04	0.04	0.04	**0.75**	0.12
A	0.02	0.00	0.13	0.03	**0.69**	0.03	0.03	0.03	**0.73**	0.11
N	0.02	0.00	0.14	0.02	**0.71**	0.03	0.03	0.03	**0.76**	0.11
E	0.02	0.00	0.15	0.02	**0.73**	0.04	**0.44**	0.04	**0.75**	0.11
C	0.02	0.00	0.16	0.02	**0.73**	0.03	0.03	0.03	**0.77**	0.10

(a)

Class	O	A	N	E	C
AC	0.00	0.74	0.76	0.30	0.31
BE	0.00	0.74	**0.76**	0.29	0.29
CO	0.00	0.78	**0.79**	0.31	0.32
HE	0.00	**0.76**	0.76	0.30	0.30
PO	0.00	**0.91**	0.90	0.44	0.43
SE	0.00	0.75	0.76	0.30	0.30
SD	0.00	0.75	0.77	0.30	0.30
ST	0.00	0.76	0.76	0.30	0.31
TR	0.00	**0.89**	0.90	0.42	0.41
UN	0.00	0.78	**0.79**	0.31	0.31

(b)

number of members in different communities may vary. We then consider communities below the calculated threshold (median of x_{scaled} value) for further analysis. In Table 2a, we can observe from the column-wise distribution of the Achievement (AC: row 1) that the Security (SE: col. 6) people find it difficult to manage in any achievement oriented group, as SE people always want to be safe and are unwilling to go against rules; whereas AC people are always very keen to achieve their goals and are ready to take risks for the same. Another interesting observation is that Traditional (TR: col. 9) people can hardly manage themselves in any other oriented group. Similarly in TR (row 9) oriented groups other people hardly join, resulting very low entropy scores in almost all the entries. Similar trend can be seen for Power (PO: row 5) groups and for the Conscientiousness (C: row 2) vs Extroversion (E: col. 3) personalities in Table 2b.

We further attempt to draw a relationship between Values vs Personality within a community. To do so, we first extract communities having less (than the average for the particular type) entropy scores on any particular Values (*resp.* Personality) type and calculate entry scores of constituent members for all the Personality (*resp.* value) traits; the relationship between Personality and Values (*resp.* Values and Personality) is reported in Table 3a (*resp.* Table 3b). We find interesting observations as reported below – people with achievement value are less neurotic; however, they tend to show extroversion and conformity, which is psychologically justifiable. We also observe that the entropy scores of benevolent people is low with conformity, which suggests that benevolent people obey rules and regulations. Another important observation is that power oriented people have a high entropy score with all other Personality types since they are very assertive and authoritative. It is quite expected that power oriented people do not mingle with others easily.

Community Detection: Personality & Values

The immediate next question one may ask would be – given a network of individuals with their Values and Personality traits provided a priori, *can we discover more accurate community structure compared to the one obtained from only network information?*

Therefore, we use the state-of-the-art algorithm, CESNA (Yang, McAuley, and Leskovec 2013), which considers both the network structure and node attributes to detect communities.

Table 4: The performance of CESNA in terms of NMI, ARI, PU and F-Score with different feature sets.

Sl. No	Feature	NMI	ARI	PU	F-score
(i)	Network information	0.57	0.61	0.65	0.41
(ii)	(i) + Values feature	0.57	0.61	0.66	0.42
(iii)	(i) + Personality feature	0.59	0.64	0.69	0.44
(iv)	All	**0.61**	**0.68**	**0.71**	**0.45**

Table 4 presents the results of CESNA for different feature sets. With all the features considered together, CESNA achieves 7%, 11.41%, 9.23% and 9.75% performance gain in terms of NMI, ARI, PU and F-score respectively compared to the case with only the network information. Therefore, the results corroborate with (Yang, McAuley, and Leskovec 2013) considering the fact that the appropriate additional information related to nodes can significantly aid to the performance of the community detection.

Conclusion

This work unfolds semantic interpretation of communities present in social networks in terms of Personality and Values of individual. We also showed how it can be leveraged to detect more accurate communities . Our future direction would be to examine demographic psycholinguistic variance of social network communities.

References

Verhoeven, B.; Daelemans, W.; and De Smedt, T. 2013. Ensemble methods for personality recognition. In *WCPR13*.

Yang, J.; McAuley, J.; and Leskovec, J. 2013. Community detection in networks with node attributes. In *ICDM*.