A Societal Sentiment Analysis: Predicting the Values and Ethics of Individuals by Analysing Social Media Content
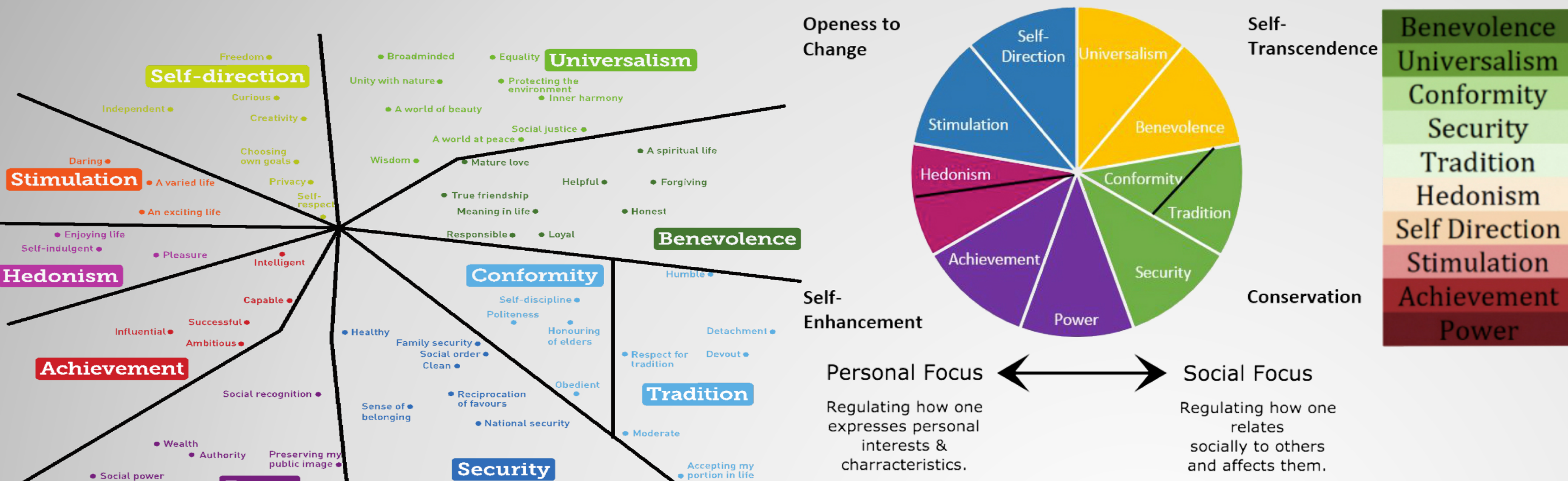
Aishwarya N Reganti ,Tushar Maheshwari, Upendra Kumar, *Tanmoy Chakraborthy & Amitava Das
Indian Institute of Information Technology (IIIT), Sri City, AP, India, 517588
*University of Maryland, College Park, USA, 20742
{aishwarya.r14, tushar.m14, upendra.k14, amitava.das}@iiits.in, *tanchak@umiacs.umd.edu

1 — Schwartz Values Definitions and Super-Classes

❖ **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.

❖ **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.

❖ **Universalism (UN):** Individuals who seek social justice and tolerance for all.

2 — Twitter Values Corpus

TABLE I: 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.

Corpus Statistics
❖ 367 users data
❖ highest no. of tweets: 15K
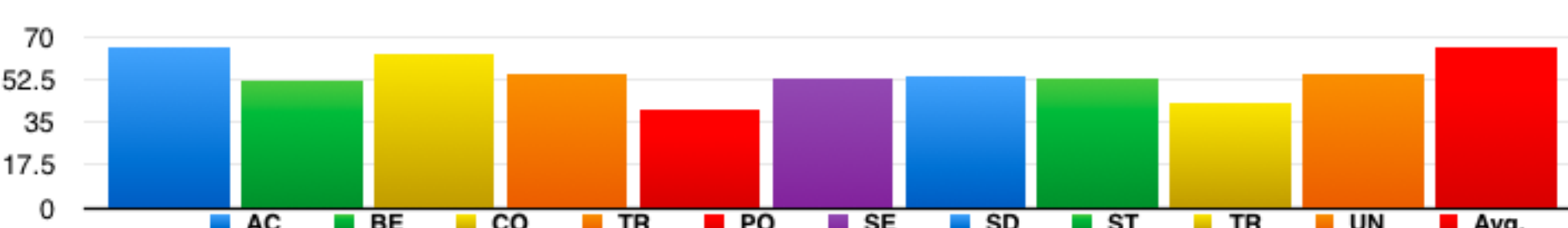❖ lowest no. of tweets: 100
❖ Avg. no. tweets 1,608

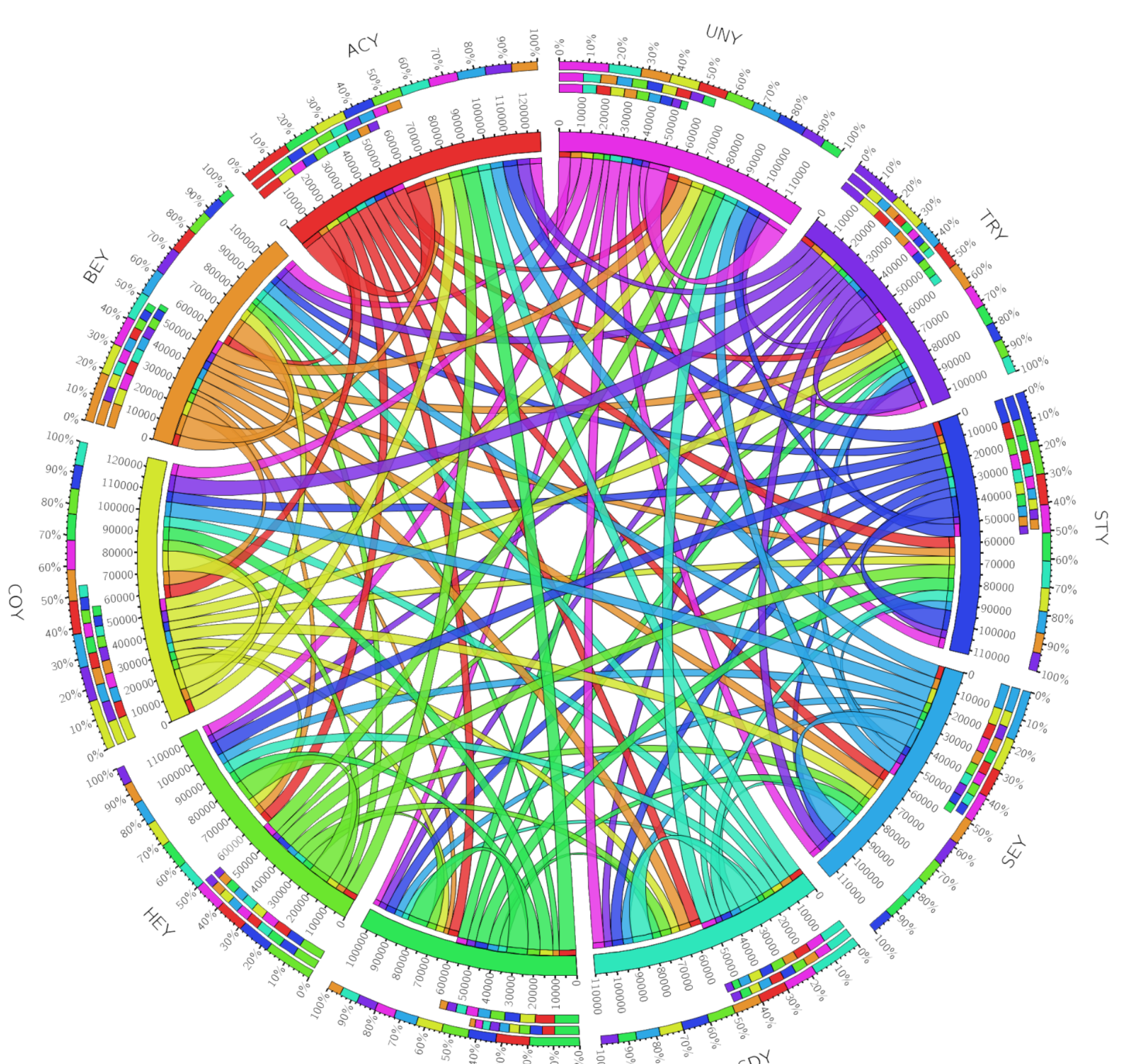**Figure 1**: Values Class Distribution in the Twitter Corpus.

Figure 2: Schwartz Values fuzziness for Twitter Values corpus using Circos. The width of each outgoing band from any value represents the degree of membership of that Value with other Values.

7 — Indian Values : Cosmopolitan Mumbai, Orthodox Delhi, Techcity Bangalore

	AC	BE	CO	HE	PO	SE	SD	ST	TR	UN	SC1	SC2	SC4	SC3
North India	37.83	15.0	100.0	15.0	100.0	45.53	100.0	100.0	100.0	100.0	100.0	100.0		33.0
South India	15.0	91.36	100.0	22.58	15.0	93.75	100.0	100.0	15.0	100.0	33.0	100.0	0	89.96
East India	18.91	98.7	100.0	55.48	45.92	100.0	100.0	100.0	70.15	100.0	63.5	100.0	66.0	97.19
West/Central India		100.0	100.0	100.0	57.97	15.0	15.0	100.0	29.09	100.0	87.22	33.0	12.55	100.0

❖ North India [ST: 100, SE:100, CON: 33, OC:100]
  ➢ [Srinagar, Amritsar, Chandigarh, Delhi, Lucknow, Varanasi, Kanpur, Indore]
❖ South India [ST: 33, SE: 100, CON: 89.96, OC: 100]
  ➢ [Bangalore, Hyderabad, Chennai, Vizag]
❖ East India [ST: 63.5, SE: 100, CON: 97.19, OC: 66]
  ➢ [Ranchi, Kolkata, Guwahati]
❖ West India [ST:87.22, SE: 33, CON: 12.55, OC:100]
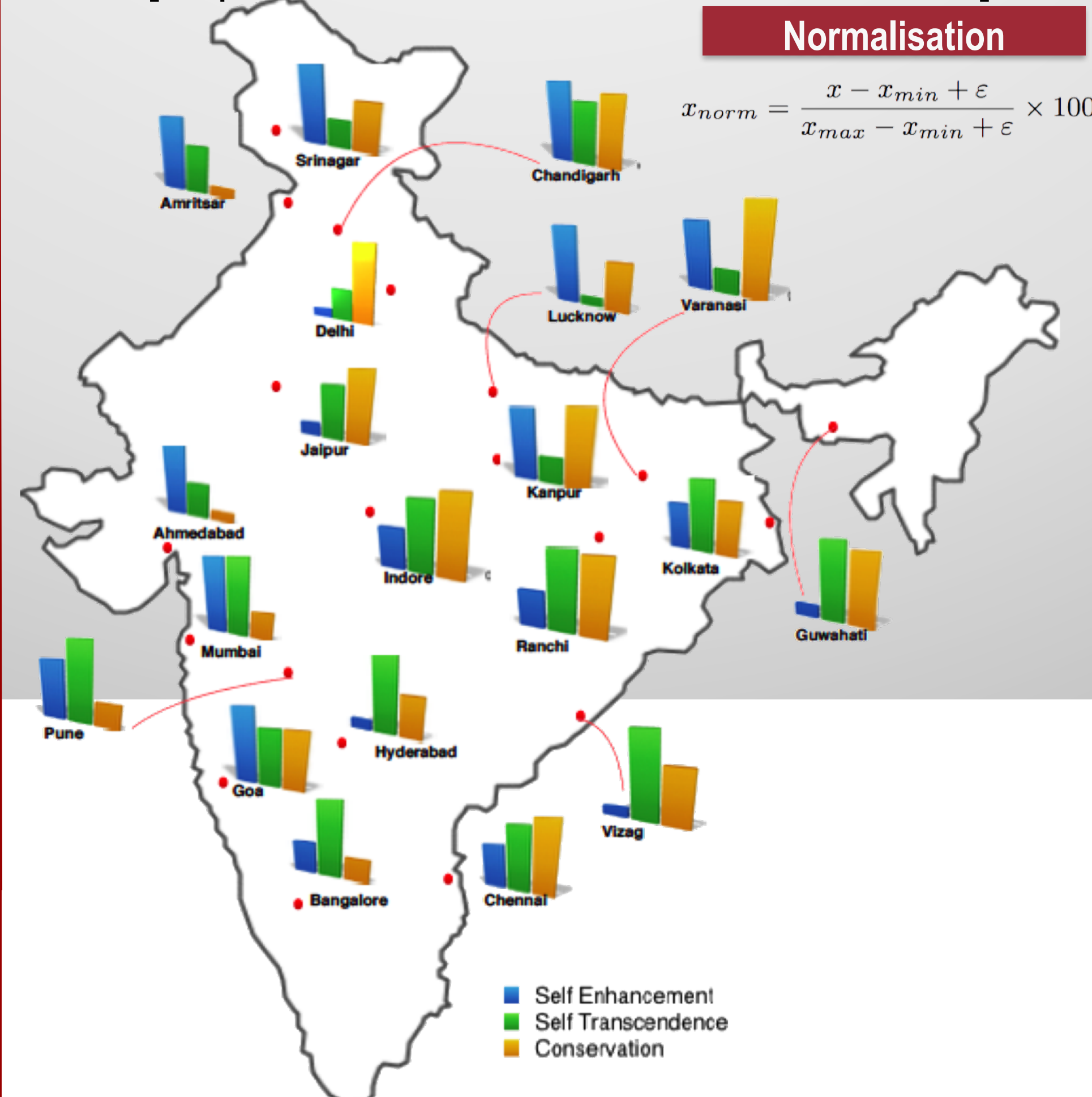 ➢ [Jaipur, Ahmedabad, Mumbai, Pune, Goa]

Normalisation

$$x_{norm} = \frac{x - x_{min} + \varepsilon}{x_{max} - x_{min} + \varepsilon} \times 100$$

Figure 3: "*Unity in Diversity*": City-Wise Indian Values: An analysis on Twitter data of 20 most populous cities of India.

3 — Features & Classifiers

❖ Linguistic Inquiry Word Count analysis
❖ N-grams (*2, 3, 4, 5*)
❖ Topic modeling (*10, 20, 50, 75, and 100*)
❖ Psycholinguistic lexica
  ➢ Harvard General Inquirer (*pleasure, pain, virtue, vice..*)
  ➢ MRC Psycholinguistic lexicon (*familiarity, concreteness, imagability*)
❖ Speech-Acts (*wh-question, yes-No question, opinion, thanking , appreciation*)
❖ Non-linguistic features (*centrality,betweenness*)

Figure 3: Best LIWC feature selection (Pearson) for each of Schwartz' ten Values.

Feature selection	AC	BE	CO	HE	PO	SE	SD	ST	TR	UN
Before Ablation	65.84	56.06	64.02	58.02	58.80	53.06	60.89	56.58	64.28	65.58
After Ablation	65.84	58.54	64.80	58.93	59.58	55.80	61.53	56.84	65.06	66.10
Number of features	52	37	65	38	54	47	46	53	39	48

Table 2: F-Scores for best feature selection on LIWC feature set for each of Schwartz' ten Values.

4 — Performance on Schwartz Values Classification

As a Multi-Class Classification Problem

Table 3: Performance of the Values model

As a Multi-Valued Regression Problem

Values	AC	BE	CO	HE	PO	SE	SD	ST	TR	UN	Avg.
PC	0.32	0.21	0.21	0.25	0.28	0.32	0.32	0.27	0.35	0.34	0.29

Table 4: Performance (Pearson) of the Values Regression model

5 — Semantic Interpretation of Communities

SNAP Dataset
https://snap.stanford.edu/

Original Dataset:
❖ Nodes: 81,306, edges: 1,768,149

We Extracted:
❖ 1,562 ground-truth communities
  ➢ only communities having <5 member, and
  ➢ Users with tweets less than 100

Cross-Entropy Relationships

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	0.01	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**	–

Table 5 : Values vs. Values

❖ Traditional people can hardly manage themselves in any other oriented group.
❖ Self-direction (SD) people find it very hard to fit into a conformity (CO) group as SD(Self-Direction) oriented people want to lead life on their own rules.

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	–

Table 6 : Personality vs. Personality
❖ Openness(O) people who are comfortable adjusting with other groups.

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

Table 7 : Values vs. Personality
❖ People with the Achievement value are less neurotic; however, they tend to show extroversion and conformity.
❖ Security, self-direction and stimulation oriented people are noticeably more inclined towards being extrovert and conform to rules.

Class	AC	BE	CO	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

Table 8 : Personality vs. Values
❖ People with an open personality are not power oriented or traditional
❖ This suggests that power and tradition oriented people are very reluctant to change themselves or not very keen to mingle with others.

6 — Community Detection

SL. No.	Features	NMI	ARI	PU	F-Score
(i)	Network Information	0.57	0.61	0.65	0.41
(ii)	(i) + value 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 9: The performance of CESNA in terms of NMI, ARI, PU and F-Score with different feature sets.

❖ CESNA along with personality and values regression features 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.