

# Generic Multilayer Network Data Analysis with the Fusion of Content and Structure

Xuan-Son Vu<sup>1</sup>, Abhishek Santra<sup>2</sup>, Sharma Chakravarthy<sup>3</sup>, Lili Jiang<sup>4</sup>,  
<sup>1,4</sup>Department of Computing Science, Umeå University, Sweden;  
<sup>2,3</sup>Information Technology Laboratory, University of Texas at Arlington, USA;

## Problem

- Analysis of complex datasets containing multiple heterogeneous features such as numeric, categorical, text-based features.
- Our goal is to adapt a Multilayer Network (MLN) approach that allows us to efficiently and flexibly analyze social network data using explicit (known or given) & implicit (derived or extracted) features of multiple datasets.

## Key Contributions

- (1): using a novel, emerging MLN approach for flexible analysis of a large complex real-world dataset.
- (2): establishing MLN's modeling benefits, flexibility of analysis, and efficiency of computation.
- (3): integrating content analysis seamlessly with structural network analysis.
- (4): extensive analysis and result validation for the social network datasets.

## Analytical Queries

(Q1) **Dominant Political Views:** How the user-declared political view (e.g., democrat, doesn't care, republican) varies across age groups in the dataset? (There are 100 political views in Facebook dataset used in this paper).

(Q2) **Relationship Status Correlation:**

- With respect to age groups, how does relationship status (e.g., single, in a relationship, and married) vary?
- How do the relationship statuses affect the personality traits of an individual? Is it different with gender?

(Q3) **Personality Trait Analysis:**

- How much of the population demonstrate contrasting personality traits (e.g., OPN and NEU)?
- How do the personality traits evolve with age? E.g., which age group of people deals better with stress?

(Q4) **Privacy Concern Correlation:**

- How does the individuals' age correlate with their comfort level of sharing personal information on social media?
- Do personality traits have a bearing on the level of privacy-concern?

## References

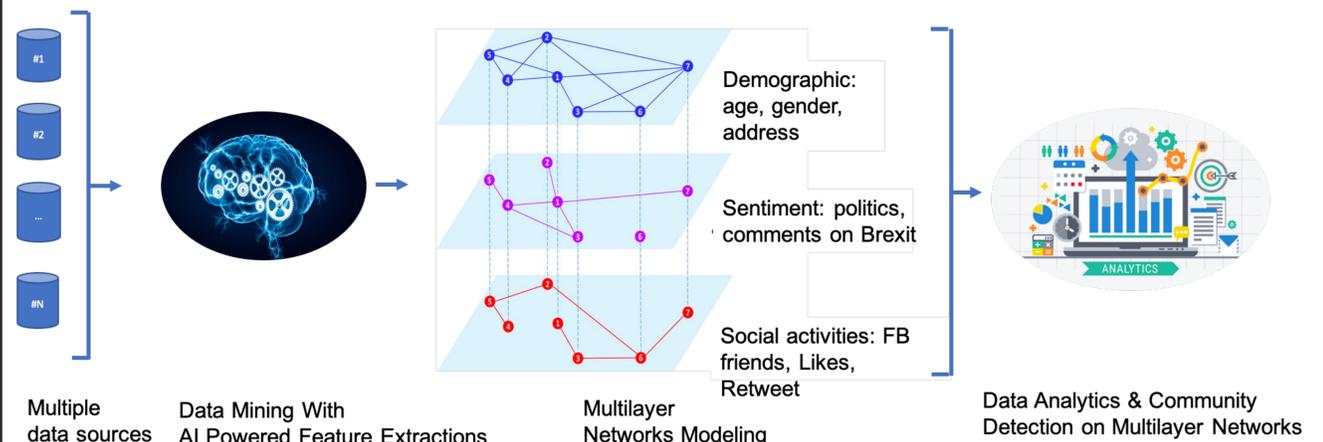
- [1] Santra, A., Bhowmick, S., Chakravarthy, S.: *Efficient community re-creation in multilayer networks using boolean operations*. In: International Conference on Computational Science, ICCS 2017. pp. 58–67 (2017).
- [2] Xuan-Son Vu, Lili Jiang: *Self-adaptive Privacy Concern Detection for User-generated Content*. Proceedings of the 19th International Conference on Computational Linguistics and Intelligent Text Processing, Hanoi, Vietnam, **best student paper award**.

## Acknowledgements

This work is supported by the STINT project funded by The Swedish Foundation for International Cooperation in Research and Higher Education.

## Multilayer Network (MLN) Analysis Model

- The model is created based on 4 datasets: (1) Demographic, (2) Political Views, (3) Personality, and (4) FB Posts. **11 layers** are created, **Privacy Concern** layer is inferred from a DNN model.



## Key Analytical Results

- **Dominant Political Views (Q1):** among the socially active youth ( $\leq 30$  years old), majority of them have the political view of "doesn't care"; above 61 years old favored republicans over democrats.

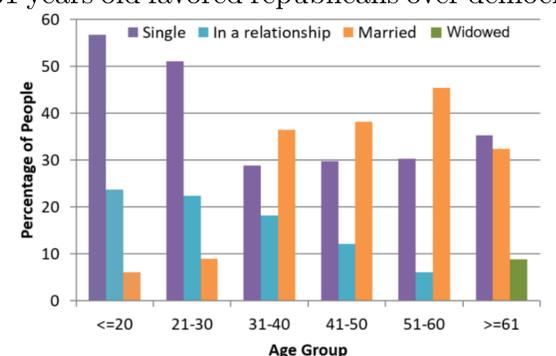
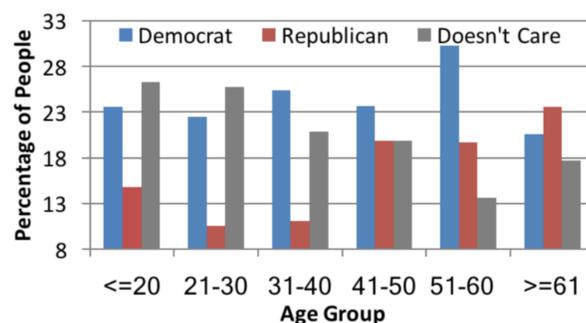


Fig. 4: Top 3 political views by age group Fig. 5: Top 3 relationship statuses by age group

- **Relationship Status Correlation (Q2):** changes in relationship status effects on the personality: S (single), R (in a relationship), M (married): married females have least OPN, and highest CON:

Trait	Males			Females		
	S(%)	R(%)	M(%)	S(%)	R(%)	M(%)
OPN (Openness to experience)	<b>55.7*</b>	54.8	47.3	53.0	55.6	43.0†
CON (Conscientiousness)	44.2†	50.2	52.7	46.1	51.7	<b>58.5*</b>
EXT (Extraversion)	45.0	<b>54.8*</b>	35.5†	49.1	50.8	44.6
AGR (Agreeableness)	42.9	50.2	31.8†	46.4	45.9	<b>50.4*</b>
NEU (Neuroticism)	36.5	26.6†	33.6	45.2	<b>53.8*</b>	48.1

- **Personality Traits Analysis (Q3):** younger lot does not deal well with stress (NEU) (Fig. 6).

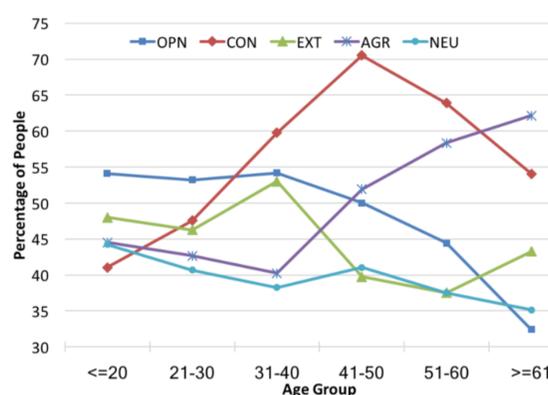


Fig. 6: Changing personality distribution with age

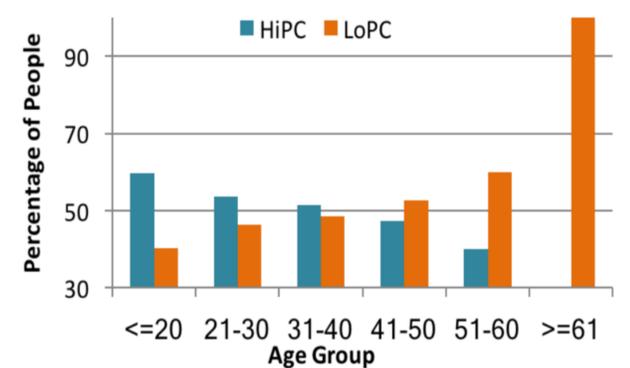


Fig. 7: Distribution of HiPC and LoPC by age group

- **Privacy Concern Correlation (Q4):** People ( $\leq 40$  years old) prefer the higher privacy level (Fig. 7), FB posts of  $\geq 41$  people contain more personal information and this increases with age.

## Efficiency Analysis

- **Efficiency Analysis:** the MLN approach with parallelism: (1) the cost of processing the most complex layer reduces  $\approx 80.4\%$ ; (2) **36%** reduction in the total time taken to answer the queries.