

Title: Towards a Set Theoretical Approach to Big Social Data Analytics: Concepts, Methods, Tools, and Empirical Findings

Ravi K. Vatrappu, Raghava Rao Mukkamala and Abid Hussain

Background: Our work is situated in the emerging fields of Data Science (Cleveland, 2001), Computational Social Science (Lazer et al., 2009) and Social Media Analytics (Vatrappu, 2013). Computational methods, formal models and software tools for big social data analytics are largely limited to graph theoretical approaches (Gross & Yellen, 2005) such as social network analysis (Borgatti, Mehra, Brass, & Labianca, 2009) informed by the social philosophical approach of relational sociology (Emirbayer, 1997). There are no other unified modelling approaches to social data that integrate the conceptual, formal, software, analytical and empirical realms (Mukkamala, Hussain, & Vatrappu, 2013).

Objective: Our objective is to present, discuss, and empirically demonstrate an alternative holistic approach to relational sociology, graph theory, and social network analysis. Our alternate holistic approach is based on associational sociology (Latour, 2005), set theory and fuzzy set theory (Ragin, 2000), and social set analysis (Mukkamala, Hussain, & Vatrappu, in press/2014).

Methods:

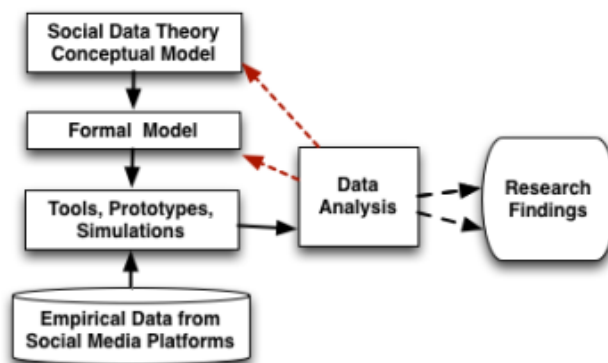


Figure 1: Overall Methodology

Our methodological approach is presented in figure 1. First, we present and discuss a theory and conceptual model of big social data. Second, we outline a formal model based on fuzzy set theory and describe the operational semantics of the formal model with real-world a social data example from Facebook. Third, we briefly present and discuss the Social Data Analytics Tool (SODATO) that realizes the conceptual model in software and provisions social data analysis based on the conceptual and formal models. Fourth, we use SODATO (Hussain & Vatrappu, 2014) to fetch social data from the facebook wall of a global brand, H&M and conduct a sentiment classification of the posts and comments. Fifth, we analyse the sentiment classifications by constructing

crisp as well as the fuzzy sets of the artefacts (posts, comments, likes, and shares) and explore correlations to the real-world outcomes of quarterly sales.

Results: SODATO was used to collect analyse facebook wall data of H&M from 01-Jan-2009 to 31-July-2013. The data corpus consists of 100,465 posts, 262,588 comments on posts, 7, 779,411 likes on posts and comments across 3,134,249 unique facebook ids/users We found statistically significant correlations between real-world business outcomes (quarterly sales) and social media activities (measures of social graph (posts, likes, comments) as well as social text (positive, negative or neutral sentiment expressions). Findings from the crisp set and fuzzy set analysis of actor and artefact sentiment reveal seasonal variation (more peaks during the spring and fall period where the fashion industry traditionally reveals new products) as well as crisis periods (for example, garment factory accidents in Bangladesh).

Conclusions: An unified approach combining social graph analysis with social text analysis as shown in the paper can help better understand the relationship between real-world events/outcomes and social media activities. Our argument is not that relational sociology, graph theory, and social network analysis are invalid or ineffective. Instead, as articulated and demonstrated in this paper, a fundamental change in the foundational mathematical logic of the formal model from graphs to sets can yield new social science insights beyond informing new formal models, computational methods and software tools.

References:

- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916), 892-895.
- Cleveland, W. S. (2001). Data science: an action plan for expanding the technical areas of the field of statistics. *International statistical review*, 69(1), 21-26.
- Emirbayer, M. (1997). Manifesto for a relational sociology. *The American Journal of Sociology*, 103(2), 281-317.
- Gross, J. L., & Yellen, J. (2005). *Graph theory and its applications*: CRC press.
- Hussain, A., & Vatrappu, R. (2014). Social Data Analytics Tool. *DESRIST 2014, Lecture Notes in Computer Science (LNCS)*, 8463(Springer), 368–372.
- Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*: Oxford University Press.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., . . . Van Alstyne, M. (2009). Computational Social Science. *Science*, 323(5915), 721-723. doi: DOI 10.1126/science.1167742
- Mukkamala, R., Hussain, A., & Vatrappu, R. (2013). Towards a Formal Model of Social Data. *IT University Technical Report Series, TR-2013-169*, https://pure.itu.dk/ws/files/54477234/ITU_TR_54472013_54477169.pdf.
- Mukkamala, R., Hussain, A., & Vatrappu, R. (in press/2014). Towards a Set Theoretical Approach to Big Data Analytics. *Proceedings of IEEE Big Data 2014, Anchorage, USA*.
- Ragin, C. C. (2000). *Fuzzy-set social science*: University of Chicago Press.
- Vatrappu, R. (2013). Understanding Social Business. In K. B. Akhilesh (Ed.), *Emerging Dimensions of Technology Management* (pp. 147-158). New Delhi: Springer.