Big Data and Fuzzy based Techniques in Analysis of Social Networks

Ali Abbasov[1], Shahnaz Shahbazova[2]
[1]Institute of Information Technology of ANAS, Baku, Azerbaijan
[2]Azerbaijan Technical University, Baku, Azerbaijan
[1]*abbasov@mincom.gov.az*, [2]*shahbazova@gmail.com*

Abstract - **Social networks – a phenomenon of an early 21st century – is perceived as a source of data generated by users about users themselves and things and dealings related to them. Because of that social networks are treated as an object of many academic, corporate and industrial research activities leading to a better understanding of individuals' behavior, (dis)likes and needs, as well as events and issues important for them. A very active involvement of individuals in social network means that every day millions of new pieces of information is generated. Analysis of this vast amount of available data requires methods and approaches taken from the domain of big data.**
The theory of fuzzy sets and systems, introduced in 1965, provides the researchers with techniques that are able to cope with imprecise information expressed linguistically. This theory constitutes a basis for designing and developing methodologies of processing data that are able to identify and understand views and judgments expressed in a unique, human way – the core of information generated by the users of social networks.
The paper tries to recognize a few important example of extracting value from social network data that can benefit from application of fuzzy set and systems methodology.

***Keywords - fuzzy sets; social network; data-driven manufacturing; statistical approaches; graph based approaches;* summarization; sentiment analysis; event detection.**

I. INTRODUCTION

More and more often social networks, or shall we say data generated by its users, is an object of research activities being conducted by variety of organization and corporations in order to extract information and knowledge about multiple aspects characterizing activities, behavior, as well as interests and likes of users and group of users. We can find multiple examples of corporations and agencies putting enormous effort to analyze and understand data that is generated by actions, interactions, and conversations involving users, as well as by the users' views and opinions on almost everything what happens in their lives and surroundings.

Varieties of techniques are used to process data generated by users in social networks: statistical approaches, graph based approaches, and many others [8], [13], [20], [21], [22], [26]. However, a human nature is present in the social networks. This means that the networks are human-like – full of imprecise relations and connections between individuals, vague terms, groups and individuals with indefinite descriptions and characteristics of interests. It seems that many aspects of social networks resemble the ones of their users.

In the light of these statements, we would like to state that techniques of processing social networks of users and groups should reflect such human facets. These techniques should be based on a human-like methodology, and the theory of fuzzy sets and systems [29] is suitable for such a purpose. Its ability to deal with ambiguous data and facts, its ability to describe things in a human manner, and its ability to handle imprecision and ambiguity make fuzzy sets [11] one of the best tools for analyzing social network [17].

II. ANALYSIS OF SOCIAL NETWORKS: TARGETED AREA

The nature of data generated by users of social networks allows us to identify multiple areas of human life, as well as industrial and corporate activities that could benefit from analyzing social network data. In the following subsections we describe just a few of them in the wide category of issues contributing to improvements in: quality of human life, manufacturing, and corporate visions and goals.

II.I QUALITY OF LIFE

The user's ability to observe and quickly react to different events – would they be positive or not – mean that analysis of data can be an important element of sophisticated and intelligent **Disaster Management** systems. An early detection of disasters, for example earthquakes or floods, would enable quick interventions that increase changes of minimizing or even mitigating the effects of disasters. Tools that allow visualization of social media data would allow for better understanding of dynamics of investigated calamities. They would allow for close, real-time monitoring of disasters and their progress, and also estimating effectiveness of preventing actions and provided help. This is closely related to another very important area that could benefit form data analysis: **Health Care**. In this case, application of big data techniques would have an enormous impact on coping with diseases – better

understanding of their origin, better knowledge about their spreading, as well as better understanding of their impact on human behavior and actions. Systems for detecting diseases and tracking them could be of great importance.

Education is yet another domain where preforming big data analysis on social network data could play an important role. For example, it is already known that analysis of posts generated by pupils and students can lead to detecting bullying and even preventing tragedies caused by such a behavior. Development of systems monitoring and analyzing variety of issues related to ways of study as well as problems and issues encountered by students would mean creation of a better environment for education. Such an environment would increase effectiveness of education systems and contribute to better-educated societies.

Analysis of social network data could also lead to a more effective **Criminal Justice** system. More information extracted from social data that is related to actions violating laws and by-laws would definitely change the way law enforcement agencies work. Ability to detect undesired behaviors as well as presence of dangerous individuals would lead to safer communities.

II.II. MANUFACTURING AND INDUSTRIAL ACTIVITIES

Facebook [32], Twitter [34], Pinterest [33] and other social sites continue to generate and provide an uninterrupted stream of data. Development of methods and approaches that analyze this data from a perspective of the users' opinions on different products and services will allow companies to identify the customers' preferences, as well as their needs and likes. This would lead to a so-called **data-driven manufacturing** – a scenario where existing and potential customers influence what is being manufactured. Analysis of the users' data would also help identifying weaknesses and strengths of manufactured goods. Social media sentiment investigations could determine if users intent to purchase specific products or if they dislike these products. All this would provide insight that can be explored and acted upon. Such analyses could also assess consumers' interest in a product before it is launched.

Introduction of tools and systems analyzing social network data would affect variety of **service and utility** companies. Such systems could change the way companies provide services to the users, and how these systems could react to changes in the demand and needs. For example, General Electric (GE) is about to release a system that uses analysis of social media to support estimation of potential problems in an electrical grid – it is called Grid IQ [35]. Many aspects of individuals' lives are being discussed on the forum of social networks therefore any company that provides services would benefit a lot from social network data. Another interesting aspect is related to a so-called geo-tagging where many aspects of collected data is location sensitive – and this alone will bring valuable information to be analyzed.

II.III CORPORATE VISIONS AND GOALS

Understanding of the users' needs and their attitude to multiple services and products offered by companies are key elements of building corporate strategies and plans for future. Also here, vast amounts of data collected during the users' activities could provide corporations with valuable information. In such areas like **advertising and marketing** any indications regarding the users' moods, attitudes and opinions expressed directly or indirectly would be able to change policies and strategies of companies. The analysis would provide companies with indicators of their social presence, their ranking, and popularity among users.

One of the most interesting and intriguing aspects of analysis of social network data could be related to strategies and operations of **insurance and financial** companies. Information – in a form of opinions, facts or evaluations – that describe the users' behaviors, patterns and rules of actions would be of great interest to insurance companies. They could adopt and customize their policies and offer insurance packages to variety of customers trying to fit their specific needs. Financial companies could use social media data to improve returns on investment – analysis of the users' discussions and posts from the point of few of sentiments and attitudes towards investment and associated with it expectations can lead to identification of trends, patterns and motivations of the users. All this would translate into better strategies and policies of investment polices of companies.

III. DISCUSSIONS AND CONCLUSION

The above section provides a just few examples of areas that can benefit from processing, analyzing and modeling social network data. At this stage we would like to foresee how synthesis of big data methods with fuzzy technologies could contribute to analysis of social networks.

Most of the work dedicated to analysis of network data is targeting structural information of networks. The interconnection and relations between actors have been the main source of information [1], [2], [3], [5], [9], [15]. Some works [4], [6], [7], [14], [19], [23], [27], [28] show an attempt to use additional information that brings different aspects of analysis. It is well known that each of the nodes/actors as well as connections/relation is associated with supplementary information. It seems very reasonable that including that information in analysis procedure and methods would increase the scope of analysis and allow looking and discovering new insights regarding the users and their behavior.

A very important aspect of analysis of network data is to look into posts, comments, tweets, or any form of text generated by the users. They contain enormous amounts of information about users: what is being currently discussed, what are moods among people, what they like, what types of things invoke positive responses, what type of things are perceived as negative. It seems critical to be armed with the ability to find answers to these questions – this would result in detecting

trends, popular and/or important topics, determine users' requirements and expectations regarding variety of items and events. In such a context, we would like to state that application of fuzzy-based approaches will enhance the capabilities of data analytical methods, will enable deeper and more semantic oriented analysis, and what is also essential will make the obtained outcomes more human-like. Fuzzy methods targeting **summarization** of texts [16], **sentiment analysis** [24], [25], **trust** inference and propagation [10], [12], and **event detection** [30][31][18], are just a few examples of research topics that constitute important issues that can be addressed with, and will benefit from application of fuzzy-based techniques and methods.

ACKNOWLEDGMENT

This scientific work was supported by Grant a Science Development Foundation under the President of the Republic of Azerbaijan, Ministry of Communications and High Technologies of the Republic of Azerbaijan in order to support projects aimed at the development of ICT for grants for 2013 for the 2nd joint ICT Contest (SDF-MCI-MQM-2 / ICT-2-2013-7 (13)).

REFERENCES

[1] Abbasov A.M. Information Boom: New Trends and Expectations, SPRİNGER, Series Title: Soft Computing: State of the Art Theory and Novel Applications (Studies in Fuzziness and Soft Computing, Editors: Yager R.R., Abbasov A.M., Reformat M.Z. and Shahbazova Sh.N. Springer; 2012, pp.1-12.

[2] S.P. Borgatti, Centrality and network flow, Social Networks, vol. 27, no. 1, pp. 55-71, 2005.

[3] J. Boyd and M. Everett, Relations, residuals, regular interiors, and relative regular equivalence, Social Networks, vol. 21, no. 2, pp. 147-165, 1999.

[4] M. Brunelli, M. Fedrizzi, and M. Fedrizzi, Fuzzy m-ary adjacency relations in social network analysis: Optimization and consensus evaluation, Information Fusion, vol. 17, pp. 36-45, 2014.

[5] Y. Cao, J. Cao, and M. Li, Distributed Data Distribution Mechanism in Social Network Based on Fuzzy Clustering, F. Sun et al. (eds.), Foundations and Applications of Intelligent Systems, Advances in Intelligent Systems and Computing 213, Springer-Verlag Berlin Heidelberg, pp. 603-620, 2014.

[6] T. Casasús-Estellés and R. R. Yager, Fuzzy Concepts in Small Worlds and the Identification of Leaders in Social Networks, A. Laurent et al. (Eds.): IPMU 2014, Part II, CCIS 443, Springer International Publishing Switzerland, pp. 37-45, 2014.

[7] S. Elkosantini and D. Gien, A dynamic model for the behavior of an operator in a company, in Proceedings of the 12th IFAC Symposium on Information Control Problems in Manufacturing, France, vol. 2, pp. 187-192, 2006.

[8] S. Elkosantini and D. Gien, Human Behavior And Social Network Simulation: Fuzzy Sets/Logic And Agents- Based Approach, in Proceedings of the 2007 Spring Simulation Multi-conference SpringSim '07, vol. 1, 102-109, 2007.

[9] L. Freeman, The Development of Social Network Analysis, Empirical Press, Vancouver, 2006.

[10] F. Hao, G. Min, M. Lin, C. Luo, and L. T. Yang, IEEEMobiFuzzyTrust: An Efficient Fuzzy Trust Inference Mechanism in Mobile Social Networks, IEEE Transactions on Parallel and Distibuted Systems, vol. 25, no. 11, pp. 2944-2955, 2014.

[11] R. Hannemanand and M. Riddle, Introduction to Social Network Methods, University of California, Riverside, 2005.

[12] S. Kim and S.Han, The Method of Inferring Trust in Web-based Social Network using Fuzzy Logic, in Proceedings of the International Workshop on Machine Intelligence Research, pp. 140-144, 2009.

[13] G. J. Klir and B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Pretience Hall, 1995.

[14] P.N. Krivitsky, M. S. Handcock, A. E. Raftery, and P. D. Hoff, Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models, Social Networks, vol. 31, 204-213, 2009.

[15] M. J. Lanham, G. P. Morgan, and K. M. Carley, Social Network Modeling and Agent-Based Simulation in Support of Crisis De-Escalation, IEEE Transactions on Systems, Man, and Cybernetics, vol. 44, no. 1, pp. 103-110, 2014.

[16] J. Liu, Fuzzy modularity and fuzzy community structure in networks, The European Physical Journal B, vol. 77, no. 4, pp. 547-557, 2010.

[17] 40 X.H. Liu, Y.T. Li, F.R. Wei, and M. Zhou, Graph-based Multi-tweet Summarization Using Social Signals, in Proceedings of COLING 2012, pp. 1699-1714, 2012.

[18] T. Matuszka, Z. Vincellér, and S. Laki, On a Keyword-Lifecycle Model for Real-time Event Detection in Social Network Data, in Proceedings of 4th IEEE International Conference on Cognitive Infocommunications, pp. 453-458, 2013.

[19] W. Pedrycz and S.-M. Chen (eds.), Social Networks: A Framework of Computational Intelligence, Studies in Computational Intelligence 526, Springer International Publishing Switzerland, 2014.

[20] M. Z. Reformat and R. R. Yager, Using Tagging in Social Networks to Find Groups of Compatible Users, in Proceedings of Join IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), Edmonton, Canada, June 24-28, 2013, pp. 697-702.

[21] Shahbazova Sh.N. Application of Fuzzy Sets for Control of Student Knowledge // Applied and Computational Mathematics, An International Journal, Volume 10, N 1, 2011. Special Issue on Fuzzy Set Theory and Applications, p.195-208.

[22] J. Scott, Social Network Analysis. A Handbook, London, Sage, 2000.

[23] T.A.B. Snijders and C. Baerveldt, A multilevel network study of the effects of delinquent behavior on friendship evolution, Journal of Mathematical Sociology, vol. 27, pp. 123-151, 2003.

[24] T.A.B. Snijders, Statistical Models for Social Networks, Annu. Rev. Soc. 2011.

[25] G. Stakias, M. Psoras, and M. Glykas, Fuzzy Cognitive Maps in Social and Business Network Analysis, M. Glykas (Ed.), Business Process Management, SCI 444, Springer-Verlag Berlin Heidelberg, pp. 241–279, 2013.

[26] D.N. Trung, J. J. Jung, L. A. Vu, and A. Kiss, Towards Modeling Fuzzy Propagation for Sentiment Analysis in Online Social Networks: a Case study on TweetScope, in Proceedings of 4th IEEE International Conference on Cognitive Infocommunications, pp. 331-337, 2013.

[27] D.N. Trung and J.J. Jung, Sentiment Analysis Based on Fuzzy Propagation in Online Social Networks: a Case study on TweetScop, Computer Science and Information Systems, vol. 11, no. 1, pp. 215-228, 2014.

[28] F. Vega-Redondo, Complex Social Networks, Cambridge University Press, 2007.

[29] R.R. Yager, Intelligent Social Network Analysis using Granular Computing, International Journal of Intelligent Systems, vol. 23, pp. 1196-1219, 2008.

[30] R. R. Yager and M. Z. Reformat, Looking for Like-Minded Individuals in Social Networks Using Tagging and Fuzzy Sets, IEEE Transactions on Fuzzy Sets, vol. 21, no. 4, pp. 672-687, 2013.

[31] L.A. Zadeh, Fuzzy Sets, Information and Control. Vol. 8, pp. 338-353, 1965.

[32] https://www.facebook.com (accessed Nov 7, 2015)

[33] https://www.pinterest.com (accessed Nov 7, 2015)

[34] https://twitter.com (accessed Nov 7, 2015)

[35] http://www.gegridsolutions.com/demandopt/Catalog/GridIQ.htm (accessed Nov 7, 2015)