International Journal of Computer Science & Communication (ISSN: 0973-7391)
Special Issue, Page84-86, December 2020 ------ available at www.csjournals.com

# Survey on Chaos identification in Social Data

**RAKSHITHA PRABHU**[1]
SJCIT, Chickballapur, Karnataka, India,

**Dr. S N CHANDRASHEKARA**[2]
C Byregowda Institute of Technology, Kolar, Karnataka, India

**Abstract:** The most important data used for the Social data analysis is cyber data. The variants in cyber data is high. Today, the information around is highly dynamic in nature. The source of the data is unknown sometimes and as well the identification is a challenge. Altogether this has created chaos in this world of internet. The information is unpredictable, unorganized in the social chaotic system. It has been observed that complex dynamic chaotic stream of information flows in the Internet. To understand this complete set of dynamics and complexity a Neural Network which can understand this chaotic data is a need. So, in this paper we are trying to understand the vast analysis done in this area of chaos analysis in Cyber Social data through Social network analysis and various machine learning methods used for data analysis.

**Key Terms**
Social Data analysis, Multi-level Social Network, Anonymous Data

## I. INTRODUCTION

Research on social network analysis has become a popular research area. With social media gaining popularity large amounts of data has been generated in everyday life. This data has been applied in various streams such as human behavior analysis, for guiding Internet advertisement efforts, crime analysis, etc. All the users in social network will have friends in their network and details of profiles. The variants of the user's taste, preferences, his or her group and other factors are described in profile functions. Each variant is given a tag. The tag is structured hierarchically making complex social networks divided into multilevel structures (person, group, community, organization, etc.).

In the field of social network analysis, simply obtaining information through processing one level of data is sometimes one-sided and superficial. The analysis results at different perspectives help people to understand the event more thoroughly. The key principles of information are understood once deeply observed the data. The critics involved, the impact it can do shall only be evident if the minute data is understood.

The structure and characteristics of data are different at each level. Based on researchers' enormous knowledge of data analysis, multilevel social network analysis process is performed comprehensively. If the algorithms are not chosen right, then the results derived are not appropriate. An easy-to-understand intelligent assistant which is unique and universal is necessary for social network data analysis. With this intelligent assistant, the sequence of information can be overlooked which can prevent any human from being a victim.

## II. LITERATURE REVIEW

*What is chaos?*

An unorganized means of information infinitely phrased, which is unintentional, confused, unordered.

*How social data is of chaos?*

The people in social network have their own network who share information with and across networks. Everyone per say is said to be different, carrying their own interests, well-being and methods. When everyone express their likes, dislikes, opinions in social world, the network people as well react to this. These information is not methodological or ordered in any means. These views turned to be meaningless in many instances. The final view of social data is Choas.

*Is Chaos theory relevant for Social Data?*

The relevance of chaos theory in social media is talked about in this article. A scientific problem is associated in phenomena of social analysis by applying chaos model. When the observations of social phenomena contain measurement errors which are generated by nonlinear dynamics ascertain considerable amount of doubt; in this case, there is a problem of external validity. Secondly, and more important, as a theory of irregular cyclical social behavior is lacking, inductive-statistical theory formation about such behavior, which is based on fitting a mathematical model of chaos to observations of social phenomena, is impossible unless additional information is used concerning the context and circumstances wherein the social phenomena occur; i.e., the internal validity of any theoretical explanation that is derived from only a fitted mathematical model (of chaos) cannot be assessed [4]. Theory-formation about irregular cyclical social behavior derived from established theoretical insights and empirical evidence is required while research into the suggestion derived from mathematical chaos theory that irregular cycles may be present during the development of social phenomena over a period, rather than fitting sophisticated mathematical models of chaos to observations of social phenomena.

*How can the Social Data be Anonymized?*

Social networks have generated huge amount of data based on its popularity. This data is available for commercial, research and many other valuable applications. On similar lines, the release of these data has alarmed the personal privacy to the fullest. Basically, there exists two set of sensitive information, one is primary and other is secondary. Primary information consists of ID and name, Secondary consists of hobbies, work, interests and so on. To protect this information the data needs to be anonymized so that it is not reachable to the other people. So, a set *K-anonymization* rule is applied for set-valued data to protect the information. But in this method the partition and search operations turned out to be expensive. Hence a proposed method of *Set_Anonymize* [1] which instead of searching the set of all values, operates on a recent search that is transacting. The method of Anonymizing a dataset is shown in Figure 1. This approach, searches for a minimum transaction and proves to be faster. However, the running time improvement and the complexity in structural data anonymizing needs to be worked upon from this work.

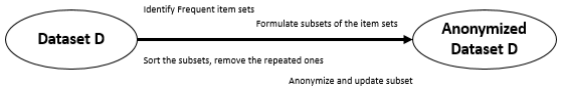


Figure 1: Anonymize dataset

*Does Multilevel Social Network Analysis need Multilevel Intelligent Assistant !*

The social network has grown to such an extent that the information does not stay a one level. This information has crossed to multi-levels and hence determining the information has been a challenge. A multi-level intelligent data processing assistant [2] is needed to understand the information in the multilevel social network. This Intelligent system is built with the ontology knowledge which features the steps involved for social network analysis, data and output requirements. The tasks and their relations are derived by the objects and properties. With this intelligent system user, can get the details process description of the inputted data along with detailed tasks and required characteristics.

This Multilevel Intelligent Assistant process the data in horizontal format. At first the raw data in the form of html, web, image, text is read. Then this raw information is combined to formulate an integrated Information which is collected from various levels of network. This Integrated information is logically segregated to a group of information which are indicative of various pointed information. This indicative information is made Tidy by removing the noise which later forms the output model. The conversion flow is shown in Figure 2.


*Figure 2 : Multilevel Intelligent Assistant Data Flow*

The different characteristic of social data can be propagated via the Multilevel Social Network Assistant to get a meaningful information. This ontology based information segregation helps an category of people to apply their data get the needed information.

*Social Network Analysis and Machine Learning Methodologies ?*

Today machine learning techniques have been proven to be useful in many of the Social Network Analysis. In this study the survey has been made on the various social network and the different algorithm that is being used on various social network. So, machine learning has been used in Fake Profile detection, Troll counting, Personality Trait detection, Depression detection and so on as described in paper [3].

These analyses have been done on major social network such as Facebook, Twitter, LinkedIn [7]. For Fake profile detection, algorithms such as SMO-PolyKernel with Decision Trees, Random Forest, SVM, ANN, PCA are being used. For Depression analysis, LDA and SVM, NB and ME, are used [6].

For mining, constructive information there are many aspects of social media analysis which can rely on machine learning techniques [8]. An in-depth analysis of a grid is required while study of Social network analysis which cannot be limited to study of social network. The need also extends in areas such as identifying the online chat-bots, identifying the emotional state of the user during publication of content and in depth analysis of the relationship between two users who are connected [9, 10].

## III. CONCLUSION

Social Data Analysis is emerging. The information shared on social media is huge, dynamic and has many variances. The above Literature review has been done to understand how the Social Data has created Choas in the Social Network and what the different methods and algorithm exists for analyzing content of the social data. With this study, it is understood that there is a need to understand Social data in a deeper sense.

## IV. FUTURE WORK

It is evident that the need of the hour is to ensure the data is rightly accommodated in Social network. Investigations across the variants that Social data needs attention. However, future study is focused on types of Chaos the Social Network deals with.

## REFERENCES

[1] Shyue-Liang Wang, Yu-Chuan Tsai, Hung-Yu Kao and Tzung-Pei Hong, "Anonymizing Set-Valued Social Data", 2010 IEEE.

[2] Man Tianxing, Nataly Zhukova1 and Georgi Tsochev (2020), "A Multilevel Intelligent Assistant for Multilevel Social Network Analysis", 2020 IEEE.

[3] Runa Ganguli, Akash Mehta, Soumya Sen (2020), "A Survey on Machine Learning Methodologies in Social Network Analysis", 2020 IEEE

[4] Chaos theory and social science: A methodological analysis, https://link.springer.com/article/10.1007%2FBF01097019

[5] L. Sweeny. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge based Systems, 10(5):557–570, 2002.

[6] P. Galán-García, J. G. De La Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, "Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying," Log. J. IGPL, vol. 24, no. 1, pp. 42–53, 2014.

[7] F. Bonchi, C. Castillo, A. Gionis, and A. Jaimes, "Social network analysis and mining for business applications," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 1–37, 2011.

[8] C. C. Aggarwal, "An Introduction to Social Network Data Analytics," Soc. Netw. Data Anal., pp. 1–15, 2011.

[9] C. S. Andreassen, "Online Social Network Site Addiction: A Comprehensive Review," Curr. Addict. Reports, vol. 2, no. 2, pp. 175–184, 2015.

[10] R. I. M. Dunbar, V. Arnaboldi, M. Conti, and A. Passarella, "The structure of online social networks mirrors those in the offline world," Soc. Networks, vol. 43, pp. 39–47, 2015.