IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⋇ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114184

Social Media Analysis

Vaishnavi D. Sonawane[1], Sheetal A. Wadhai[2]

Department of Computer Engineering, Universal College of Engineering and Research, Pune[1,2]

Abstract: Social media is very important factor in analysing modern society as a whole, their values, norms, and behaviours, as being a part of our everyday life. This study is oriented towards analysing social media in order to allow users to create their own preferences to follow (analyse) a specific social media source. The web application has been developed to allow a user to follow specific Facebook accounts and categorize the Facebook posts on those accounts based on the user defined taxonomies. Results of this study are various reports generated from the Facebook posts and their statistics that are clustered based on the user defined taxonomies. The benefit of this project is that any user can track in real time when people are talking about some topic, and it enables anyone to have better insight about society as a whole, their values, norms, what they find interesting, and many other things. This tool is also useful for different companies to track the user feedback on social networks for their products.

Keywords: Social media, people, society, generation.

I. INTRODUCTION

The study of social networks is a new but quickly widening multidisciplinary area involving social, mathematical, statistical, and computer sciences (see Burt, Minor, & Associates, 1983, for application in diverse social environments; in the latter sciences, see Wassermann & Faust, 1994, and especially for the field of economics, see Dutta & Jackson, 2003). It has its own parameters and methodological tools. In this book, we intend to show how graph-theoretic and statistical techniques can be used to study some important parameters of global social networks and illustrate their use in social science studies with some examples of real-life survey data. We hope our illustrations will provide ideas to researchers in various other fields as well.

1) Social media is defined as web-based and mobile-based Internet applications that allow the creation, access and exchange of user-generated content that is ubiquitously accessible (Kaplan and Haenlein 2010). Besides social networking media (e.g., Twitter and Facebook), for convenience, we will also use the term 'social media' to encompass really simple syndication (RSS) feeds, blogs, wikis and news, all typically yielding unstructured text and accessible through the web.

2) Social media is especially important for research into computational social science that investigates questions (Lazer et al. 2009) using quantitative techniques (e.g., computational statistics, machine learning and complexity) and so-called big data for data mining and simulation modeling (Cioffi-Revilla 2010).

3) This has led to numerous data services, tools and analytics platforms. However, this easy availability of social media data for academic research may change significantly due to commercial pressures. In addition, as discussed in Sect. 2, the tools available to researchers are far from ideal. They either give superficial access to the raw data or (for non-superficial access) require researchers to program analytics in a language such as Java.

II. CONCEPT OF A SOCIAL NETWORK

The term social network refers to the articulation of a social relationship, ascribed or achieved, among individuals, families, households, villages, communities, regions, and so on. Each of them can play dual roles, acting both as a unit or node of a social network as well as a social actor. Kinship is a very common example of an ascribed relationship, while some common examples of an achieved relationship are those that are established in the course of regular interaction in the processes of daily life and living, cultural activities, and so on, such as one household requesting help, support, or advice from another; ties of friendship or choice of individuals to spend leisure time together; and preferences in marriage. Incidentally, a relationship can also be negative—for instance, hostility or conflict as opposed to friendship or alliance and alienation versus mutuality or integration. In this book, we will focus on positive relationships. Again, much of what we will discuss is based on sociological data, but it can also be used to study demographic and economic processes such as migration from one region to another, value of any type of economic (e.g., postal money order or trade) exchange between regions, volume of flow of goods between countries, flow of traffic between different places, and so on. Thus, the units of a “social network” can be different, no doubt, such as individuals, families, households, and rural or urban areas, according to the relationship under consideration. But there is a common feature—namely, whatever the type of units we study, a specific dyadic relationship exists or does not exist between the members of any pair of them. Furthermore, if the relationship exists between a pair of For analytical purposes, a social network is conceptualized as a digraph (or a graph if the relationship has no direction). Digraph diagrams may be drawn to instantly provide direct mapping of ties showing their clustering as well as scatteredness. In a digraph, we call a unit—whether an individual, a family, a household, or a village—a vertex or node. A tie between two nodes indicates the presence of the relationship connecting them. Absence of a tie indicates absence of the relationship. A tie with a direction is called an arc, and a tie without direction is called an edge. One could also note the value or volume of flow as the weight of a tie and thus obtain a network that would then be a weighted digraph. Since the structure of the same network can be visually perceived differently depending on the manner in which a diagram is drawn, it is necessary to eliminate the bias in visual perception in order to draw an inference about the structure of a network from a digraph diagram. This visual bias is eliminated if we take recourse to numerically measure some of the selected important characteristics of a network and draw inference from there. For the sake of simplicity, we will concentrate on social networks showing only the presence (1) or absence (0) of the relationship. We also assume that ties have directions. Later, in Chapter 6, we will indicate, citing reciprocity as an illustration, how social network analysis can be extended to the case when the 0–1 restriction is dropped and there are nonnegative weights associated with the ties.

III. SOCIAL NETWORK ANALYSIS

Search for a Theoretical Base in Sociological Theories of Generalized Social Exchange Behavior—A Brief Interlude Social network analysis (SNA) means analyzing various characteristics of the pattern of distribution of relational ties as mentioned above and drawing inferences about the network as a whole or about those belonging to it considered individually or in groups. Beginning its journey as a descriptive metaphor, social network, in the course of the past few decades, has, as a parallel to the theories of market exchange, carved out a position for itself in the realm of theories and methodology for the study of society (Collins, 1988). Although its theoretical premise seems to be very close to market theories, it does not consist of looking for a best bargain in the case of an utilitarian exchange of goods and services. Rather, as a matter of generalized social exchange, it conceptualizes exchange not only in terms of economic interest but also of reciprocal role expectations as well as value orientations, social norms, and obligations. (See Homans, 1961, and Blau, 1964, for explication of the basic ideas relevant to understanding the rationale of the workings of social network, and see Turner, 1987, for a comprehensive discussion of social exchange and exchange network theories.) These attributes have made network theory more comprehensive and flexible enough to accommodate both asymmetric and symmetric relations as its natural elements. Hence, while social network theory does not deny the role of traditionally used a priori structural-functional concepts and categories in social research such as family, kinship, caste and ethnic groups, status groups, class, strata, and organization, it sees the actors and their roles and positions in a real-life situation rather in the light of the crystallization of patterns of interactions among individuals (Laumann, 1966; Wellman, 1988). This has also been discussed in detail in the context of Indian society (Srinivas & Beteille, 1964). Social network theories do not consider individuals as forming a mechanical aggregate but as an organic whole where the constituent elements are connected among themselves as well as with the others through a mosaic of ties based on interactions, directly or indirectly, at various domains such as social, economic, political, and the like. This enables a social network to be quite flexible to include the ties of the relationship of a social actor, which exist in ground reality even if those fall outside the boundaries of traditional social categories and derive appropriate ways to incorporate them in theoretical and methodological structure befitting the dynamics in social reality. Thus, SNA, unlike conventional social science methodologies, is rooted in the fact that the social universe does not consist of an aggregate of mutually independent social actors. On the contrary, they exist in a system of interlinkages and interdependence, creating and structuring ties among themselves. (Incidentally, Berkowitz, 1982, has encapsulated the concept of SNA in the wider perspective of structural analysis.)

IV. PRELIMINARIES

Methods of SNA, in fact, have evolved in an ad hoc manner according to the needs of the topic (Mitchell, 1969). Even then, the methodology of SNA shows a pattern on the whole. It is largely bifocal in the sense that we can broadly classify the methods as leading to local measures, which analyze the network attributes with respect to individual units or dyads, and the global measures, which study the characteristics of the network considered as a whole. The two types of measures are not unrelated. Rather, the latter can be obtained from the former in a few instances by some sort of aggregation, as in the case of density or reciprocity. It may also be noted that the usual statistical methods of data analysis and inference, such as measures of central tendency and dispersion, are applicable mainly in case of the former

DOI: 10.17148/IJARCCE.2022.114184

derived from personal attributes. In contrast, analysis of the global measures generally remains problematic as we will see later in the case of fragmentation, level of hierarchy, reachability, and so on. By their very nature, for such characteristics of a network, either there is no local measure or the global measure cannot be obtained by aggregation of the local measures. Vertices and arcs and their counts are the basic data used for analysis of a social network. These are used to obtain the values of various parameters of the network. We now give a few such parameters of a social network, discuss their social meaning and significance, and show how one can apply them in not only describing a social situation but also studying its dynamicity. Last, we conclude by describing, with the help of digraphs, how SNA, applied to actual data on a village community at two points of time, provides insight into its structural dynamicity.

We emphasize that in social network analysis, we study dyadic relation (which involves a pair of nodes), whereas in the usual statistical or socioeconomic surveys, one studies one or more attributes of a single node such as income, educational status, age, family size, and so on, which are assumed to be statistically independent. But the data of social network blatantly violate this assumption. For this reason, the usual statistical techniques may not always be applicable to SNA. Moreover, whereas the analytical forms of the first few theoretical moments give a reasonably good picture of a statistical distribution, there seems to be a large number of features of a social network, each of which can vary independently of the others and may not be amenable to statistical study only through their moments. Even if the measures are amenable to statistical analysis, it requires an extremely complex exercise to derive exact statistical formulae for estimation of these measures, especially when one is considering global measures. Besides, a whole social network is a unique case and, as such, has not been drawn at random from a pool of social networks. Hence, the usual mode of drawing statistical inference is also not valid. Again, global characteristics of a social network, even its outdegree and in-degree sequences, cannot be assumed to be necessarily normally distributed; they follow exponential or power law, mostly in a finite range. Hence, it requires selection of appropriate nonparametric statistical tools for SNA data.

V. NEED FOR STANDARDIZATION

For the purpose of comparing different social networks, one has to carefully standardize the parameter in question to get a measure. To illustrate this, consider two social networks N1 and N2. Suppose the vertices represent persons, and a tie from u to v indicates that u goes to v to spend (some of) his or her leisure time. How do we find the answer to the following question: Which of the networks is more cohesive? By cohesion, we mean that the actors in a network are bound closely by ties of interaction. For example, suppose N1 is a network with 10 vertices and 20 ties, and N2 is a network with 100 vertices and 200 ties. Which network is more cohesive, or are they equally cohesive? Even though the second network has a larger number of ties, we cannot conclude that it is more closely tied up. Let us look at another network to illustrate this better. Suppose N3 that has 5 vertices and 20 ties. Clearly, N3 is more cohesive than N1 since everybody goes to everybody else in N3, whereas in N1, on average, a person goes to only 2 out of the other 9. Thus, one has to standardize the number of ties in terms of the number of vertices properly before using it to compare two networks. How do we do this? The density of the network, defined as $m/n(n-1)$, provides a natural measure (Berkowitz, 1982, pp. 45–46). Here, and in what follows, n and m denote, respectively, the number of vertices and the number of ties in the network. Note that $n(n-1)$ is the maximum possible number of ties given that n is the number of vertices. Hence, the measure is $m/n(n-1)$. Density lies in the range of 0 to 1, irrespective of the number of vertices, and can be used to compare two networks. Clearly, the density is 20/90 = 22.2%, 200/9,900 = 2.02%, and 20/20 = 100% in N1, N2, and N3, respectively. Although density as defined above is a good measure of cohesion (i.e., how closely the vertices are tied up in the network), what about its validity when n is large as in the network N2? The implicit assumption in the

VI. CONCLUSION

As discussed, the easy availability of APIs provided by Twitter, Facebook and News services has led to an 'explosion' of data services and software tools for scraping and sentiment analysis, and social media analytics platforms. This paper surveys some of the social media software tools, and for completeness introduced social media scraping, data cleaning and sentiment analysis.

Perhaps, the biggest concern is that companies are increasingly restricting access to their data to monetize their content. It is important that researchers have access to computational environments and especially 'big' social media data for experimentation. Otherwise, computational social science could become the exclusive domain of major companies, government agencies and a privileged set of academic researchers presiding over private data from which they produce papers that cannot be critiqued or replicated. Arguably what is required are public-domain computational environments and data facilities for quantitative social science, which can be accessed by researchers via a cloud-based facility.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⋇ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114184

REFERENCES

- Adhikari, B. P. (1960). Construction of socio-economic models for planning. Eastern Anthropologist, India, pp. 84–94.
- Bandyopadhyay, S., & von Eschen, D. (1981). The conditions of rural progress in India. Submitted as a report to the Canadian International Development Agency.
- Bandyopadhyay, S., & von Eschen, D. (1988). Villager failure to cooperate: Some evidence. In D. W. Attwood & B. S. Baviskar (Eds.), Who shares cooperatives and rural development? (pp. 112–145). Oxford, UK: Oxford University Press.
- Bandyopadhyay, S., & von Eschen, D. (1991). Agricultural failure: Caste, class and power in rural West Bengal. In D. Gupta (Ed.), Social stratification (pp. 353–368). Oxford, UK: Oxford University Press.
- Bandyopadhyay, S., & von Eschen, D. (1995). Electoral communism and destruction of cooperation in West Bengal. In B. S. Baviskar & D. W. Attwood (Eds.), Finding the middle path (pp. 293–322).
- Boulder, CO: Westview. Banfield, E. C. (1958). Moral basis of a backward society. Glencoe, IL: Free Press.