

The Usage of Information Technology Tools for Social Network Analysis



by Cosmin TĂNASE

An Honorary Member of the Romanian Distribution Committee

Abstract

Nature has shown many different ways of creating life. The essential thing about life is that it requires great cooperation. We see cooperation starting at an atomic level through cells forming complex organisms. As these structures get complicated, cooperation also gets complicated. It does not require so much to build stable structures at an atomic level but at higher levels, things are much more complicated. Living beings have created tribes, cliques, states, etc., to survive. Maybe it is not wrong to say that the ‘survival of the fittest’ is strongly tied to the ability of creating strong communities. These forms of living together ensure the viability of the species. Different species created different forms of living together. As one of the most sophisticated species, the human species also has built different structures, like tribes, societies, etc. What makes human beings different from other living beings is that they possess the ability to examine the social networks and structures built and think about them. Besides, the knowledge and intellect of studying them can shape them and as well as they get shaped by them.

Keywords: Machine Learning, Connections, Opinion Mining, Computational Tools, Analysis, Networks

JEL Classification: C88; D83; M31; M37; O33

Social networks have always been pervasive in our daily lives but with the prevalent use of information technologies in modern societies, social networks evolved to become social media networks. Social media is shaping societies in many profound ways. It is imperative to create better tools and methods to understand them. We need newer approaches and utensils to create satisfactory documentation and visualization of these networks to understand ‘human behavior’.

Although it is a simple and somewhat clear truth, it should be stated again that social media networks are networks. The science dealing with the study of patterns of connections is network science. Network science deals with all kinds of patterns of connections in physical and also in social occurrences. Tools, methods and methodologies of network science can be applied to local area networks in computer engineering, to swarm behavior of *Sturnus vulgaris*, electricity networks, city or country maps, neural networks, etc. A network consists of entities or things which are called nodes, vertices, agents, items or entities and, in case of social networks, also people. They can represent people, workgroups, teams, institutions, countries, cities, etc. They can also represent things like Web pages, videos, documents and so on.

Considering all kinds of connections among a group of people and things, one can define social networks as being created from such a kind of collection of connections. From that point of view, it is evident that social networks are a subset of networks. These kinds of networks are created whenever people interact with other people, institutions or artifacts. This interaction may be direct, e.g., one person sending an email or a letter to another person or it may be indirect, e.g., broadcasting a video on YouTube. The connections between the vertices are called edges, ties, links, relationships or connections. They can represent a friendship, a relationship, a transaction, some shared attributes like students of a class, members of a club, etc. Edges can be directed or asymmetric. Such kinds of edges have a clear starting vertex, an origin and a clear end vertex, a destination. Their visual appearance in a graph is mostly an arrow from one vertex to another vertex. A line without an arrow in a graph is called an undirected, symmetric or mutual edge. Directed edges can be reciprocated or not, e.g., if one person sends a letter to another person, the receiver can answer this letter with another letter or not. Another property of an edge is its weight; an edge can be weighted or unweighted. Unweighted edges are also called binary edges because they only show the existence or non-existence of an edge. In contrast to them, weighted edges include some values that are related to that edge. These values show the strength of a connection. Commonly, for the analysis, edges are stored and represented in matrices or edge lists.

In many cases social networks are considered from an individual's point of view. In these cases, the individual is called ego and the connected people are called alters. When networks include only people connected to a specific ego, they are called egocentric networks. Networks are also examined by their completeness. A full or complete network contains the subset of people or entities who match some interest or attribute and includes information about the set of connections among them all. Networks connecting the same type of entities are called unimodal.

In applications like Facebook, Twitter or web sites, like Imdb.com or Amazon.com, there are connections not only between people but also between people and documents, like pictures, videos, Web pages, etc. In such cases where the vertices can be of different types, the network is called multimodal. The analysis of multimodal networks is cumbersome and their handling is more difficult than unimodal networks. For the sake of better and meaningful analysis, such kinds of multimodal networks are often transformed into simpler unimodal networks. The same definitions and thoughts also apply to the edges; in many practical applications, edges are considered to be only of one type, like two people who send each other e-mails or messages. In such cases, the networks show a uniplex

relation symbolized as a single connection between the entities. In fact, the interaction and relationship between people can be of much broader complexity, e.g., a woman can be the wife of a man but also his manager at work. For such kinds of relationships networks with different type of edges could be constructed, although they are not very common. Such kinds of networks with multiple connection types are called multiplex networks. Initially the measurement of social networks was based on simply counting the connections. With the increase of computational power and strong computational tools, more sophisticated and complex measures have been developed in network analysis. Some basic concept of the measurement methods can be summarized as follows:

- **Density:** Density can be defined as the division of existing relationships by the total number of possible relationships which could be presented in a network.
- **Centralization:** Centralization as an aggregate metric designates the “amount to which the network is centred on one or just a few important nodes”.
- **Degree centrality:** Degree centrality is a “simple count of the total number of connections linked to a vertex”. It is regarded as a sort of ‘popularity measure’. Although it measures popularity, it should be noted that it does not differentiate between quality and quantity.
- **Degree:** “Degree is the measure of the total number of edges connected to a particular vertex.”
- **Distance:** The distance is given as the minimum number of neighbour-to- neighbour hops from one vertex to another. In the context of social media analytics, the vertices are considered as people and the distance is measured between people who are not neighbors.
- **Geodesic distance:** The shortest path between two people.
- **Structural hole:** A ‘structural hole’ is a term for recognizing a missing bridge.
- **Closeness centrality:** Closeness centrality measures each individual’s position in the network. It captures the average distance “between each vertex and every other vertex in the network”. If we assume that nodes can only pass messages to their existing connections or can only influence them, a low closeness centrality indicates that this node which is considered to be people is directly connected or ‘just a hop away’. In contrast, nodes in distant or peripheral spots can have high closeness centrality scores. This case designates that these nodes/vertices/people need a high number of hops to connect to these peripheral locations, which is also an indication of more connections.

With the advent of information technologies and computers, social analytics also changed. The immense computational capacities and power of computers is both a curse and a blessing for social analytics. Now it is possible to visualize and analyze amounts of data very fast. On the other hand, the pervasiveness of information technology devices means the increase in data and also raises the complexity of networks. Today we are speaking about ‘big data’ but technologies, like the Internet of Things, and approaches like Industry 4.0 will push the ‘big’ into the direction of ‘huge’. The good

news, for example about IoT, is that data is and will be generated in machine-readable form so that automated analysis becomes easier after being freed of the burdens of data collection and preparation. Artificial intelligence, data mining, machine learning and all related computer science tools promise a better handling of social media networks and a better understanding of human interactions.

The drastic shifts from read-only to read-write access to the Web lead the people to interact with each other through social media networks, like Wikis, online forums, communities, etc. Due to this, user-generated content through social media platforms is increasing tremendously. Specifically, Web-based data of the form—opinionated text, reviews of products and services has been one of the most contributing factors in social big data.

Sentiment Analysis

Sentiment analysis is a relatively new field of research. It has blossomed rapidly since the 2000s. Even though there is some earlier work on this subject, research concerning sentiment analysis has surged up as a result of the social platforms, such as Twitter and Facebook getting prevalent. Among the terms that refer to the same or similar fields of research are sentiment mining, review mining, opinion mining, opinion extraction, affect analysis, subjectivity analysis, emotion analysis, etc. The essential issues in sentiment analysis are to identify how sentiments are expressed in texts and whether the expressions indicate positive (favorable) or negative (unfavorable) opinions toward the subject. Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities, such as products, services, organizations, individuals, issues, events, topics and their attributes. Some concepts are closely related to sentiment analysis. Among them are opinion, view, belief, conviction, persuasion and sentiment. **Sentiment analysis** is essentially a document classification task. Machine learning is one of the best-known techniques of classification. In general terms, machine learning is the technique whereby computers make generalizations resting on inductive inference strategies. More specifically, machine learning is learning a hypothesis that converges with a set of example input-output pairs to the function that maps the inputs to the outputs.

Another crucial question that needs to be answered when doing sentiment analysis is: Which learning attributes to use? An answer to this question can be given by detecting which attributes are most informative for any given learning task. A common assumption in the field of sentiment analysis is that adjectives and adverbs are the most informative words and, hence, the words of other syntactic categories can be ignored. The sentential positions of words, the syntactic structures of sentences, the semantic features of words and sentences and the discursive properties of texts can also be used as learning attributes when doing sentiment analysis with machine learning. However, the extraction of attributes of these kinds largely necessitates to do natural language processing.

Some researchers maintain that machine learning is sufficient as a technique to do Natural Language Processing (NLP). It is a fact that almost every contemporary approach to NLP involves a learning-based element, however, we still hold the view that linguistic knowledge bears non-negligible importance to this field, and hence, to sentiment analysis. The core work of rule-based NLP is to

transform texts in natural language to general-purpose linguistic structures, such as phonemes, morphemes, syntactic categories, tree-structured representations of phrasal structures and logic-based representations of meaning.

There are many ways in which rule- and learning-based techniques can be combined in sentiment analysis. Texts to be analyzed often need to be pre-processed for various reasons. Before feeding the text to the classifier, it might, for instance, be desirable to remove uninformative characters from it, to tokenize it into words, or to analyze its words into morphemes. Any of these tasks can be accomplished by using rule-based methods. Words are often used as learning features in sentiment analysis applications. Still, sometimes phrasal structures might be informative for a learning algorithm. It is true however that advanced machine learning techniques have cleaned up expert-engineered representations from the peripheral areas of NLP, namely those concerning phonetics and text/discourse pragmatics. But learning-based algorithms do not perform satisfactorily when the task is more abstract, such as detecting phrasal structures or conducting reasoning processes. Parsers based on syntactic knowledge or logic-based reasoners can handle such abstract tasks to some extent. Paradigms, like inductive logic programming, can serve to combine learning- and rule-based approaches when doing sentiment analysis.

CONCLUSIONS

Life is social and species can only survive within a well-defined structure. Humans also created different social structures. The study of these structures is not only an intellectual amusement but also vital for the species. With the rise of information technologies, the creation and the analysis of social network analysis reached another quality, and not only is the quality affected, but also the quantity of social networks. The good news about the extensive usage of information technologies is that all data created is in a machine-readable form and we have now the strong computational powers to analyze them. The bad news is that it turns out to be a vicious circle; we have the ability to analyze big data much faster, but we also have the power to create much more data. With new technologies it will be not only the human creating the data but also machines will contribute on a large scale. In a time where data is created very fast but also destroyed and forgotten, the need to understand human relations still exists and becomes more important. New approaches are needed and deep learning looks like a promising direction to handle the upcoming complexity and break the vicious cycle.

References

- [1] Arif, T. 2015. The mathematics of social network analysis: metrics for academic social networks, *International Journal of Computer Applications Technology and Research*, 4(12): 889–893
- [2] Eisenstein, J. 2019. *Introduction to Natural Language Processing*, The MIT Press.
- [3] Freeman, L.C. 2004. *The Development of Social Network Analysis. A Study in the Sociology of Science*. Vancouver BC North Charleston
- [4] Hansen, D.L., Shneiderman, B., Smith, M.A. and Himelboim, I. 2020. *Analyzing Social Media Networks with NodeXL. Insights from a Connected World*, Morgan Kaufmann Publishers.

- [5] Jiang, Y., Meng, W. and Yu, C. 2011. Topic sentiment change analysis. Machine learning and data mining in pattern recognition, MLDM 2011. Lecture Notes in Computer Science, 443–457.
- [6] Jurafsky, D. and Martin, J.H. 2020. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson.
- [7] Pathak, R.A., Agarwal, B., Pandey, M. and Rautaray, S. 2020. Application of deep learning approaches for sentiment analysis. pp. 1–31
- [8] Zhai, S. and Z.M. Zhang. 2016. Semisupervised autoencoder for sentiment analysis. pp. 1394–1400. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16)

