Social Network Mining – an Overview

Marenglen Biba
Editor in Chief. *International Journal of Social Network Mining*
www.inderscience.com/ijsnm

The growing use of the Internet has led to the birth of networked interaction environments such as social networks. In these kinds of networks, intensive and complex interactions take place among many different entities, leading to huge information networks with outstanding business potential. However, the exponential growth in the number of interacting nodes in these networks has given rise to important challenges for the machine learning and data mining community. Most social networks have exceeded the dimensions for which it is feasible to perform a thorough and accurate analysis with traditional data mining methods.

Traditionally the area of machine learning has dealt with data which were usually not networked. More recently, the learning and mining community has started facing difficult challenges in upgrading previous traditional methods and algorithms in accordance with the new scenario which is being created both by the extremely rapidly growing amount of available data coming from social networks, and by the inherently complex nature of the data produced by rich social interactions. Social network mining is a growing research area which aims at bringing together researchers from different fields such as machine learning, data mining, artificial intelligence, optimisation, graph theory, networks, mobile computing and other areas, with the goal of attacking important problems that the birth of social networks has brought into the scientific arena. In the following we present some areas of growing interest within social network mining that deserve the attention of future research efforts due to their high complexity.

The dimension problem is of critical importance to the success of the analysis of social networks. The exponential space to explore for a thorough analysis of social networks calls for contributions from well-established fields that have for a long time dealt with searching combinatorial spaces, such as the areas of research on metaheuristics and stochastic local search, sub-areas of artificial intelligence dealing with search, optimisation and constraint programming, and other related fields. In all of these areas the classical problem addressed has been the exploration of huge spaces of solutions. Social networks have presented the same old problem for these areas but under a new perspective which is related to the complexity of the data at hand. Now the algorithms have to handle complex spaces created by social interactions which, being based on human nature, are often highly unpredictable and hard to understand.

Recently, researchers in the above areas have started to adapt their methods and algorithms to the analysis of social networks. However, more joint interdisciplinary effort is required from different research actors towards a unifying framework for social network mining where the process of search of huge spaces of solutions is well-defined together with new features introduced by social networks.

Another approach to addressing the dimension problem is trying to compress social networks in a way that the compression process does not lead to loss of crucial information such as that used for neighbour queries which are typical in social network mining. The compression process requires contributions from old subfields of computer science such as data structures, graph theory and computational theory. The challenge is to be able to compress the network while not only preserving the original information, but also keeping at the same time the complexity of the operations for mining the network without decompressing it.

A problem closely-related to the large-scale nature of social networks is the necessity to perform aggregation operations. These are often very helpful in analysing social networks when trying to extract interesting patterns from social interactions. However, dealing with aggregation in tables with terabytes of data is prohibiting if the analysis has to be useful for fast decision making. This gives rise to interesting research challenges with contributions needed not only from the machine learning and data mining

communities, but also from other areas of computer science such as databases, data warehouses, knowledge representation and compilation, and other related fields.

Recently, a problem which has had much research dedicated to it is link prediction. This problem is defined as discovering relationships among entities in a social network which can be directly or indirectly connected. The difficulty introduced by social networks is related to the multi-relational nature of the data and the uncertainty which is due to noisy, missing or partially observed variables. In this context, joint contributions are required from fields that have long dealt with multi-relational data (or learning from structured data) and statistical methods that are able to deal with uncertainty in a robust manner. Growing research areas such as statistical relational machine learning have started building models for dealing with both the structural complexity of the domains involved and the inherent uncertainty that characterises all of the environments where social interactions take place intensively. These models, mainly based on probabilistic graphical models and relational languages (or logic-based approaches), are among the most promising ones to deal with the complex problems in social network mining. However, the large-scale application of statistical relational learning to social network mining calls for further research efforts in this direction.

A distinguishing feature of social networks is that often in these kinds of information networks, multiple phenomena diffuse in competition with one another. Handling competitive diffusion in large-scale networks is not an easy task and this problem needs further research effort in order to handle networks with millions of vertices. Typically these problems may be translated into satisfiability tasks for finding the most probable state among many potential candidates. Providing a scalable solution to this problem leads to the need for contributions from researchers working on tractable inference, stochastic search, knowledge compilation or theory of approximation.

One of the most interesting operations in a social network is finding the most influential persons in the network. This task has outstanding marketing value since the most influential entities can be used for a highly targeted campaign for new products. However, the problem of discovering the most influential node has proved to be NP-hard and exact algorithms are not appropriate. This leads to a growing interest in approximation methods and algorithms that are able to produce solutions with proven approximation guarantees.

Mining dynamic behavioural patterns is a highly interesting task in social networks due to its relevant business value. Understanding the behaviour of users in a social network might open powerful channels towards marketing products in a more personalised manner. However, most of the time, social interactions correspond to infrequent and hard-to-detect interaction patterns which at the same time also show meaningful features. From a machine learning and computational point of view this is equivalent to discovering periodic or near-periodic subgraphs in dynamic social networks. Efficient, effective and scalable methods are required to solve this problem and this requires contributions from areas that have long dealt with the analysis of dynamic networks.

Explicit interactions correspond with actions which can be identified as having known actors producing these, and often the nature of these interactions is well-understood. However, social actors are involved in multiple relationships simultaneously which will not necessarily correspond to direct known interactions or patterns. Often, operations such as tagging or posting may lead to the construction of a graph which is to be considered as implicit. Discovering patterns created implicitly is a challenging task in knowledge discovery and social networks have pushed this challenge towards new frontiers due to the many different forms of interactions and the heterogonous nature of the actors involved. Contributions from areas of temporal and spatial data mining are welcome for this exciting new problem in social network mining.

From the point of view of mining and analysis of social networks, it is also necessary to develop efficient sampling methods which are tractable and uniform. This is important if a representative sample of large social networks is needed in order to provide a practical estimation of the properties of the social network. The two approaches for analysing the social network, i.e. collecting direct data and sampling the network, are both important for a thorough analysis of the properties of the network and for mining interesting

patterns on the network. Graph sampling techniques are very helpful for tackling this problem and represent an important area of investigation for future research.

Monitoring the evolution of a network of entities is considered to be an important task not only from a computing point of view, but also from a business perspective. This is due to the demand for capturing changes in users' attitudes towards a certain product or service. However, discovering the dynamics underlying the evolution of social networks requires a multidisciplinary effort from several research areas. Mining dynamic networks in the presence of millions of nodes, with sometimes partially observed entities or noisy data, makes the task of mining the evolution and drift of social networks one of the most challenging.

Finally, developing methods for analysing and mining social networks also has to take into account the privacy protection of social network users. Privacy-preserving data mining has successfully dealt with the privacy problem for traditional scenarios, but social networks have brought such a variety of interactions and therefore typologies of data that it has become crucial to develop ways in which such rich interactions can be anonymised when mined and analysed.

To conclude, social network mining is a growing, exciting area of research that has in front of itself a long way to go, with the contribution of many research fields. The broad range of problems and challenges in mining social networks calls for powerful new methods and algorithms towards new frontiers in understanding the exciting phenomenon of social networks.

The recently published *International Journal of Social Network Mining* (*http://www.inderscience.com/ijsnm*) is a global journal in the field of social network mining. Its Editorial Board is a balanced mixture of both practitioners in real social networks and researchers creating new theoretical models for analysis and mining. *IJSNM*, with an Editorial Board that represents the five continents and many countries, has a broad appeal whether contributors work in business, government or academia, regardless of nation of practice.