A Survey on Tie Strength Estimation Methods in Online Social Networks

Isidoros Perikos¹ and Loizos Michael¹,²
¹Open University of Cyprus, Nicosia, Cyprus
²CYENS Center of Excellence, Nicosia, Cyprus

Keywords: Tie Strength, Social Networks, Social Relations, Social Interaction.

Abstract: Social networks constitute an important medium for social interaction where people communicate and formulate relationships in a way similar to what they do in real life. The analysis of the users’ relationships in social networks can lead to new insights into human social behavior. Tie strength constitutes a core aspect of social relationships, which represents the importance of a relationship and the closeness of individuals. Understanding the key features of tie strength in social networks can assist in formulating more efficient user-centric services. This survey paper examines the advances in the area of the analysis of tie strength in social networks. We study the dimensions of tie strength and review the key predictive features for each dimension. We then undertake a comparative study of methodologies to model tie strength and examine the key findings. Finally, we discuss open issues and challenges in specifying tie strength.

1 INTRODUCTION

During the last years, the advancement of social networks has completely redefined the way that we conceive our social relationships and has created the sensation of having broken the barriers of time and geography that are limiting people’s social world (Pappalardo et al., 2012). With the rising expansion of the social networks the capacity of the people to interact, communicate and network has been greatly increased (Liberatore and Quijano-Sanchez, 2017). Social networks create new online environments where social relationships can not only map and develop preexisting relationships that are established face-to-face in the physical world, but can also facilitate the development of new relationships that may exist and evolve only in the world of the social network (Arnaboldi et al., 2015).

Social networks constitute environments where social ties among individuals are developed and have become a predominant medium for social interaction that have changed completely the way of human communication. People connect and formulate relationships in social media in a similar way they do in the real world (Dunbar et al., 2015). Actually, the formulation and the development of social relationships is actually what makes social media ‘social’ (Gilbert and Karahalios, 2009).

Although individuals have different types of relationships with other ones that differ in kind and closeness, social networks do not distinguish and diversify them. So, they treat all connections the same, even if one relationship refers to a trusted friend and another to a total stranger. In social networks, the online connection between users is often defined as a ‘friendship’ or ‘follow’ or ‘connected’ manifestation and a connection between two individuals is considered to either exist or not. So, different types of friendships best friends, total strangers and long acquaintances are all grouped under the label of ‘friend’, ‘follow’ or ‘connection’. In this regard, social networks do not make distinctions between best friends, that are the relationships that one trusts, and mere acquaintances and so all the relationships are uniformly labeled. However, some social relationships and connections are stronger than others. It is natural for people to have not only friends but also best friends and also to distinguish friends from acquaintances (Jones et al., 2013). Social scientists highlighted this point of the social connections and researchers utilize the expression of the tie strength to pertain to this concept (Granovetter, 1973; Marsden and Campbell 2012).

Tie strength is a key concept in social networks associated with the value which is placed by individuals on their relationships referring to the
general sense of closeness with another individual (Granovetter, 1983). So, relationships in social networks can be measured with the currency of the tie strength (Gilbert and Karahalios, 2014), a concept introduced by Granovetter (1973) which has become the metric to measure the relationships in social media. In this context, two types of ties are specified, strong ties and weak. In general, when the sense of closeness between two individuals is strong, a strong tie is defined and, in the same regard, when it is weak, a weak tie is defined (Granovetter, 1983). Strong ties are considered to exist with people one trusts, and that their social circles can highly overlap with his own. Weak ties are mainly considered acquaintances.

Measuring and predicting tie strength, and moreover, understanding the factors that drive tie strength, has been an expanding area of interest in social sciences, with increasing utility in the analysis of social networks (Mattie et al., 2018). Analyzing and predicting tie strength in social networks can lead to new insights into human social behavior and assist in designing novel user-centric services (Arnaboldi et al., 2013). So, the analysis of the social relationships and the accurate specification of the tie strength is highly desired.

In the context of this work, we examine the advent of the last decade in predicting tie strength in social networks. The survey reviews the elements and the characteristics of tie strength and categorizes the research works with respect to the dimensions that are involved in the modelling of the relationships’ tie strength. The remainder of the paper is structured as follows. Section 2 examines the dimensions and manifestations of tie strength in social networks while Section 3 examines the predictive variables that can be extracted from interaction data and associates the variables to dimensions. Section 4 examines prediction methods and models for tie strength estimation, categorizes them according to the dimension of tie strength that are considered and reviews the main findings regarding the components of the tie strength and the good predictors. Section 5 examines the utilization in a wide spectrum of social network analyses. Section 6 examines challenges and open research directions. Finally, Section 7 concludes the work.

2 DIMENSIONS OF TIE STRENGTH

Tie strength constitutes a factor of high importance in the analysis of social networks and it is considered to be a complex factor that is hard to accurately estimate. The main reason for this is that tie strength is a multidimensional factor where different forms and levels of interaction need to be considered. Tie strength is highlighted to have many dimensions and different manifestations. Granovetter (1973), in his landmark work on The Strength of Ties, specified four main dimensions for tie strength: the time spent connecting and interacting with others, the emotional intimacy, the intensity and the reciprocal services. After that, three additional dimensions were proposed and the list of dimensions was extended with the proposal of the emotional support (Wellman and Wortley, 1990) and the social distance (Lin et al., 1981), and the structural topology of the social network (Xiang et al., 2010).

Each dimension captures different elements of the social relationships. The dimension of time captures the duration and the frequency of the communication. In general, the more frequent and higher the interaction between a pair of individuals is, the stronger the sentiment of friendship and tie people feel (Luarn and Chiu, 2015; Mathews et al., 1998). Strong tie is bound up with the constant and frequent communication and the amount of time can promote other dimensions too (He et al., 2012).

The intensity represents the recognition of entities producing emotions that stresses on the cognition of others. It is relative to the absolute strength and individuals with highly intensive relationships is expected to spend more time with each other, that is greater than individuals with relationships that are less intensive (Kwon et al., 2013).

The intimacy concerns the affection between two individuals and acts as a sense of security and reliance (He et al., 2012). It is stated that intimacy relationships are willing to talk with open mind and demonstrate great support and recognition. It necessitates considerably more commitment and presumably a greater amount of positive affect between each other (Lewis et al., 2008).

The dimension of the reciprocal services represents the different forms of communication and the services utilized in interaction. An important parameter to develop a relationship is revenue that can be measured by the cost and the profit including energy, time emotion, and others. Social networks can reduce the cost of the social activities (He et al., 2012). Strong ties can easily share the information and the resources they possess and also they can provide access to information circulating in their dense network. So, strong tie includes more reciprocity services in exchanges (Granovetter, 1983).
The emotional support dimension represents a tie on an emotional level and concerns cases of discussions and advice offering on personal and family problems, something that can indicate a strong tie between the users (Gilbert and Karahalios, 2009). The dimension refers to providing messages that involve emotional content, re-assuring that the one is valuable and care about. Strong ties provide powerful emotional support that unites to face challenges and overcome crises.

Social distance is also highlighted to influence tie strength and factors like gender, race, socioeconomic status, education and political views and affiliation and can affect the tie strength development between individuals. Research studies have indicated that strong ties are more common between individuals of the same age, interests and who share certain life activities (Gilbert and Karahalios, 2009; Luarn and Chiu, 2015).

3 PREDICTIVE VARIABLES FOR TIE STRENGTH

The dimensions have facilitated the definition and the quantification of possible factors and predictive variables of tie strength (Mattie et al., 2018). These variables derive from the social information in the networks that relate to the profiles of individuals as well as to the interaction with their peers and which will be used as predictors of tie strength between two individuals. Table 1 categorizes the predictive variables used in research works in the literature. The predictive variables are mapped into the seven dimensions of the tie strength. Given that different social networks provide their users with different means of interaction, some variables generalize to any network, while some other may be specific to a number of social networks (Mattie et al., 2018). For example, photo tags variables and check-in denoting that individuals appear together in photos. User social profiles and interaction activities with their peers in a social network need to be analyzed in order to identify relative variables that can be quantified and be used in order to infer tie strength between individuals.

In the dimensions of time, time since first communication measures the length of the connection while the time since last communication captures the recency. The frequency is a proxy for the volume of the interaction between two individuals.

In the dimension of intensity, exact communication aspects and messages are measured like the number of the messages exchanged, the posts, comments, likes. The variables of this dimension rely heavily on the characteristics and the communication means of each social network.

In the intimacy dimension, intimacy words measure the topics of the messages exchanged while the relationship status captures specific types of relationships that may be denoted by the users such as married with each other, family members, etc. Common appearances in photos is another
measurement of the intimacy counting photographs that the two individuals appear together and the common check-ins measures places that they have been together.

In the dimension of reciprocal services, common applications measure the services and the applications that both the user and the friend share. The same stands for the links exchanged where URLs passed between two users can be indicative of the reciprocal services both use (Gilbert and Karahalios, 2009).

In the dimension of structural topology, variables capture aspects of the network structure and the groups that the individuals belong to. So, common groups variable measures the groups that the two individuals belong to, while the overlapping networks capture the social circles, organizations and networks like universities and companies that both individuals are members. Mutual friends can also indicate clues for tie strength and having mutual friends can foster relationship development (Adamic and Adar, 2003). Structural variables can be measured by the interests individuals have in common and the normalized TF-IDF of the interests too.

In the dimension of emotional support, deep analysis of text messages and interaction of the individuals aims to specify emotional support indicators and predictive variables like positive emotional words and negative emotional words. Dictionaries and linguistic resources like LIWC can provide indicative information about the categories of the words and the context messages.

In the social distance dimension, variables measure the age difference, the education discipline and level of the individuals, the political view and the occupation status. The identity information of the profiles as well as the language (location) beliefs (philosophy, political view) are used to measure the social distance of the individuals (Luarn and Chiu, 2015).

### Table 1: Categorization of variables.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Main Predictive Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Time since first communication&lt;br&gt;Time since last communication&lt;br&gt;Frequency of communication</td>
</tr>
<tr>
<td>Intensity</td>
<td>Communication aspects with friend&lt;br&gt;Messages exchanged&lt;br&gt;Comments exchanged&lt;br&gt;Likes</td>
</tr>
<tr>
<td>Intimacy</td>
<td>Relationship status&lt;br&gt;Appearances together in photos&lt;br&gt;Common check-ins&lt;br&gt;Intimacy words in the communication</td>
</tr>
<tr>
<td>Reciprocal Services</td>
<td>Common applications&lt;br&gt;Links exchanged by wall posts</td>
</tr>
<tr>
<td>Structural Topology</td>
<td>Common groups&lt;br&gt;Mutual friends</td>
</tr>
<tr>
<td></td>
<td>TF-IDF of interests&lt;br&gt;Listed in overlapping networks&lt;br&gt;Betweeness&lt;br&gt;Centrality</td>
</tr>
<tr>
<td>Emotional Support</td>
<td>Text analysis and specification of emotional context and emotional words. Specification of positive and negative contexts</td>
</tr>
<tr>
<td>Social Distance</td>
<td>Age difference&lt;br&gt;Occupation difference&lt;br&gt;Education difference&lt;br&gt;Political difference&lt;br&gt;Religion difference</td>
</tr>
</tbody>
</table>

4 METHODS FOR MODELLING TIE STRENGTH

In the literature, several works study tie strength with the aim to formulate predictive methods to estimate it. The first direction for the estimation of tie strength was indicated by Granovetter (1973) who stated that tie strength can indeed be quantified and that the strength of a tie is a probably linear combination of the amount of time, the intimacy, the intensity and the reciprocal services (Granovetter, 1973). During the last years, research works have examined different methods based on the characteristics of the particular research domain examined, the aims of the study and the access to the predictive variables that can be calculated in the domain examined. Typically, users are involved and participate in surveys in order to label the nature of each one of their social relationships.

Methods to obtain the ground truth data of the tie strength of relationships in a social network is a process of special attention. In the literature, two main stream of approaches prevail. The first approach which is the most common is to survey the users who accepted to participate in the study and collect feedback about the strength of their relationships with their friends (Jones et al., 2013).
The second approach is to use trusted networks in order to determine strong ties (Rotabi et al., 2017). An example is the use of telephone network which is considered to be a trusted network and the tie strength between two individuals who have phone contacts is determined as strong one.

Kahanda and Neville (2008) examined the nature and the dimensions of the relationship strength in Facebook and utilized a set of characteristics like marital status, gender, topological features like user connectivity, graph of friendship, shared posts on the wall to specify special friends. The authors applied a supervised learning approach and the results of the study concluded that the network transactional features like shared posts on the wall are the most prominent in predicting tie strength.

Gilbert and Karahalios (2009) defined indicators of tie strength that are specific to Facebook users and formulate a regression predictive model which reports 85% accuracy for the classicization of binary tie strength. The authors conclude that the dimension of intimacy makes the greatest contribution to tie strength specification and that educational difference strongly assists in the prediction of tie strength with tie strength diminishing as difference grows.

Pappalardo et al. (2012) present a model to estimate tie strength that is based strongly on variables of the network topology dimension and the intensity. The authors describe a quantitative measure of tie strength that is domain-independent and can be generalized and applied to any social network. The findings indicate that the strength of a tie is strictly related to the number of interactions among the people involved and that it is also related to the number of different contexts in which those connections take place.

Arnaboldi et al. (2013) present a linear model which estimates tie strength in Facebook. The authors’ model takes into account variables from the dimensions of time, network structure, intensity and intimacy and describing different aspects of user interaction. The liner model reports quite good performance and has accuracy higher than 80%. The work indicates that the recency of contact is the most indicative predictor of tie strength.

Jones et al. (2013) determine tie strength from users’ behavior in Facebook. The authors’ model takes into account variables from the dimensions of time, intimacy, reciprocal services, intensity, structure topology and social distance. The authors in their study surveyed Facebook users asking them to specify their closest friends, a piece of information that is used as ground truth. An additive logistic regression model was formulated which achieved an accuracy of 84% on the context of the study. The authors report the frequency of online interaction was the most indicative information for strong ties.

Servia-Rodriguez et al. (2014) present a model to assess strength and classify it within four categories of social spheres. The model assesses tie strength by taking into account users’ interactions and predictive features from the dimensions of intensity, reciprocal services, intimacy and structural topology. The work also points out the importance of using information from as many social networks as possible in order to avoid losing data in estimation of the tie strength.

<table>
<thead>
<tr>
<th>Work</th>
<th>Time</th>
<th>Intensity</th>
<th>Intimacy</th>
<th>Reciprocal Services</th>
<th>Emotional Support</th>
<th>Structure Topology</th>
<th>Social Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Granovetter, 1973)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Kahanda and Neville, 2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Gilbert and Karahalios, 2009)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>(Pappalardo et al., 2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Arnaboldi et al., 2013)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>(Jones et al., 2013)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>(Servia-Rodriguez et al., 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Liberatore and Quijano-Sanchez, 2017)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>(Mattie et al., 2018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>(Stolz and Schlereth, 2020)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ureña-Carrion, 2020)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
The reason for this is the fact that people usually possess accounts in different social platforms and it is possible not to interact with their peers with the same frequency in all of them.

Liberatore and Quijano-Sanchez (2017) present a computational model for tie strength that is independent of the domain of the social network. In the context of the work, users were requested to participate and define their relationships. The authors analyzed their personal and friendship interaction data and a linear model was examined. Predictive variables from the seven dimensions were examined and authors in their study indicate that relying exclusively and solely on intensity or intimacy may not be enough to efficiently calculate the tie strength.

Mattie et al. (2018) present a method for tie strength estimation that is based mainly on the structure dimension. The bow tie framework is proposed which consists of a focal tie and all actors connected to either or both of the two focal nodes on either side of the focal tie. The authors utilize machine learning techniques as well as regression methods and study which variables are most useful in predicting tie strength. The results show that the more the friend that two individuals share the stronger their tie is and also that geographical location can increase the tie strength of individuals.

Stolz and Schlereth (2020) present an approach for the prediction of tie strength that takes into account predictive variables from network, social distance and intensity dimensions. The proposed approach relies heavily on ego network structures that are user connections and interlinkage among them. The authors include also user similarity variables such as the matching gender and language of the users. The precision of the authors’ method in identifying all observed strong ties is 45%. The work also indicates that individuals react stronger to suggestions that are made of a close friend compared to the suggestions made by an acquaintance.

Ureña-Carrion et al. (2020) study how communication events and contact time can be an indicative predictor for tie strength. The authors study tie strength through the four main dimensions of Granovetter’s theory. The results of the authors’ work indicate that the number of days and hours with contracts are quite indicative variables in the estimation of tie strength. Also, the time of the first and last communication can provide indicative information of tie strength and perform better than the communication-intensity variables.

5 IMPACT OF TIE STRENGTH ON SOCIAL ANALYSES METHODS

Tie Strength is a prevalent feature in social network research and various studies are examining the impact of tie strength in various procedures. Understanding of the tie strength is essential in order to study the dynamics of social behaviors in network as well as the relationships of the users in it (Arnboldi et al., 2013). Tie strength can to affect career advancement and the word-to-mouth propagation of information (Mittal et al., 2008). It is highlighted that the tie strength has a general impact on behavior outcomes and intentions in the social network contexts (Ureña-Carrion et al., 2020). Specifically, strong ties are the ones that are most likely to transit norms and behavior change (Kim et al., 2015) and the identification of strong ties in a social network can assist in focus targeting these relationships in a wide spectrum of positive interventions that can have great multiplier effects as they spread from an individual to another (Jones et al., 2013). Research studies have indicated the impact of tie strength on decision making and that strong ties can greatly affect users in opinion seeking as well as in adoptions of stances (Stolz and Schlereth, 2020). The information about the tie strength and the knowledge of the social dynamics that affect and contribute to tie strength is reported to increase the efficiency of link prediction in social networks (Mattie et al., 2018). Tie strength estimation in the context of online social networks can assist in more efficiently detect communities in social networks. Tie strength is also assistive in modeling information diffusion in social networks (Bakshy et al., 2012). The literature suggests that novel information comes mainly from weak ties. Weak ties can provide access to novel information that is pieces of information that are not circulating in the dense network formulated by strong ties (Gilbert and Karahalios, 2009). Weak ties might be more frequently to convey new pieces of information, so even though someone is somebody you don't interact much with, and you might have less in common with, they might connect you to a part of the world that you don't normally have access to, and so they can still be very important in accessing information that might not be redundant with people that you often interact with. So, weak ties can play a very important role in information diffusion and to provide access to novel pieces of...
information. Moreover, of importance is also the bandwidth that a connection possesses.

6 CHALLENGES AND OPEN ISSUES

Although the methods in the literature achieve quite promising performance in predicting tie strength there are many challenges and open issues to be addressed. A first challenge concerns the detection and the proper handling of salient users. Most of the methods in the literature merely focus on active users and the salient users will be inappropriately considered as acquaintances due to their inactiveness (Li et al., 2018).

Another challenge concerns the emotional support dimension of tie strength and the fact that emotional factors are not directly measurable and are quite hard to quantify (Arnaboldi et al., 2013; Gilbert and Karahalios, 2009). The more efficient measurement of indicative variables from the emotional dimension could greater impact tie strength prediction.

The social distance dimension needs special handling too. An analytic framework for the fine-grained measurement of the social distance. Most of the existing methods do not properly capture the social distance as it derives from differences in education, in political views and so the specification of an analytical and fine-grained framework could be assistive in measuring social distances and reflect the diversity in the population.

Furthermore, a fine-grained estimation of the tie strength will also be a next. Most of the existing works model and estimate tie strength on a binary scale where strong and weak ties are specified between the users in a social network. Although research studies of the previous years have provided advances and valuable insights into the dimensions and the predictive variables, the analytical and numerical specification of tie strength constitutes a direction for research (Mattie et al., 2018).

Reproducibility issues constitute a great challenge too since most of the approaches in the literature utilize so many features and predictive variables from social networks (Facebook, Twitter) that their models are almost impossible to reproduce mainly because of recent APIs restrictions (Liberatore and Quijano-Sanchez, 2017). In addition, in the literature, there is a lack of public benchmark data sets for the study of methods and for systematically appraising the performances of models. The availability of such benchmark, pseudonymized or anonymized datasets for the fair assessment of new methods and models is a next important step too.

Kin relationships need special attention too. Research studies indicate that kin relationship remain stable over time even if they interact rarely or even not at all (Roberts and Dunbar, 2011). So, such relationships need special handling in modeling user interactions and formulating models for tie strength prediction. The proper identification of kin relationships and the formulation of social models to handle them, could further enhance the performance of prediction models and provides another direction for further research.

7 CONCLUSIONS

Tie strength constitutes a primitive challenge in the domain of social networks and the modeling and prediction of the tie strength of users’ relationships has attracted increased research interest. This paper presents the advent of the last years in predicting tie strength in social networks and examines methods for the estimation of tie strength. The dimensions and the manifestations of the tie strength are studied and state-of-the-art research works in tie strength estimation are examined with respect to the different tie strength dimensions. Meta-analyses on the results and the findings of the studies in the literature are performed. Last but not least, main challenges and open issues in designing methods for the estimation of tie strength are discussed.

ACKNOWLEDGEMENTS

This work was supported by funding from the EU's Horizon 2020 Research and Innovation Programme under grant agreements no. 739578 and no. 823783, and from the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation, and Digital Policy.

REFERENCES


Granovetter, M. (1973) The Strength of Weak Ties. The American Journal of Sociology, 78(6), 1360–1380.


