Computer Communications 110 (2017) 59–74

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

What do we really need to compute the Tie Strength? An empirical study applied to Social Networks

F. Liberatore*, L. Quijano-Sanchez

UC3M-BS Institute of Financial Big Data, Universidad Carlos III de Madrid, Madrid, Spain

ARTICLE INFO

Article history:
Received 13 October 2016
Revised 29 May 2017
Accepted 3 June 2017

Keywords:
Tie Strength
Relationship Modeling
Social Media Networks
Network Applications

ABSTRACT

Most of existing network-based decision-support systems, such as recommender systems, require knowing users' social context and, thus, the strength of their interactions. However, previous studies related to the usage and estimation of *tie strength* either assume that this parameter is given or use a computational model of their own. The amount, variety and domain specific information required to apply these models makes the reproducing and reusing of existing results extremely costly or utterly impossible. In our research, we show empirically the relative importance of different social variables for the computation of the *tie strength* and propose a computational model independent of the Social Networks' domain. Our experiments are based on a dataset obtained from a survey that involved more than 100 participants and comprised more than 500 social ties. The dataset is the first publicly available dataset to explicitly include *tie strength* measures.

1. Introduction

With the raising expansion of information technologies known as Social Media (SM), our capacity to interact, collaborate and network has highly and rapidly increased [1]. Research in a number of academic fields has shown that SM can leverage the way many problems are solved [2–5]. The main reason is that SM can offer new insights and innovative means by targeting information more effectively [6]. Proof of this is the recent use of different social measures in decision-support systems, such as recommender systems, where it has been proven that the use of SM information along with some specific measures, like *tie strength* estimations, can be used to aid their users in decision-making processes [7,8]. It is on this measure of *tie strength* -the importance of the social relationship between two individuals [9]- that this paper is focused on.

In the last decades, the academic interest on *tie strength* has substantially grown both in model design [10–14] and in decision-support systems that use or could benefit from its computation, in the area of recommender systems [8,15,16], fraud detection [17] or viral marketing [18]. Social Network (SN) users post on their profiles a huge amount of personal information (likes and interests, photos, etc) that can be analyzed to compute their *tie strength* with other users [9,19]. One of this papers' goals is to study the potential of using SNs to extract knowledge than can be used to compute *tie strength*. Different SNs provide their users with different technical features to interact. Although we may find similar interaction facilities among them or, at least, used for the same purposes, this fact makes very difficult (and sometimes impossible) the task of obtaining all the social predictors required by the different existing *tie strength* definitions [8–10,20] and, as a consequence, to devise a general model to compute *tie strength*. A simple solution could be directly asking users to rate the *tie strength* with their contacts [7,21,22]. However, the tasks of tagging and rating are sometimes found tedious and can generate resentment [8,23], hence, decreasing the systems' usability. Besides, in the case of tools without a public interface or Big SNs, asking users to directly rate their *tie strength* with all their contacts is unaffordable or simply unrealistic, a fact that should be taken into account when designing *tie strength* estimation.

When needing to compute *tie strength* other researchers have given several different definitions according to their research domain, needs and access to the predictors that compose it. For example, it has been affirmed that *tie strength* could be estimated by the communication reciprocity [24], by the possession of at least one mutual friend [25], with recency of communication [26] or with the interaction frequency [10,27]. This heterogeneous and unsystematic definitions make the reutilization of others' conclusions and/or models very difficult. Against this background, this paper aims to perform a thorough analysis of the social predictors that can be used to compute this measure. Also, their importance and

* Corresponding author.
E-mail addresses: fliberat@inst.uc3m.es, federico.liberatore@urjc.es (F. Liberatore), laraquij@inst.uc3m.es (L. Quijano-Sanchez).

http://dx.doi.org/10.1016/j.comcom.2017.06.001
0140-3664/

their strength of association is studied, providing guidelines on how to abstract their concept to ensure a feasible and satisfactory computation of the *tie strength*. Hence, presenting a methodology independent of the SN from which social factors can be estimated and a set of conclusions that can be reused by other researches. Our aim is to propose a general model of *tie strength* that could be applied to most contexts. Besides, we provide a public dataset obtained from a survey that involved more than 100 participants and thoroughly analyzed more than 500 social ties. This is the first public dataset to explicitly include *tie strength* measures.[1] We hope that it will be a relevant contribution to researchers in the field and encourage many to pursue further investigations in this subject matter. Finally, we show how the model proposed and the insights drawn in the analysis can be used to obtain an estimation of *tie strength* in a financial network comprised of clients of a financial institution and their operations and relationships. The estimated strength of the tie between clients finds application in Customer Relationship Management operations, such as identifying influencers to recommend financial products.

In summary, the contributions of this paper are the following: (1) Measuring the strength of association between the *tie strength* and several SN variables (Section 5.1). (2) Analyzing the relevance of the proposed variables by exploring different approaches to compute the *tie strength* and studying their estimation precision (Section 5.3). (3) Testing other's *tie strength* proposals, their applicably, efficiency and limitations (Sections 2 and 5.4). (5) Introducing a practical example of how to reapply the results of this paper's study in a financial network (Section 5.5).

The remainder of this paper is structured as follows. The next section shows some previous works related with our research topic. A description of the research questions raised and answered in this piece of research is given in Section 3. Section 4 introduces the details of the novel dataset. Next, in Section 5 different proposals of *tie strength* models and a variable analysis are presented. A comparison with other literature's models is illustrated in Section 5.4. The case study on a financial network is the subject of Section 5.5. Finally Section 6 concludes the paper with insights and future research guidelines.

2. Literature review

The most widely regarded definitions of *tie strength* is Granovetter's [27]: "The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding) and the reciprocal services which characterize the tie." This definition has provided a base for many studies that have made use of the concept [35] and has served as starting point for several researchers [10,12,20] . Thus, this research uses this seminal work as a baseline, analyzes each of these four components and studies if the best way to compute the *tie strength* is indeed their linear combination. This starting point and not more recent proposals [9–11] has been chosen following Petroczi *et al.*'s [20] justification for the correctness of Granovetter's approach and avoiding unreproducible focuses, that are either unquantifiable models [13,14,24–26,30–33] or domain specific models [8–11,20]. As illustrated in Table 1, our approach is the only quantitative model that does not have these limitations.

Regarding the four components described by Granovetter: the *amount of time* is a measurement of the duration of a tie between two nodes; the *intensity* is defined as the degree, amount of strength or force that something has (Webster's dictionary); the *intimacy* is defined as the state of being in a very personal or private relationship (Webster's dictionary); and concerning the *reciprocal services*, the term "reciprocal" (of a pronoun) indicates that action is given and received by each subject (Collins dictionary), that is, actions carried out in common between two nodes in an SN.

With these four dimensions as a guide, Gilbert and Karahalios [10] identified 74 Facebook (FB) variables as potential predictors of *tie strength*. On the other hand, Burt [30] proposed that *tie strength* could be modeled by structural factors such as the network topology or informal social circles. Xiang et al. [14] proposed an unsupervised model to distinguish strong from weak ties based on profile similarity and interaction activity. Lin et al. [31] stated that *tie strength* is mainly influenced by social distance, manifested by factors such as socioeconomic status, education level or political affiliation. Recently, Rodríguez et al. [11] have classified *tie strength* within four different types of social spheres computed through a set of several factors extracted from FB and Twitter, while Arnaboldi et al. [9] have presented quantitative linear models to estimate *tie strength* from a set of FB variables. Quijano-Sánchez et al. [8] proposed a non-intrusive method to compute *tie strength* by automatically analyzing users' FB profiles as opposed to other works [33] that needed to explicitly ask for the data that conforms the *tie strength*. They concluded that to move from theory [25,27,30,31] to practice [8–10,20] it is important to note that the factors used to compute *tie strength* are not easy to quantify and are limited by the capabilities of the API from which you extract them. Also, Hossmann et al. [32] showed, through two datasets obtained from both FB and Twitter, that *tie strength* is coupled with mobility and communication. In this line, Socievole et al. [13] performed an analysis showing that, in general, FB variables are strongly related to *tie strength*. Finally, Pappalardo et al. [34] present a quantitative measure of *tie strength* that, although it has not been validated against real *tie strength* measures, represents a SN domain independent approach. Albeit theoretically sound, their model needs as input social network variables such as the cardinality of the neighborhood of all the actors involved or the dimension relevance [36], that due to privacy issues or domain restrictions may not be available in other designed applications or researches, therefore limiting its practical applicability.

As illustrated in Tables 1 and 2, the heterogeneous, unsystematic and domain dependent definitions of *tie strength* make the reutilization of others' conclusions and/or models very difficult. Petroczi et al. [20] affirmed that Granovetter's four indicators are the actual components of *tie strength*, whereas contextual contingencies (communication reciprocity [24], possessing at least one mutual friend [25], recency of communication [26] or social distance [31]) are predictors. Predictors are related to *tie strength* but are not components of it. This paper focuses on the components and how to identify a SN domain independent predictor for each of them. Besides, although for the last 30 years many attempts have been made to find valid indicators and predictors of *tie strength* (see Table 2), Table 1 shows how most of these studies' results [13,14,24–26,30–33] are based on nominal data or binary indicators and, hence, they are not suitable for quantitative analysis. That is, most of the studies so far attempt to simply use and apply former knowledge on *tie strength* rather than try to actually measure these ties [37]. On the other hand, those that do propose quantitative results [8–11,20,34] do not provide a concrete model or, in other cases, a way to quantify or abstract to other contexts the predictors that they use. Also, these works providing quantitative results do not make publicly available their datasets. Without this data it is impossible to reproduce their results. Additionally, these studies make use of specific SN attributes and, therefore, their *tie strength* computation can not be extrapolated to other networks. Hence, in order to unify *tie strength* definitions and avoid the constant creation of new application specific models, this pa-

[1] Lewis *et al.* [28] provided a public dataset about FB users without any information regarding the *tie strength*. Also, MIT Human Dynamics Lab provided a public dataset [29] regarding mobile data and social dynamics of several communities, again with no specific *tie strength* measures.