Tie Strength Analysis: New Metrics and Open Problems

Michele A. Brandão and Mirella M. Moro

[1]Universidade Federal de Minas Gerais – Belo Horizonte – MG

{micheleabrandao,mirella}@dcc.ufmg.br

Abstract. *The study of social ties has lead to building rigorous models that reveal the evolution of social networks and their dynamism. In this context, a central aspect is the strength of ties, which allows the study of the roles of relationships. Here, besides analyzing the strength of co-authorship ties, we also present a set of metrics and algorithms to measure such strength.*

1. Introduction

Initial studies of social networks have emphasized the importance of properly measuring the strength of social ties to understand social behaviors [Granovetter 1973, Newman 2001]. Also, the study of social ties is fundamental for building rigorous models that reveal the evolution of social networks (SN) and the dynamics of social exchange [Aiello et al. 2014]. More recently, analyzing tie strength has allowed to investigate the roles of relationships including ranking for influence detection [Freire and Figueiredo 2011], as well as its influence in communication patterns [Wiese et al. 2015] and team formation [Castilho et al. 2017].

One of the first notable studies covering tie strength was published by Granovetter [1973]. He presents the importance of weak ties in SNs for various aspects, such as the spread of information. Since then, the strength of ties has been studied in different contexts with distinct goals, e.g., [Brandão and Moro 2015, Lopes et al. 2011, Silva et al. 2014]. However, few studies have addressed the strength of ties in temporal social networks [Dasgupta et al. 2008, Karsai et al. 2014, Kostakos 2009, Nicosia et al. 2013]. In such temporal context, studying the strength of ties allows to identify patterns of relationship over time, to detect aspects that influence it, to determine a limit of relationship in a period, among other relationship perspectives.

Tie strength may be measured by a combination of the amount of time, the cooperation intensity and the reciprocal services that characterize the tie [Granovetter 1973, Rana et al. 2014]. Such strength may also be measured by using the *neighborhood overlap* metric (also known as topological overlap) [Easley and Kleinberg 2010], a numerical quantity that captures the total number of collaborations between the two ends of each edge. This metric has been used for uncovering the community structure and measuring tie strength among others [Brandão and Moro 2015, Easley and Kleinberg 2010, Vaz de Melo et al. 2015]. In this work, neighborhood overlap is the base for developing new tie strength metrics and algorithms. Hence, this paper summarizes the main contributions of [Brandão 2017] as follows.

- An analysis of how nine topological properties affect the strength of co-authorship ties when measured by neighborhood overlap (Section 3.1).

- Four case studies on problems of measuring the tie strength with only neighborhood overlap and absolute frequency of interaction. A new metric to measure tie strength in non-temporal social networks, called *tieness*, accompanied by a nominal scale (Section 3.2).
- An analysis of how tie strength is defined and varies over time. We do so by: (1) improving an existing algorithm (RECAST [Vaz de Melo et al. 2015]) as named fast-RECAST; (2) introducing a new algorithm (STACY – *Strength of Ties Automatic-Classifier over the Years*) that automatically classifies ties strength in temporal co-authorship social networks, and (3) deriving a computational model from STACY, called *temporal_tieness*. Surprisingly, our results show that: most ties (even the strong ones) tend to perish over time; real co-authorship social networks have more weak and random ties than strong and bridge ties; and STACY is able to better identify strong ties than fast-RECAST (Section 3.3).
- A description of future directions and open problems (Section 4).

2. Brief Related Work over Tie Strength

Tie strength in social networks has been addressed with diverse goals such as measuring the strength of weak ties [Granovetter 1973], co-authorship ties [Brandão and Moro 2015], contact ties [Wiese et al. 2015], friendship ties [Zignani et al. 2016] and work ties [Castilho et al. 2017]. Such studies contextualize the importance of measuring tie strength in an appropriate manner: relationships play different roles and should be distinctly qualified as well through (for example) their strength. Indeed, studies show that the strength of ties has large impact at micro-macro levels in the network, depending on their weight, and influence the patterns of communications [Brandão and Moro 2015, Granovetter 1973, Zignani et al. 2016].

Tie strength can be calculated by considering topological and/or semantic properties. Topological properties capture structural features on the graph of the network [Brandão and Moro 2017a]. For instance, Brandão and Moro [2015] use neighborhood overlap to measure tie strength in co-authorship networks. Then, semantic properties catch non-structural features of nodes and edges. For example, Gilbert and Karahalios [2009] define a relationship as weak or strong on Facebook by considering features from the interaction history, e.g. days since first or last communication time, and inbox messages. Finally, combining both is also possible as done by Zignani et al. [2016].

Regarding semantic properties, one of the most relevant is the temporal aspect, which usually challenges traditional methods applied in static networks. Although the many research efforts in investigating social networks, the combination of tie strength and temporal aspects has not been largely explored yet. For example, Dasguptaet al. [2008] use tie strength associated with time to demonstrate its influence in operators network, whereas Karsai et al. [2014] use it to characterize the impact of time-varying and heterogeneous interactions on rumor spreading. Both studies consider the temporal evolution of ties strength, but do not propose a new metric that includes time. On the other hand, Kostakos [2009] and Nicosia et al. [2013] propose network properties that consider temporal aspects and show how such processing differs from those for static networks.

A related problem is how to define what strong and weak ties are in temporal networks. For instance, Karsai et al. [2014] consider both the amount of interactions and the time of the interactions to define the strength of ties. Then, strong ties are time repeated and frequent interactions among pairs of individuals, whereas weak ties occur only occasionally. In a different manner, Nicosia et al. [2013] define two nodes i and j as strongly connected if they are in a not symmetric relation (i is temporally connected to j but not vice-versa), whereas they are weakly connected if in a symmetric relation (both i is temporally connected to j, and j is temporally connected to i).

3. Tie Strength Analysis

In this section, we describe the main aspects related to the results of this work. Specially, we summarize a characterization of neighborhood overlap when it is used to measure tie strength (Section 3.1), an analysis of problems of measuring tie strength using solely neighborhood overlap or co-authorship frequency, and a new tie strength metric to non-temporal SN (Section 3.2), and an analysis of tie persistence and transformation in temporal SN by using an existing algorithm and a new one (Section 3.3).

3.1. Characterizing Tie Strength when Measured by Neighborhood Overlap

We built non-temporal co-authorship networks from three areas to quantify the impact of properties on tie strength (neighborhood overlap) [Brandão and Moro 2015]. Characterizing neighborhood overlap shows that its average value is around 0.2, i.e. the networks are formed more by weak ties. Also, our analyses show that the Granovetter's theory governs the three networks and how topological properties are affected by removing weak and strong ties. Then, we have studied how nine topological properties (edge betweenness, co-authorship frequency, closeness, eccentricity, clustering coefficient, number of triangles, weight degree, eigenvector and page rank) impact on neighborhood overlap in non-temporal co-authorship social networks from three areas: computer science, medicine and sociology. We have also evaluated each property for increasing the quality of a regression model. Out of them, the clustering coefficient and edge betweenness were related to neighborhood overlap in the three networks. Such result is trivial, because of the definition of neighborhood overlap. However, the most important contribution is discovering other properties related and non-related to ties strength, and whether the relations are linear or not. Such study can help to improve the quality of systems whose design considers ties strength, and to better understand the reasons for a tie being strong or weak.

3.2. Measuring tie strength in non-temporal SN

In non-temporal social networks, neighborhood overlap and absolute frequency of interaction (a.k.a. co-authorship frequency or edge weight) have been largely used to measure the strength of ties. However, through empirical analyses, we identified four main problems with using solely neighborhood overlap and co-authorship frequency to measure tie strength [Brandão et al. 2016, Brandão and Moro 2017b]: *(Case 1)* when a pair of collaborators does not have any common neighbor, neighborhood overlap will be zero; *(Case 2)* when determining if two collaborators are from the same community or not, co-authorship frequency considers only the absolute frequency of interaction; *(Case 3)* when there is little collaboration between a pair of collaborators and manyn common neighbors, neighborhood overlap and co-authorship frequency will present opposite results; and *(Case 4)* when the results are extreme values, neighborhood overlap may not represent the reality.

Then, we presented a new metric called *tieness* [Brandão et al. 2016, Brandão and Moro 2017b], which has relatively low computational cost and can be applied to other social networks types (since tieness is a topological feature). Also, the definition of tieness comes with a nominal scale that allows to identify when a tie is weak or strong and if it links researchers from different communities or not. The main limitation to such a new metric is that the social network must have nodes collaborating with each other. We have performed empirical studies by considering the networks from three different areas of expertise (Computer Science, Medicine and Physics). Overall, our analyses showed that tieness provides more distinct values through the ties than neighborhood overlap and absolute frequency of interaction. Such distinction is important to better compare how strong (weak) a tie is regarding another one. We also observed similar behavior through the three different research areas. Furthermore, all the four co-authorship social networks are dominated by the presence of weak ties. This happens because most pairs of researchers have low amount of shared neighbors and small co-authorship frequency of interaction. Therefore, tieness is able to classify as strong ties only pairs of researchers with very high neighborhood overlap and co-authorship frequency.

3.3. Measuring tie strength in temporal SN

The concept of tie strength is well understood and analyzed for static networks, but little is known about it when applied to temporal networks. Therefore, we characterized the strength of ties in temporal networks by measuring the persistence and the transformation of ties over time [Brandão et al. 2017]. In order to do so, we built four temporal co-authorship social networks considering three real publications datasets. We also proposed fast-RECAST, a parallel and faster version of an existing algorithm (RECAST – Random rElationship ClASsifier sTrategy **??**) that classifies edges into four classes of relationship according to their level of tie strength. Moreover, we propose STACY (Strength of Ties Automatic-Classifier over the Years), a parallel and fast algorithm that classifies the ties into eight different and more specialized classes. By grouping the edges into these classes, we were able to quantify the dynamism of tie strength over time.

Regarding results, the link persistence analysis reveals strong ties and bridges tend to persist over the years more than weak and random ties. Overall, this supports our initial hypothesis that strong ties persist more than the others. Then, STACY was able to find strong ties that persist more than those found by fast-RECAST. The results of fast-RECAST also show a different pattern for co-authorship social network from APS when the data is divided in 80% and 20%. In this experimental setting, the proportion of strong and bridge ties from the past to the present is very high compared to other social networks. Moreover, the link transformation analysis by using fast-RECAST and STACY revealed that most ties tend to disappear over time. This may occur due to the co-authorships nature, e.g., researchers tend to publish with students during a period and when the students graduate, they finalize the process of publishing together.

Finally, by using STACY, we defined a new computational model called *temporal_tieness* and a range of values for each class. Thus, tie strength can be computed with low computational cost when compared to fast-RECAST and STACY.

4. Future Directions and Open Problems

During this research, we have identified open questions and directions as follows.

Expanding the study to other collaboration social networks. The approaches proposed in this work to measure tie strength can be applied to other collaboration networks (for example, GitHub and Stack Overflow). We have already started to work on this direction by considering different ways to measure the strength of social coding collaboration on GitHub [Alves et al. 2016]. One of these metrics is *tieness*: while most metrics consider only the network topology, tieness is able to better differentiate the relationships by considering distinct weights associated to the edges.

Using qualitative research to evaluate tie strength. So far, we have evaluated the strength of ties by analyzing Granovetter's theory in non-temporal social networks and link persistence/transformation in temporal social networks. Another direction is asking for users to analyze if they agree or not with their relationships strength generated by our new approaches. Doing so, we would able to build a ground-truth to evaluate our new tie strength metrics and algorithms.

Evaluating tie strength methods by comparing with synthetic data. One of the main problems of working on social networks is the absence of a ground-truth to evaluate the results. A possible solution is to build a synthetic dataset that represents a completely random and/or perfect social network (allowing to compare the results from real networks with the synthetic ones). However, creating a realistic dataset has many challenges related to topologies, data distributions, correlations, attribute values, and so on.

Clustering analyses and evaluation. Due to the common nature of clusters in social networks, an interesting application is to evaluate clustering algorithms through network topological and semantic metrics. Initial results (fully reported in [Brandão 2017]) are promising but still need more evaluation by considering different algorithms.

Differentiating the α parameter of each property in *temporal_tieness*. So far, we have evaluated the range of values of *temporal_tieness* for each class by considering $\alpha = 1$. Although *temporal_tieness* is able to directly identify *strong*, *weak* and *random* ties, we can study how to better configure such parameter for each topological property. Thus, it would allow to classify ties in all the eight classes.

Adding other social network features to STACY. Our new algorithm considers three topological properties to classify tie strength (edge persistence, neighborhood overlap and co-authorship frequency). The main advantage of considering these metrics is that they are free of context. Nonetheless, adding other properties and extending STACY to different domains are still possible directions.

Group recommendation. Another interesting and open issue is group recommendation. Overall, our hypothesis is that tie strength among researchers helps to understand the importance their relationships, which may improve collaboration recommendation. Thus, a possible application of this work is to use our metrics associated with clustering algorithms to recommend groups of people to another person.

5. Conclusion

A property related to social ties is the strength of ties, which we have thoroughly studied here over co-authorship networks. Besides our analyses, we have also introduced *tieness* (a metric that better distinguishes tie strength of non-temporal co-authorhip networks), *STACY* (a temporal classifier for strong ties), and *temporal_tieness* (a new computational to directly classify tie strength over temporal co-authorship social networks with low computational cost). Finally, we presented open questions and directions for future work.

Acknowledgements. Work partially funded by CNPq and FAPEMIG, Brazil.

References

Aiello, L. M., Schifanella, R., and State, B. (2014). Reading the source code of social ties. In *ACM WebSci*, pages 139–148, New York, USA.

Alves, G. B., Brandão, M. A., Santana, D. M., da Silva, A. P. C., and Moro, M. M. (2016). The strength of social coding collaboration on github. In *SBBD*, pages 247–252, Salvador, Brazil.

Brandão, M. A. (2017). *Tie Strength in Co-Authorship Social Networks: Analyses, Metrics and a New Computational Model*. PhD thesis, PPGCC, Universidade Federal de Minas Gerais, Brazil.

Brandão, M. A., Diniz, M. A., and Moro, M. M. (2016). Using topological properties to measure the strength of co-authorship ties. In *BRASNAM*, pages 199–210, Rio de Janeiro, Brazil.

Brandão, M. A. and Moro, M. M. (2015). Analyzing the strength of co-authorship ties with neighborhood overlap. In *DEXA*, pages 527–542, Linz, Austria.

Brandão, M. A. and Moro, M. M. (2017a). Social professional networks: A survey and taxonomy. *Computer Communications*, 100:20 – 31.

Brandão, M. A. and Moro, M. M. (2017b). The strength of co-authorship ties through different topological properties. *Journal of the Brazilian Computer Society*, 23(1):5.

Brandão, M. A., Vaz de Melo, P. O. S., and Moro, M. M. (2017). Tie strength persistence and transformation. In *AMW (to appear)*, Montevideo, Uruguay.

Castilho, D., de Melo, P. O. V., and Benevenuto, F. (2017). The strength of the work ties. *Information Sciences*, 375:155–170.

Dasgupta, K. et al. (2008). Social ties and their relevance to churn in mobile telecom networks. In *ICDE*, pages 668–677, Cancun, Mexico.

Easley, D. and Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.

Freire, V. P. and Figueiredo, D. R. (2011). Ranking in collaboration networks using a group based metric. *Journal of the Brazilian Computer Society*, 17(41):255–266.

Gilbert, E. and Karahalios, K. (2009). Predicting tie strength with social media. In *Procs. of SIGCHI*, pages 211–220, New York, USA.

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

Karsai, M., Perra, N., and Vespignani, A. (2014). Time varying networks and the weakness of strong ties. *Scientific Reports*, 4(4001).

Kostakos, V. (2009). Temporal graphs. *Physica A: Statistical Mechanics and its Applications*, 388(6):1007–1023.

Lopes, G. R., Moro, M. M., da Silva, R., Barbosa, E. M., and de Oliveira, J. P. M. (2011). Ranking strategy for graduate programs evaluation. In *ICITA*, pages 56–64, Sydney, Australia.

Newman, M. E. (2001). Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review E*, 64(1):016132.

Nicosia, V. et al. (2013). Graph metrics for temporal networks. In Holme, P. and Saramäki, J., editors, *Temporal Networks*, pages 15–40. Springer Berlin Heidelberg.

Rana, J. et al. (2014). The Strength of Social Strength: An Evaluation Study of Algorithmic Versus User-defined Ranking. In *ACM SAC*, pages 658–659, Gyeongju, Korea.

Silva, T. H. P., Moro, M. M., Silva, A. P. C., Meira Jr, W., and Laender, A. H. F. (2014). Community-based endogamy as an influence indicator. In *JCDL*, pages 67–76, London, UK.

Vaz de Melo, P. O. S. et al. (2015). Recast: Telling apart social and random relationships in dynamic networks. *Performance Evaluation*, 87:19–36.

Wiese, J. et al. (2015). You never call, you never write: Call and sms logs do not always indicate tie strength. In *CSCW*, pages 765–774, Vancouver, Canada.

Zignani, M., Gaito, S., and Rossi, G. P. (2016). Predicting the link strength of newborn links. In *WWW, Companion Volume*, pages 147–148, Montreal, Canada.