Tie Strength Persistence and Transformation

Michele A. Brandão, Pedro O. S. Vaz de Melo and Mirella M. Moro

Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
{micheleabrandao, olmo, mirella} @dcc.ufmg.br

Abstract. A tie is a link between two persons in a social network. Here, we analyze tie strength in temporal co-authorship social networks by measuring ties persistence and transformation over time. Surprisingly, most ties tend to perish over time. Also, weak and random ties are more present in real co-authorship networks than bridges and strong ones.

Keywords: Social Networks, Tie Strength, Temporal Graph

1 Introduction

Time is a fundamental factor when characterizing the nature and the strength of relationships, as acquaintances might become friends and vice-versa. Such time-varying ties may be modeled as a temporal social network (SN), or temporal graph, where each node is a person and an edge connects two nodes in time t if they share any relationship at t. However, most studies focus on static aggregated graphs [1,3], in which the type and the class of the edges are invariant. If in static graphs such temporal aspects are aggregated, and therefore hidden, in temporal graphs they come naturally, serving as an appropriate model for dynamic SNs.

Nonetheless, computing temporal SNs properties and their time-varying behavior is very challenging, as their values evolve. Hence, concepts and metrics to analyze static networks must be adapted and extended to time-varying networks. Tie strength (a.k.a. strength of the ties) is one of those concepts, which is originally defined as a merge of the time of relationship, the emotional force, the intimacy, and the reciprocal services that link (through a tie) people [4].

Here, our goal is to verify if current definitions of tie strength hold for temporal networks. To do so, we analyze the dynamism of tie strength by observing link persistence and link transformation over time. This goal is divided in two research questions. First, *how is tie strength defined for temporal networks?* We consider a strong tie characterizes interactions likely to appear in the future, whereas a weak tie occurs sporadically. Second, *how much does tie strength vary over time?* Nicosia et al. [5] claim that if two nodes are strongly (or weakly) connected in a time t_1, they also will be strongly (or weakly) linked at t_2, where $t_2 > t_1$. Here, we challenge such claim in the context of temporal co-authorship SNs. Also, studies observe edge features as good indicators of tie strength, e.g., edge persistence [5,6] and topological overlap [2,6]. Here, we analyze the dynamism of tie strength by observing the dynamics of four edge classes composed

Description	# nodes	# edges
D. Articles	837,583	2,935,590
D. Inproceedings	945,297	3,760,247
PubMed	443,784	5,550,294
APS	180,718	821,870

Table 1: Description of the datasets used to build the co-authorship SNs.
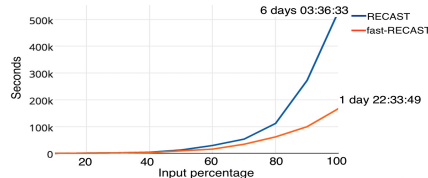

Fig. 1: The performance of RECAST and fast-RECAST for PubMed.

of edge persistence and topological overlap[1]. These properties represent the regularity of interaction and the similarity between people in a relationship.

2 Analyses of Persistence and Transformation

In order to proceed, we first need to formally define a model for temporal social networks. Therefore, we associate a start time and a duration to each co-authorship. Then, a temporal co-authorship social network is modeled as a graph $G_k(\mathcal{V}_k, \mathcal{E}_k)$ in which time is discretized into steps of duration δ[2], and k is the time step in which a co-authorship (encounter) occurs. The set of nodes $\mathcal{V}_k$ is formed by all network nodes in a co-authorship during the k-th time step, and the set of edges $\mathcal{E}_k$ is composed of co-authorships during the same time step. A time-varying representation of the SNs can be defined by a temporal accumulation graph $G_t(\mathcal{V}_t, \mathcal{E}_t)$, where $G_t = G_1 \cup G_2 \cup ... \cup G_t$, in which t is the last time step. Then, $\mathcal{V}_t$ and $\mathcal{E}_t$ are the set of all nodes and edges in the SNs, respectively, in the time step 0 to t. Since G_t accumulates all co-authorships from the datasets and evolves over time, such aggregate graph contains social and random encounters. Thus, a random version G_t^R of the temporal aggregated graph G_t is necessary to analyze the patterns of such SN. For this model to work, it requires a definition of tie strength in temporal SNs and an algorithm that implements it.

Definition of tie strength. Given a temporal graph $G_k(\mathcal{V}_k, \mathcal{E}_k)$, where k is the time step in which a co-authorship occurs, a tie (i, j) is likely to be strong if it is present in G_k for most values of k. On the other hand, the tie (i, j) is likely to be weak if it is present in G_k for just a few values of k.

Implementation of the algorithm. One contribution of this work is to modify an existing algorithm called RECAST [6] to measure the strength of ties in large temporal SNs. We chose RECAST because it is the only one that defines different classes to the tie strength in temporal networks. Such algorithm was originally applied in relatively small mobile networks to classify users' wireless interactions differentiating random interactions from the social ones (friends – called as strong, bridges and acquaintances – called as weak). It implements the model previously described by building both G_t and G_t^R. The construction of G_t^R increases the complexity of RECAST to $O(t \times (|\mathcal{V}_t| + |\mathcal{E}_t^R|))$.

---

[1] Technical Report at http://www.dcc.ufmg.br/~mirella/projs/apoena
[2] Here, we consider a duration of $\delta = 1year$.

Then, we propose to apply a multiprocessing Pool module from Python[3] in such step of RECAST in order to reduce its complexity. We call this novel, multiprocessing algorithm as *fast-RECAST*. The idea is that more than one random event graph G_t^R is built at a time in a multi-core computer. Thus, the new computational cost is $O(\frac{t}{p} \times (|\mathcal{V}_t| + |\mathcal{E}_t^R|))$, where p is the number of processes. We also add a multiprocessing Pool module from Python to call the functions to compute the edge persistence and topological overlap from the aggregated graphs. Both features are computed in parallel and asynchronously.

Dataset. To analyze tie strength persistence and transformation, we consider three publication datasets: DBLP, PubMed and APS, as collected in September 2015, April 2016 and March 2016, respectively. Considering these datasets, we build four co-authorship SNs whose main statistics are in Table 1.

Time Performance. In order to show that *fast-RECAST* performs better than RECAST, we measure the execution time of both algorithms in a laptop with 8 GB 1600 MHz DDR3 of memory and 2.5 GHz Intel two Core i5 of processor. The operation system is Mac OS X El Capitan version 10.11.6. Figure 1 presents the execution time in seconds of *fast-RECAST* and RECAST. Note that we present the results only for PubMed dataset, because it is the largest one.

Tie strength persistence results. In order to analyze the persistence over time, we divide the networks into two time windows, which from now on we call *past* and *future*[4]. We apply *fast-RECAST* in the past and then, verify if the edges of each class (*strong*, *bridge*, *weak* and *random*) continue to be in that same class in the future. To do that, we split the networks into two time windows and in two ways. First, we split the networks into a time window comprising 80% of the initial timestamp (*past*) and a time window comprising 20% of the final timestamp (*future*). Second, we divide the networks into time windows of 70% (*past*) and 30% (*future*). For both 80-20% and 70-30%, strong ties and bridges tend to persist over the years more than weak and random ties. Moreover, we emphasize the differences in the results of the APS network in the 80%-20% and 70%-30% partitions. In the first partitioning, the proportion of strong and bridge ties from the past to the present is very high, whereas in the second partitioning such proportion is lower. This result may indicate that the co-authorship social network from APS changes more through the years than the other networks. Another possibility is that physics researchers do not change very much the level of co-authorship with their collaborators over time, and this is a pattern of more recent researchers (note that 80% of data consider more recent co-authorships than 70%). We leave for future work further analyses of such insights.

Tie strength transformation results. We now evaluate the amount of ties from a class in the past that continues in the same class (or changes) in the future, i.e, tie strength transformation analyses. To avoid any kind of bias in the process of classifying the edges, here we divide the temporal co-authorship social net-

---

[3] Multiprocessing with python: `docs.python.org/2/library/` `multiprocessing.html`

[4] One may see the present as the timestamp between these two time windows

works into two time windows of 50% of the timestamp. We apply *fast-RECAST* in both parts and then we analyze the link transformation through the classes. Surprisingly, we cannot see ties classified as *weak* and *random* in DBLP articles and DBLP Inproceedings. This indicates that the features (edge persistence and topological overlap) of these social networks have high (or *social*) values. Furthermore, most ties from the past tend to disappear in the present, especially the *bridges*. This result may be explained by the nature of co-authorships, as researchers collaborate during a period towards a common goal and then, start to collaborate with others. This also reinforces the theory of Granovetter that weak ties are the ones that connect different communities [4], which is the case of the bridge edges. Furthermore, we observe similar behavior between PubMed and APS, and most ties tend to disappear, especially the bridges and random ties. Without disappeared links, most strong and weak ties become weak or random. Surprisingly, the weak ties are the ones that keep more in the same class, compared to the others in both networks.

3 Conclusion

We analyzed ties strength dynamism in temporal SNs. We built four temporal co-authorship SNs considering three real publications datasets, and proposed *fast-RECAST*, a parallel version of an existing tie classification method. The resulting link persistence analysis reveals that strong ties and bridges tend to persist more than weak and random ties. This supports our hypothesis that strong ties persist more than others. The results also show a different pattern for 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 SN. Also, the link transformation analysis revealed most ties tend to disappear over time. As future work, we plan to investigate the patterns discovered here and to modify *fast-RECAST* to better capture tie strength.

References

1. Brandão, M.A., Moro, M.M.: Affiliation influence on recommendation in academic social networks. In: Procs. of AMW. pp. 230–234 (2012)
2. Brandão, M.A., Moro, M.M.: Analyzing the strength of co-authorship ties with neighborhood overlap. In: Procs. of DEXA. pp. 527–542 (2015)
3. Castilho, D., Vaz de Melo, P.O., Benevenuto, F.: The strength of the work ties. Information Sciences 375, 155–170 (2017)
4. Granovetter, M.S.: The strength of weak ties. The American Journal of Sociology 78(6), 1360–1380 (1973)
5. Nicosia et al, V.: Temporal Networks, chap. Graph Metrics for Temporal Networks, pp. 15–40. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
6. Vaz de Melo et al, P.O.: Recast: Telling apart social and random relationships in dynamic networks. Performance Evaluation 87, 19–36 (2015)