

# Using Topological Properties to Measure the Strength of Co-authorship Ties

Michele A. Brandão, Matheus A. Diniz, Mirella M. Moro

Departamento de Ciência da Computação  
Universidade Federal de Minas Gerais – Belo Horizonte – MG – Brazil

{micheleabrandao,matheusad,mirella}@dcc.ufmg.br

**Abstract.** *Studying the strength of ties in social networks allows to identify impact at micro-macro levels in the network, to analyze how distinct relationships play different roles, and so on. Indeed, the strength of ties has been investigated in many contexts with different goals. Here, we aim to address the problem of measuring ties strength in co-authorship social networks. Specifically, we present four case studies detailing problems with current metrics and propose a new one. Then, we build a co-authorship social network by using a real digital library and identify how the strength of ties relates to the quality of publication venues when measured by different topological properties. Our results show the best ranked venues have similar patterns of strength of co-authorship ties.*

## 1. Introduction

The strength of ties in social networks has been explored in different contexts, such as information diffusion, analysis of patterns in communication logs, and so on. Specially, analyzing how strong a tie is allows to investigate the different roles of relationships, identify impact at micro-macro levels in the network [Burt 2010] and its influence in the patterns of communications [Wiese et al. 2015]. Despite the importance of analyzing the strength of ties, there are not many studies on evaluating how to measure the strength of ties in co-authorship social networks, where nodes are researchers and edges (ties) represent co-authorship between them. Studying the strength of co-authorship ties may reveal how its behaviors relate to research. Thus, any application based on co-authorship patterns may benefit. For instance, this study may help to improve Ductor [2015]’s work, which uses co-authorship to investigate if relationships lead to higher productivity.

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]. Such strength may also be measured by using the *neighborhood overlap* metric [Brandão and Moro 2015, Easley and Kleinberg 2010], a topological property that captures the total number of collaboration between the two ends of each edge and identifies edges forming bridges in a community (set of nodes that are densely connected). The advantages of using such metric are: its simplicity to compute, the possibility to identify if ties are bridges or not, and the consideration of the neighbors to calculate the ties strength (then allowing to analyze how a tie is in the social network, for example, isolated or not).

Another metric that has been largely used to measure the intensity of co-authorship between ties is *weight* (the number of publications between pairs of researchers) [Silva et al. 2014, Yan et al. 2012]. Besides the simplicity to calculate such topological property, another advantage is the representation of the exact frequency of

collaboration between ties. However, we find a few problems in both metrics that complicate their sole use to measure the strength of co-authorship ties. The existence of such problems suggests the metrics should be considered together and with other social network (SN) properties to better measure ties strength.

Considering the problems of neighborhood overlap and weight to capture alone the strength of co-authorship ties, this work aims to propose a new metric, called *tieness*, that helps to define a tie as weak or strong. Note the goal of *tieness* is not to replace neighborhood overlap and weight, but being an additional feature. Another goal of this work is to study the relationship of the three metrics with the quality of publication venues. We measure such quality considering Qualis rankings<sup>1</sup>, because most publications venues are classified according to such ranking [Laender et al 2008]. Such study may reveal another feature that may be used to measure the strength of co-authorship ties.

Indeed, after discussing related work (Section 2), we present the contributions of this paper, summarized as follows:

- We discuss four case studies where neighborhood overlap and weight alone have problems to measure the strength of ties. Also, we show the relationship between both metrics in a real dataset – DBLP<sup>2</sup> (Section 3).
- We propose a new metric called *tieness* that is a combination between a modification in neighborhood overlap and weight (Section 4).
- We describe a nominal scale to *tieness* based on the values of a modified neighborhood overlap and weight. Such nominal scale allows to identify when a tie is weak or strong and if it links researchers from different communities or not (Section 5).
- We verify the existence of a relation between the quality of a publication venue and the strength of ties when measured by different metrics (Section 6).

## 2. Related Work

There are many studies that address the ties strength [Brandão and Moro 2015, Goulas et al. 2014, Granovetter 1973, Wiese et al. 2015]. They contextualize the importance of our study to use a different SN to corroborate previous insights, such as distinct relationships play different roles, ties have large impact micro-macro level in the network depending on their strength, the influences in the patterns of communications, and so on.

Specifically, the strength of ties has been analyzed in different social networks. For example, Pappalardo et al. [2012] propose a definition of tie strength by measuring the interaction between two individuals over three different social channels: Facebook, Twitter and Foursquare. The existence of interactions on different channels indicate a greater chance of a stronger tie. Moreover, Gilbert and Karahalios [2009] classify friendships in Facebook as strong or weak based on variables obtained through an interaction history, such as inbox messages exchanged, days since first or last communication, and so on. Also using Facebook, Kahanda and Neville [2009] reach 87% accuracy in distinguishing strong and weak ties by mapping four different categories of features: transactional (such as picture postings and groups), network-transactional (considers the interaction between a pair of users and the overall interaction of these two users with the remaining users), topological (e.g., node degree and number of shared neighbors), and attribute

<sup>1</sup>Qualis: <https://qualis.capes.gov.br/>

<sup>2</sup>DBLP: <http://dblp.uni-trier.de/>

based features (such as gender and interests). Furthermore, McGee et al. [2011] study if the geographic distance influence the strength of ties among users in Twitter. In order to do so, the authors consider users' friends, followers and recent tweets.

All of these methods require an interaction through the history (messages on timeline, tweets, etc) to build the predictive model (or to measure the strength of ties). Furthermore, Wiese et al. [2015] shows that the accuracy of methods based only on interaction history may be misleading. They classify strong ties through phone calls and SMS data, and managed to achieved 91.6% of accuracy even though only half of the strong ties were correctly classified. This happens because of the low proportion of strong ties regarding weak ties. Also, Zignani et al. [2016] classify ties in Facebook as interactive (strong) or non-interactive (weak) at their creation time, i.e, links without interaction history. In order to do so, the authors consider topological features, interaction-graph features and temporal features in supervised learning classifiers. Thus, it is important to explore and develop features different from interaction history.

Additionally, some datasets from social networks do not offer data on interaction among people (nodes). For example, in academic social networks, which have been largely studied [Cheng et al. 2014, Digiampietri and Maruyama 2014, Machado et al. 2013], the only data available is about the collaboration between different authors or features from authors or/and publications. This indicates the need of better topological features. In this context, we propose a new topological feature that helps to measure ties strength in co-authorship social networks.

### 3. An Analysis of Neighborhood Overlap and Weight

We have identified four cases where the existing metrics neighborhood overlap and weight cannot be solely used to measure strength of ties. They are defined based on empirical analysis of co-authorship social networks, as detailed next.

**Case 1: A pair of collaborators without any common neighbor.** One of the problems of using only neighborhood overlap to measure the strength of ties is when an author has a high frequency of collaboration with another author, but they do not have any *common neighbor*. In this case, the neighborhood overlap is zero, which does not represent the reality. Figure 1 exemplifies this case. Another problem here is that neighborhood overlap and weight present contradictory results. Analyzing neighborhood overlap, the pair  $AC$  is a bridge and the strength of co-authorship is very weak. At the same time, weight may indicate that such tie is not very weak. Therefore, considering both metrics may help to better analyze how strong a tie is.

**Case 2: Determining if two collaborators are from the same community or not.** One of the problems of measuring the strength of ties using only the weight is that such metric provides a simple vision of the relationship. It is not possible to know if the relationship is intra a community or not. This case is exemplified by Figure 2. Since ties with low weight may be intra a community and ties with high weight may be inter communities, the problem is that using only weight, we cannot verify Granovetter's theory [Granovetter 1973] (ties are weak when they serve as bridges in the network by connecting users from different groups, and strong when they link individuals in the same group).

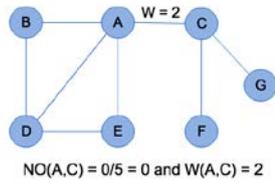


Figure 1: Case 1 – No common co-author.

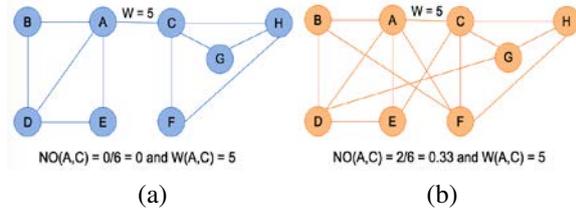


Figure 2: Case 2 – No representation if the tie is inside a community.

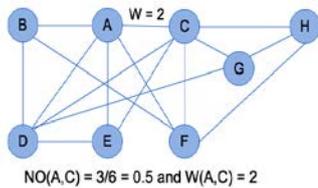


Figure 3: Case 3 – Many common co-authors.

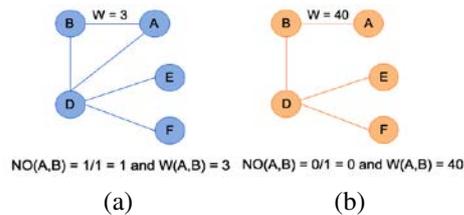


Figure 4: Case 4 – Results very small or too high.

**Case 3: Little collaboration between a pair of collaborators and plenty of common neighbors.** In this case, the neighborhood overlap and weight give values with opposite meaning, i.e. high neighborhood overlap and low weight. Thus, it is hard to identify if the tie would be weak or strong. Certainly, it depends on the analysis of the context. However, following Granovetter’s theory, such tie should be strong. Figure 3 gives an example of this case.

**Case 4: Results with extreme values.** Here, the problem is the neighborhood overlap or weight has extreme values that may not represent the reality. Figure 4a shows a maximum value to neighborhood overlap, because the edge is part of a triad. Nevertheless, the weight of the same edge is very small, which means that the tie is not necessarily very strong. Figure 4b presents a similar situation, but when the weight is very high and neighborhood overlap has the minimum value (zero). In this case, defining a tie as weak or strong based on only one of the metrics may provide a misleading interpretation.

Given all such problems, developing a new metric is necessary. In order to do so, we initially analyze the relationship between neighborhood overlap and weight in DBLP, downloaded in September 2015. We consider publications from articles and inproceedings with Qualis rankings at levels A1, A2, B1, B2, B3, B4 and B5, and from the year 1964 to 2015. Then, we build a co-authorship SN using such dataset whose features are: 145,604 authors and 208,808 links between them from 1,096,797 publications.

We now analyze the relationship between both on DBLP, as shown in Figure 5. The horizontal and vertical lines for axis x and y represent the frequency of neighborhood overlap and weight in each value, respectively. Most edges have neighborhood overlap lower than 0.50 and weight smaller than 15. Also, the correlation coefficient between the two variables is small for the three coefficients (Kendall, Spearman and Pearson).

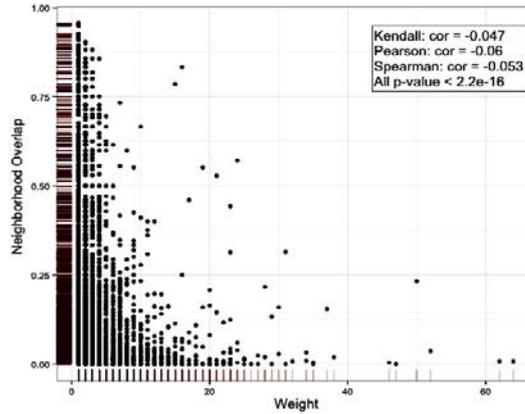


Figure 5: DBLP: Neighborhood overlap versus weight.

Therefore, neighborhood overlap and weight are monotonically and linearly independent.

Both metrics are important to measure the strength of ties and they capture different characteristics of the SN. In the next steps of this work, we intend to consider both as features to a computational model to automatically classify the strength of ties.

#### 4. Tieness: A New Metric to Measure the Strength of Ties

Given the problems generated by using neighborhood overlap and weight alone to measure tie strength, in this section we introduce a new metric called *tieness*. Specifically, tieness results from a combination between a modification in neighborhood overlap (entitled modified neighborhood overlap) and weight, as shown by Equation 1.

$$tieness = \frac{|(A_{c_1} \cap A_{c_2}) + 1|}{|A_{c_1} \cup A_{c_2}| - (A_1 \text{ or } A_2 \text{ themselves})} \text{ weight} \quad (1)$$

where  $A_{c_1}$  represents the co-authors of researcher  $A_1$ , and  $A_{c_2}$  the co-authors of  $A_2$ . Note that we sum one at the numerator of neighborhood overlap to indicate that there is a link between  $A_1$  and  $A_2$ . This solves the problem when a pair of authors does not have any co-author in common. Also, we emphasize for unweighted social networks the tieness metric is equal to the modified neighborhood overlap.

Regarding computation time cost of tieness, the operations with the highest time cost are intersection ( $O(A_{c_1} + A_{c_2})$ ) and union ( $O(\min(A_{c_1}, A_{c_2}))$ ) using hash tables. Thus, the time complexity of the tieness is  $O(\max(A_{c_1}, A_{c_2}))$  (Big O notation property:  $O(\min(A_{c_1}, A_{c_2})) + O(A_{c_1} + A_{c_2}) = O(\min(A_{c_1}, A_{c_2}) + A_{c_1} + A_{c_2}) = O(\max(\min(A_{c_1}, A_{c_2}), A_{c_1}, A_{c_2}))$ ) [Cormen 2009].

Also, the overall level of tieness in a network is measured by the average of the tieness of all the edges:

$$\overline{tieness} = \frac{1}{|E|} \sum_{i=1}^{|E|} tieness_i \quad (2)$$

Table 1: Tieness for each case study and an extra case study representing the situation when NO and weight are in accordance.

Case	Image	NO	Weight	Tieness
Case 1: A pair of researchers without any common neighbor	Figure 1	0	2	0.4
Case 2: Determining if two researchers are from the same community or not	Figure 2a	0	5	0.83
Case 2: Determining if two researchers are from the same community or not	Figure 2b	0.33	5	2.5
Case 3: Little collaboration between a pair of researchers and a plenty of common neighbors	Figure 3	0.5	2	1.33
Case 4: Results with extreme values	Figure 4a	1	3	6
Case 4: Results with extreme values	Figure 4b	0	40	40
Regular Case: NO and weight in agreement	Figure 3 with $w = 12$	0.5	12	8.0

*Note: NO is neighborhood overlap.*

where  $|E|$  is the number of edges in the social network. Also, the time complexity of the algorithm to measure the overall tieness is  $O(|E| \max(A_{c_1}, A_{c_2}))$ .

In order to understand how tieness represents ties in SN, Table 1 shows tieness' values for each case study. In Case 1, tieness gives a small value that indicates the presence of interactions (opposite of neighborhood overlap). However, analyzing only the final result of tieness for Cases 1, 2 and 3 is not enough to identify if a pair of researchers is intra a community or not. Also, regarding Case 4, the tieness is the same as weight when neighborhood overlap is zero and  $2 \times \text{weight}$  when neighborhood overlap is one. In the Regular Case, when neighborhood overlap and weight are in accordance indicating that a tie is strong, tieness also provides a high value that may represent a strong tie. Such results cannot be used to identify if the tie belongs to a community and if it is a bridge or not. Indeed, an advantage of using our new metric is the values of ties strength are more different one from another, then allowing to better differentiate the strength of a tie from another. Also, we can consider the value of the modified neighborhood overlap and weight separately to evaluate the final result of tieness. Thus, the definition of a nominal scale is necessary to identify when a tie is weak or strong, and intra a community or not.

## 5. A Nominal Scale to Tieness

Here, we define a nominal scale to tieness by comparing the modified neighborhood overlap and weight. In doing so, we follow concepts discussed by Easley and Kleinberg [2010]: a weak tie has a small neighborhood overlap and a strong tie has a large one. Also, we define a nominal scale to neighborhood overlap comparing with weight.

In order to do so, Figure 6 shows the ECDFs (Empirical Cumulative Distribution Function [Lewis and McKenzie 1988]) and quartiles for neighborhood overlap, weight, modified neighborhood overlap and tieness. The analysis of ECDFs shows that neighborhood overlap and weight provide many repeated results to the strength of ties. For neighborhood overlap, 50% of data are equal to 0. In a similar manner, 50% of data have weight equal to 1. On the other hand, the modified neighborhood overlap and tieness provide different results for each quartile. Such difference among the quartiles may indicate

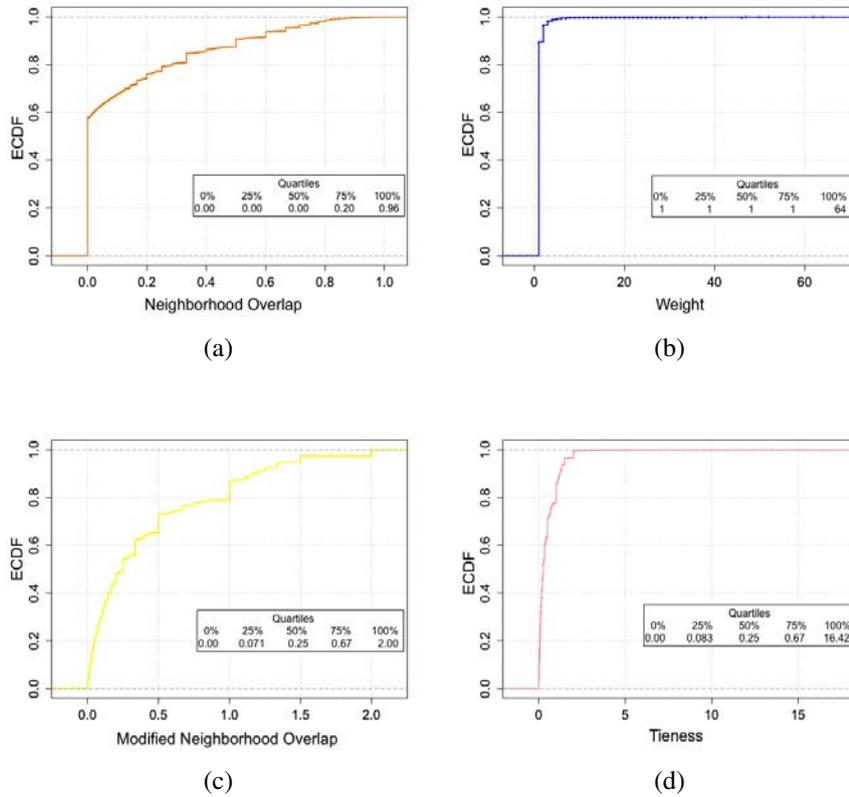


Figure 6: ECDF of each metric. In this scenario, modified neighborhood overlap and tieness metrics have more distinct values through the quartiles.

that tieness better represents the strength of ties than neighborhood overlap and weight, because it better distinguishes the relationship between nodes.

Here, we consider the values of quartiles to define a nominal scale, i.e., to identify when a tie is weak or strong, and if it connects different communities or not. Equation 3 shows the nominal scale to tieness based on the quartiles. Note for an unweighted social network, such scale is also valid because modified neighborhood overlap has the same value as tieness to the second and third quartile.

$$\begin{cases} \mathbf{weak}, & tieness \leq 0.25 \\ \mathbf{moderate}, & 0.25 < tieness < 0.67 \\ \mathbf{strong}, & tieness \geq 0.67 \end{cases} \quad (3)$$

To validate the proposed nominal scale, we verify if Granovetter's theory governs the social network and the strength of ties with such values. In summary, his theory claims the network tends to be more disconnected when weak ties are removed (i.e., the number of connected components tends to increase). Hence, we analyze the number of connected components in the network after removing weak and strong ties.

Table 2 presents the number of edges and connected components after removing weak and strong ties. Also, we show results when the strength of ties is mea-

Table 2: Number of connected components when weak and strong ties are removed from the social network.

State of the SN	# edges	% edges	# c. components	# times changed
<b>Original</b>	208,808	–	14,445	–
<b>Tieness - weak ties removed</b>	100,264	48.01	87,453	6.05
<b>Modified NO - weak ties removed</b>	95,706	45.83	90,864	6.29
<b>Tieness - strong ties removed</b>	152,529	73.07	34,605	2.4
<b>Modified NO - strong ties removed</b>	155,911	74.67	33,305	2.3

Note: c. components is connected components.

Table 3: DBLP sub-networks according to Qualis rankings.

SN	#nodes	% DBLP nodes	#edges	% DBLP edges	AvgWeight	AvgNO	AvgTieness
<b>A1</b>	18,429	12.66	21,622	10.35	1.15	0.094	0.25
<b>A2</b>	31,248	21.46	47,658	22.82	1.12	0.206	0.594
<b>B1</b>	47,101	32.35	56,889	27.24	1.18	0.101	0.47
<b>B2</b>	16,045	11.02	18,405	8.8	1.08	0.146	0.586
<b>B3</b>	14,576	10.01	17,240	8.26	1.04	0.213	0.7
<b>B4</b>	12,328	8.47	14,971	7.2	1.13	0.17	0.66
<b>B5</b>	3,939	2.7	4,345	2.08	1.07	0.136	0.6

Note: SN is social network and NO is neighborhood overlap. Weight, NO and Tieness are the average value on the SN.

sured by tieness (weighted SN) and modified neighborhood overlap (considering the SN as unweighted). According this table, when weak ties are removed, the number of connected components is higher than when removing strong ties. Also, the number of removed edges is larger when weak ties are removed. Indeed, the larger number of connected components may be explained by the larger removal of edges. However, we compare the proportion of the number of connected components by the number of edges for tieness and modified neighborhood overlap: for tieness, the proportion is  $87,453/100,264 = 34,605/152,529 \Rightarrow 0.87 = 0.22$ , and for modified neighborhood overlap, the proportion is  $90,864/95,706 = 33,305/155,911 \Rightarrow 0.94 = 0.21$ . The analysis of such proportions shows that the number of connected components per edge is greater when weak ties are removed. Thus, the nominal scale is valid. Also, as the removal of weak ties (defined according to the nominal scale) breaks the connected components of the network. Hence, tieness is indeed able to identify when a tie connects different communities or not.

## 6. The Strength of Ties and Publications' Quality

Here, our goal is to answer the following research question: is there any relation between the quality of a publication venue and the strength of ties when measured by different metrics? How are the values of the three metrics different from each other when considering the Qualis rankings at levels A1, A2, B1, B2, B3, B4 and B5? Such analysis provides another feature that may influence the strength of ties. Thus, we may be able to define which relationship is more important to a researcher. For instance, if a researcher has two co-authors  $a$  and  $b$ , and the publications with  $a$  are solely in high quality venues and with  $b$  in low quality venues – such situation could happen when  $a$  is a senior researcher and  $b$  is still a undergrad or grad student, who is still publishing in small venues.

Hence, using DBLP dataset, we have built a sub-network for each Qualis ranking. For example, a sub-network of researchers considering only publications in A1, then in

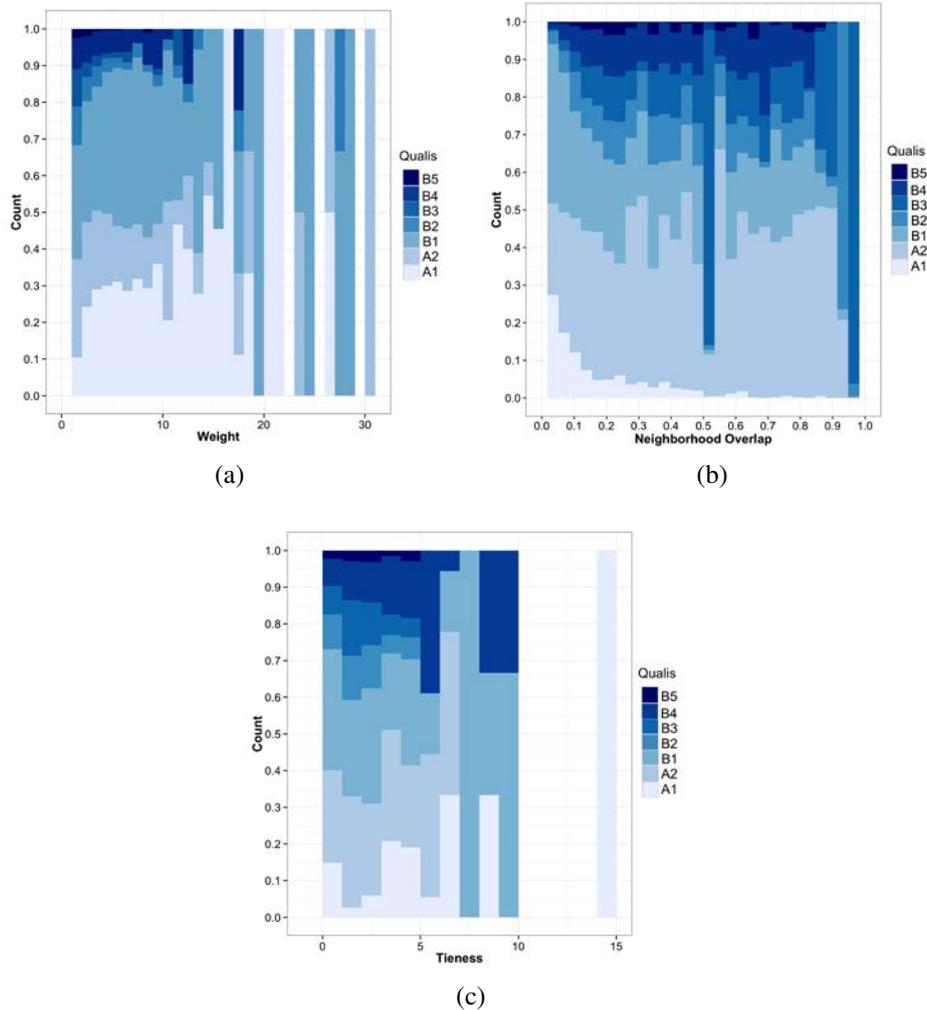


Figure 7: Proportion of co-authorships in venues grouped by Qualis rankings.

A2, and so on. Table 3 presents the main characteristics of each network: number of nodes, percentage of nodes from DBLP social network, number of edges, percentage of edges from DBLP SN, average weight of the edges in the sub-network, average neighborhood overlap and average tieness. Also, we applied a t-test (a statistical examination to compare two distributions [Jain 1991]) for the means of weight, neighborhood overlap and tieness. Most means are significantly different at 95% confidence level. The exceptions are the difference between weight mean of A2 and B4 sub-networks, and B2 and B5. In both cases, the confidence interval includes zero. Indeed, Table 3 reveals that co-authorships from publications in B1 have the highest weight average, and the co-authorships have the highest values to neighborhood overlap and tieness in B3. Additionally, B3 has the smallest weight, and A1 the lowest neighborhood overlap and tieness. Also, we observe that the values of tieness through the sub-networks are more different than the values of neighborhood overlap and weight. Thus, tieness might better represent the overall strength of the ties in the social network.

In order to compare the frequency of the strength of ties, we plot a histogram for each metric. Figure 7 contrasts the social network of each Qualis rankings. Specially,

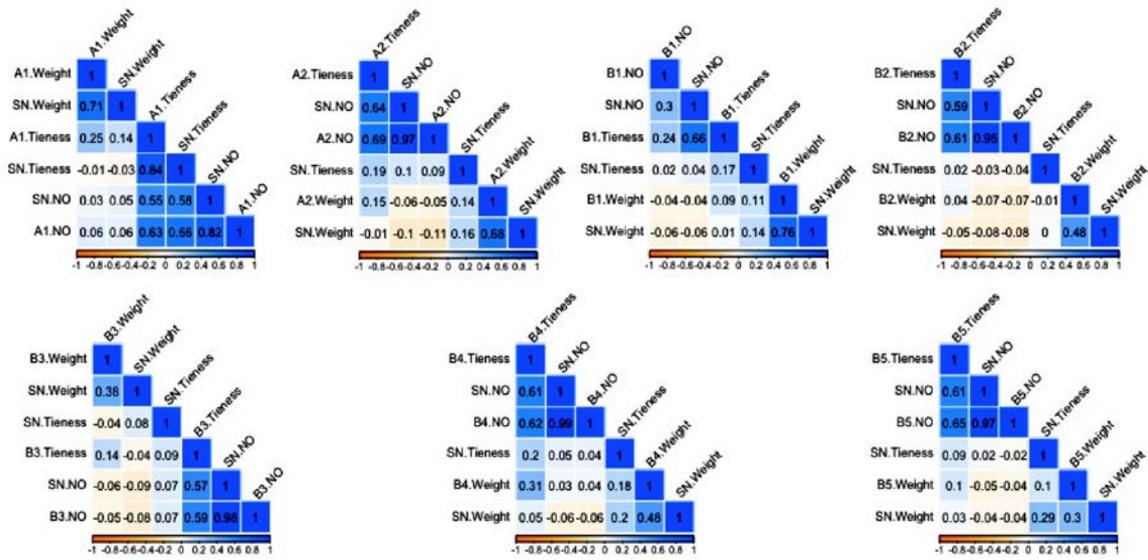


Figure 8: Correlation matrix based on Spearman's rank correlation coefficient for each metric in DBLP full social network and sub-networks.

Figure 7a shows pairs of researchers with weight higher than 20 are only in publications from A1 to B2. In addition, the proportion of co-authorships with large weight is greater to publications in classes A1 to B1 than classes B2 to B5. Figures 7b and 7c also present the difference between the strength of ties through the Qualis rankings. Also, the maximum value for tieness is only present in A1 social network (there are only 14 pairs of researchers with tieness larger than 10). Although the size of the sub-networks are different, such results indicate different behavior of co-authorship among Qualis rankings.

In this context, we also analyze the monotone correlation between weight, neighborhood overlap and tieness following Cohen [1988]'s conventions (*very large* when is greater than 0.7, *large* when is within [0.5;0.7), *moderate* when is within [0.3;0.5), *small* when is within [0.1;0.3), and *insubstantial* for any value smaller than 0.1). Figure 8 presents the correlation matrix for each Qualis rankings sub-network. Note that Figure 8 shows the value of each metric for pairs of researchers in the sub-networks (for example, A1.Tieness, B2.NO) and the equivalent value for the same pairs in the full social network of DBLP (for instance, SN.Weight, SN.Tieness and SN.NO). Hence, the correlation between SN.Weight and SN.NO in the A1 sub-network is different from A2, because each network considers the total value of the pairs of researchers in the full SN only if it exists in its sub-network.

Additionally, Figure 8 shows that changing the Qualis ranking reveals dramatic shifts in correlation. For instance, more metrics in A1 sub-network are correlated than in the others. Also, the weight in A1, A2 and B1 sub-networks are very largely and largely correlated with the weight of the full social network, whereas the weight in the remaining sub-networks are moderately correlated. Such result indicates that co-authorships from A1, A2 and B1 contribute more for the weights of the full social network. Regarding neighborhood overlap, we observe that most sub-networks have such metric largely or very largely correlated with neighborhood overlap of the full SN. The exception is B1 sub-network in which such correlation is moderate. This is expected since publications

in any Qualis ranking contribute to turn the full social network more connected. Considering tieness, only in the A1 sub-network such metric is very largely correlated with tieness of the full SN. In the other sub-networks, such correlation is small or insubstantial. This result may indicate that the weight and modified neighborhood overlap of pairs of researchers with publications in class A1 contribute more to the whole tieness of pairs of researchers. In addition, we note that there is monotone correlation between the tieness of the sub-network and the tieness of the full DBLP SN when there is also correlation between tieness and neighborhood overlap in the full SN. Such result may indicate whether the total value of neighborhood overlap of pairs of researchers in the sub-networks is not correlated to the total value of tieness, then the tieness also depends from other pairs of researchers from other sub-networks.

In summary, we observe that there are differences in the strength of ties when it is measured by neighborhood overlap, weight and tieness in each Qualis rankings. Our results show the best ranked venues (A1, A2 and B1) have similar patterns of strength of co-authorship ties – differently from the other Qualis rankings (B2, B3, B4 and B5).

## 7. Concluding Remarks

We identified some problems with using solely neighborhood overlap and weight to measure the strength of co-authorship ties. Then, we presented a new metric to measure such ties strength, called tieness, which has relatively low computational time cost and can be applied to other social networks types (since tieness is a topological feature). Also, we described a nominal scale to such metric that is based on the values of a modified neighborhood overlap and weight (since weight is part of the equation to compute the tieness). Such nominal scale allows to identify when a tie is weak or strong and if it links researchers from different communities or not. In addition, we checked the existence of a relation between the quality of a publication venue and the strength of ties when measured by neighborhood overlap, weight and tieness, and how the values of the three metrics are different from each other considering the publications' quality. Finally, our study showed that tieness provides more different values through the ties than neighborhood overlap and weight. This allows to better compare how strong a tie is regarding another one. We also found that high quality publications tend to have stronger ties than others.

As future work, we intend to consider neighborhood overlap, weight, tieness, publications' quality, temporal aspects and other topological properties as features to a computational model to automatically define the strength of co-authorship ties.

**Acknowledgments.** The authors thank the research funding agencies CAPES and CNPq.

## References

- Brandão, M. A. and Moro, M. M. (2015). Analyzing the strength of co-authorship ties with neighborhood overlap. In *Procs. of DEXA*, pages 527–542, Linz, Austria.
- Burt, R. S. (2010). *Neighbor networks: Competitive advantage local and personal*. Oxford University Press.
- Cheng et al., C.-B. (2014). Study of scientific collaborations in the intelligence and security informatics research community by social network analysis. In *Procs. of CSBC – BraSNAM*, Rio de Janeiro, Brazil.

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Second edition.
- Cormen, T. H. (2009). *Introduction to algorithms*. MIT press.
- Digiampietri, L. and Maruyama, W. (2014). Predição de novas coautorias na rede social acadêmica dos programas brasileiros de pós-graduação em ciência da computação. In *Procs. of CSBC – BraSNAM*, pages 243–248, Rio de Janeiro, Brazil.
- Ductor, L. (2015). Does co-authorship lead to higher academic productivity? *Oxford Bulletin of Economics and Statistics*, 77(3):385–407.
- Easley, D. and Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.
- Gilbert, E. and Karahalios, K. (2009). Predicting tie strength with social media. In *Procs. of SIGCHI*, pages 211–220, New York, USA.
- Goulas et al., A. (2014). The strength of weak connections in the macaque cortico-cortical network. *Brain Structure and Function*, pages 1–13.
- Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78(6):1360–1380.
- Jain, R. (1991). *The Art of Computer Systems Performance Analysis: techniques for experimental design, measurement, simulation, and modeling*. John Wiley & Sons.
- Kahanda, I. and Neville, J. (2009). Using transactional information to predict link strength in online social networks. *ICWSM*, 9:74–81.
- Laender et al, A. H. F. (2008). Assessing the research and education quality of the top brazilian computer science graduate programs. *ACM SIGCSE Bulletin*, 40(2):135–145.
- Lewis, P. and McKenzie, E. (1988). *Simulation methodology for statisticians, operations analysts, and engineers*, volume 1. CRC press.
- Machado, M., Andrade, R., and Serpa, R. (2013). Teambuilder: Uso de mídias sociais para a colaboração de grupos na organização de tarefas. In *Procs. of CSBC – BraSNAM*, Rio de Janeiro, Brazil.
- McGee, J., Caverlee, J. A., and Cheng, Z. (2011). A geographic study of tie strength in social media. In *Procs. of CIKM*, pages 2333–2336, New York, USA.
- Pappalardo, L., Rossetti, G., and Pedreschi, D. (2012). How well do we know each other? detecting tie strength in multidimensional social networks. In *Procs. of ASONAM*, pages 1040–1045, Washington, USA.
- Silva et al., T. H. P. (2014). Community-based endogamy as an influence indicator. In *Procs. of JCDL*, pages 67–76, Piscataway, USA.
- Wiese et al., J. (2015). You never call, you never write: Call and sms logs do not always indicate tie strength. In *Procs. of CSCW*, pages 765–774, New York, USA.
- Yan et al., R. (2012). To better stand on the shoulder of giants. In *Procs. of JCDL*, pages 51–60, New York, USA.
- Zignani, M., Gaito, S., and Rossi, G. P. (2016). Predicting the link strength of “newborn” links. In *Procs. of WWW*, pages 147–148, Switzerland.