Affiliation Influence on Recommendation in Academic Social Networks

Michele A. Brandão and Mirella M. Moro

Departamento de Ciência da Computação, Universidade Federal de Minas Gerais
micheleabrandao@dcc.ufmg.br, mirella@dcc.ufmg.br

Abstract. Social networks have been the focus of many studies, from communities' identification to link prediction. Here, we propose a method based on researchers' institution affiliation for predicting links in a collaboration social network. Initial experiments show that considering the institution affiliation aspect, the set of recommendations is more accurate and concise, leading to a more efficient result.

1 Introduction

A social network (SN) is a collection of individuals (or organizations) that have relationships in a certain context, *e.g.* friendship and co-authorship [8]. Those networks have been studied for over two decades in order to analyze the interactions between people and detect patterns in such interactions [1]. Understanding the mechanisms by which a SN evolves is a fundamental question that is still not well solved [4]. Many methods have been proposed for different aspects of SN analysis. Link prediction is one of such methods that may be applied in other functions, such as those in recommender systems. In this work, we study the recommendation problem in order to suggest new links in the network.

Specifically, we focus on a unique type of SN, the *academic social networks*, which are formed by researchers and their connections (given by paper and patent co-authorships, for example). In this research world, recommending new links may help a researcher to form new groups or teams, to search for new collaborations when writing a grant proposal and to investigate different research communities. Moreover, a recent work shows that research groups with well connected academic SN tend to be more prolific [6]. Discovering new links in this scenario is not a trivial task, because the social proximity has different interpretations in which the institutions, the connection between people and the academic context (*e.g.* affiliation and research area) must be considered.

Related Work. SN analysis has become important for academic research communities (such as mathematics [1]) to understand their own characteristics and behavior [2]. Specifically, co-authorship networks are an important class of academic networks and have been analyzed in search for different characteristics and behavioral patterns of scientific collaborations [5].

Given the researchers organized in a network, recommending people with whom a researcher may collaborate is a way of predicting links. Regarding link prediction, in [4], the authors evaluate topological measures (e.g., Jaccard coefficient) for classifying co-author collaborations. In [3], the author uses a topological measure to describe links occurrence. For instance, in [7], recommending collaborations considers each co-author as a link in an academic SN.

Contributions. Given an academic social network (with researchers connected to their co-authors) and the fact that well connected networks are more prolific, this work aims to recommend collaborations by predicting links between researchers. The novelty is: the recommendation function focuses on the researchers affiliations by prioritizing persons from institutions with which the researcher has already collaborated. The experimental evaluation shows that, when compared to the state of the art method, our function does indeed return more accurate and concise set of possible collaborators.

2 Recommending Collaborations

Overview. Social Networks are formed by *actors* and their *relational ties* [9]. The importance of a relationship between its actors may be defined by a weight measure. In this paper, we use an academic social network, in which two researchers are connected if they have co-authored a paper [9]. The final goal is to recommend new collaborations over the academic network, which is mapped to predicting links in a social network. We also explore how institution affiliations affect the relationship between researchers. Specifically, we present a new recommendation (or prediction) function that identifies collaborations that may be intensified as well as new collaborations that can be formed. The novelty relies in the function that considers the researchers' institution affiliation aspect.

Existing Recommendation Function. Different aspects of the SN may be considered for defining a recommendation function. One of the most relevant is the *weight* associated to the relationships. Each weight is important because it reflects the link's semantics, instead of just the network topological feature. In other words, the weight semantics represents rich information from the SN and their connections. Determining such weights is a great challenge and closely related to the type of data the network models. For instance, in [7] (which also uses a co-authorship network), the weight is defined by three different metrics: *cooperation (Cp)*, *correlation (Cr)* and *social closeness (Sc)*. *Cr* and *Sc* are combined to form an single metric Cr_Sc (weighted average) with weights w_{Cr} and w_{Sc} that define the importance of each one.

Affiliation Based Function. Here, we propose to consider the institution affiliation aspect in the relational ties' weighting metrics. As byproduct, having an institution-oriented weight provides more information to the SNA, such as assisting in the search for collaborations with different institutions and analyzing the influence of the cooperation with an institution upon the collaborations. Hence, we introduce the *affiliation index (Affin)* that represents the new weight. For any given pair of researchers $\langle i, j \rangle$, $Affin_{ij}$ is defined by Equation 1.

$$Affin_{i,j} = \frac{NPI_{i,j}}{NT_i} \quad (1)$$

where $NPI_{i,j}$ is the number of papers of researcher i co-authored with people from j's institution, and NT_i is the total number of papers authored by i. *Affin* follows the natural intuition that an institution is more important to an author, if he has collaborated with someone from that institution, and hence is more likely to contact other researcher in the same institution.

This way, we build upon the existing recommendation function from [7] by adding *Affin* to it. Then, for each pair of researchers, the relationship among *Affin*, *Sc*, *Cp* and *Cr* establishes the necessity (or not) of having more academic interaction between them (in order to improve the overall connection of the academic social network). We combine *Affin* and *Sc* to establish a single metric *Affin_Sc* defined by Equation 2.

$$Affin\_Sc_{i,j} = \frac{w_{Affin}.Affin_{i,j} + w_{Sc}.Sc_{i,j}}{w_{Affin} + w_{Sc}} \tag{2}$$

where given an academic network with the authors i and j, $Affin\_Sc_{i,j}$ is a weighted average, w_{Affin} and w_{Sc} weights determine, respectively, the importance of the metrics *Affin* and *Sc* to the resulting value. Hence, the weights may be used for emphasizing either the affiliation or the social closeness.

In order to equally consider $Affin\_Sc_{i,j}$, *Cp* and *Cr* indexes, we use degrees to represent ranges of values that are possible: "high", "medium" and "low". The actual values for the ranges may follow a linear scale (for example, $low < 33\%$ and $high > 66\%$). Equation 3 shows the combination of the indexes and their recommended actions: "Initiate_Collaboration" and "Intensify_Collaboration".

$$r_{i,j} = \begin{cases} Initiate\_Collab, & \text{if } (Cp_{i,j} = 0) \wedge \\ & (Affin\_Sc_{i,j} > \mathit{threshold}); \\ Intensify\_Collab, & \text{if } (Cp_{i,j} \in \mathit{low}) \wedge \\ & ((Affin_{i,j} \in \mathit{medium}) \vee (Affin_{i,j} \in \mathit{high})) \wedge \\ & ((Cr_{i,j} \in \mathit{medium}) \vee (Cr_{i,j} \in \mathit{high})); \end{cases} \tag{3}$$

where pairs of researchers with zero $Cp_{i,j}$ and nonzero $Affin\_Sc_{i,j}$ (we choose "low" degree as threshold) are recommended to create a collaboration; and pairs with "low" $Cp_{i,j}$, "medium" or "high" $Affin_{i,j}$, and "medium" or "high" $Cr_{i,j}$ are recommended to intensify their collaborations.

Example. Figure 1 exemplifies the use of Equation 3. It presents five researchers (from A to E) from different Brazilian institutions. For instance, considering the pair of researchers $\langle A, B\rangle$, there is no recommendation because $Affin\_Sc_{A,B} <$ "low" (moreover, $Affin_{A,B} = 0$ indicates that there has never been a collaboration between the researchers' institutions). On the other hand, $\langle B, C\rangle$ is recommended to initiate collaboration, because $Cp_{B,C} = 0$ and $Affin\_Sc_{B,C} >$ "low". One observation when comparing *Affin* to previous work, the method in [7] recommends to initiate collaboration in both cases.

Datasets. To evaluate *Affin*, we have applied the recommendation function to a real social network (from *CiênciaBrasil*[1]), and compared its results to the state of the art . Usually, precision is evaluated by actual users who verify whether the recommendations make sense. Here, we take a different approach. In order

[1] CienciaBrasil: http://pbct.inweb.org.br

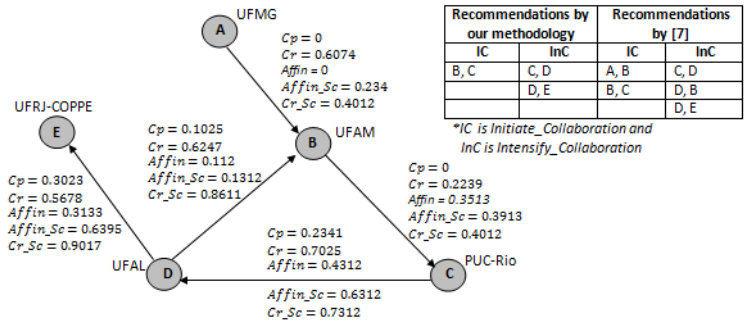
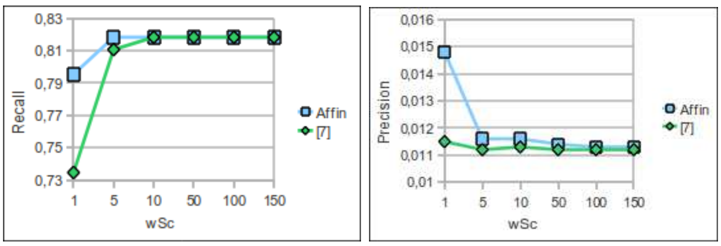


Fig. 1. Example of determining new collaborations within a partial Social Network



(a) Recall (b) Precision

Fig. 2. Effect of w_{Sc} variation on recommendation results

to have a fair comparison, we follow the evaluation given by [7] and consider the network in two moments: until 2007 and until 2010. The baseline for measuring precision and recall is given by the new collaborations that appeared between 2008 and 2010. By not considering the users' opinion, as [7], we expect to have small precision, giving more importance to the recall results.

Initial Results. Our method and [7] rely on the value of w_{Sc}. Hence, Figure 2 presents (a) Recall and (b) Precision obtained when varying the w_{Sc} values and using a fixed $w_{Affin} = 1$ (and $w_{Cr} = 1$). These results show that the precision and recall stabilize around $w_{Sc} = 10$. Then, the default value of weight w_{Sc} chosen for the remainder evaluations is 10.

Table 1 shows the number of relevant recommendations retrieved (underlined), the total number of recommendations retrieved (in parentheses), precision and recall for $Affin$ and [7] in the recommendation of new collaboration. The results show that varying the weight w_{Affin} and w_{Cr} (*i.e*, increasing the importance of the metrics *Affin* and *Cr* to the resulting value), *Affin* is more accurate and the number of relevant recommendations retrieved is greater than the state of the art [7]. Furthermore, when increasing the importance of *Affin*, the number of relationships recommended is reduced. Similarly, the results in Table 2 show that $Affin$ is also more accurate than [7] in the recommendation of intensify collaborations.

Table 1. New collaborations ($w_{Sc} = 10$)

Weight	Initiate	
w_{Affin}/w_{Cr}	**Affin**	**[7]**
1	108 (9314)	108 (9545)
precision	1.16%	1.13%
recall	81.82%	81.82%
5	107 (8653)	101 (9180)
precision	1.23%	1.10%
recall	81.06%	76.52%
10	105 (7117)	97 (8415)
precision	1.48%	1.15%
recall	79.54%	73.48%
15	97 (6131)	86 (7720)
precision	1.58%	1.11%
recall	73.48%	65.15%

Table 2. Intensify Collaborations

Method	Relevant/ Retrieved	Preci- sion	Recall
Affin	94 (223)	42.15%	71.21%
[7]	80 (196)	40.82%	60.60%

3 Conclusion

This paper introduced a new function for recommending collaborations in an academic social network. Its novelty relies on considering the institution affiliation aspect (given by a metric called $Affin$) with cooperation, correlation and social closeness metrics. Our experiments show the new function can reduce the number of recommendations and is more accurate than the state of the art. We are currently refining the recommendation function by evaluating other metrics that can improve even further the results. We are also working on a more through experimental analysis considering different networks.

Acknowledgments. This work was partially funded by CAPES, CNPq, Fapemig and InWeb, Brazil.

References

1. Barabasi, A. Linked: The new science of networks. Basic Books, 2002.
2. Ding, Y. Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. Journal of Informetrics 5(1), 187 - 203, 2011.
3. Huang, Z. Link prediction based on graph topology: The predictive value of generalized clustering coefficient. In: LinkKDD, USA, 2006.
4. Liben-Nowell, D., and Kleinberg, J. The link prediction problem for social networks. In: CIKM, USA, 2003.
5. Liu, X., et. al. Co-authorship networks in the digital library research community. Inf. Process. Manage. 41(6), 1462 - 1480, 2005.
6. Lopes, G.R., et. al. Ranking Strategy for Graduate Programs Evaluation. In: ICITA, Australia, 2011.
7. Lopes, G. R., et. al. Collaboration recommendation on academic social networks. In: WISM, Canada, 2010.
8. Monclar, R. S., et. al. Using Social Networks Analysis for Collaboration and Team Formation Identification. In: CSCWD, Switzerland, 2011.
9. Newman, M. E. J. The Structure and Function of Complex Networks. *SIAM Review* 45(2), 167 - 256, 2003.