Using Link Semantics to Recommend Collaborations in Academic Social Networks

Michele A. Brandão, Mirella M. Moro
Univ. Federal de Minas Gerais, Brazil
{micheleabrandao|mirella}@dcc.ufmg.br

Giseli Rabello Lopes
Pontifícia Univ. Católica do Rio de Janeiro, Brazil
grlopes@inf.puc-rio.br

José P. M. Oliveira
Univ. Federal do Rio Grande do Sul, Brazil
palazzo@inf.ufrgs.br

ABSTRACT

Social network analysis (SNA) has been explored in many contexts with different goals. Here, we use concepts from SNA for recommending collaborations in academic networks. Recent work shows that research groups with well connected academic networks tend to be more prolific. Hence, recommending collaborations is useful for increasing a group's connections, then boosting the group research as a collateral advantage. In this work, we propose two new metrics for recommending new collaborations or intensification of existing ones. Each metric considers a social principle (homophily and proximity) that is relevant within the academic context. The focus is to verify how these metrics influence in the resulting recommendations. We also propose new metrics for evaluating the recommendations based on social concepts (novelty, diversity and coverage) that have never been used for such a goal. Our experimental evaluation shows that considering our new metrics improves the quality of the recommendations when compared to the state-of-the-art.

Categories and Subject Descriptors

H.2 [**Database Management**]: Miscellaneous

General Terms

Experimentation, Measurement

1. INTRODUCTION

A social network (SN) is a collection of individuals (or organizations) that have relationships in a certain context, for example, friendship, politics and co-authorship. Social networks have been studied for over two decades in order to analyze the interactions between people and detect patterns in such interactions [3]. Social Network Analysis includes patterns and principles that are defined by social theories, such as homophily, proximity, contagion, etc. These patterns, principles and models provided by SNA can assist in exploring and predicting the individuals' behavior.

In this context, many methods have been proposed for different aspects of SNA, including viral marketing and link prediction [1, 22]. Furthermore, in such an online context, link prediction may also be mapped to link recommendation; so, instead of *inferring* future connections, it also allows to *suggest* new ones [21].

Among all types of social networks, our focus is on those where social links are given by research (or academic) ties. For example, an academic tie exists among people from the same research group and co-authors. Within those, *co-authorship social networks* are formed by researchers and their connections given by publication and patent collaborations. In this research-oriented world, recommending or predicting new links may help a researcher to form new groups or teams, to search for collaborations when writing a grant proposal, to improve the quality of communication in the network and to investigate different research communities. Also, recent work shows that research groups with well connected co-authorship SN tend to be more prolific [16].

Discovering new links in this scenario is not a trivial task. As pointed out by [15], when recommending new friendships in a traditional SN, the number of friends in common can be used to estimate the social proximity between users. On the other hand, in the academic context, social proximity has different interpretations in which the social connection between people and their academic background (e.g., geographic location and research area) must be considered. Specially, we are interested in the social proximity between researshers that is defined by social theories. Therefore, our focus is discovering how the homophily and proximity principles represented respectivelly by the institutional affiliation and the geographic location of the researchers (link semantics in the SN) increase the quality of the recommendations and influence in the collaboration.

Another problem is how to evaluate the recommendations. Common metrics for evaluating recommendations, such as precision and recall, practically do not explore any particular feature of the SN. Therefore, we employ SNA-based concepts for evaluating the recommendations from the SN perspective (which makes sense because the recommendations were defined from the social perspective as well).

The contributions of this work are summarized as follows: analysis of approaches for link prediction and recommendation of collaboration (Section 2); definition of two new metrics that consider social principles, called *Affin*[1] and *GLI*, and two recommendation functions to recommend collaborations considering links semantics (Section 3); description

[1]An initial version of this metric with a superficial experimental evaluation was briefly presented as a poster in [4].

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

of three metrics (*novelty*, *diversity* and *coverage*) for analyzing the quality of the recommendations (Section 4); an experimental evaluation using two real SN and comparison to the state-of-the-art (Section 5). We also combine our metrics to the state-of-the-art to increase the accuracy of its recommendations.

2. RELATED WORK

In this section, we discuss related work grouped in recommendation in social networks and social principles on link prediction. Then, we emphasize the main contributions of our work in the presence of such state-of-the-art.

Recommendation in Social Networks. Existing recommendation approaches suggest items (product, music, hotel, club) and people (friends, co-workers, lovers) to users in different settings such as e-commerce, online dating, social networks and employment websites. Among all of those, our work focuses on the social network setting.

The approaches presented in [9] and [11] recommend relevant items based on information extracted from SN. Recommending collaborations differ significantly from recommending items. Indeed, people-to-people recommendation must consider different aspects from the social connections [10, 15, 21]. Regarding people recommendation, a node similarity measure and an algorithm to recommend friends in SN is presented in [21], and a novel system for providing users with recommendations of people to invite into their explicit enterprise SN in [10]. Finally, the authors in [15] present a new methodology for recommending collaborations in academic social networks. These works are related for making people-to-people recommendation, but [15] differs from others due the kind of relationship recommended.

The requirements to recommend friends is different from recommending people to work with. For example, in [5] users are recommended to others based on similarity measures as *taste* (whom they like) and *attractiveness* (who likes them). However, this form to measure similarity cannot be applied in the academic setting, because it is not possible to infer how a researcher likes (or not) another.

Social Principles in Link Prediction. Different social principles may influence on predicting links. For instance, recent studies show that the homophily principle can improve link prediction [1, 22]. In [1], the authors developed an unsupervised model to estimate links' strength based on users' similarity and interaction activity. In [22], many measures considering the homophily principle in the human mobility have been explored to predict links. Others, as [19] and [22], predict new links in a social network considering both the homophily and the proximity principles (both use mobile phone to capture users trajectory). On the other hand, no work that uses proximity principle for predicting links in *academic* social network has been found so far.

Recommendation of Collaborations in Social Networks. Recommendation of collaboration is a specific recommendation problem in which two individuals are recommended to work together. For doing relevant recommendations, it is necessary to consider aspects that influence collaboration relationships. For example, in CORALS (Collaboration Recommendation on Academic Social Networks) [15], a weight represents each relation between researchers and is defined for the measures: *cooperation* (Cp, how much the two researchers have collaborated), *correlation* (Cr, how similar the areas of the researchers are) and *social closeness* (Sc, a normalized variant of the shortest path metric). Cr e Sc are combined to form a single, weighted average measure. Furthermore, the *cooperation* between authors a and b is a value in the range [0,1] defined by the ratio of the number of papers that a has co-authored with b by the total number of a's papers. The *correlation* is defined by an equation that considers the researchers publications area and the vector space model (VSM) to compute the values between each pair of co-authors in the network.

Discussion on Contributions. This work aims to investigate how our new metrics based on affiliation and geographic location information influence in recommending new links (or intensify existing ones) between researchers in an academic SN. The new metrics follow theoretical mechanisms (homophily and proximity principles) that have been used to explain the creation, maintenance and dissolution of SN [7]. These metrics explore weights and how different features on the SN (e.g., links semantics) affect the relationship between researchers. Determining such weights is a great challenge, because they should be closely related to researchers profile, the type of data and the network model.

The work more related to our is [15] (CORALS), whose emphasis is also on recommending collaborations in academic SN and considering the researchers publications area. Our work differs from CORALS, specially for considering social theories in the definition of the weights: homophily principle in *Affin*, and proximity principle in *GLI*. Moreover, the experimental evaluation of CORALS considers only the accuracy of the recommendations, whereas we also consider novelty, diversity and coverage.

3. RECOMMENDING COLLABORATIONS USING LINK SEMANTICS

Social Networks are formed by *actors* (people) and their *relational ties* (links) [18]. The importance of a relationship between its actors may be defined by a weight measure. Each weight is relevant because it reflects the link semantics, instead of just the network topological feature; i.e., the weight semantics provides rich information from the SN and its connections. We use an academic SN in which two researchers (actors in the network) are connected if they have co-authored a publication [18]. Although we focus on publication co-authorship, our metrics may be easily extended to work on similar relationships such as writing patents, editing books, proceedings and so on. The final goal is to recommend collaborations (new or intensification) over this network, which is mapped to predicting links in a SN.

Let $\mathcal{T}$ be a set of *target researchers* (i.e., the researchers that are going to receive the recommendations) and $\mathcal{R}$ the universe of researchers that will be evaluated (i.e., the researchers considered for the recommendation). Given a graph built from all researchers ($\mathcal{R} \cup \mathcal{T}$) and their connections (e.g., defined by their co-authorships), in which each link is associated to a set of weight values $\mathcal{W}$ that represents the semantics of the network (e.g., cooperation, number of papers co-authored and so on). The weights are combined to form a metric $\mathcal{M}$, which is then employed by a recommendation function. The recommendation function $f(\mathcal{T}, \mathcal{R}, \mathcal{M})$ evaluates the two input sets according to the metric and

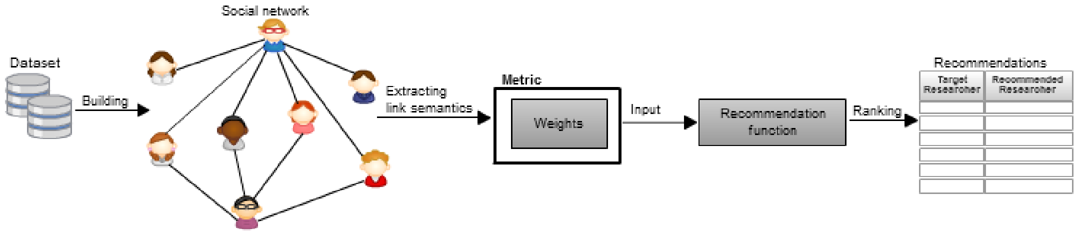


Figure 1: Framework to recommend collaborations.

returns a ranked list of recommended pairs $\langle t, r \rangle$ that maximizes the f value.

Figure 1 shows the framework for generating the recommendations. First, a SN is built from the existing datasets. The link semantics of the social relations define the weights. Then, the weights may be further elaborated in order to define the metric that composes the recommendation function. Finally, the recommendation function returns the ranked pairs of researchers. The hardest part of defining a recommendation function is choosing a proper metric. Next, we present both new metrics followed by a shared example.

3.1 Affin - Affiliation Metric

Affin is a metric that considers the homophily principle for recommending collaborations. This principle is derived from the institutional affiliation of the researchers and defined by the *affiliation weight* $Affin_{i,j}$, which represents the link semantics for any given pair of researchers $\langle i, j \rangle$ according to Equation 1,

$$Affin_{i,j} = \frac{NPI_{i,j}}{NT_i} \tag{1}$$

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 already collaborated with someone from that institution; hence, it is more likely to contact other researcher in the same institution.

Recommending based solely on the researchers' affiliations is not enough, because it disregards the history of the researchers' collaborations. Therefore, we propose it be combined with existing metrics in order to be more useful for the recommendation function. A more complete way of considering the affiliation aspect is combined with cooperation, social closeness and correlation aspects (from CORALS [15]). This combination allows to consider different characteristics between researchers in the recommendation function. The final goal is to have a recommendation function that is able to consider affiliation and the existing metrics in order to provide a better result and improve the overall connection of the academic social network.

Following [15] that combines correlation and social closeness, we combine $Affin_{i,j}$ and $Sc_{i,j}$ to establish a single weight $Affin\_Sc_{i,j}$ 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 a network with authors i and j, $Affin\_Sc_{i,j}$ is a weighted average, w_{Affin} and w_{Sc} weights determine, respectively, the importance of $Affin_{i,j}$ and $Sc_{i,j}$ to the resulting value. Hence, the weights may be used for emphasizing either the affiliation or the social closeness; i.e., allowing to emphasize the homophily in different ways.

In order to equally consider $Affin_{i,j}$, $Sc_{i,j}$, $Cp_{i,j}$ and $Cr_{i,j}$, *Affin* uses degrees to represent ranges of values: "high", "medium" and "low" that may follow a linear scale (e.g., *low* < 33% and *high* > 66%). Equation 3 shows the *recommendation function* that combines the metrics and returns two recommended actions: "*Initiate_Collaboration*" (InC) and "*Intensify_Collaboration*" (IntC),

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

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

We have performed an experimental evaluation combining *Affin* with the existing metrics in different ways and the results show that Equation 3 generates better results (due to lack of space, this prior results are not presented here).

Finally, it is also important to notice that *Affin* is more complete than its predecessor CORALS, because it regards the homophily principle. Moreover, 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.

3.2 GLI - Geographic Location Information Metric

GLI is a metric that follows the proximity principle. The theoretical mechanisms of this principle (that considers the influence of distance in the relationships) can be captured in the SN's links. In order to measure the physical proximity between pairs of researchers, we introduce the *geographic location weight* $GLI_{i,j}$ that considers the geographic location information for any given pair of researchers $\langle i, j \rangle$ defined by Equation 4,

$$GLI_{i,j} = distance(GC_i, GC_j) \tag{4}$$

where GC_i and GC_j represent the geographical coordinates of the researchers i and j institutions, respectively; and *distance* is the selected function to compute the distance between researchers locations. In this work, a geographical coordinate is that of the city where is a researcher's institution. The data with geographic location was gathered from

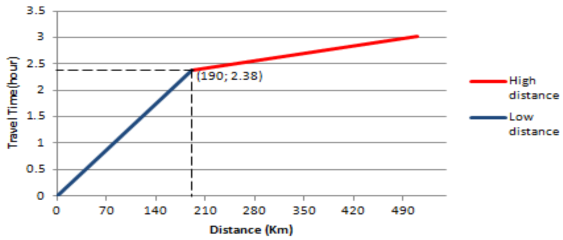


Figure 2: Distance versus Travel Time.

Wikimapia[2] and stored in a PostgreSQL[3] database. This DBMS (database management system) was chosen because it has an open source spatial database extension denominated PostGIS[4]. This extension provides operators and functions to manipulate geographic data. *ST_Distance* is one of these functions that was selected to calculate the distance between researchers.

In order to define a qualitative scale, we are interested in the *travel time* that covers the distance (represented by $GLI_{i,j}$) between researchers. It allows to specify how far two researchers are from each other. Thus, given a pair of researchers $\langle i, j\rangle$, the *travel time* is defined by Equation 5,

$$\begin{cases} \textit{if } GLI_{i,j} < \textit{190 Km}, & \Delta t_{i,j} = \frac{GLI_{i,j}}{80(Km/h)} \\ \\ \textit{else}, & \Delta t_{i,j} = \frac{GLI_{i,j}}{500(Km/h)} + 2h \end{cases} \tag{5}$$

where $\Delta t_{i,j}$[5] represents the *travel time weight.*

Equation 5 was defined considering that there is no flight when the distance is less than $190Km$ (because it is very short). Using land transports, the velocity is approximately $80Km/h$ that indicates a travel time approximately 2 hours. For longer distances, greater than or equal to $190Km$, the air transport compensates, due to reduction of the travel time. Moreover, $500Km$ of flight is approximately 1 hour, with more 1 hour to arrive and to leave the airports, the travel time would be 3 hours or less.

Figure 2 shows that there is an intersection between high and low distance in $190Km$ and 2.38 hours, because that it was defined that researchers are near when the *travel time* is less than 2.5 hours and far from each other when the *travel time* is greater than or equal to 2.5 hours. This defines a qualitative scale: "near" < 2.5 and "far" ≥ 2.5.

Equation 6 shows the *recommendation function* that combines $\Delta t_{i,j}$, $Cp_{i,j}$ and $Cr_{i,j}$ and its recommended actions:

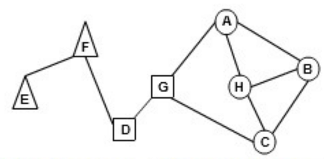
$$r_{i,j} = \begin{cases} InC, & \text{if } (Cp_{i,j} = 0) \wedge \\ & (\Delta t_{i,j} \in \{near\}); \\ IntC, & \text{if } (Cp_{i,j} \in \{low, medium\}) \wedge \\ & (\Delta t_{i,j} \in \{near, far\}) \wedge \\ & (Cr_{i,j} \in \{medium, high\}); \end{cases} \tag{6}$$

where pairs of researchers with zero $Cp_{i,j}$ and "near" $\Delta t_{i,j}$ are recommended to create a collaboration; and pairs with "low" or "medium" $Cp_{i,j}$, "near" or "far" $\Delta t_{i,j}$, and "medium" or "high" $Cr_{i,j}$ are recommended to intensify.

Note that the *travel time* was not combined with the social closeness as the *affiliation weight* in Equation 2. Prior experiments showed that combining them retrieved a larger set of recommendations without increasing the number of relevant recommendations.

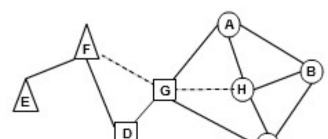
3.3 Example of using Affin and GLI

Figure 3 shows an example of using the new metrics and recommendations functions. Considering an academic social network in Figure 3(a) in which collaborations can be recommended to initiate or to intensify. In this SN, nodes with similar form belong to the same institution, and the weights of each relation are described in the table below. In order to simplify the explanation, the weights refer to only one direction, for example, the relation of A and B, but not B and A (depending of the direction, the weights vary). The *Affin* and *GLI* metrics are then applied to make this SN more connected.



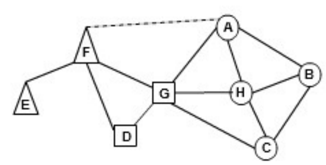
Relation	$Cp_{i,j}$	$Cr_{i,j}$	$Affin_{i,j}$	$Affin\_Sc_{i,j}$	$\Delta t_{i,j}$
A,B	0,67	0,4	0,82	0,74	0
A,H	0,23	0,33	0,82	0,74	0
H,C	0,77	0,52	0,64	0,53	0
B,C	0,85	0,70	0,78	0,61	0
G,A	0,84	0,81	0,49	0,34	5
G,C	0,92	0,55	0,49	0,34	5
G,D	0,88	0,71	0,73	0,70	0
F,E	0,13	0,57	0,80	0,75	0
F,D	0,79	0,62	0,47	0,43	6
B,H	0,67	0,86	0,78	0,61	0

(a) Original network



Relation	$Cp_{i,j}$	$Cr_{i,j}$	$Affin_{i,j}$	$Affin\_Sc_{i,j}$	$\Delta t_{i,j}$
G,H	0	0,43	0,49	0,34	5
F,G	0	0,57	0,47	0,43	6

(b) Recommendation by *Affin*



Relation	$Cp_{i,j}$	$Cr_{i,j}$	$Affin_{i,j}$	$Affin\_Sc_{i,j}$	$\Delta t_{i,j}$
F,A	0	0,62	0	0,24	0,81

(c) Recommendation by *GLI*

Figure 3: Example using *Affin* and *GLI* metrics.

Figure 3(b) presents the recommendation generated by *Affin*. This metric considers not only the relation between pairs of researchers, but also the relation of each researcher with other researchers from the same institution with which the former has already collaborated. Thus, the pairs of researchers $\langle F, G\rangle$ and $\langle G, H\rangle$ are recommended to initiate collaboration, because there is no cooperation between them

[2]Wikimapia: `http://wikimapia.org`
[3]PostgreSQL: `http://www.postgresql.org/`
[4]PostGIS: `http://www.postgis.org`
[5]Considering the International System of Units (meter)

($Cp_{i,j} = 0$) and $Affin\_Sc_{i,j}$ is greater than "low". In other words, F has collaborated with researchers from G's institution, and G with researchers from H's institution.

Likewise, Figure 3(c) shows the recommendation made by the *GLI* metric. In this case, the relation is established (or not) considering the physical distance and the travel time. Thus, the pair of researcher $\langle F, A\rangle$ is recommended to collaborate, because there is no cooperation between them ($Cp_{i,j} = 0$) and the weight ($\Delta t_{i,j}$) is "near".

Regarding the recommendation to intensify collaboration, the pairs of researchers $\langle A, H\rangle$ and $\langle F, E\rangle$ are recommended by *Affin* and *GLI*. Both relations have weights that satisfy the two metrics ($Affin_{i,j} \in \{medium, high\}$ and $\Delta t_{i,j} \in \{near, far\}$). Moreover, the researchers of the two pairs are correlated, i.e., they work in similar research areas.

4. EVALUATION METRICS

Evaluating the quality of recommendations and the effectiveness of recommendation functions are very difficult tasks, mainly for two reasons [8]: *(i)* different algorithms may have different performance on different datasets, and *(ii)* the goals for which an evaluation is performed may differ. Many studies focus on evaluating the accuracy, such as [13] and [15]. Having a high accuracy is important, but *insufficient* to ensure the quality of the recommendations [8, 20]. Therefore, it is important to consider a large number of metrics to analyze different aspects in the evaluation of the recommender systems (specially, in the context of recommendations).

Among a large number of metrics discussed in [20], accuracy, *novelty*, *diversity* and *coverage* are more appropriate to evaluate recommendation of collaborations. Metrics such as confidence, trust and utility are not appropriate, because prior information about researchers preferences is necessary but beyond our reach. Next, we detail each metric and show how they are employed for evaluating the recommendation results.

Accuracy. The accuracy of most recommender systems is evaluated according to precision and recall. However, calculating these metrics for a recommender algorithm presents some problems [2, 12]. First, these metrics require knowing whether each resulting item is relevant. In general, it is very difficult to define the item relevance. Second, there is a small number (in general) of relevant items in a items set. Third, it is necessary to consider resulting (recommended) items that are selected from a much larger set.

Therefore, the focus of this paper is on recall because: *(i)* in general, the networks are very sparse and the total number of possible links is large (as shown in Section 5); *(ii) Affin* and *GLI* metrics aim to make networks more connected, as opposed to totally connected; and *(iii)* high recall indicates that the metrics provide correct recommendations. Just to give an idea of result size, the average number of recommendations for each researcher is 176 in *CiênciaBrasil* and 22 in DBLP (details of these datasets will be presented in Section 5).

This decision (of focusing on recall) is also emphasized over the literature. Specifically, the work in [17] presents many examples of situations where high recall (and low precision) are useful, including: a commercial Web search engine like *Google* that reports more than 10^9 Web pages to a query with the word "software" and the effort involved in looking at a page is so low that users do not mind examining false results; and Huang et al won the "best paper" at the 2006 IEEE Requirements Engineering conference with a data mining method exhibiting precision *approx* 0.25 (even with low-precision, the analysis of results suggest that the proposed classification algorithm can effectivelly detect several different types of non-functional requirements) [6].

Novelty. *New recommendations* are indications of items that users do not know and would not know in absence of a recommender algorithm [8]. The *novelty* metric aims to quantify the "novel" characteristic in a recommendation list [8]. In order to compute this metric, we have adapted the idea proposed in [8] for the setting of academic SN.

Given a set of target researchers $\mathcal{T}$, a recommendation list $\mathcal{L}$ and the total number of target researchers n. First, we calculate the frequency $\Im_r$ of each recommended researcher r, where $r \in \mathcal{L}$. This frequency represents the popularity degree of the researchers, i.e., researchers with high frequency are likely to be known. In this case, we consider that the less popular a recommended research, the most probable he/she is unknown to a target researcher. Then, we take the median f_m as a central tendency metric to represent the frequencies (following the proposal from [8]). Finally, the frequency median f_m of the recommended researchers is divided by the total number of target researchers n. Hence, it allows to see the distribution of the frequency median in relation to target researchers.

The resulting value represents the *novelty* in a recommendation list. The *novelty* metric varies in the range [0,1], in which values near zero represent greatest novelty and the opposite when approaching one.

Diversity. In some cases, suggesting a set of similar items may not be useful for users. For example, considering the recommendation of researchers to collaborate. Presenting a list with 10 researchers, all from the same research area, may not be as useful as recommending researchers from different areas. This follows the intuition that researchers from the same research area probably already know each other.

The most explored method to measure *diversity* in a recommendation list is using the intra-list similarity metric [20]. We use this method based on the approach presented in [23], which evaluates traditional recommender systems. In addition, some changes have been made in this approach to evaluate recommendations of collaborations.

Given a set of target researchers $\mathcal{T}$, a recommendation list $\mathcal{L}$ and the total number of target researchers n. Equations 7 and 8 describe how to calculate the *diversity* using the intra-list similarity for any pair of researchers $r_1, r_2 \in \mathcal{L}$.

$$ILS(\mathcal{L}_t) = \frac{\sum_{r_1 \in \mathcal{L}} \sum_{r_2 \in \mathcal{L}, r_1 \neq r_2} Cr(r_1, r_2)}{2} \quad (7)$$

$$Diversity(\mathcal{L}) = \frac{\sum_{t \in \mathcal{T}} ILS(\mathcal{L}_t)}{n} \quad (8)$$

Equation 7 defines how to measure the similarity between recommended researchers in a recommendation list for each target researcher. In general, this similarity is defined by *Pearson's correlation* or *cosine distance* [23]. However, here we use the correlation among researchers Cr (defined in [15]), which represents the semantics of the SN relations (links), to calculate this similarity. Finally, Equation 8 measures the *diversity* of the recommendation list regarding all

Table 1: Information about the networks.

Information	*CiênciaBrasil* 90%	*CiênciaBrasil* 10%	*DBLP* 90%	*DBLP* 10%
Period in years	2000-2009	2009-2011	1971-2011	2011-2012
Total of publications	11,598	1,289	9,583	1,064
Publications average by researcher	34.11	3.79	15.24	1.69
Number of co-authorship relations	454	75	517	105

target researchers, where high values indicate low diversity. The values of *diversity* are not provided in any specific range. Thus, after computing this metric for different databases, the values are linearly normalized within $[0, 1]$.

Coverage. In recommender systems, coverage is represented by a metric that computes how unequally different the recommended items are to users [20]. Two different metrics of this distributional inequality are *Gini index* (GI) and *Shannon Entropy* (SE) [20]. Here, we compute such metrics through equations based on the work presented in [20].

Given a recommendation list $\mathcal{L}$, its total number of recommended researchers n and its total number of different recommended researchers d, Equations 9 and 10 compute the Gini index.

$$\mathcal{P}_r = \forall r \in \mathcal{L}, \frac{frequency(r)}{n} \tag{9}$$

$$GI(\mathcal{L}) = \frac{1}{d} \sum_{i=1}^{d} (2i - d - 1) S_{\mathcal{P}_{(i)}} \tag{10}$$

Specifically, Equation 9 calculates the proportion of each recommended researcher r in the recommendation list. Following the approach presented by [20], the set of proportion $\mathcal{P}$ is sorted (increasing order) as $S_{\mathcal{P}}$. Finally, in Equation 10, *Gini index* is computed according to approach defined by [20]. The index is zero when all researchers are recommended equally often, and one when a single researcher is always recommended.

In this work, we are interested in recommendations with *Gini index* near zero, because it means that each researcher receives recommendations according to his/her characteristics (affiliation, geographic location and similar researcher area). If all researchers receive the same recommendations, the recommendation metric may be wrong. Finally, Equation 11 shows how to measure Shannon Entropy.

$$SE(\mathcal{L}) = -\sum_{i=1}^{d} \mathcal{P}_{(i)} \log \mathcal{P}_{(i)} \tag{11}$$

The proportion of each recommended researcher is computed the same way that *Gini index* (Equation 9). The entropy is zero when a single researcher is always recommended, and $\log d$ when d researchers are recommended equally often (remember that d is the total number of distinct recommended researchers in the recommendation list). Similarly to *Gini index*, here we are interested in a recommendation list with many different researchers, i.e., Shannon Entropy near $\log d$.

5. EXPERIMENTS AND RESULTS

This section first details the datasets employed in our experimental evaluation (Section 5.1) and then presents the evaluation results (Section 5.2).

Table 2: New collaborations - Recall

Network	Affin	GLI	CORALS	CORALS+Affin
CiênciaBrasil	0.8533	0.6666	0.7733	0.8533
DBLP	0.8571	0.7647	0.8571	0.8571

5.1 Dataset Details

The experiments were performed using two real SN that were building from *CiênciaBrasil*[6] (considering 340 Computer Science researchers) and DBLP datasets (considering 629 researchers from 45 Brazilian institutions).

It is possible that only a subset of the researchers' publications is represented in the SN, when researchers may have other publications that are outside the datasets. However, given the coverage of both datasets in terms of conferences and journals, we believe that the most relevant part of the researchers publications is reflected in the datasets and is enough for providing good recommendations. Finally, the focus of this paper is in comparing the results across different evaluation metrics, not the absolute results themselves.

The social network of each datasets was divided in two parts (based on the concept of *split* [2]): 90% of the data as validation set, and the remaining 10% for testing. The first part (the large percentage of the data) was explored to create the researcher profile and the social network. The second, smallest part is the testing one, which means that it contains the expected results a recommender system should provide. Both parts also follow the time interval distribution, where the first part considers publications prior to the second part. In other words, the second part represents the "future" of the first one, and hence allows us to see what recommendations would be more useful based on the data of the first part.

Table 1 describes the splits from both datasets and their SN. It is clear that each SN is sparse (less than 460 relations of co-authorships from a possible total of 57,630 for *CiênciaBrasil*, and less than 520 from 197,506 for DBLP network). Consequently, there are many possible results for a recommendation function to consider (approximately, 57,170 for *CiênciaBrasil* and 196,986 for DBLP).

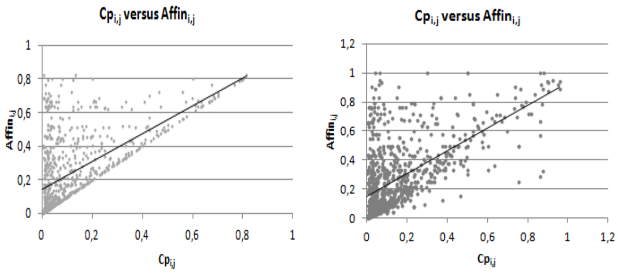
Furthermore, we compare the results of *Affin* and *GLI* with *CORALS*. *CORALS* builds the SN for each dataset considering the publications of all researchers with one relevant difference: it includes researchers correlated by researchers area and some level of social closeness. On the other hand, both *Affin* and *GLI* will consider the same universe of researchers that *CORALS* plus the researchers correlated by affiliation when building their SN. In order to provide a better comparison, we have combined *CORALS* and *Affin* in a new metric, called *CORALS+Affin*, that works on a network built as *CORALS* including all researchers correlated by research area, social closeness and affiliation. Note that we did not consider combining *CORALS* and *GLI*

[6]CiênciaBrasil: `http://pbct.inweb.org.br` [14]

Table 3: New collaborations - *Novelty* and *Diversity*

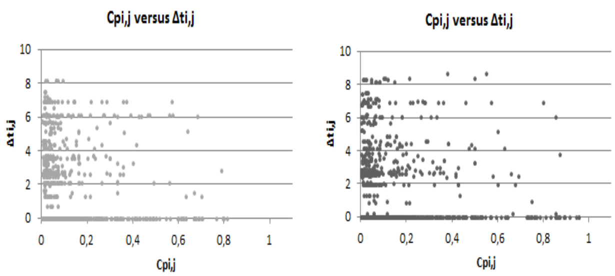
	Affin		GLI		CORALS		*CORALS+Affin*	
Network	**Novelty**	**Diversity**	**Novelty**	**Diversity**	**Novelty**	**Diversity**	**Novelty**	**Diversity**
CiênciaBrasil	0.139	0.75	0.1233	0.0	0.137	1.0	0.139	0.78
DBLP	0.124	0.96	0.044	0.0	0.124	1.0	0.124	1.0

Note: the higher the values, the worse the results.



(a) CiênciaBrasil (b) DBLP

Figure 4: The (clear) relation between Affin and Cooperation for *CiênciaBrasil* and *DBLP*.



(a) CiênciaBrasil (b) DBLP

Figure 5: The (non-existant) relation between travel time and Cooperation for *CiênciaBrasil* and *DBLP*.

because, as shown in the next section, there is no relation between cooperation and location.

5.2 Results and Discussion

We have grouped our experimental results as follows. First, we present the results when the metrics are used to recommend new collaborations. Then, we show the results when the metrics are used to recommend "intensifiable" collaboration (i.e., those existing collaborations that can be further intensified).

Recommending New Collaborations. Table 2 presents the experiments recall. The results show that in *CiênciaBrasil* and DBLP, *Affin* performs better than *GLI* and CORALS. The recall of *CORALS+Affin* is equal to *Affin*, because *Affin* considers just adds affiliation to the original *CORALS*. Indeed, a complementary result is illustrated in Figures 4(a) and 4(b). It shows that affiliation and cooperation ($Affin_{i,j}$ and $Cp_{i,j}$) are directly related. This fact explains why *Affin* provides more accurate recommendations.

Regarding the geographic location, *GLI* presents the worst accuracy results. For better understanding, the graphics in Figure 5 show that intensifying cooperation or improving *travel time* ($Cp_{i,j}$ and $\Delta t_{i,j}$) are not related. This is clear when observing that there are pairs of researchers (points) indicating high cooperation in high *travel time* and low cooperation in low *travel time*.

Table 3 shows the results to the *novelty* and *diversity*, in which the values in parentheses represents the *diversity* normalized in $[0, 1]$ (note that zero and one is only a representative value to compare the metrics). In both SN, *GLI* provides recommendations with more *novelty* and *diversity*. Furthermore, *Affin* presents the second best value for *diversity* and the same result that *CORALS+Affin* for *novelty*.

Table 4 shows that *GLI* generates a recommendation list with more unequally different researchers and presents the best results for GI and SE in both SN. *Affin* presents the second best result in *CiênciaBrasil* and the worst in DBLP. *CORALS+Affin* shows results better than CORALS for *coverage* in both networks, because *CORALS+Affin* considers more researchers than CORALS in the recommendation, which increases the difference between them.

The comparative analysis of Tables 2, 3 and 4 shows that even *GLI* presenting the worst results for accuracy, it presents the best ones for *novelty*, *diversity* and *coverage*. The reasoning for such results is as follows. Each target researcher receives recommendations considering similarity criteria (e.g., homophily or proximity principles); and increasing the number of recommended researchers (in this work, it increases the accuracy) also improves the similarity between them; hence, decreasing the *novelty* and *diversity*. Moreover, the number of researchers is finite, which means that the greater the number of recommended researchers, the less different they are in the resulting recommendation list; thus, the lower the *coverage*.

In general, *coverage* is related to *diversity* and *novelty*, because the more researchers considered, the more diverse is the recommendation list. Then, if there is more diversity, there is also more novelty.

Recommending "Intensifiable" Collaborations. As previously discussed, besides recommending new collaborations, we also work on recommending (existing) collaborations that can be further intensified. In order to evaluate the results of such recommendations, we consider only accuracy. Note that the other evaluations do not apply for intensifiable collaborations, because *novelty*, *diversity* and *coverage* make more sense when evaluating new collaborations.

Table 5 presents the results for accuracy of the recommendations to intensify collaborations. *GLI* shows recommendations with the best recall (for both networks), which is justified because it distinguishes researchers with "near" and "far" *travel time*, increasing the number of relevant results. *Affin* presents the second best recall (for the two social networks). This shows that *Affin* improves the accuracy of the recommendations. Finally, CORALS and *CORALS+Affin* present the same results.

Table 4: New collaborations - *Coverage*

Network	Affin		GLI		CORALS		*CORALS+Affin*	
	Gini I.	Shannon E.	Gini I.	Shannon E.	Gini I.	Shannon E.	Gini I.	Shannon E.
CiênciaBrasil	0.416	4.93	0.385	5.19	0.445	4.85	0.424	4.92
DBLP	0.492	5.214	0.473	5.46	0.490	5.217	0.490	5.218

Table 5: Intensify collaborations - Recall

Network	Affin	GLI	CORALS	CORALS+Affin
CiênciaBrasil	0.8831	0.9805	0.7467	0.7467
DBLP	0.7714	0.9518	0.7619	0.7619

6. CONCLUDING REMARKS

This paper presented two new metrics for recommending collaborations in an academic SN and analyzed how these metrics influence in the resulting recommendations. Given a recommendation function, the hardest part is to define which metric should the function rely upon when producing the results. The base of our work is to consider the social aspects when recommending collaborations to researchers. Specifically, we rely in considering the institutional affiliation aspect (*Affin*) and the geographic localization information (*GLI*) of all researchers in the network. We have also proposed evaluation metrics that consider novelty, diversity and coverage.

We have performed an extensive evaluation considering two real datasets and compared our metrics with the state-of-the-art. The results show that using *Affin* leads to an improvement in accuracy of the recommendations. Even though using the *GLI* presents the worst accuracy in the recommendation to initiate collaboration, it has a very positive impact when recommending intensifiable collaborations. Regarding *novelty*, *diversity* and *coverage*, *GLI* presents the best results and *Affin* the second best. Overall, the new metrics generate recommendations with more quality than the state-of-the-art (CORALS).

Ongoing work. We are currently working on evaluating datasets from other areas.

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

7. REFERENCES

[1] L. M. Aiello et al. Friendship prediction and homophily in social media. *ACM TWeb*, 6(2), 2012.

[2] R. A. Baeza-Yates and B. A. Ribeiro-Neto. *Modern Information Retrieval - the concepts and technology behind search*. Pearson Education Ltd., 2011.

[3] A.-L. Barabasi. *Linked: The New Science of Networks*. Perseus Books Group, 2002.

[4] M. A. Brandão and M. M. Moro. Affiliation Influence on Recommendation in Academic Social Networks. In *Procs. of AMW*, 2012.

[5] X. Cai et al. Learning collaborative filtering and its application to people to people recommendation in social networks. In *Procs. of ICDM*, 2010.

[6] J. Cleland-Huang et al. The Detection and Classification of Non-Functional Requirements with Application to Early Aspects. In *Procs. of RE*, 2006.

[7] N. S. Contractor, S. Wasserman, and K. Faust. Testing Multitheoretical, Multilevel Hypotheses about Organizational Networks: An Analytic Framework and Empirical Example. *Acad. Manag. Review*, 31(3), 2006.

[8] F. Fouss and M. Saerens. Evaluating Performance of Recommender Systems: An Experimental Comparison. In *Procs. of WI-IAT*, 2008.

[9] J. Freyne et al. Social networking feeds: recommending items of interest. In *Procs. of RecSys*, 2010.

[10] I. Guy, I. Ronen, and E. Wilcox. Do you know?: recommending people to invite into your social network. In *Procs. of IUI*, 2009.

[11] J. He and W. W. Chu. A Social Network-Based Recommender System (SNRS). Technical Report 090014, UCLA, 2010.

[12] J. L. Herlocker et al. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1), 2004.

[13] Z. Huang. Link Prediction Based on Graph Topology: The Predictive Value of Generalized Clustering Coefficient. In *Procs. of LinkKDD*, 2006.

[14] A. H. F. Laender et al. CiênciaBrasil-The Brazilian Portal of Science and Technology. In *Procs. of SEMISH*, 2011.

[15] G. R. Lopes et al. Collaboration Recommendation on Academic Social Networks. In *ER Workshops*, 2010.

[16] G. R. Lopes et al. Ranking Strategy for Graduate Programs Evaluation. In *Procs. of ICITA*, 2011.

[17] T. Menzies et al. Problems with Precision: A Response to "Comments on 'Data Mining Static Code Attributes to Learn Defect Predictors' ". *IEEE Trans. SE*, 33(9), 2007.

[18] M. E. J. Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45(2), 2003.

[19] D. Quercia and L. Capra. FriendSensing: recommending friends using mobile phones. In *Procs. of RecSys*, 2009.

[20] G. Shani and A. Gunawardana. Evaluating Recommendation Systems. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*. Springer, 2010.

[21] P. Symeonidis, E. Tiakas, and Y. Manolopoulos. Transitive node similarity for link prediction in social networks with positive and negative links. In *Procs. of RecSys*, 2010.

[22] D. Wang et al. Human mobility, social ties, and link prediction. In *Procs. of KDD*, 2011.

[23] C.-N. Ziegler et al. Improving recommendation lists through topic diversification. In *Procs. of WWW*, 2005.