Relatedness Measures between Conferences in Computer Science – a Preliminary Study Based on DBLP*

Suhendry Effendy
effendy@comp.nus.edu.sg

Irvan Jahja
jahja@comp.nus.edu.sg

Roland H.C. Yap
ryap@comp.nus.edu.sg

School of Computing
National University of Singapore
13 Computing Drive, Singapore

ABSTRACT

A large percentage of the research in computer science is published in conferences and workshops. We propose three methods which compute a "relatedness score" for conferences relative to a pivot conference, usually a top rated conference. We experiment with the DBLP bibliography to show that our relatedness ranking can be used to help understand the basis of conference reputation ratings, determine what conferences are related to an area and the classification of conferences into areas.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*information filtering*; K.4.m [**Computers and Society**]: Miscellaneous

Keywords

bibliographic database; publication analysis; ranking; social network

1. INTRODUCTION

It is common in computer science for research to be published in conferences or workshops. In this paper, we will use the term *conference* to refer to either a conference or a workshop. Thus, the range of research published in conferences is large and diverse, covering most of computer science. Some conferences are mainly concerned with the research in a particular area but other conferences may span more areas which may be related or quite distinct. Conferences are also dynamic, with new conferences created over time or changes/mergers in existing ones.

An interesting question is to study what can be learned about the structure of conferences in computer science and their relationships with each other. To illustrate, we investigate how to answer questions such as the following:

> What conferences are highly related to one another? Are conferences which have a broader set of areas more related to more conferences than narrowly focused conferences? A ranking of conferences by reputation is a complex and consensus-based and manual process. Can automatic methods help with this process?

We propose some simple methods as an initial attempt to help analyse conferences by how related they are to some selected conferences, the pivot. We apply our methods to analyse the DBLP computer science bibliography. We show in our evaluation that the relatedness ranking produced can correlate well with the reputation ranking from CORE (Australian Computer Research and Education conference ranking). For example, our evaluation of our preliminary experiments (see Appendix for conference abbreviations) shows that VLDB is closer to ICDE than ICDT which shows the differences between "system versus theory" conferences. We also see that conferences in Cryptography are closely related to each other since the community is tightly knit, [11] find that authors in Cryptography collaborate intensively. However, theory conferences, e.g., STOC and FOCS are also related to cryptography conferences. Thus, we see that the basic methods proposed here are able to give relatedness rankings for conferences which correspond well with intuition. We were surprised how well it supports the CORE ratings and also has good consistency with the conference rankings from Microsoft Academic Search.

1.1 Related Work

There are many works which analyse scientific collaboration networks such as the early work of [12]. Most works [4, 6, 7, 8, 11, 12, 13, 14] analyse the structure of scientific research primarily with respect to the author collaboration graph. However, we are concerned with a graph/network of the conferences. DBLP has been used as a dataset [6, 7, 8, 11] in many of the studies. Bird et al. [6] mention various data cleaning issues when processing the DBLP bibliographic data, we encounter the same as well as other issues as we also make use of the CORE conference ranking list.

Many papers [4, 6, 12, 13, 14] show that the collaboration network of authors has characteristics of a small world graph. The author network has been shown to have a single giant component [4, 12, 13] and extensive bibliometric statistical analysis in [7] shows this for a subset of conferences which they selected to be Database venues in DBLP. However, our conference graph is not a small world network and resembles a dense graph. Rather than having many compo-

*We acknowledge the support of grant T1 251RES1207.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2579035.

nents dominated by a single giant component, we find it to be just a single component.

While most papers are focused on analysis of authors, there are some related similarities to our approach. We use weighted graphs in our analysis, similar to Newman [14], but our weights are based on the set and multiset Jaccard index. We compare with conference reputation and make use of highly reputable conferences, this is also done by [6, 11], however, they manually construct their list. We instead show that choosing the top-k relatedness ranking of conferences produced by our approaches correlates well with conference reputation. The difference in this paper is that rather than analysing conferences in the aggregate, e.g. statistical analysis, we want the rankings produced to help in comparing conferences with each other. For example, this could be a human conference ranking which may still need to be subjective but could be informed by objective processes.

2. ANALYSING CONFERENCES

2.1 Preliminaries

A bibliographic database, such as DBLP, contains the details of papers such as title, authors, etc., published at a conference, journal or a workshop.[1] In this paper, we will focus on conferences and workshops and will simply use *conference* to refer to either a conference or a workshop.

Papers are published at a conference c where each paper has a set of authors $\{a_1, a_2, \ldots\}$. This information can be modelled as a bipartite graph $G_{c,a}$ where for each paper p in conference c_i, there is an edge from c_i to each author a_j in p. We use the graph with only conference vertices, G_c which is a projection of $G_{c,a}$ as follows – graph G_c has edge (c_i, c_k) if $G_{c,a}$ has edges (c_i, a_j) and (c_k, a_j), i.e. a_j has a paper in both c_i and c_k [15].

The graph is weighted, each edge (c_i, c_j) has weight $w_{i,j}$. In addition, each vertex i also has a weight denoted by x_i. The function $\alpha(c)$ returns a set (respectively, multiset) of authors who have written a paper in conference c.

2.2 Measuring the Relatedness Between Conferences

Our goal is to define a "relatedness measure" for conferences. Rather than doing this between every pair of conferences, we propose that it makes more sense to make use of a reference point which we call the *pivot*. Intuitively, the pivot p is meant to be the set of most important conferences for an area in computer science. Ideally, the number of conferences in a pivot should be small, possibly just one. For example, in databases, a reasonable choice is SIGMOD or VLDB. The idea is that it is easier to employ a relative measure than an absolute one and adding the notion of an area allows us to ask some interesting questions (see Sec. 4).

We now define a number of approaches for measuring relatedness of a conference to its pivot. Each of these approaches has an intuitive rationale with a simple definition. The goal of all these algorithms is to assign a value to each conference, denoting the similarity or relatedness of the conference to the given pivot (the larger the value, the more similar the conference is). We denote the value for conference c_i by x_i.

[1] The distinction between a conference and a workshop is not always clear. For example, ALENEX and PEPM.

2.2.1 *Relatedness by Direct Jaccard Index (D)*

The Jaccard index $J(A, B)$ is used as a measure of similarity between two sets, A and B, and is defined as $\frac{|A \cap B|}{|A \cup B|}$. In the conference graph G_c, each vertex c_i is associated with a set (respectively, multiset) of authors. We assign weight $w_{i,j}$ to every edge (c_i, c_j) in G_c such that $w_{i,j} = J(\alpha(c_i), \alpha(c_j))$. Furthermore, we define $w_{i,j} = 0$ if there is no edge (c_i, c_j).

In our simplest approach, we view $w_{p,i}$ as giving the direct similarity between pivot p and conference i, thus, we define x_i to be the maximum weight of edges connecting the pivot(s) to i:

$$x_i = \max_{c_p \in \text{pivot}} w_{p,i}$$

In the case when the pivot is just one conference, $x_i = w_{p,i}$. The definition of x_i can be used with the authors in conference i modelled as a set or a multiset. In our experiments, we find the multiset treatment to be better.

This definition only works for conferences which are adjacent to p. However, we will see in the experiments that it does surprisingly well. We now turn to measures which go beyond direct edges on p.

2.2.2 *Relatedness by Random Walk (RW)*

In the second approach, the idea is to view the relatedness between conference i and j as a probability $w'_{i,j}$, for example, it may represent that probability that an author publishing a paper in conference i will also publish one in conference j. This is a natural extension of the Jaccard approach to compare conferences which are not adjacent to the pivot.

We model this as a random walks starting from the pivot where x_i is the expected number of times vertex i is visited in a walk up to length L. Let the transition probability of visiting j from i be $w'_{i,j}$ which is obtained by normalizing the Jaccard index as $w'_{i,j} = w_{i,j}/\omega_i$ for $i \neq j$, otherwise 0, and $\omega_i = \sum_{j \neq i} w_{i,j}$. Note that $w'_{i,j}$ is not symmetric unlike $w_{i,j}$ so the graph can be viewed as being directed.

We compute the vector X containing the vertex value x_i as follows. Let $E(s)$ be a vector at time step s and W' be a matrix with transition probabilities $w'_{i,j}$. We define the expected value of X from the random walks of up to length L from pivot p as:

$$E(s) = W' \times E(s-1)$$

$$E(0)_i = \begin{cases} 1, & \text{if } c_i \in \text{pivot} \\ 0, & \text{otherwise} \end{cases}$$

$$X = \sum_{s=0,\ldots,L} E(s)$$

Note that when $L = 1$, this approach reduces to relatedness by Direct Jaccard index as described in Sec. 2.2.1. We propose that L should be short to focus the similarity measure on the pivot otherwise the pivot may not have much effect. We would also argue that all other things being equal, we would prefer conferences which are closer to the pivot than conferences further away, which can be achieved using a shorter L.

2.2.3 *Relatedness by Pivot Aggregation (PA)*

In our third approach, the basic idea is that a conference is considered to be similar to the pivot if it is also similar to other conferences that are similar to the pivot. In addition,

Table 1: CORE Conference List & Ranks

Rank	A*	A	B	C	Total
Count	61	186	319	373	939

we incorporate the observation that some authors only publish in only one conference. In the extreme case, an author may have published a single paper. In such cases, we think that the similarity should also be decreased by a certain factor, which we call the *damping factor* (analogous to the damping factor in PageRank algorithm [9, 10]). We model this with a linear programming formulation where a conference's similarity value, x_i, is proportional to the weight of edges to its neighbours in G_c as follows:

$$\begin{aligned} \text{maximize: } & \sum x_i & \\ \text{subject to: } & x_i = 1 & c_i \in \text{pivot} \\ & x_i \leq (1 - d \cdot f_i) \cdot \sum_{j \neq i} \frac{x_j \cdot w_{i,j}}{\omega_i} & c_i \notin \text{pivot} \\ & 0 \leq x_i \leq 1 & \end{aligned}$$

In the above, ω_i is the sum of the outgoing weights of a vertex as in Sec. 2.2.2, f_i denotes the fraction of $\alpha(c_i)$ that publish papers exclusively in conference c_i

$$f_i = \frac{|\{a_j | a_j \in \alpha(c_i), i \neq k \rightarrow a_j \notin \alpha(c_k)\}|}{|\alpha(c_i)|}$$

and $0 < d \leq 1$ is a discount factor.

We briefly discuss the necessity of the damping factor which is $(1 - d \cdot f_i)$. Without the term $(1 - d \cdot f_i)$, for example, if either $d = 0$ (discount all) or $f_i = 0$ (no authors publishing exclusively in c_i) for all i, the linear program can be satisfied by setting $x_i = 1$ for all i which makes this formulation useless. In our dataset, we find for all conferences $f_i > 0$ (which is not surprising) with an average value of f_i of about 18.4%. A major contributor to the value of f_i are authors who write only one paper (and by extension, only publish in one conference). This is in line with the observation from Elmaciouglu, et al. [7] which suggests that the number of authors who publish single paper is high. We can argue that even if $f_i = 0$ for a new conference, it can be expected in future to have single conference authors.

The discount factor serves to modify the effect of the damping factor. Through experiments with various values of d, we found that choosing $d = 0.1$ gives good results. When $d = 1$, the similarity value of all non-pivot conferences become too low to have significant impact on their neighboring conferences (consequently, the conferences are ranked solely by the weight of the edge connecting it to the pivot).

3. DATASET

DBLP [1] is a publicly available computer science bibliographic database. We downloaded the data from DBLP on 23 December 2013 and extracted only papers published in conferences. In addition, we employ the conference list from Computer Research and Education (CORE) [2] which also classifies conferences by their reputation, dividing conferences into ranks A*, A, B and C. While such a reputation ranking is admittedly subjective, there is correlation between the CORE ranking and other conference rankings. For the purposes of this paper, we employ the CORE rankings in two ways: (a) we use the A* ranked conferences as pivot; and (b) we used the ranking as one way of evaluating our experiments. The A* conferences are those which flagged as flagship conferences by CORE.

The complete CORE list contains conferences from many disciplines (denoted by their FoR-Field of Research). As we focus on computer science, we extract entries where FoR relates to computer science areas, i.e. 8*, 1005, and 1006. This can include engineering conferences which might be outside computer science, e.g. device physics, but this is not a problem as we then only select conferences which are in both DBLP and the selected CORE list.

Pre-processing and cleaning of the data is needed. For example: (a) some conferences are journals, e.g., papers in VLDB from 2008 onwards are published in the PVLDB journal; (b) inconsistent conference naming; (c) multiple entries for the same conference; and (d) name-abbreviation mapping differences between DBLP and CORE. The data cleaning process cannot be fully automated and thus requires non-trivial manual effort. For our preliminary experiments, we have not checked all conferences and have prioritized data cleaning on conferences with rank A*, A and B in CORE. Intersecting the cleaned CORE and DBLP dataset gives 939 conferences – the number of conferences in the various ranks are given in Table 1.

Table 2: Dataset and Graph Statistics

	All	A*
# of conferences	939	61
# of papers	712,780	99,346
# of authors	546,726	95,882
Minimum degree	52	334
Maximum degree	921	906
Median degree	598	757
Average degree	569.91	732.93
# of edges	267,573	–
Diameter	2	–
Average shortest path	1.39	–

We created the weighted conference graph G_c, the statistics of the graph and papers are given in Table 2 where the "All" column is on all conferences in the graph while the "A*" column is only on the A* conferences. The constructed graph is quite dense with ∼60% edges of the edges of a complete graph. Dense graphs can be expected from projection even if the original bipartite graph is not dense [15]. The high degree also leads to the average shortest path being as small as 1.39. This suggests that using only the geodesic distance to measure the similarity between two conferences is not a good idea (all conferences will be "similar"). We see that as the A* degree is high, the Direct Jaccard approach is able to rank a large percentage of the conferences.

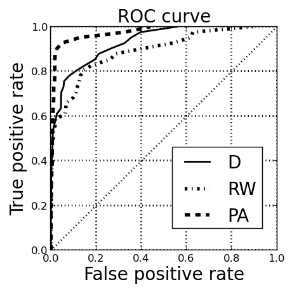

Figure 1: ROC curve of the 3 methods evaluated on [7]

4. PRELIMINARY EXPERIMENTS

We show experiments on three different computer science areas to evaluate our methods. The following areas are chosen because they have different characteristics: Database, Cryptography, and Web. The corresponding pivot selected are VLDB (Database), WWW (Web), and CRYPTO (Cryptography). In order to understand the evaluation, *familiarity with conferences in these areas is needed.* This is because such evaluations are necessarily subjective. As there is no ground truth, we use some proxies which are probably subjective or biased in some fashion, nevertheless, they have some broad use in the computer science community.

In this section, we distinguish between "ranking" which is simply a score produced by an algorithm and "rating" which is the reputation of a conference (our reputation rating is from CORE). Our three methods give a score. The top-k scores give a ranking of the k conferences having the highest x_i values relative to the pivot. We abbreviate our approaches as D (Direct Jaccard Index), RW (Random Walk) and PA (Pivot Aggregation). We treated conference authors as multisets which gave slightly better results than sets. The walk length for RW was 4.

Some questions which our relatedness measures can help to answer are the following. What are some of the relevant conferences in the area? What is the correlation between ranking and rating? Are the rankings of broad conferences different from more narrowly focused conferences?

We compare our results with the rankings given by Microsoft Academic Search [3], also known as *Libra*. We use their Database and WWW conference classification (called "field of study" in Libra). We did not do this for Cryptography as Libra combines Cryptography into a broader Security and Privacy area. It is important to note that the scores (thus, ranking) from Libra are based on citations which has more information than what we use, namely, the bibliographic entries of papers.

Our D approach is only applicable for conferences that are connected to the pivot. The number of conferences which do not have a valid Jaccard index are as follows – VLDB:48, WWW:55, and CRYPTO: 327. The number is higher for Cryptography probably because it is a more focused and tighter community than Database and WWW.

Database

Database is chosen being a moderately broad area where the data management can be combined with theory and systems. Table 3 shows our ranking and also Microsoft Academic Search for the Database area. VLDB was used as the pivot, as it is rank 1 in Microsoft Academic Search. The table shows the top 10 ranked conferences from our approaches and Libra, i.e. the union of the top 10 conferences. The conference name is superscripted with its rating from CORE. The table is sorted by its rank in the PA approach. The columns M-FR and M-C are from Libra ranked with Field Rating (M-FR, a variant of H-index) and number of citations (M-C). The superscript in the M-FR and M-C columns is the respective score. In Libra, KDD is classified under Data Mining and WWW under itself, hence they do not appear in the ranking. The last two rows count the number of top 10 ranked conferences which are rated A*, and A* or A.

Table 3: Ranking with VLDB as pivot

Conference	D	RW	PA	M-FR	M-C
VLDB ^{A*}	1	1	1	1^{154}	1^{121079}
SIGMOD ^{A*}	2	2	2	2^{117}	3^{67032}
ICDE ^{A*}	3	3	3	3^{104}	2^{67380}
EDBT ^{A}	4	4	4	8^{52}	10^{11687}
CIDR ^{A}	11	18	5	23^{33}	21^{5638}
PODS ^{A*}	6	6	6	4^{100}	5^{45232}
SSDBM ^{A}	8	10	7	19^{34}	17^{6441}
CIKM ^{A}	5	5	8	7^{67}	6^{28623}
WEBDB ^{C}	24	29	9	16^{39}	20^{5701}
ICDT ^{A}	14	13	10	11^{48}	12^{11559}
DASFAA ^{A}	7	7	12	28^{26}	27^{4003}
KDD ^{A*}	9	8	13	MINING	MINING
ER ^{A}	13	14	18	9^{48}	9^{12742}
WWW ^{A*}	10	9	19	WWW	WWW
DEXA ^{B}	12	12	21	18^{35}	8^{14662}
ISWC ^{A}	45	53	57	6^{69}	7^{22144}
AFIPS ^{C}	63	307	59	10^{48}	10^{11841}
HICSS ^{A}	169	254	474	5^{69}	4^{62798}
# top 10 A*	6	6	4	4	4
# top 10 A*/A	10	10	9	9	9

When interpreting the ranking, it is important to remember that it is not a reputation rating, rather, it is an ordering by score to obtain a relatedness measure. We can see that the top 10 is all A*/A except for PA. The rankings are also close to the union of the top 10 M-FR and M-C: D and RW have 2 outside and PA has 4 outside. Libra does not classify KDD and WWW under Database. It is unclear how Libra does classification. In addition, some conferences are listed in multiple areas, e.g. ISWC as Database and WWW. It may not be so simple to (automatically) classify the area(s) of a conference, for example, why does Libra classify KDD only as Data Mining. Our approach, on the other hand, shows the expected link between Database with Data Mining and WWW. Area connection is illustrated with ISWC (Semantic Web), this is ranked 45 and lower in our approaches but very high in M-FR (6) and M-C (7). We would argue that ISWC is further away from VLDB and core database unlike the ranking in Libra (this is reversed with WWW as the pivot).

PODS (6) and ICDT (10-14) are ranked lower and the ranking is consistent with Libra. However, PODS and ICDT tend more towards the theoretical aspects of Database. Selecting PODS as the pivot (not shown), the ranking in all our approaches is PODS (1) and ICDT (2). This explains our results, as VLDB has less of a theory focus, it is reasonable that PODS and ICDT are considered further away than

Table 4: Ranking with WWW as pivot

Conference	D	RW	PA	M-FR	M-C
WWW ^{A*}	1	1	1	1^{110}	1^{60781}
WSDM ^{B}	3	5	2	19^{19}	21^{1499}
SIGIR ^{A*}	4	3	3	IR	IR
CIKM ^{A}	2	2	4	IR	IR
ISWC ^{A}	6	12	5	2^{69}	2^{22150}
Hypertext ^{A}	13	30	6	3^{63}	3^{16319}
KDD ^{A*}	5	4	7	MINING	MINING
ECIR ^{B}	14	22	8	IR	IR
VLDB ^{A*}	7	8	9	DB	DB
SIGMOD ^{A*}	10	9	10	DB	DB
WEBDB ^{C}	29	44	11	6^{39}	8^{5701}
ICDE ^{A*}	9	6	12	DB	DB
ICDM ^{A*}	8	7	14	MINING	MINING
EDBT ^{A}	12	10	15	DB	DB
WISE ^{A}	11	11	29	11^{28}	9^{4788}
IMC ^{A}	40	51	39	4^{56}	4^{10304}
ICWS ^{A}	18	21	61	7^{37}	5^{8197}
USITS $^{-}$	-	-	-	5^{48}	6^{8002}
INET $^{-}$	-	-	-	8^{32}	11^{3322}
ECHT $^{-}$	-	-	-	9^{31}	13^{3006}
AH $^{-}$	-	-	-	10^{30}	10^{3361}
WI $^{-}$	-	-	-	12^{25}	7^{6262}
# top 10 A*	7	7	5	1	1
# top 10 A*/A	9	9	8	5	6

Table 5: Ranking with CRYPTO as pivot

Rank	D	RW	PA
1	CRYPTO^{A*}	CRYPTO^{A*}	CRYPTO^{A*}
2	EUROCRYPT^{A*}	EUROCRYPT^{A*}	EUROCRYPT^{A*}
3	ASIACRYPT^{A}	ASIACRYPT^{A}	TCC^{A}
4	TCC^{A}	PKC^{B}	ASIACRYPT^{A}
5	PKC^{B}	TCC^{A}	FSE^{B}
6	CT-RSA^{B}	CT-RSA^{B}	PKC^{B}
7	FSE^{B}	FSE^{B}	SACRYPT^{B}
8	SACRYPT^{B}	SACRYPT^{B}	ANTS^{B}
9	CCS^{A*}	ACNS^{B}	CT-RSA^{B}
10	STOC^{A*}	ISW^{B}	SCN^{B}
11	FOCS^{A*}	SCN^{B}	ICITS^{C}
12	FC^{A}	CCS^{A*}	CHES^{C}
13	SCN^{B}	FC^{A}	INDOCRYPT^{B}
14	CHES^{C}	INDOCRYPT^{B}	PAIRING^{C}
15	ISW^{B}	ICISC^{B}	FC^{A}
16	ACNS^{B}	STOC^{A*}	WEWORC^{C}
17	INDOCRYPT^{B}	FOCS^{A*}	MMICS^{C}
18	ICISC^{B}	CHES^{C}	ICISC^{B}
19	ICALP^{A}	ICICS^{B}	CARDIS^{B}
20	COCO^{A}	CANS^{B}	WAIFI^{C}
# A*	5	5	2
# A*/A	10	8	5

more system conferences. Thus, our results can be tuned by changing the pivot.

CIDR is an interesting result. As it is about innovative ideas, many influential ideas appear in CIDR. This is reflected in the PA rank (5) while D (11) and RW (18) is lower. However, CIDR is much lower in M-FR (23) and M-C (21).

HICSS is interesting, it is ranked 5 and 4 in M-FR and M-C. However, HICSS is a general conference with many tracks from different areas rather than specifically Database. Libra, however, classifies HICSS only as a Database conference. We rank HICSS very low, lower than 169, which seems intuitively better than the M-FR and M-C result. A similar issue arises with AFIPS which is ranked 10 in Libra.

We also compared against another subjective classification. An extensive study of the Database community in DBLP is done in [7]. In addition, they handpick a list of conferences in Database in DBLP from 1968-2003. We compare our results against their manually constructed list (see Appendix) by using our ranking as a binary classifier where all conferences ranked l or higher are classified as Database. We then evaluate the performance of this classifier using the ROC curve [5].[2] Figure 1 depicts the ROC curve for all of our approaches. We see the cutoff l works well as a classifier with the PA approach doing much better.

WWW

The WWW area was chosen as it is broader than Database and is related to many areas. Furthermore, several Database conferences can be regarded as relevant to the WWW area. Thus, it is interesting to compare the Database conferences with WWW as the pivot. Table 4 compares the results. The last five conferences are in Libra but not in CORE, thus, we did not compare them.

The results are more diverse than with Database – many of the conferences that are highly ranked are not classified under WWW in Libra. This is not surprising with a broad area like WWW, and furthermore, Libra has separate classifications for the areas of Database, Data Mining and Information Retrieval. Our results are reasonable as the ratings of most of the top 10 are A*/A, and topic wise the conferences have an intersection with WWW.

Information Retrieval conferences (SIGIR, CIKM) are ranked high. Notice that SIGIR does not appear in the Libra Database ranking while CIKM does. This is because Libra classifies CIKM under three areas: Data Mining, Information Retrieval and Database. We see that ISWC is closer to WWW with WWW as the pivot, but further away, with VLDB as the pivot.

Cryptography

Cryptography is chosen as a more tightly focused area. Table 5 depicts the top 20 results of all three approaches. All the three approaches produce similar results for the top rankings which correlates well with the ratings. We notice that PA has fewer A*/A rated conferences, however, a closer look shows that in terms of relevancy, all the conferences are indeed the most relevant to Cryptography. Whereas D and RW also include conferences which are broader, for example, CCS is an A* security conference but it includes any security topic and less focused on Cryptography. We also see a similar effect with STOC and FOCS which are A* Theory conferences, relevant but not focused on cryptography. Finally, RW and D have similar results - having the same set

[2]The ROC curve plots true positive versus false positive rate – we vary l to obtain the graph. The area under the ROC curve is the probability that a randomly picked Database conference being ranked higher than a non-Database one.

of conferences for their top 18 results. This may be because D approach is a special case of RW.

5. CONCLUSION

We propose some simple approaches which can be used to score the relatedness of a conference with respect to a specified pivot. This can be used to score or rank all conferences in a bibliographic database. By using DBLP and the CORE conference ratings, we show that the ranking produced by our methods can reveal the implicit connection between conferences. The rankings are consistent with CORE conference ratings and the classification also accords well with the ranking of conferences into areas by Microsoft Academic Search which uses citation data in addition to publication data.

It is surprising that the simple methods can already work quite well. Perhaps this is due to the special nature of the conference graph which is quite dense. Our approaches are an initial attempt and can be tuned further to take into account the nature of conference publications. In future work, we intend to take features like conference size into account and also investigate the application to investigating the relationship between conferences and journals.

6. REFERENCES

[1] The DBLP Computer Science Bibliography. http://www.informatik.uni-trier.de/~ley/db/

[2] Computing Research and Education. http://core.edu.au/index.php

[3] Microsoft Academic Search. http://academic.research.microsoft.com

[4] A. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. In SIGKDD, 2006.

[5] A. P. Bradley The use of the area under the ROC curve in the evaluation of machine learning algorithms. In Pattern Recognition, 1997.

[6] C. Bird, E. Barr, A. Nash, P. Devanbu, V. Filkov, and Z. Su. Structure and dynamics of research collaboration in computer science. In SDM, 2009.

[7] E. Elmacioglu, and D. Lee. On six degrees of separation in DBLP-DB and more. In SIGMOD, 2005.

[8] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In KDD, 2006.

[9] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: bringing order to the web. In Technical Report 1999-0120, Computer Science Department, Stanford University, 1999.

[10] M. Bianchini, M. Gori, and F. Scarselli. Inside PageRank. In ACM Trans. Internet Technology, 5, 2005.

[11] M. Biryukov, and C. Dong. Analysis of computer science communities based on DBLP. In ECDL, 2010.

[12] M. E. J. Newman. The structure of scientific collaboration networks. In Proc. of the National Academy of Sciences, 98(2):404-409, 2001.

[13] M. E. J. Newman. Scientific collaboration networks: I. Network construction and fundamental results. In Physical Review E, 64, 016131, 2001.

[14] M. E. J. Newman. Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. In Physical Review E, 64, 016132, 2001.

[15] M. Latapy, C. Magnien, and N. D. Vecchio. Basic notions for the analysis of large two-mode networks. In Social Networks, 30(1):31 - 48, 2008.

APPENDIX

Conference Abbreviation & Names

Abbreviation	Conference Name
ACNS	Intl. Conf. on Applied Cryptography and Network Security
AFIPS	American Federation of Information Processing Societies
AH	Adaptive Hypermedia and Adaptive Web-Based Systems
ANTS	Algorithmic Number Theory Symp.
ASIACRYPT	Intl. Conf. on the Theory and Application of Cryptology and Information Security
CANS	Intl. Conf. on Cryptology and Network Security
CARDIS	Smart Card Research and Advanced Application Conf.
CCS	ACM Conf. on Computer and Communications Security
CHES	Workshop on Cryptographic Hardware and Embedded Systems
CIDR	Conf. on Innovative Data Systems Research
CIKM	ACM Intl. Conf. on Information and Knowledge Management
COCO	IEEE Symp. on Computational Complexity
CRYPTO	Advances in Cryptology
CT-RSA	The Cryptographer's Track at RSA Conf.
DASFAA	Database Systems for Advanced Applications
DEXA	Intl. Conf. on Database and Expert Systems Applications
ECHT	European Conf. on Hypertext
ECIR	European Conf. on Information Retrieval
EDBT	Extending Database Technology
ER	Intl. Conf. on Conceptual Modelling
EUROCRYPT	Intl. Conf. on the Theory and Application of Cryptographic Techniques
FC	Financial Cryptography and Data Security Conf.
FOCS	IEEE Symp. on Foundations of Computer Science
FSE	Intl. Workshop on Fast Software Encryption
HICSS	Hawaii Intl. Conf. on System Sciences
Hypertext	ACM Conf. on Hypertext and Hypermedia
ICALP	Intl. Colloquium on Automata Languages and Programming
ICDE	Intl. Conf. on Data Engineering
ICDM	IEEE Intl. Conf. on Data Mining
ICDT	Intl. Conf. on Database Theory
ICICS	Intl. Conf. on Information and Communications Security
ICISC	Intl. Conf. on Information Security and Cryptology
ICITS	Intl. Conf. on Information Theoretic Security
ICWS	International Conf. on Web Services
IMC	Internet Measurement Conf.
INDOCRYPT	Intl. Conf. on Cryptology in India
INET	Internet Society Conf.
ISWC	International Semantic Web Conf.
ISW	Information Security Conf.
KDD	ACM Intl. Conf. on Knowledge Discovery and Data Mining
MMICS	Mathematical Methods in Computer Science
PAIRING	Intl. Conf. on Pairing-based Cryptography
PKC	Intl. Conf./Workshop on Practice and Theory in Public Key Cryptography
PODS	ACM SIGMOD-SIGACT-SIGART Conf. on Principles of Database Systems
SACRYPT	Selected Areas in Cryptography
SCN	Conf. on Security and Cryptography for Networks
SIGIR	ACM Intl. Conf. on Research and Development in Information Retrieval
SIGMOD	ACM Special Interest Group on Management of Data
SSDBM	Intl. Conf. on Scientific and Statistical Data Base Management
STOC	ACM Symp. on Theory of Computing
TCC	Theory of Cryptography Conf.
USITS	USENIX Symp. on Internet Technologies & Systems
VLDB	Intl. Conf. on Very Large Databases
WAIFI	Workshop on the Arithmetic of Finite Fields
WEBDB	Intl. Workshop on the Web and Databases
WEWORC	Western European Workshop on Research in Cryptology
WI	Web Intelligence
WISE	Web Information Systems Engineering
WSDM	ACM Intl. Conf. on Web Search and Data Mining
WWW	Intl. World Wide Web Conf.

List of Databases Conferences from [7]

ADB, ADBIS, ADBT, ADC, ARTDB, Berkeley Workshop, BNCOD, CDB, CIDR, CIKM, CISM, CISMOD, COMAD, COODBSE, CoopIS, DAISD, DANTE, DASFAA, DaWaK, DBPL, DBSEC, DDB, DDW, DEXA, DIWeB, DMDW, DMKD, DNIS, DOLAP, DOOD, DPDS, DS, EDBT, EDS, EFIS/EFDBS, ER, EWDW, FODO, FoIKS, FQAS, Future Databases, GIS, HPTS, IADT, ICDE, ICDM, ICDT, ICOD, IDA, IDC(W), IDEAL, IDEAS, IDS, IGIS, IWDM, IW-MMDBMS, JCDKB, KDD, KR, KRDB, LID, MDA/MDM, MFDBS, MLDM, MSS, NLDB, OODBS, OOIS, PAKDD, PKDD, PODS, RIDE, RIDS, RTDB, SBBD, SDMSIAM, Semantics in Databases, SIGMOD, SSD, SSDBM, SWDB, TDB, TSDM, UIDIS, VDB, VLDB, WebDB, WIDM, WISE, XP, XSym.