# The Layered World of Scientific Conferences

Michael Kuhn and Roger Wattenhofer

Computer Engineering and Networks Laboratory, ETH Zurich, Switzerland
{kuhnmi,wattenhofer}@tik.ee.ethz.ch

**Abstract.** Recent models have introduced the notion of dimensions and hierarchies in social networks. These models motivate the mining of small world graphs under a new perspective. We exemplary base our work on a conference graph, which is constructed from the DBLP publication records. We show that this graph indeed exhibits a layered structure as the models suggest. We then introduce a subtraction approach that allows to segregate layers. Using this technique we separate the conference graph into a thematic and a quality layer. As concrete applications of the discussed methods we present a novel rating method as well as a conference search tool that bases on our graph and its layer separation.

## 1 Introduction

Imagine your research has drifted into a field unfamiliar to you, and you do not know where to publish. In such situations it is helpful to have a better understanding of the world of computer science conferences. Throughout this paper we will explore this world and present an application that bases on this exploration and seeks to assist people that are in situations as described before.

The starting point of our research are recent findings in the context of social networks. These findings emphasize the fact that nodes in natural graphs are interconnected for different reasons, such as common interests, close geographic distances, or family relations in case of friendship networks. Based on the researchers' social network we will introduce a similarity measure for conferences and setup a conference graph. Similar as in friendship networks or the world-wide-web, edges in this graph are caused by different reasons—we will refer to them as the layers of our graph. Such reasons surely are area of research, but maybe also the quality, geographic location, or the community behind the conference. Throughout this paper, we demonstrate that and how it is possible to isolate some of these layers in the case of the conference graph. In particular, we will show that:

- The social network behind conferences provides a good measure to relate conferences to each other.
- This measure consists of a thematic as well as a quality component—the major layers of our graph.
- The thematic layer can be identified by a mere analysis of publication titles.
- The quality layer can be partly isolated by subtracting the thematic component from the overall relationship.

As a result of the layer separation, it becomes possible to explore the conference graph under different points of view. We introduce a novel idea for conference rating based on the quality layer of the graph. Afterwards, we present a collaborative conference search website that demonstrates the advantages of having independent notions of the thematic scope and the quality of a conference. It offers different ways to search for conferences and can be fine-tuned to match the quality and deadline restrictions of an author and thus greatly assists researchers finding themselves in situations as described in the very beginning of this paper.

## 2 Related Work

This section briefly reviews relevant literature in the context of our work, namely the mining of bibliometric data and social networks.

The analysis of publication records has been an active field of research for a long time. Clearly, one of the most attractive goals for publication database mining is automated conference and journal rating. Garfield's pioneering work in 1972 [1], which describes the use of citation analysis for this purpose, initiated a long—and still ongoing—controversy. On one hand, many authors point out the wide variety of problems of the citation indexing approach [2–4]. On the other hand, citation analysis is presumably still the best method to automatically rate scientific conferences and journals. Other measures that are used to indicate a venue's relevance are the acceptance rate as well as time delays, such as turnaround time, end-to-end time, or reference age [5]. It seems that the community behind a conference has so far not been taken into account for automated rating. We believe that this criterion should not be neglected and provide an idea to fill this gap.

The rating of venues is not the only motivation for research on bibliometric data. Other insights have been gained from publication databases. One closely related aspect is the characterization of authors (rather than venues). Various measures, such as closeness [6–10], betweenness [6, 8, 9], or AuthorRank [9] have been evaluated in this context. Also, many studies analyze the evolution of different properties [6, 7, 10–12]. For us, the publications of Lee et al. [12] and Smeaton et al. [10] are of particular interest, as they study the topical changes within a single conference over the years. Thereby they show that the analysis of publication titles, keywords, and abstracts is sufficient to extract the thematic scope of a venue—a fact that we will take advantage of.

Another perspective to looking at the thematic scope of venues is presented in [6]. By considering a common author of two venues an indicator for thematic similarity, a weighted graph is constructed that interrelates the most important conferences in the field of database research. We improve on this measure by incorporating some means of normalization and show that the thematic proximity is only one aspect contained in this weight.

Newer studies on social networks emphasize that many of these graphs exhibit some sort of social dimensions [13–16]. They state that there exist different catalysts for friendships, such as geography, family ties, or occupation. An observation that Killworth and Bernard [17] already made in 1978, when examining the different reasons by which a starter in a Milgram-like experiment would choose the next hop. Their findings show that most decisions are based on the geographic location and the occupation of the target. This result agrees with the findings of Dodds et al. [13] in a recent Internet-based small-world experiment. Based on this evidence, Watts et al. [16] developed a graph model based on different social dimensions. We will show that similar dimensions can also be found in our conference graph and refer to them as layers.

## 3 The Conference Graph

This section describes how the publication records of DBLP[1] can be used to generate a graph that interconnects scientific conferences. The graph construction bases on the social network behind these conferences. We basically assume, that the more common authors two conferences have, the more related they are. To avoid overestimating the similarity of massive events—they naturally have a large number of common authors—we improve on this idea by incorporating a normalization method: Consider two conferences, $C_1$ and $C_2$, that contain a total of $s_1$ and $s_2$ publications, respectively. Further, assume that there are $k$ authors $A_i$ $(i = 1, ..., k)$ that have published in both places and that author $A_i$ has $p_{i,1}$ publications in conference $C_1$ and $p_{i,2}$ publications in conference $C_2$. We can now define the similarity $S(C_1, C_2)$ between $C_1$ and $C_2$ as follows:

$$S(C_1, C_2) = \sum_{i=1}^{k} \min\left(\frac{p_{i,1}}{s_1}, \frac{p_{i,2}}{s_2}\right)$$

Applying this similarity measure to all pairs of conferences results in the desired graph. The required information for this graph was extracted from the DBLP bibliographic repository. Any publications that appeared in a scientific conference between 1996 and 2006 have been taken into account. To reduce the amount of data, edges of extremely low weight that do not significantly contribute to the connectivity have been removed. To give a more concrete idea of the structure of this graph, Table 1 lists the 10 top edges for some sample conferences.

## 4 The Layers

This section introduces the idea of *layers* as the building blocks of our graph. These layers reflect, as we will see, different catalysts for edges. We will have a closer look at two of these layers, namely the thematic and the quality layer, throughout this section.

[1] `http://dblp.uni-trier.de`

| KDD | | AAAI | | ECAI | |
|---|---|---|---|---|---|
| ICDM | 0.69 | IJCAI | 0.76 | IJCAI | 0.53 |
| SDM | 0.58 | ATAL | 0.37 | KR | 0.29 |
| PKDD | 0.45 | ICML | 0.33 | ATAL | 0.27 |
| PAKDD | 0.40 | AGENTS | 0.32 | AAAI | 0.26 |
| ICML | 0.37 | AIPS | 0.31 | AI*IA | 0.24 |
| DMKD | 0.37 | ECAI | 0.26 | JELIA | 0.22 |
| CIKM | 0.36 | KR | 0.25 | ECSQARU | 0.21 |
| SIGMOD | 0.36 | UAI | 0.25 | CP | 0.19 |
| ICDE | 0.35 | CP | 0.23 | IEA/AIE | 0.19 |
| VLDB | 0.33 | FLAIRS | 0.20 | KI | 0.19 |

**Table 1.** The 10 strongest links to the conferences KDD, AAAI and ECAI

Proximity in the conference graph is not purely defined by the thematic similarity of venues as a careful look at Table 1 reveals. *ECAI* is, for example, typically said to be thematically closer to *AAAI* than *ATAL*, *ICML*, or *AGENTS*, which appear earlier in the *AAAI* top-10 list. We conclude that authors choose conferences not only because of the topic it covers. Other properties, such as quality, geographic location, or the community behind a venue also influence the author's decision. In fact, we believe that it is a weighted combination of all these factors that leads to a submission at a certain place. Exactly this combination is reflected by the conference graph presented in the previous section. The graph consists of different *layers*, where each layer represents one of these factors. This idea is illustrated in Figure 1.

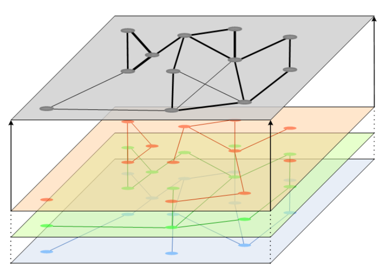

**Fig. 1.** The total graph can be seen as the sum of its layers.

### 4.1 The Thematic Layer

Clearly, the thematic scope of a venue has a significant impact on its relationship to other venues. In the following, we present a technique that bases on publication title analysis and measures the thematic similarity of conferences. It thus allows

| KDD | | AAAI | | ECAI | |
|---|---|---|---|---|---|
| mining | 0.051 | learning | 0.013 | reasoning | 0.011 |
| data | 0.013 | planning | 0.012 | learning | 0.010 |
| discovery | 0.013 | robot | 0.010 | qualitative | 0.009 |
| clustering | 0.013 | reasoning | 0.008 | planning | 0.008 |
| association | 0.010 | knowledge | 0.007 | knowledge | 0.008 |
| sigkdd | 0.010 | search | 0.007 | logics | 0.008 |
| kdd | 0.009 | agent | 0.006 | logic | 0.008 |
| frequent | 0.009 | constraint | 0.006 | ecai | 0.008 |
| rules | 0.009 | ai | 0.006 | constraint | 0.007 |
| discovering | 0.009 | reinforcement | 0.006 | diagnosis | 0.007 |

**Table 2.** The 10 best matching keywords to KDD, AAAI and ECAI together with their TF-IDF score.

to define the *thematic layer*, which is surely an ingredient of the social similarity measure, as a majority of authors mostly work in only one area and therefore submit papers to thematically similar venues.

For each conference, we have extracted all the titles from DBLP and applied the well-known *term frequency - inverse document frequency (TF-IDF)* method (see [18] for some theoretic background) to identify the most relevant keywords. The TF-IDF score for a document increases proportionally to the number of occurrences of the keyword in the document (TF). However, words that have a high overall frequency are penalized (IDF). In our context, a document corresponds to a venue and the words stem from publication titles. Consequently, a document consists of all titles in a venue and the complete corpus consists of all venues in DBLP.

Once a score has been applied to all the keywords that appear in the conference's collection of titles, the scope of the conference can easily be estimated by looking at the most relevant terms. Table 2 shows some examples.

Using the keyword-lists seen before, we have implemented a simple algorithm that estimates the thematic relationship between two venues. It takes the top-50 keywords of each conference, and counts the number of keywords appearing in both lists, resulting in a score from 0 to 50 for each pair of conferences.[2]

Applying the thematic similarity function to each pair of venues results in a weighted undirected graph—the *thematic layer* of our graph. The corresponding neighborhood lists for our sample conferences are shown in Table 3.

### 4.2 The Quality Layer: Filtering by Subtraction

Section 2 briefly discussed the problem of conference rating and its difficulties. For computer sciences, the Citeseer Impact List tries to estimate the impact

[2] Surprisingly, this simple comparison function achieved slightly better results than the more commonly used cosine similarity approach.

| KDD | | AAAI | | ECAI | |
|---|---|---|---|---|---|
| ICDM | 26 | IJCAI | 37 | IJCAI | 29 |
| PKDD | 23 | ECAI | 27 | AAAI | 27 |
| PAKDD | 21 | FLAIRS | 22 | ICTAI | 22 |
| SDM | 20 | ICTAI | 21 | KI | 21 |
| Dis. Science | 20 | AIPS | 17 | FLAIRS | 20 |
| DMKD | 18 | Can-AI | 16 | Can-AI | 19 |
| ADMA | 17 | IEA/AIE | 16 | IEA/AIE | 18 |
| ISMIS | 17 | PRICAI | 15 | PRICAI | 18 |
| IDA | 15 | Aus-AI | 15 | KR | 16 |
| IDEAL | 15 | KI | 14 | Aus-AI | 16 |

**Table 3.** The 10 closest neighbors to KDD, AAAI and ECAI in the thematic layer, together with their thematic score.

of venues based on citation analysis. Further, many researchers maintain hand-made lists that distinguish between tier-1, tier-2, and tier-3 conferences. Even though hand-made lists suffer from a subjective bias and citation analysis from other weaknesses (recall Section 2), tier-1 conferences typically have a high impact and, contrariwise, tier-3 conferences get low scores in the Citeseer list. We will refer to similarly classified conferences as conferences of similar quality.

Comparing the neighborhood tables for the total graph (Table 1) and the thematic layer (Table 3) shows, that the total graph is *not* purely defined by the thematic correlation of conferences. Looking at the total graph, an interesting observation is that conferences often considered to be of high quality (such as *KDD* and *AAAI*) tend to have other high quality conferences in their proximity. In contrast, the number of lower-tier conferences in the proximity of *ECAI*, which is mostly classified as tier-2, is significantly higher. This observation is illustrated in Table 4 that uses the impact value of the Citeseer Impact List[3] to classify the conferences.

We conclude that a single author tends to publish not only in venues of similar topic, but also in venues of similar quality, meaning that our graph contains a second major layer—the *quality layer*.

The observation that thematically weaker related nodes in a conference's proximity tend to be closer in quality suggests that the quality layer can be extracted using the information about the total graph and the thematic layer. In the following we will introduce a *layer subtraction approach* to demonstrate that such a layer separation can indeed be achieved. The approach bases on the assumption that the total graph is a linear combination of the single layers. As a result of the observations in the previous section we assume that the major layers of the conference graph are the thematic layer $t$ and the quality layer $q$. This also matches our experience when selecting a conference: We make sure the publication matches the call for papers and we try to submit at a conference of reasonable quality. Other factors, such as geographic location, play a minor

[3] `http://citeseer.ist.psu.edu/impact.html`

| AAAI Total | | AAAI Thematic | | ECAI Total | | ECAI Thematic | |
|---|---|---|---|---|---|---|---|
| IJCAI | 1.10 | IJCAI | 1.10 | IJCAI | 1.10 | IJCAI | 1.10 |
| *ATAL* | *1.51* | ECAI | 0.69 | KR | 1.76 | AAAI | 1.49 |
| *ICML* | *2.12* | FLAIRS | N/A | *ATAL* | *1.51* | ICTAI | 0.25 |
| *AGENTS* | *1.00* | ICTAI | 0.25 | AAAI | 1.49 | KI | 0.41 |
| AIPS | 1.53 | AIPS | 1.53 | *AI*IA* | *0.26* | FLAIRS | N/A |
| ECAI | 0.69 | Can-AI | 0.26 | *JELIA* | *0.72* | Can-AI | 0.26 |
| *KR* | *1.76* | IEA/AIE | 0.09 | *ECSQUARU* | *0.38* | IEA/AIE | 0.09 |
| *UAI* | *N/A* | PRICAI | 0.19 | *CP* | *1.04* | PRICAI | 0.19 |
| *CP* | *1.04* | Aus-AI | 0.16 | IEA/AIE | 0.09 | KR | 1.76 |
| FLAIRS | N/A | KI | 0.41 | KI | 0.41 | Aus-AI | 0.16 |

**Table 4.** The 10 closest neighbors to AAAI (left) and ECAI (right) in the total graph and the thematic layer, together with the Citeseer impact value. Note that for AAAI, conferences in the total graph neighborhood that are not present in the thematic layer list (*italic*) all have relatively high impact value. The impact value of such conferences in the neighborhood of ECAI is considerably lower.

role in the decision. These factors (including noise) are thus subsumed into a remainder layer $r$. Consequently the total edge weight $S$ becomes to $S = \alpha_1 \cdot t + \alpha_2 \cdot q + \alpha_3 \cdot r$, for some weights $\alpha_i$, with $\alpha_1, \alpha_2 \gg \alpha_3$. Neglecting $\alpha_3$ and setting $\alpha_2 = 1$ ($\alpha_2$ can be chosen arbitrarily as it only results in a scaling of $q$) allows to extract the quality layer $q$ as

$$q \approx S - \alpha_1 \cdot t,$$

Note that the validity of the linear combination assumption greatly depends on the characteristics of the weight functions in the different layers. In [19] Fernandez et al. presented the idea of score distribution normalization for aggregation purposes. They suggest to shape the histograms of the independent score functions to match the "ideal" distribution prior to merging them by linear combination. For simplicity we assume a uniform weight distribution for both, the total as well as the thematic scores.

Observe that the subtraction approach generally allows to extract one out of $L$ layers of a graph, if the remaining $L-1$ layers are known. It seems that such a layered structure can often be observed—recall Section 2 and also think of recommendation systems that often build on similar co-occurrence structures as our graph. We thus believe that the layer-subtraction approach might be a valuable preprocessing step in various data-mining settings.

The next sections discuss how the quality of the filtering can be estimated by producing a conference rating and thereby show some evidence of the correctness of the proposed subtraction approach.

### 4.3 Interpolation Based Conference Rating

The proximity of a conference in the quality layer is supposed to contain mostly conferences of similar quality. This observation immediately leads to the idea of conference rating by interpolation: Provided some initial ratings are known, the tier of a conference can be estimated by looking at its proximity in the quality layer. Initial ratings can be retrieved from manually created lists (we use the one found at `www.ntu.edu.sg/home/assourav/crank.htm` and refer to it as *CS Rating List*) as well as from Citeseer's impact list. We have further introduced the *Citeseer Tier List*, which assigns a tier (1, 2, or 3) to each conference in the Citeseer Impact List. The borders between tiers have been chosen such that the number of incorrectly rated conferences with respect to the CS Rating List becomes minimal. The best that can be achieved is an error rate of 38.8%, which indicates how difficult the task of conference rating is.

We have then defined a heuristic to rate a conference $C_0$ as follows:

1. For all conferences in the CS Rating List or the Citeseer Tier List, set the initial rating to the value found in the lists. In case of conflicts, the CS Rating List is treated with priority. For any conference not in the lists, set the initial rating to *unrated*.
2. Overwrite the initial rating of $C_0$ with *unrated*. This step avoids that the rating function is biased towards the initial value.
3. Take the 30 shortest edges $e_i$ adjacent to $C_0$ in the total graph, together with their values $S_i$ and $t_i$. For all these edges, calculate $q_i = S_i - \alpha_1 \cdot t_i$ (for some value of $\alpha_1$) and sort by $q_i$. We will call the resulting list the *filtered neighborhood list* of $C_0$: $N_f(C_0)$.
4. For the first 5 entries $C_j$ ($j = 1..5$) in $N_f(C_0)$, calculate $N_f(C_j)$.
5. Return the median of all the rated conferences found within the first 5 entries in all the lists $N_f(C_j)$ ($j = 0..5$) as the rating of $C_0$.

Note that this conference rating method is in some sense natural. Many people would judge a venue based on people participating in it (or leading it). This information is implicitly contained in the total graph which forms the basis of the rating heuristic.

The quality of the heuristic can be estimated by comparing the calculated ratings to those found in the CS Rating List (which is presumably the most accurate list we dispose of). The optimal value of $\alpha_1$ was scanned for by exhaustive search over some reasonable interval.[4] This is illustrated in Figure 2 which plots the error rate of the rating function with respect to the CS Rating List for different values of $\alpha_1$. The figure clearly shows that the subtraction approach reduces the number of incorrect ratings and suggests that the optimal value of $\alpha_1$ is somewhere between 0.5 and 1.

Arguing with error rates beyond 40% might at the first glance seem suspicious. However, the fact that approximately 75% of the input values (namely those that originate from the Citeseer Tier List) exhibit an error rate of approximately 40% themselves relativizes the high error rate produced by our algorithm.

[4] Note that optimizing for $\alpha_1$ using regression by comparing to a "quality relationship" between two conferences is likely to fail, as this quality relationship is very vague (i.e. can take only the values 0, 1, and 2).

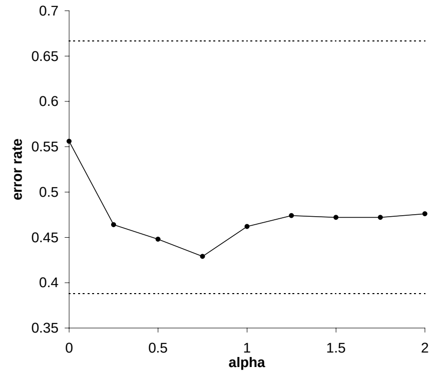


**Fig. 2.** The fraction of incorrectly rated conferences using our rating function versus the value of $\alpha_1$. The dotted lines indicate the error rates for a random guess (0.667) and for the Citeseer Tier List (0.388), which is in about the best we can expect to reach as most of the initial rating values stem from this list.

Ignoring all the conferences rated as tier-2 either by the algorithm or the CS Rating List shows that the errors are not random. Dividing the number of conferences rated as tier-1 instead of tier-3 (and vice versa) by the number of conferences the algorithm rates as tier-1 or tier-3 results in an error rate of around 6.4% without thematic filtering and of 2.6% for the optimal value of $\alpha_1$. (Note that dividing by the number of tier-1 and tier-3 conferences in the list would overestimate an algorithm that tends to rate conferences as tier-2.)

These low values show three things:

1. The total edge weight is clearly influenced by the quality of conferences. This supports the assumption that the thematic and the quality layer are the two main layers of the graph.
2. The success of extracting the quality layer by subtraction of the thematic layer as shown in Figure 2 is confirmed.
3. Most of the around 43% of errors are minor errors. That is, they are wrong by only one tier. Severe errors are rare,they make up less than 3%.

**Remark:** The rating heuristic was developed for two reasons: To provide a complete rating list for the conference search application presented next, and to demonstrate the effect of subtraction filtering. It is thought as a proof-of-concept algorithm that neither has a strong mathematical foundation nor provides any guarantees on the results.

## 5 ConfSearch

In this section we will show that the previously discussed conference graph and its separation into different layers can directly be applied for confer-

| $\beta_q = 0.0$ | $\beta_q = 0.5$ | $\beta_q = 1.0$ |
|---|---|---|
| PKDD | KDD | KDD |
| KDD | ICDE | ICDE |
| INFOVIS | PKDD | ICDM |
| ICDM | ICDM | VLDB |
| Web Intelligence | Web Intelligence | Web Intelligence |
| PAKDD | INFOVIS | PKDD |
| ICDE | VLDB | DMKD |
| ICDM | DMKD | SDM |
| JSAI Workshops | SDM | INFOVIS |
| DaWaK | PAKDD | DASFAA |

**Table 5.** The results for the search query "social graphs data mining" for different quality weights (controlled by the parameter $\beta_q$).

ence search. For this purpose we have developed a website that is able to suggest conferences together with their most important attributes (try it at `http://www.confsearch.org`). The application offers four different search types:

- *Keyword Search*: Search by keywords provided.
- *Related Conference Search*: Explore the proximity of a given conference in the conference graph and return the closest neighbors.
- *Author Search*: Search for the places a given author publishes most often.
- *General Search*: A weighted combination of the above search methods.

For all search types the application allows to sort the results by deadline, a criteria that has a considerable impact when deciding for one or the other venue. Motivated by the success of Wikipedia-like services, we follow a collaborative approach to gather conference deadlines as well as locations and website URLs. Our application can be seen as an improvement on the many lists with conference deadlines found in the Internet today: We basically cover the whole area of computer science and augment the typically static lists with sophisticated search options.

The *keyword search* bases on a score $s_{ij}$ for each keyword-conference pair (where only keywords appearing in the query are considered), which is a slightly modified variant of the TF-IDF value presented in Section 4.1. Next, the scores $s_{ij}$ of the conference-keyword pairs are combined to a single value $S_i^*$ per conference $C_i$ using the *p-norm* method introduced by Salton et al. [20]. The final score $S_i$ results from the quality adjustment of $S_i^*$ controlled by a user-settable parameter $\beta_q$: $S_i = S_i^* \cdot f(Q)^{\beta_q}$. The function f(Q) is defined on a per query basis to smoothly adapt to the different score distributions for different queries. The quality part $Q$ is estimated using the heuristic presented in Section 4.3. Table 5 presents a keyword search example and the effect of quality filtering.

The *related conference search* operates directly on the conference graph. We simply return the closest nodes around a conference in terms of path-length.

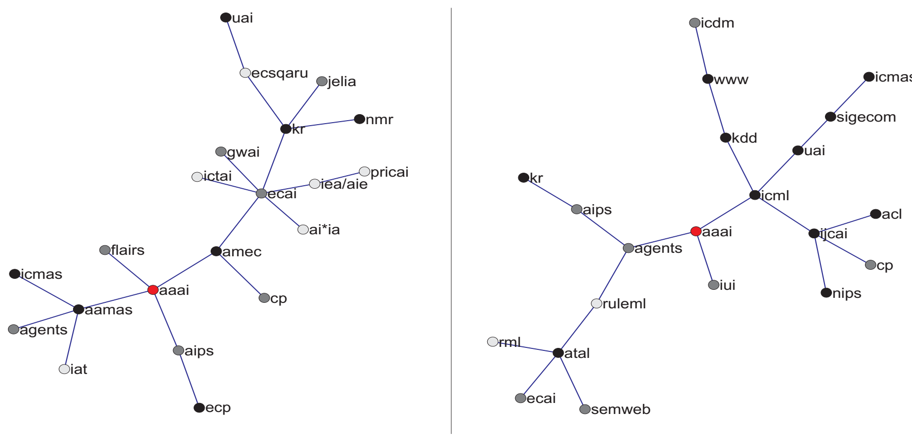


**Fig. 3.** The minimum spanning tree around *AAAI* in the thematic layer (left), and the quality layer (right). Darker nodes refer to higher tier venues.

Again, a user settable parameter allows to control whether the thematic or the quality aspect should be emphasized. A visualization of the *AAAI* neighborhood in the thematic and the quality layer can be found in Figure 3. The increased amount of high quality nodes (dark) in the *AAAI*'s "qualitative proximity" indicates that *AAAI* itself is also likely to be of high quality. The search option on one hand allows to browse the conference graph and on the other hand might prove extremely helpful if you look for alternative places to submit, after a reject, for example, or because a deadline does not fit.

## 6 Conclusion

Throughout this paper we have provided evidence for recent small-world models using the real-world data of a scientific conference graph. We have shown that this graph indeed consists of layers and demonstrated that these layers can be effectively combined. The combination assumption has led to the subtraction approach for layer segregation which provides an attractive preprocessing step when mining graphs. In our setting it was used to accentuate the different aspects of conference relations.

We have seen that the conference graph consists of two major layers—the thematic layer and the quality layer. We have then presented a novel rating method for scientific conferences that operates on the quality layer of the graph.

The separation of the two layers further builds the basis of the conference search application presented in the last section. Exploring the thematic layer allows to retrieve venues matching a user query. The sorting of the retrieved venues can then be adjusted using the information gained from the quality layer and thereby effectively fitted to the user's needs.

## References


1. Garfield, E.: Citation Analysis as a Tool in Journal Evaluation. Science (178) (1972) 471–479
2. Linde, A.: On the pitfalls of journal ranking by impact factor. Eur J Oral Sci **106**(1) (1998) 525–526
3. MacRoberts, M.H., MacRoberts, B.R.: Problems of citation analysis: A critical review. JASIS **40**(5) (1989) 342–349
4. Scharnhorst, A., Thelwall, M.: Citation and hyperlink networks. Current Science **89**(9) (2005)
5. Snodgrass, R.: Journal relevance. SIGMOD Rec. **32**(3) (2003) 11–15
6. Elmacioglu, E., Lee, D.: On Six Degrees of Separation in DBLP-DB and More. SIGMOD Rec. **34**(2) (2005) 33–40
7. Nascimento, M.A., Sander, J., Pound, J.: Analysis of SIGMOD's Co-Authorship Graph. SIGMOD Rec. **32**(3) (2003) 8–10
8. Newman, M.E.J.: Who is the best connected scientist? A study of scientific coauthorship networks. Physical Review E **64**(1) (2001) 016132–1–016132–7 Scientific collaboration networks. Part II. Shortest paths, weighted networks, and centrality.
9. Liu, X., Bollen, J., Nelson, M.L., de Sompel, H.V.: Co-authorship networks in the digital library research community. Inf. Process. Manage. **41**(6) (2005) 1462–1480
10. Smeaton, A., Sodring, T., McDonald, K., Keogh, G., Gurrin, C.: Analysis of Papers from Twenty-Five Years of SIGIR Conferences: What Have We Been Doing for the Last Quarter of a Century? SIGIR Forum **36**(2) (2002)
11. Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. Physica A **311**(4) (2002) 590–614
12. Lee, B., Czerwinski, M., Robertson, G., Bederson, B.B.: Understanding Research Trends in Conferences using PaperLens. In: CHI '05: CHI '05 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM Press (2005) 1969–1972
13. Dodds, P., Muhamad, R., Watts, D.: An experimental study of search in global social networks. Science **301**(5634) (2003) 827–829
14. Goussevskaia, O., Kuhn, M., Wattenhofer, R.: Layers and Hierarchies in Real Virtual Networks. In: IEEE/WIC/ACM International Conference on Web Intelligence (WI), Silicon Valley, California, USA. (2007)
15. Kleinberg, J.: Small-World Phenomena and the Dynamics of Information. In Dietterich, T.G., Becker, S., Ghahramani, Z., eds.: Advances in Neural Information Processing Systems 14, Cambridge, MA, MIT Press (2002)
16. Watts, D.J., Dodds, P.S., Newman, M.E.J.: Identity and Search in Social Networks. Science **296** (2002) 1302–1305
17. Killworth, P.D., Bernard, H.R.: The reversal small-world experiment. Social Networks **1**(2) (1978) 159–192
18. Robertson, S.: Understanding Inverse Document Frequency: On theoretical arguments for IDF. Journal of Documentation (5) (2004) 503–520
19. Fernández, M., Vallet, D., Castells, P.: Probabilistic Score Normalization for Rank Aggregation. In: ECIR. (2006) 553–556
20. Salton, G., Fox, E.A., Wu, H.: Extended Boolean information retrieval. Commun. ACM **26**(11) (1983) 1022–1036