

# Author's Paper Similarity Prediction based on the Similarity of Textual References to Visual Features

Mostafa Alli

Dept. of Computer Science and Technology, Tsinghua University, Beijing 100084, China

**Keywords:** Paper Prediction, Common Coauthors, Textual Reference, Visual Feature, Table, Figure, Boolean Function, Sensitivity, Scientific Stop Word.

**Abstract:** In this paper we introduce a mechanism to find similar papers of an author, based on the author's previous<sup>1</sup> publications. In other words, since the author(s) of a paper are more likely to publish similar work(s) to their paper, we use this intuition to seek related papers based on the visual similarity of those papers. The visuality here is the figures and or tables that are commonly used by authors to describe their method structure and/or the result of their experiments. Since similar works of authors are focused on solving similar problems as well as developing and improving similar techniques, we noticed that comparing these visual features among their publications would help to spot most similar papers of those authors. We call our method, *Similarity of Textual References to Visual Features* which means, we compare parts of content of any two arbitrary papers that have references to any figures and/or tables. In our experiment we show that how we can use this similarity together with other factors of a paper to form a *Boolean function* which helps to build an indexation for papers based on the number of their authors. In this way, we omit time consuming process of papers' content determined analysis, such as, textual content analysis, building coauthor network, citation network etc. In addition, our Boolean function has the ability of adjusting level of *Sensitivity*. If we want to achieve higher accuracy of similar papers, the Boolean function needs to be enabled for more<sup>2</sup> conditions.

## 1 INTRODUCTION

Ultimate recommending similar papers will be a lot of help and will reduce the effort, time and the chance of missing related publications for researchers. As a solution, there are numerous application to suggest related papers based on a user profile (Hong et al., 2013a)(Hong et al., 2013b), citation data (Bogers and van den Bosch, 2008)(Ma et al., 2008), a combination of page rank algorithm (Brin and Page, 1998) and citation network (Nykl et al., 2014) etc. Nonetheless, these techniques may not work well since this is shown (Vellino, 2009)(Vellino, 2010) that using page rank algorithm values will not improve the similarity measurement as well as a user profile technique needs a huge set of data and documents to work with, and a content-based filtering for paper suggestion is time consuming since it goes through a whole text (He et al., 2010). In addition, this can be argued that

a citation oriented system deliberately ignores two facts ,i.e, a recent and similar paper has few-or even no- citation score, while, a paper with broad focus but less similarity, such as survey papers, has gained much more citations. Study (Pohl et al., 2007) shows that a paper has very low citation within its first 2 years of its publication. That means, a paper, at least, needs 2 years time to be started to be seen by researchers. However, this is not the only drawback of citation-based scheme. Such systems suffer from *Matthew Effects* (Stanovich, 1986) which is, rich gets richer and poor gets poorer. In other words, the paper with more citation gets more attention and then gets more cites and one with less cites, will be ignored for some more time. There is another shortcoming of a citation-based and similar systems which is the coverage. According to (Good et al., 1999), the coverage of a recommender system is a critical factor for its accuracy. Since a citation based system will ignore a significant number of items(papers) (He et al., 2010), the accuracy would be decreased significantly too.

Although in a recent study (Sayyadi and Getoor,

<sup>1</sup>Or inversely, future published work, depending to the paper selection.

<sup>2</sup>Or even all

2009) authors tried to adjust some of the shortcomings of citation-based systems such as *Matthew Effects* which is also called *Slow start* for citation count, however, their assumptions for doing such are still representing the Matthew effect issue, e.g., *Good research papers are written by researchers with high-reputations, Important articles are cited by many important articles.* etc.

In our previous work (Alli, 2015) we introduced a paper similarity technique which takes into account the summarization version of each paper. The work was motivated for the fact that a full-text comparison is aggressive and slow (He et al., 2010). Moreover a citation-based system requires complex NLP techniques and needs a long time to obtain enough citation scores for meaningful similarity analysis (Pohl et al., 2007).

However, we still think that using either of existing techniques<sup>3</sup> will not guarantee to take into account the previous publications of the author(s) of the input paper. Despite the fact that there are efforts made to consider this issue, but we believe they are not so accurate.

A context-aware recommender system (He et al., 2010) is introduced which considers a set of factors to build up a group of paper candidates for future analysis. One of the factors that they consider is previous publications of an author. However, there is no selection policy and all of an author's papers will be added to the candidate set. Consequently, the previous publications of other authors of candidate set will be added and so on. This can be clearly seen that this *Blind* technique will add loads of unnecessary papers and even can lead to infinite loop of paper addition.

Similarly, a key-phrased based system (Sugiyama and Kan, 2010) is proposed to recommend papers based on the *users' interests* by considering his or her previous publication records. To do so, the system builds up a user profile of researchers based on their publication record as either senior or junior researchers. However, this needs to build a profile for each user which takes extra time and needs to gather additional data. In addition, such system has even a much bigger problem. There is no necessity for someone to narrow down his/her researching field to one or few topics. It will be so common that someone changes his/her working direction after an accomplishment or achievement. Authors has mentioned that they would put more weight(near to 1) for recent papers and less weight(near to 0) for older papers to fix this problem. Nonetheless, this is still possible that an author is already finished his/her

<sup>3</sup>including our previous work

work with some recent publication and has started a new topic which includes no publications record yet or a work can be followed after a period of time.

As a compromise, we are motivated to specifically build a system that *intellectually* finds similar papers of an author regarded to his/her previous/fututre<sup>4</sup> We believe that the writer of a research paper is more likely to publish one or more related or similar paper based on his/her own paper's topic. To find out that which publications must be regarded as similar ones, we evaluated our method in section 4 where in first experiment we interrogated the feasibility of the textual references to visual features and then, using this methodology, in second experiment, we investigated 3 different factors in an *Author Previous/Future Publication Similarity Prediction System*, i.e, The placement of author in his/her paper, the year of publication and the number of common coauthors. Accordingly, in third experiment, we developed a *Boolean function* which will be used for indexing papers of an author based on the significant value of those factors, if there is any.

The rest of the paper is organized as follow. In section 2, we give a review of related work, in section 3, we briefly explain our proposed approach, and finally in section 5, we gave a conclusion to this work.

## 2 RELATED WORK

A coauthor similarity system (Han et al., 2013)(Sun et al., 2011) is where the system tries to find the similarity between a user's interests and an author based on their research concerns. Such system may or may not try to recommend papers based on this similarity, however, in case of recommendation, the system will retrieve many irrelevant papers especially when predicted authors are working on a wide range of topics.

J.Tong et al (Tang et al., 2007)(Tang et al., 2008) proposed to mine social network of researchers. The aim of this work is to collect researchers information from web, based on the researcher name and then, mining the semantic part of the retrieved information in order to extract useful information. This leads to build a profile for each researcher which contains their information, such as publications, affiliations, emails etc. In addition, this system provides name disambiguation when there are several researchers with same name or same abbreviate name. For name disambiguation task, this system which is called *ArnetMiner*, applies a constraint policy which maps 6

<sup>4</sup>Depending on the selected paper's date of publication.

constraints, i.e., *CoOrganization*, *CoAuthor*, *Citation*, *CoEmail*, *FeedBack from users* and  $\tau$ -*CoAuthor*, into one single authors publications to see if the information belongs to same person or they are actually different persons. However, this system lacks the actual paper recommendation, especially based on the author's publications.

Q.He et al (He et al., 2010) proposed a context-aware system to recommend citation for a citation placeholder. The method will retrieve a large body of papers based on some factors such as papers with most title and abstract similarity, papers that have similar authors with original paper etc. The time is one of the most drawbacks of this proposed system due to the fact that it needs 50-100 seconds time for every new candidate. Moreover, they do not consider any constraint for authors' paper when they add them to the candidate set. Consequently, there will be a large number of unnecessary papers that are added to the candidate set.

H.Chen (Chen et al., 2011) introduced a recommender system which uses coauthoring network to suggest authors with similar interests. It receives a query from user and by using CiteSeerX database, it builds a coauthor network to make collaboration recommendation. Nevertheless, this system will not suggest any similar paper to the users.

### 3 PROPOSED FRAMEWORK

In contrast to the previous researches in paper similarity calculation in order to find similar papers based on ones publication records, we believe that since the author(s) of a paper are more likely to publish similar work to their previous works, we use this intuition to seek related papers based on the visual similarity. The concept of visuality here is any visual elements, e.g., *Table and Figures*, that are commonly used by authors to describe their proposed framework and/or the result of their experiments. Due to the fact that similar works of coauthors are focused on solving similar problems as well as developing and improving similar techniques, we realized that comparing these visual features among their published researches will be a great help for finding similar works of those authors. Nonetheless, the task of image processing for detecting similar figures and tables is a time consuming task and requires a high resource capacity. In addition, some figures, such as flowcharts that consist similar geometrical objects, are visually similar but are actually different in nature. Consequently, we decided to use a smarter way to tackle this obstacle. We called this method, *Similarity of Textual*

*References to Visual Factors*, that is, we compare the part of content of any two arbitrary papers that has a reference to any figure and/or table. This similarity can be both visually similar figures/tables with similar textual references to those figures and tables (Buyukkokten et al., 2001a; Buyukkokten et al., 2001b) or just similar textual references to those visual features (Tang et al., 2009; Wang et al., 2010).

To spot such similar textual references, we use the *String tokenizer* to divide the textual content of a research paper into *tokens*. Tokens here are a set of characters that are surrounded between two dots.

As we mentioned previously, we will compare the similarity of those parts of any two papers that are referencing to a visual feature. For detecting these *special* references, we used 3 *Regular expressions Regex* that can demonstrate the common way that authors refer to any table and figure. This regexes are illustrated in Listing 1:

Listing 1: Regular expression used for detection of Textual references to visual features.

```
String pattern="\\>(*[fF]igure\\s*\\d*[:.]*.*";
String pattern2="\\>(*[tT]able\\s*\\d*[:.]*.*";
String pattern3="\\>(*[fF]ig\\s*\\d*[:.]*.*";
```

Accordingly, to prevent misinterpreting, we managed to delete English stop words. We introduce a new set of stop words called, *Scientific stop words* which refers to those words that are commonly used in research papers, e.g., *Model*, *Proposed*, *Chart*, etc.. In evaluation section, in our first experiment, we show that this similarity comparison results significantly different values for comparisons between similar and dissimilar papers of common coauthors. To make our method even more useful, we investigated the impact of three factors, i.e., *Papers' Publication Year*, *Author's Position* and *The Number Of Common Coauthors*, in the result of similarity measurement in our second experiment. To make the coauthors factor more general and accurate, we introduce a coauthor ratio and use it instead of the absolute number of common coauthor in our third experiment in order to create a *Boolean function* which can be used to index one's publication. The ratio formula is illustrated in Equation 1:

$$CoauthorRatio = Ratio_{CandidatePaper} / Ratio_{PivotPaper} \quad (1)$$

Where the *Pivot paper* indicates the input paper and the numerator is equal to:

$$N_{CommonCoauthors} - 1 / N_{TotalAuthorsInCandidatePaper} - 1 \quad (2)$$

And denominator is equal to:

$$N_{CommonCoauthors} - 1 / N_{TotalAuthorsInPivotPaper} - 1 \quad (3)$$

The subtraction in both Equation 2 and 3 demonstrates the existence of the first author of the pivot paper that obviously appears in all of his/her publications. For example, if the pivot paper has 4 authors in total and an arbitrary paper of first author of the pivot paper, denotes as  $P_i$ , has total number of authors of 6 which 3 of them are same with the pivot paper, the ratio of candidate paper and pivot papers are equal to  $Ratio_{CandidatePaper} = 3 - 1/6 - 1 = 0.4$  and  $Ratio_{PivotPaper} = 3 - 1/4 - 1 = 0.66$  and hence the final coauthor ratio of  $P_i$  will be equal to 0.60.

## 4 EVALUATION

To evaluate our method, we studied 3 experiments. In first experiment, we provided a manually selected dataset of 45 papers to show that how the similarity of textual references to visual features can help to indicate one's similar publications. In second experiment, we show that how this similarity merit might be dependent to other factors of a paper, i.e., Author's position, Publication year and Coauthor ratio, and finally, in last experiment, we show that how we can generalize this method into a *Boolean function* which its job is to index one's publication.

### 4.1 Experiment 1

As we stated previously, researchers tend to use similar visual features in their similar works. These visual features are significantly similarly mentioned in similar articles of common authors compare to the rest of their publications. To illustrate this, we managed to run an experiment over 45 scientific papers, grouped in 15 categories. Each category consists of 3 papers from similar coauthors, which 2 of them are similar papers and one of them is not. We show that the similar papers in each pair, are significantly using similar references to their visual features compare to the dissimilar one. To show the significant effect of our proposed method, we first extracted the textual references to visual features of each paper. To do such, we used a *String Tokenizer* to divide the whole contextual part of paper into blocks that are surrounded by two dots. Accordingly, we remove all the English language stop words and common words that researchers usually use in their work to mention a table or figure, such as *Model, Proposed, Chart etc.* We call these type of words, *Scientific stop words*. Using the Cosine Similarity,

we obtained the similarity of these references for each 3 groups for each category, once between first group and second group, calling it result A<sup>5</sup> and once between first and third group calling it result B<sup>6</sup>, and once between second and third group calling it result C<sup>7</sup>. By running ANOVA test once between A and C and once between B and C, the results are significant (P-value=0.000 and 0.001 respectively). In other hands, this can be said that the visual similarity of authors' papers based on the textual references are significantly different between similar and dissimilar ones.

### 4.2 Experiment 2

In this experiment, our aim is to use the results of previous experiment, investigating the relations between papers of an author based on the 3 factors, i.e., *Author's position on the paper, Publication's year and Number of common coauthor(s)*. To do such, we selected an arbitrary author with 41 papers publication as well as 7 journal articles. We categorized this data set based on the three factors that we stated previously. By choosing a paper in the year of 2007, we examined our visual similarity comparison. Since the data set is highly unbalanced, we decided to use *General Linear Model (GLM)*<sup>8</sup> to seek for significance in the observed data. The result of GLM tests are illustrated for each category at Table 1, 2 and 3. By looking at *Sig* column of each categories' GLM tests results, we can see that the author's position in his/her other publications will not give any significance impact on similarity based on the textual references to visual features<sup>9</sup>. To make it more precise for the other two significant factors that how much they effect the observed similarities value, we can obtain the *Effect size* from the values of *Sum of Squares (SS)* from Table 1 and 2. There are two ways of measuring the effect size: either of *Eta square* or *Partial eta square*. However, this is suggested (Levine and Hullett, 2002) to use *Eta square*<sup>10</sup> since it will give a more precise value<sup>11</sup>. The formula for computing the *Eta square* is illustrated in Equation 4.

<sup>5</sup>Similarity value between two similar papers

<sup>6</sup>Similarity value between first similar paper and the dissimilar paper

<sup>7</sup>Similarity value between second similar paper and dissimilar paper

<sup>8</sup>Not to confuse with *Generalized Linear Models*

<sup>9</sup>Please note that this factor can be significant for another author or even for a different paper of same author.

<sup>10</sup>Or *Omega square, Epsilon square*

<sup>11</sup>Although it will reduce the value of effect size.

$$\eta^2 = \frac{SS_{treatment}}{SS_{treatment} + SS_{error}} \quad (4)$$

Where *SS* stands for *Sum of Squares*. By using the values of *SS* reported in Table 1 and 2, we can calculate the Eta square values for publication year and common coauthors factors which are 0.444 and 0.392 respectively. Since the eta square is an analogous of r-square, this means that papers based on their publication year and number of common coauthors are similar at the rate of 44.4% and 39.2% respectively.

Table 1: GML tests' results for number of common coauthors category.

Source	Sum of Squares(III)	F	Sig
Corrected Model	1692.075	11.943	0.000
Number of Common Coauthors	1692.075	11.943	0.000
Error	2621.097		
Total	10139.385		

Table 2: GML tests' results for publication year category.

Source	Sum of Squares(III)	F	Sig
Corrected Model	1918.761	2.924	0.37
Publication Year	1918.761	2.924	0.37
Error	2994.400		
Total	10139.385		

Table 3: GML tests' results for author's position category.

Source	Sum of Squares(III)	F	Sig
Corrected Model	334.641	0.736	0.574
Author's Position	334.641	2.924	0.574
Error	3979.521		
Total	10139.385		

We can make it more accurate that in which years are the most similar papers and by how many common coauthors, we can achieve the most similar papers regarding to the pivot paper. Consequently, we can apply two *Contrasts* that GLM provides, i.e, *Helmert* and *Deviation Contrast*. The former is where the effect of each category of independent variable is compared to the mean of the subsequent category<sup>12</sup> while the later is where the effect of each category<sup>13</sup> is compared with the grand mean.

By performing the Helmert contrast upon publication year category, we see a significant result only for the group 7 which stands for the year 2009. That means the year 2009 has a significance effect

<sup>12</sup>Except the last one, due to the fact that there will be no subsequent category to compare with.

<sup>13</sup>Except one of them which by default is the last category.

(P-value=0.009) compare to the subsequent category, which are the years 2005-2008. although there is no other significant result, we still can not be sure that only the year 2009 is the year that we should look for similar papers, however we know that this year, possibly together with its subsequent categories, carry a significant effect on the similarity. To make sure about this, we perform a deviation contrast to see the effect of each category. The result shows a significant effect on category 7 (P-value=0.001) and 9 (P-value=0.40), which are 2009 and 2007 respectively. In other hand, it means that, based on the paper that we selected from year 2007, we only need to consider year 2009 and 2007 in order to retrieve most similar papers.

Since the pivot paper has only 3 coauthors, we applied a *Repeated contrast* which compare effect of each category with its adjacent one. In this way, we can see that how increase of number of common coauthor would effect similarity measurement. The result defines that there is no significant effect by increasing the number of common coauthor from 0 to 1 ( $P_{value} = 0.092$ ) but there is almost significant effect on similarity when the number of common coauthors increases from 1 to 2 ( $P_{value} = 0.052$ ). We can conclude that based on the similarity of textual references to visual features, in our experimental settings, the most similar papers of the first author of pivot paper, published by 2007, are those paper that are published in year 2009 and 2007 and/or with 2 common coauthors.

### 4.3 Experiment 3

The aim of this experiment is to use the results of two previous experiments in order to build a *Boolean function* which its task is to index papers of an author. To do such, we selected 2 arbitrary authors. By selecting two random papers of each, we used the stated method to seek for significance results in any of the 3 factors, author's position, publication year and coauthor ratio. Any significant value for each factor will be added into the Boolean function under an specific ID. Since showing all of the conditions of Boolean function would take up a lot of space and confuse the reader, for being concise we only demonstrate a part of it as Boolean functions are illustrated in the Equation 5.

$$\left\{ \begin{array}{l} A_3P_4I_{19} = \omega(4 - i)[\lambda(2^{\omega(|1.2 - \text{ratio}|)}) \vee \\ \lambda(2^{\omega(|1 - \text{position}|)}) \vee \\ 2^{\omega(|3 - \text{position}|)} \vee \\ 2^{\omega(|4 - \text{position}|)}] \\ A_2P_5I_{18} = \omega(5 - i)[\lambda(2^{\omega(|0.28 - \text{ratio}|)})] \\ A_2P_6I_{27} = \omega(6 - i)[\lambda(2^{\omega(|0.66 - \text{ratio}|)}) \vee \\ \lambda(2^{\omega(|21 - \text{year}|)}) \vee \\ 2^{\omega(|22 - \text{year}|)}] \\ A_3P_8I_{30} = \omega(8 - i)[\lambda(2^{\omega(|2.31 - \text{ratio}|)}) \\ \vee \lambda(2^{\omega(|19 - \text{year}|)}) \\ \vee 2^{\omega(|21 - \text{year}|)} \\ \vee 2^{\omega(|22 - \text{year}|)} \\ \vee 2^{\omega(|23 - \text{year}|)}] \end{array} \right. \quad (5)$$

$$\omega(x) = \begin{cases} 1 & x=0 \\ 0 & \text{else} \end{cases} \quad (6)$$

$$\lambda(x) = \begin{cases} 1 & x = 1 \\ 0 & \text{else} \end{cases} \quad (7)$$

Where  $A_iP_jI_k$  is a unique ID for paper  $k$  of author  $i$  with total number of  $j$  authors. The  $\omega(x)$  function checks the significant value of each of 3 introduced factors. Note that there might be a case that a paper had no significant value for any of factors. That means the paper has no considerable similarity with other publication of its first author. In addition,  $\omega(x)$  takes care of matching correct papers in terms of number of total authors to correct Boolean function. Definition of  $\omega(x)$  is illustrated in Equation 6. *ratio* is the common coauthor ratio, calculated by Equation 1 and *year* stands for the year category of the selected paper which is 1 for 1988 and 27 for 2015. Each  $\lambda(x)$  returns 1 when  $x = 1$ . In other word, this function's condition satisfies whenever any of significant factor of a paper gets enabled. The *Logical or* symbol,  $\vee$ , indicates that as soon as either of condition satisfies, the corresponding paper will be regarded as a significantly similar paper. In order to reach a better result, we can order each factors based on its effect size. Steps of calculating effect size is demonstrated in section 4.2. On the other hand, we can change the logical or into the *Logical and* symbol  $\wedge$  only between each  $\lambda$  function and not within them. This means, the Boolean function is true if and only if all conditions satisfy together.

All Boolean functions of author  $A_1$  can be summarized in Equation:

$$\sum_{j=3}^n \sum_k^m A_1P_jI_k \quad (8)$$

where  $n$  indicates the maximum number of authors that in the publication record of author  $A_1$  appears. For any particular  $P_j$  there might be one or more  $I_k$  that are not necessary valid for all  $j$  and  $k$ . The summation starts from 3 since according to the common coauthor ratio, for papers with 1 authors, the ratio will be infinite<sup>14</sup> and for papers with 2 authors, the Equation 3 will be equal to either 1 or 0. When it is equal to 0, the value for Equation 1 would be equal to infinite.

## 5 CONCLUSION AND FUTURE WORK

In this paper, we first introduced a novel method to *intellectually* select one's similar paper from his/her other publications. Then we showed the effectiveness of our method and accordingly, we came into a Boolean function which is based on the papers' characteristics, i.e. *Author's position*, *Publication year* and *Coauthor ratio*. However for a fully functional system for paper recommendation that is not only limited to an authors publication record, we can combine this system together with our previous work (Alli, 2015) to build a fast and accurate recommender system in scientific scholarly field, specialized in computer science.

## REFERENCES

- Alli, M. (2015). Papers similarity based on the summarization merits. In *2nd International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC2015)(Accepted)*.
- Bogers, T. and van den Bosch, A. (2008). Recommending scientific articles using citeulike. In *Proceedings of the 2008 ACM Conference on Recommender Systems, RecSys '08*, pages 287–290, New York, NY, USA. ACM.
- Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117.
- Buyukkokten, O., Garcia-Molina, H., and Paepcke, A. (2001a). Accordion summarization for end-game browsing on pdas and cellular phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '01*, pages 213–220, New York, NY, USA. ACM.

<sup>14</sup>The denominator in Equation 3 will be equal to zero

- Buyukkokten, O., Garcia-Molina, H., and Paepcke, A. (2001b). Seeing the whole in parts: Text summarization for web browsing on handheld devices. In *Proceedings of the 10th International Conference on World Wide Web, WWW '01*, pages 652–662, New York, NY, USA. ACM.
- Chen, H., Gou, L., Zhang, X., and Giles, C. L. (2011). Collabseer: a search engine for collaboration discovery. In *Proceedings of the 2011 Joint International Conference on Digital Libraries, JCDL 2011, Ottawa, ON, Canada, June 13-17, 2011*, pages 231–240.
- Good, N., Schafer, J. B., Konstan, J. A., Borchers, A., Sarwar, B., Herlocker, J., and Riedl, J. (1999). Combining collaborative filtering with personal agents for better recommendations. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference Innovative Applications of Artificial Intelligence, AAAI '99/IAAI '99*, pages 439–446.
- Han, S., He, D., Brusilovsky, P., and Yue, Z. (2013). Coauthor prediction for junior researchers. In Greenberg, A., Kennedy, W., and Bos, N., editors, *Social Computing, Behavioral-Cultural Modeling and Prediction*, volume 7812 of *Lecture Notes in Computer Science*, pages 274–283. Springer Berlin Heidelberg.
- He, Q., Pei, J., Kifer, D., Mitra, P., and Giles, L. (2010). Context-aware citation recommendation. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pages 421–430.
- Hong, K., Jeon, H., and Jeon, C. (2013a). Advanced personalized research paper recommendation system based on expanded userprofile through semantic analysis. *International Journal of Digital Content Technology and its Applications (JDCTA)*, 7(15):67–76.
- Hong, K., Jeon, H., and Jeon, C. (2013b). Personalized research paper recommendation system using keyword extraction based on userprofile. *Journal of Convergence Information Technology (JCIT)*, 8(16):106–116.
- Levine, T. R. and Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, 28(4):612–625.
- Ma, N., Guan, J., and Zhao, Y. (2008). Bringing pagerank to the citation analysis. *Inf. Process. Manage.*, 44(2):800–810.
- Nykl, M., Jezek, K., Fiala, D., and Dostál, M. (2014). Pagerank variants in the evaluation of citation networks. *J. Informetrics*, 8(3):683–692.
- Pohl, S., Radlinski, F., and Joachims, T. (2007). Recommending related papers based on digital library access records. In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '07*, pages 417–418, New York, NY, USA. ACM.
- Sayyadi, H. and Getoor, L. (2009). Futurerank: Ranking scientific articles by predicting their future pagerank. In *In Proc. of the 9th SIAM International Conference on Data Mining*, pages 533–544.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. In *Reading Research Quarterly*, volume 22.
- Sugiyama, K. and Kan, M. (2010). Scholarly paper recommendation via user's recent research interests. In *Proceedings of the 2010 Joint International Conference on Digital Libraries, JCDL 2010, Gold Coast, Queensland, Australia, June 21-25, 2010*, pages 29–38.
- Sun, Y., Barber, R., Gupta, M., Aggarwal, C., and Han, J. (2011). Co-author relationship prediction in heterogeneous bibliographic networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 121–128.
- Tang, J., Sun, J., Wang, C., and Yang, Z. (2009). Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09*, pages 807–816, New York, NY, USA. ACM.
- Tang, J., Zhang, D., and Yao, L. (2007). Social network extraction of academic researchers. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 292–301.
- Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., and Su, Z. (2008). Arnetminer: Extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08*, pages 990–998, New York, NY, USA. ACM.
- Vellino, A. (2009). Recommending journal articles with pagerank ratings. In *In Recommender Systems 2009*.
- Vellino, A. (2010). A comparison between usage-based and citation-based methods for recommending scholarly research articles. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–2.
- Wang, C., Han, J., Jia, Y., Tang, J., Zhang, D., Yu, Y., and Guo, J. (2010). Mining advisor-advisee relationships from research publication networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10*, pages 203–212, New York, NY, USA. ACM.