Journal of Universal Computer Science, vol. 18, no. 8 (2012), 993-1021
submitted: 21/9/11, accepted: 14/12/11, appeared: 28/4/12 © J.UCS

Establishing Knowledge Networks via Analysis of Research Abstracts

Mahalakshmi G. Suryanarayanan, Dilip S. Sam
(Department of Computer Science and Engineering
College of Engineering, Guindy, Anna University, Chennai, India
mahalakshmi@cs.annauniv.edu, sdilipsam@gmail.com)

Sendhilkumar Selvaraju
(Department of Information Science and Technology
College of Engineering, Guindy, Anna University, Chennai, India
ssk_pdy@yahoo.co.in)

Abstract: The extraction and propagation of knowledge inherent in a social network environment is demanding higher significance in research. The knowledge hidden within a social network would be easier to be comprehended if provided in a collective form. In the field of scientific research, such presentation of appreciated knowledge evolved from research communities would aid researchers. In this paper, we propose the evolution of a knowledge network from the information available in digital bibliographic repositories like DBLP [DBLP]. The most important characteristic of this knowledge network would be the comprehension of the proficiency of the scientist in the perspective of an area of research. This is achieved by categorizing the research articles published by an author into specific domains. The quality of the research articles are ascertained by analysing the abstracts within the domain. This analysis is used to determine the quality of the research article in terms of originality, relevancy and thereby, the impact of the article with respect to a research area. This quality measure provides knowledge on the impact of the scientist on the research community is arrived at as a cumulative entity. This knowledge helps in the evolution of the knowledge network from the social network of a research community.

Keywords: Fuzzy Cognitive Maps, Knowledge Networks
Categories: H.3.3, H.3.7

1 Introduction

A knowledge network is an exceptional one within social network that the actors in the knowledge networks are related to academic research locale. On the contrary, social network is one where people interact, communicate, share things about any aspect. Knowledge network varies from social network where actors sounding better in their research oriented aspect are inter-related to form a specific group. The categorization of 'who knows who knows what' is likely to be achieved via knowledge networks [Noshir and Peter, 02]. The ability to gain knowledge from others is not simple as this knowledge does not exist readily in a single place. Social networking elements like blogs and discussion forums are not valued much due to their improper standards. The academic related information/knowledge lags in these

elements. Though semantic social network exhibit relationship among people, it does not provide certain features like author bonding, knowledge transition of authors, domain specific author quality, levels of author contribution, author centrality, etc. Hence researchers obviously trust standard bibliographic repositories like DBLP [DBLP], Cite seer [Citeseer] for better knowledge acquisition. This implicitly demands for quality of authors present in those repositories. The feasibility for the transition of social network to knowledge network is not simple because knowledge network simply possess academic oriented aspects whereas social network elements are not restricted to any particular aspect.

Hence evolution of knowledge network becomes the need of the hour for researchers to pursue their work in an intelligent way. The actors join or leave a knowledge network on the basis of tasks to be accomplished, and their levels of interests, resources, and commitments. The links within the knowledge network are also likely to change on the basis of evolving tasks, the distribution of contributions that the author made, or changes in the actors' cognitive knowledge network. The various levels of contribution of actors can be easily identified via knowledge networks confined to a particular domain. The knowledge of a researcher shall be weighted according to their previous research contributions. In this context, the research abstracts of a researcher is analysed for determining one's capacity as a researcher. The following section briefs about the beliefs and misconceptions about analysing a researchers' capacity by following citation based impact metrics.

With academic universities granting doctoral research admissions on a major scale and with the academic rules being laid on researchers support publication in reputed journals, paper publication is now emerging as a pretty good business and gradually losing its charm in the recent past. The technological advancements and open source policies have equally had a negative impact on promoting high quality research. With voluminous research papers at the desk, a novice researcher is often found misled during the early stages of research. This result in identification of incorrect, known faulty research problems and the time and energy of researchers are not utilised in the right direction. Hence there is a need to identify quality researchers for the benefit of the research community. Quality arises from, not the quantity of research contributions but from the research originality. Therefore, we believe that detecting the originality of scientific research abstracts will contribute more towards preserving the quality of scientific research.

2 Related Work

2.1 Co-Authorship Networks Representations

Co-Authorship networks are usually represented as undirected graphs. In these, the vertices are the authors and an edge exists between two authors if they have co-authored a research publication. Erdős number was initially used as an index for mathematicians taking into consideration their collaboration with the mathematician Paul Erdos. Paul Erdos had published around 1500 mathematical articles and the proximity to Erdos was used as a measure for the prominence of a mathematician. However, the approach had limitations as an independent author might be lost in the network, as authors not connected to Erdos (having no co-authorship chain) had their

Erdos number as undefined. This led to a bias in the calculation of the prominence of a mathematician. Similar networks were used in the movie industry, known as the Bacon number.

Multi-Weighted co authorship networks [Evelyn et. al., 10] tried to add weightage to the co-authorship networks based on the type of publishing, books carrying the most weightage and conferences the least. This approach improves the co-authorship network by improving the simple co-authorship network. However, the problems of individual authors being left out of the network persists in this approach too.

Combination of co-authorship and content similarity information to improve the co-authorship network based on [Varlamis et. al., 10] solve the problem of individual authors being left out by analysis of content, thereby authors with similar research interests can be linked. This can be done despite the absence of a co-authorship chain between them. Iraklis Varlamis uses power graphs to represent the co-authorship networks. Other approaches include Hyper-graphs, Evolutionary graphs [Gupta et. al, 11]. Evolutionary graphs can be a useful representation to perform spatial-temporal analysis of the co-authorship network.

2.2 Mining Research Communities

Extracting information from bibtex data for potential research use has been the focus of data mining and information retrieval research. Evolving research communities which are made up of authors, representing different research groups, that are linked with different type of relations has the concept of social network in it. Hence, viewing and understanding the research relationship between the nodes of the network is an essential part of social network analysis.

Various Social Network Analysis (SNA) methods [Wasserman and Faust, 94] exist to analyse citation based research networks. Community Mining [Girvan and Newman, 02; Newman, 03] has received considerable attention over recent years. Identifying the connections existing between the nodes of the communities with nodes sharing similar properties with each other is very interesting yet challenging task [Osmar et. al., 07]. Our idea is to find potential collaborating researchers by discovering communities in an author-centric research network. However, we tend to differ from the formation of research communities of Osmar et.al. [07].

In community mining, the closeness of related concepts is usually measured by ‘relevance score’. For this measure, relationships between the entities need to be identified. With the possibility of multiple, multi-level and multi-variant relationships between the nodes of research communities, quantifying a relevance score would be more approximate or would be done under varying assumptions. Euclidean distance or Pearson correlation [Wasserman and Faust, 94] could be used for such purpose. Since social networks could be modelled as graphs, usage of traditional graph algorithms such as spectral bisection method [Pothen et. al., 90] which is based on Eigen vectors, or Kernighan-Lin algorithm [Kernighan and Lin, 70] which greedily optimizes the number of internal and interface level community edges suffer from graph bisection problems. The decision on when to stop the graph bisection is of prime importance. Hierarchical clustering [Hastie, 09] could be a better bet, however, if nodes of the communities are not close to one another, then forming the clusters would be a major problem.

Random walk approach [Martin and Carl, 08; Backstrom, 11] is widely used to determine the relevance score between the entities of a community network. Another variation of Random walk approach, called Random walk with restart (RWR) [Osmar et. al., 07] is used by considering the traditional random walk with a restart probability. Using this iterative random walk algorithm, the relevance score is computed for recommendation of potential research collaborators. In addition, analysing the co-author relation might reveal interesting results [Nascimento et. al., 03; Smeaton et. al., 02]. However, a community discovery to recommend potential collaborating researchers should not end up with directing only a colleague or fellow researcher as a collaborator. Co-authorship information is something which is directly available with the bibliographic data. Rather deriving other implicit information about the researchers would be more difficult because of the volume of bibliographic data.

DBLife [http://dblife.cs.wisc.edu/], DBConnect [Osmar et. al., 07] and Libra are some projects experimented over DBLP for evolving heterogeneous information networks. They provide related researchers and related topics to a given researcher. With research community primarily interested in knowing relevant conferences, similar authors and interesting research topics and call for papers, DBConnect [Osmar et. al., 07] provides more accurate recommendation to research collaborators. To extract the research topic DBConnect [Osmar et. al., 07] is careful in assuming that the authors of a similar conference could possibly research in different yet tangential and/or perpendicular research themes. Therefore, research topic is determined by DBConnect [Osmar et. al., 07] extracting titles from DBLP [http://www.informatik.uni-trier.de/~ley/db/] and abstracts from Citeseer [http://citeseerx.ist.psu.edu/]. However, the quality of research abstracts and the potential relevance of research abstracts to the corresponding titles and /or the rest of the paper is of a major concern. Therefore, the accuracy of relevance score is diminishing without the 'quality' perspective.

With the only available short-text, the title of the paper, it is hard to extract the correct research topic of the paper. In addition, the practice of researchers naming their paper with metaphoric / unrelated words / acronyms / question phrases will worsen the level of accuracy. The possible solution recommended by Osmar et al. [07] is to implement a hierarchy of topical words. In this paper, we have the objective of performing 'quality based' community discovery to recommend 'authentic' and not popular potential collaborating researchers. The idea is to bring the knowledge of a researcher as a prime component in the information network. Therefore, we tend to name it as 'knowledge networks'. This knowledge is hidden in massive links of the research network [Han, 09] and for the same reason link mining of research network only would lead to identifying knowledge out of the research community.

Here, we propose an analysis of single researcher impact factor by measuring the position of researcher among the research community. Since the community evolves time to time, the position of researcher is also looked upon with respect to the research period, and the domain of research. To assess the knowledge of the researcher, content analysis of one's research articles is essential. To lead a novice researcher into the nucleus of knowledge networks, assessing the research impact and productivity on quality (and not of popularity) perspectives becomes mandatory. Therefore, we attempt at evolving a research network where every researcher finds some place with defined values. We believe that instead of following the substitutive

measure like IF, an integrative system of research evaluation would be more appealing to the research community. The following section discusses the idea behind the evolution of knowledge networks.

2.3 Fuzzy Cognitive Maps

Fuzzy Cognitive Map (FCM) is an extension of cognitive maps which represent knowledge similar to the hippocampus. FCMs are cognitive maps within which the relations between the concepts are fuzzy. This enables better representation and interpretation [Pena, 05] of semantics while using the FCM. FCMs are applied across various domains including business, economics, robotics, expert-systems and project management [Pena, 05] [Aguilar, 03] [Jose Salmeron, 10].

Manjula and Simaan [07] propose the development of FCM in two steps: The identification of concepts, which is followed by the identification of causal relationships among these concepts. The edge weights between the concepts in the FCM can be assigned based on combinations of concepts in matrix format [Kosko, 86].

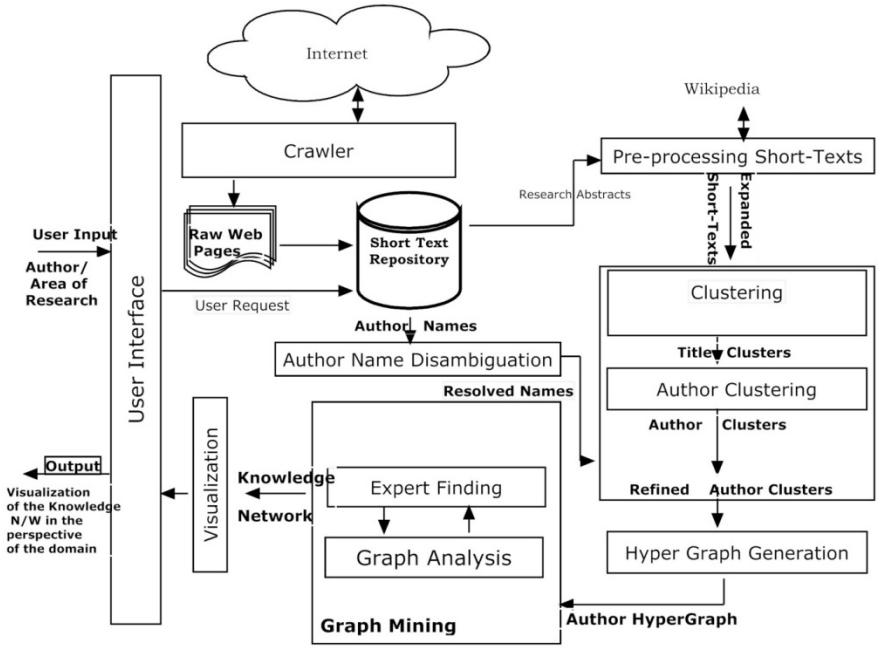


Figure 1: Evolution of Knowledge Network

3 Evolution of Knowledge Networks

This system aims in providing the user with a comprehensible network, conceived by the knowledge extracted from the content available in the digital bibliographic

repositories. This knowledge network will supply the user with inherent relationships among authors, research based upon the area of interest. In this system, a subject-specific search in the repository would return the relationship between authors and between papers based on the relevance of the release to the subject of interest. This will include the extraction of short-texts available in the digital bibliographic sites like DBLP [DBLP] to find key-phrases and arrive upon the degree of relation between the paper and the subject of interest to the user, thereby building a hyper graph. This system can be enhanced by content retrieval of papers from other related websites, analysis of abstracts. Upon applying reasoning to the hyper graph, the relationship between authors, research limited to the subject of interest is provided to the user.

3.1 Crawler

The crawler uses a modified KPS algorithm [Guan and Wong, 99]. Pages from digital bibliographic sites in the Internet are crawled for publications. From the extracted pages, information available about the journal, authors and the title / abstract of the article are extracted. The crawler is implemented as an independent agent to perform incremental crawling [Rosy et. al., 10]. Incremental crawling would ensure the freshness of the information.

3.2 Author Name Disambiguation

Authors' name disambiguation grows complex with the amount of information available to the system. The problem compounds many factors including the same author using different names, typographic errors, authors sharing the same name etc. This system provides an adapted K-way spectral clustering [Han et. al., 05] method by indexing authors and deals with ambiguity by heuristics which include the order in which the first, middle and last names appear in the journal. The ambiguities will be resolved by co-authorship and the chronological information which would be available in the bibliographic sites.

3.3 Clustering by Conceptual Similarity

To enhance the clustering of titles / research abstracts, they are expanded initially. This enrichment would contribute towards better accuracy in clustering. Enrichment of short-texts is carried out by augmenting terms from the Wikipedia[Banerjee 07].Once the enrichment of short-texts is completed, the quality of short-texts i.e. research abstracts need to be determined. The quality of abstracts is first determined from the plagiarism perspective. i.e. every research abstract is compared with every other research abstract in the corpus via conceptual similarity. The articles that do not meet the threshold quality are not considered in the generation of the clusters. The number of clusters is determined dynamically so that the documents within the clusters have maximum silhouette.

3.3.1 Ontological approach to detecting Document Similarity

In text mining, ontological approach is used as a specific characteristic to detect document similarity. Ontology can serve as repository of all concepts in a domain.

This knowledge pertaining to a domain can be harnessed in detecting similarity between documents belonging to the same domain. A domain ontology which is constructed offline can be used to detect idea/concept based similarity between documents belonging to that domain. This is achieved by analysing the underlying text across the ontology and later, by extracting the concepts and the respective relations along with their hierarchy. The level of plagiarism detected depends on the following factors: (1) the efficiency of algorithms involved (2) the exhaustiveness and reliability of the underlying ontology. Therefore, an exhaustive and reliable ontology for a domain is useful in detection of similarity of research abstracts.

In the context of applying ontology to detect similarity, the term 'similarity index' is often used in the literature. In addition, the semantic comparison and evaluation of ontologies are also the need of the hour. James Z. Wang et al [10] proposed a method to measure the ontological similarity with semantic sets obtained from the two ontologies. The semantic set is obtained based on the edge-term relationship. Comparison of text documents to detect similarity from domain ontology is done by extracting concepts from the documents. The terms are then used to evolve sub ontology possessing various levels using graph theory. Each level of the graph is compared to determine the similarity index [Vladimir and Asle, 03].

Batet et al [Sanchez et. al.,11], proposed a distance function to calculate the similarity of words using ontology. Based on the distance between the super-concepts present in the ontology, a Super Concept-Distance (SCD) metric is proposed to detect semantic similarity between the word pairs. The main drawback of SCD is that the minimum path length between the parental concepts alone are determined, thereby eliminating other taxonomic concepts present in the ontology. In addition, the various levels of ontological structure possessing parental level, siblings and leaf nodal level are evaluated to determine the similarity index. Pruning of ontology may increase the performance of semantic similarity detection [Lee and Das,11]. However pruning a full ontology for identifying the same class, sub-class and non-sibling class is always not feasible. The reason may be that ontologies are preferably user-defined and customised and therefore they tend to lack knowledge in such inner levels. In sharing knowledge available from semantic web, similarity determination among sets of concepts is attempted with the help of ontology [Davis et. al.,04].

In this context, a metric for measuring the similarity between a concept and the sets of concepts follows Dijkstra's algorithm from RDF representations of ontology [Roy et. al.,09]. Various metrics such as Bouquet, Kuper, Scoz and Zanobini's metric, Castano, Ferrara, Montanelli, and Racca's metric, Haase, Siebes, and van Harmelen's metric, Rada, Mili, Bicknell, and Blettner's metric, Wu and Palmer's metric, Hirst and St.Onge's metric and Leacock and Chodorow's metric are also discussed briefly in the literature [Roy et. al.,09]. However, they lack novelty in determining the similarity among the ontological concepts. Yet another methodology is proposed by Rajesh et al [08] to measure the semantic similarity between the terms present in the document. The terms extracted from the document are used to obtain the related concepts from ontology thereby following 'set spreading' technique. The distance between the directly related concepts and indirectly related concepts are measured initially. Finally the 'Mean' obtained from the similarity is used to obtain the final similarity index value. This algorithm does not support measuring the semantic strength of underlying concepts.

Another technique to measure the semantic similarity between texts is based on subset ontology. Here, the initial ontological structure is constructed offline. Subset ontology is evolved based on the significant terms extracted from the sentences. Each level of subset ontology and the initial ontology are compared to determine the similarity based on the commonalities and difficulties [Rekha and Sasikumar, 10].

3.3.2 Using Fuzzy Cognitive Map (FCM) in detecting Document Similarity

Fuzzy Cognitive Map (FCM) is a mental map or mind map which is often used in representing the relationships between concepts. FCM is a directed graph with concepts as nodes and causality as edges [Pena et. al., 05].Concept maps are generally evolved with the guidance of human knowledge. In a similar manner, Fuzzy Cognitive Maps are evolved from our offline ontology. As suggested by Manjula and Simaan [07] the development of FCM typically includes two steps: The identification of concepts, which is followed by the identification of causal relationships among these concepts.

Therefore, as an initial step, we have constructed an offline ontology based on valid concepts from ACM classification system [ACM, 98]. More than 1500 concepts and the respective concept details are present in the ontology specifically in computer networks and related domain. Figure 2 represents the snapshot of an offline ontology built using Protégé editor [PROTÉGÉ, 11]. Protégé is an open source tool explicitly used to build offline ontology. The concepts are retrieved from the standard ACM system based on proper hierarchical aspects as stated. The parent, child and sibling relationships are hence established based on this hierarchy among the concepts. These ontological concepts are further transformed into fuzzy concepts in order to obtain FCM. This transformation is achieved by exploring the relationships to assign weights.

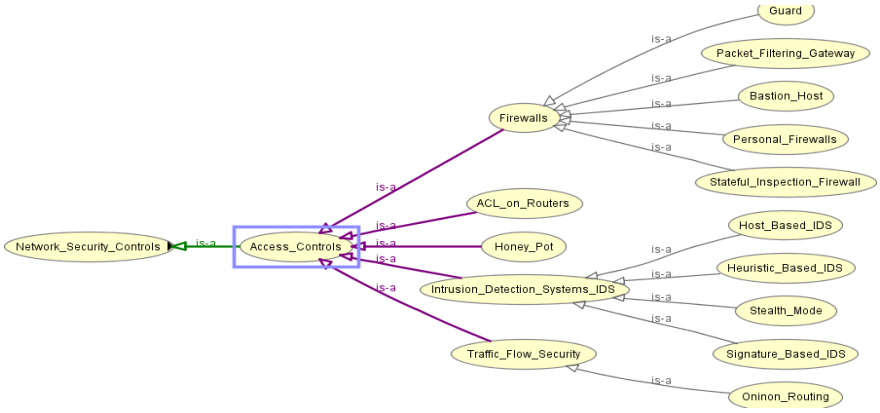


Figure 2: Snapshot of offline ontology depicting the conceptual relationships in 'networks security' – a sub-domain of our domain: 'computer networks'.

Figures 2 and 3 show the difference between FCM relationship and Ontological relationship among concepts belonging to network domain. As proposed by Kosko, the edge weights are assigned among concepts based on combinations of concepts in

matrix format [Kosko, 86] in case of FCM. The following section briefs on the construction of FCM.

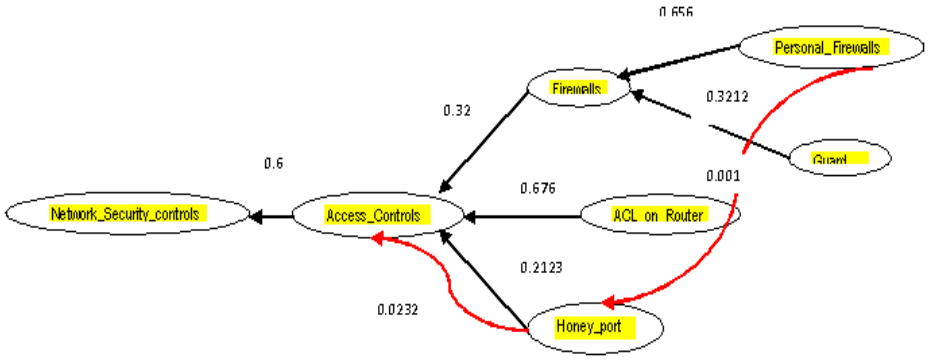


Figure 3: Snap shot of Fuzzy Cognitive Map depicting representing conceptual relationship among networks security

3.3.3 FCM Construction

The underlying text is pre-processed for removal of stop words. Later, quality terms are selected for further processing. The term quality is measured from the slight modification of the traditional metric introduced by [Salton and McGill, 83],

$$q(t) = \sum_{i=1}^{n} f_i^{\,2} - \frac{1}{n}\left[\sum_{i=1}^{n} f_i\right]^2 \qquad \text{- eq. 1}$$

In Eq 1, *q (t)* is the term quality, *f* is the frequency, and *n* is the total number of terms in the document. The term quality is thus detected for each term based on the term distribution in the text document.

The terms with high quality are clustered together by using the k-means clustering algorithm [Moore, 01], [Kanungo et. al., 02]. The k value is chosen in random to conclude the number of clusters by using Eq 2.

$$k = \sqrt{\frac{N}{2}} \qquad \text{- eq. 2}$$

where N is the total number of terms in the in the document for all the quality is determined using their distribution and k is the number of clusters. The centroid value is determined and each data point (here the quality terms) finds out which centre it is closest to. After this the term joins that particular cluster [Moore, 01]. This is repeated for all the quality terms to fall under a cluster. In our implementation, the top two clusters are considered predicting those terms in the clusters are with better quality thereby neglecting other terms in the remaining clusters.

The individual terms obtained from the clusters are given as input to the offline generated ontology and the neighbourhood concepts are extracted from the constructed ontology. The extracted concepts thus form the Ontology Set (Ontoset) [Alzahrani and Salim, 10]. The Onto sets of text documents thus obtained are merged to form a matrix with concepts of Onto set1 in rows and those in Onto set2 in columns. The relationship between the concepts are measured as the edge value where concepts as nodes. The metric used by Makoto et al [04] is considered and modified accordingly to compute Ontosets similarity. The edge value between the nodes is measured by Eq 3.

$$Sim_{(Ontoset1,Ontoset2)} = \frac{|A \bigcap B|}{|A \bigcup B|} \qquad \text{- eq. 3}$$

where A and B are the concepts of the Ontosets 1 and 2.

The similarity is computed for all the concepts of Ontoset1 and 2, thereby contributing towards determination of the document's similarity. The abstracts are grouped into clusters based on the conceptual similarity index and a hyper-graph is generated with details of authors in for every cluster.

3.4 Expert Finding

The hyper-Graph generated will have clusters of authors based on the domain. Mining the graph would provide the user with a comprehensible network. Partitioning the hyper-graph is done using the spectral hyper-graph partitioning algorithm [Zhen and Jiang,10]. Further analysis of hyper-graph will lead to finding the nucleus of the domain of research. This would enable the system to recommend the highest contributor of the underlying research domain.

The system aims to provide the user with knowledge network on the requested area of research. The user would be able to understand the changes in the knowledge network with respect to time. This could be used to recommend the author with the most significant contribution in a specific research area in any chosen timeframe. The author with the most significant contribution and impact in the research area is established as the nucleus of the research area. This information about the nucleus of the research area would enable the user with an enhanced co-authorship network which incorporated knowledge on a research area.

The set of publications in the Table 1 were hand-picked to demonstrate the outcome of the system. These research articles are from the research areas: 'reasoning', 'computer networks' and 'web search personalization'. While selecting the publications, it was ensured that the authors spanned more than one domain. This was done as scientific research is often multi-disciplinary and authors of a scientific publication can be associated with diverse research areas. In the publications considered, T.V. Geetha has co-authored in all the three aforementioned research areas. Figure 4 represents the co-authorship network for publications listed in Table 1.

G. S. Mahalakshmi and T. V. Geetha, A Mathematical Model for Argument Procedures based on Indian Philosophy, International Conference on Artificial Intelligence and Applications, part of the 24th Multi-Conference on Applied Informatics, Innsbruck, Austria, February 13-16, 06
G. S. Mahalakshmi, T. V. Geetha: Requirements Elicitation by Defect Elimination: An Indian Logic Perspective. IJSSCI 1(2): 73-90 (09)
G. S. Mahalakshmi, T. V. Geetha: Argument-based learning communities. Knowl.-Based Syst. 22(4): 316-323 (09)
G. S. Mahalakshmi, T. V. Geetha: An Indian logic-based argument representation formalism for knowledge-sharing. Logic Journal of the IGPL 17(1): 55-76 (09)
M. Harish, N. Anandavelu, N. Anbalagan, G. S. Mahalakshmi, T. V. Geetha: Result evaluation strategies for peer selection in P2P. Bangalore Compute Conf. 08: 26
G. S. Mahalakshmi, S. Sendhilkumar, P. Karthik: Automatic Reference Tracking with On-Demand Relevance Filtering Based on User's Interest. PReMI 07: 349-356
S. Sendhilkumar, G. S. Mahalakshmi: Context-based citation retrieval. IJNVO 8(1/2): 98-122 (11)
G. Aghila, D. Manjula, T. V. Geetha: Automatic Generation of Description Logic Representation of Documents. IC-AI 03:874-879
D. Manjula, G. Aghila, T. V. Geetha: Document Knowledge Representation using Description Logics for Information Extraction and Querying. ITCC 03:189-3
S. Sendhilkumar, T. V. Geetha: Challenges in Personalization for General Information Search. IICAI 09:1598-1617
S. Sendhilkumar, T. V. Geetha: Web Search Using Personalized User Conceptual Index. IICAI 05:17-1728
G. S. Mahalakshmi, M. Karpagaraj, T. V. Geetha: An Indian logic-based knowledge-sharing architecture for virtual knowledge communities. IJNVO 6(2): 177-8 (09)

Table 1: BibTex from DBLP for a sample set of authors

The number of papers co-authored is reflected in the thickness of the edges in the graph. Clustering and subsequent hyper-graph mining would evolve a knowledge network that would enable the user to see the co-authorship network in the perspective of an area of research.

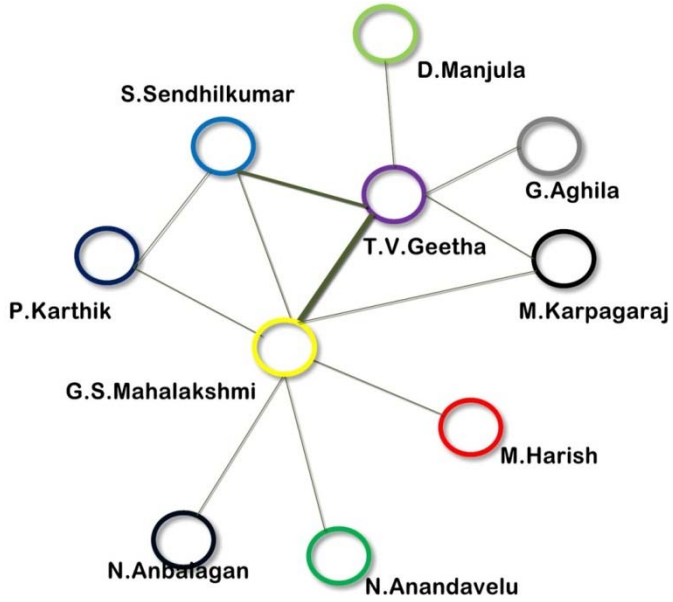


Figure 4: Co-authorship Network for the publications in Table 1

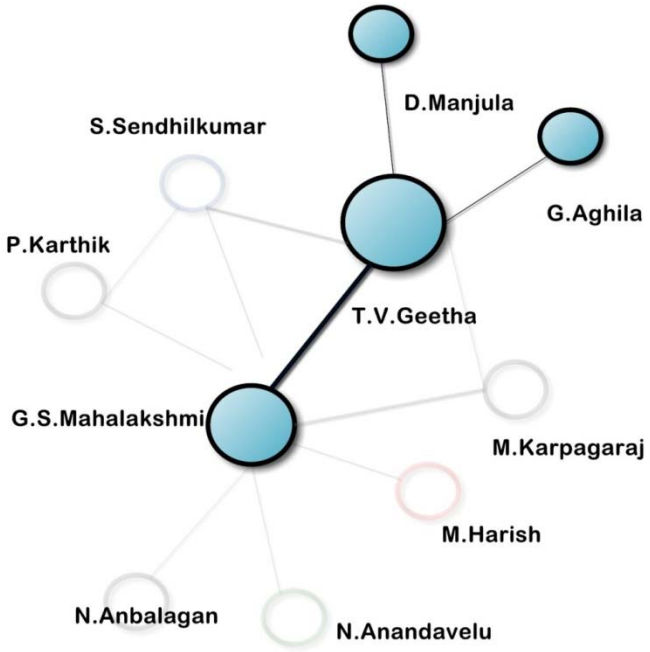


Figure 5: Knowledge Network; Domain: Reasoning; Nucleus: T.V.Geetha

In this example, the outcome of the system for the research area 'reasoning' would be similar to Figure 5. The authors who have contributed to the domain 'reasoning' form the knowledge network. The vertices are the authors and the edges represent the co-authorship. The author who has contributed the most to the research area in terms of originality and relevance in his publications would be the nucleus of that knowledge network In the list of publications in Table 1, T.V.Geetha is identified the nucleus in the knowledge network for the time-frame 05- 11, the research area being reasoning. Limiting the knowledge network to a specific research area enables the user with improved accuracy and accumulated knowledge on the contribution of scientists in the respective area of research.

4 Experimental Methodology

Text documents of research publications on 'computer networks' were used as a test bed for our work.1000+ research articles were collected manually. The abstracts of these research articles were extracted for analysis. The quality assessment of research abstracts is done in two dimensions.

1. The originality of research abstract.
2. The relevance of research abstract across the respective research publication

The ontology required for the FCM based similarity calculation contains concepts based on the standard ACM classification system.

4.1 Measurement of Originality and Relevance

We define relevance, as the similarity of abstract with the respective paper, and originality as the inverse of similarity measure obtained over the research publication corpus. The similarity between the research abstracts is calculated using the FCM based method explained in section 3.3.2. Similarity between two abstracts is measured by comparing their respective FCMs. For every individual abstract considered, its FCM is compared with every other article in the test bed. The similarity score of the considered abstract is obtained as the average of all the similarity values obtained by comparing the abstract with all the abstracts in the test bed. The research article with the least similarity score is taken to be the most original.

Relevance of an abstract to the research article is calculated by comparing the abstract to the body of the research article. This too is done by the FCM based similarity measure in which the FCM of the abstract to its body is calculated.

4.2 Calculation of Quality of the Author in the Research Area

Once the originality score of an author in the research area and the relevance between the abstract and the body of his publications have been ascertained, the quality of the author with respect to the research area in question is calculated with the two dimensions. The author with the highest quality is considered to be the nucleus.

5 Results and Discussion

5.1 Relevance of Research Abstracts

Abstracts carry the idea of any research article. Therefore, the completeness and originality of idea should reflect as well in research abstracts. In this context, the relevance of research abstract is computed via the similarity of research abstract to the rest of the respective research paper, and is measured using Dice-coefficient, Jaccard-coefficient and the Cosine similarity measures and compared with the performance of FCM (refer figure 6).

Dice suggested very little similarity between the abstract and the content of the article. FCM too suggested very little similarity. Further analysis was attempted on the grounds of normalizing the similarity values against the average similarity obtained for each method. Documents with values above the average value (8.86%) of Dice- coefficient measure were considered to be similar and others not similar. The same was done for the other three methods. For precision and recall, a similarity result was considered to be true if it agrees with one or more of the other similarity measures. From the precision and recall values (refer figure 7), it is understood that the Dice-similarity measure agrees more on dissimilarity between the abstract and the article with Jaccard and FCM-based methods As Dice- coefficient is based on common terms, the other similarity measures agree with the result of Dice. All four measures gave below-average similarity values for 810 documents. The precision and recall values of FCM-based method are average as this was calculated against methods which primarily focus on the common-terms and term frequencies for similarity measurement. From these measures, it can be seen that the FCM based method provides a different perspective to the similarity measure than the other three.

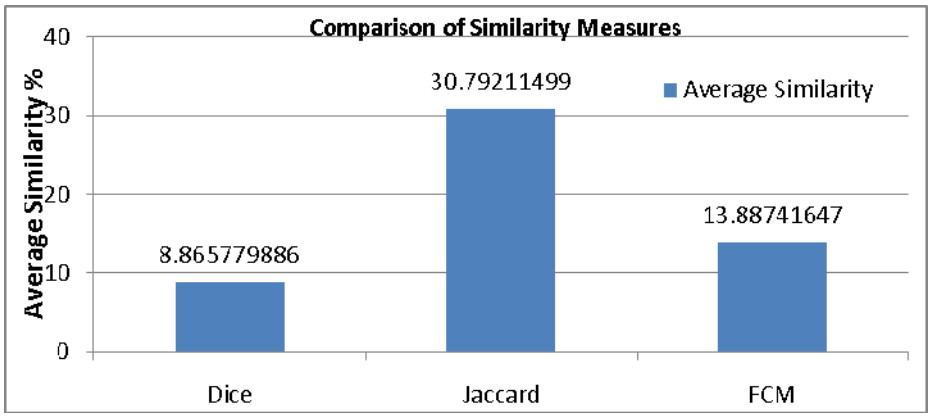
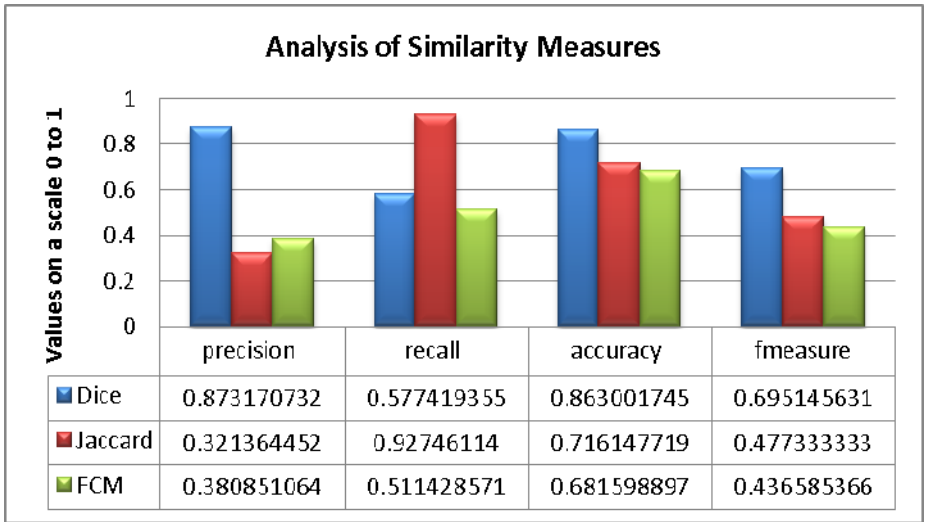


Figure 6: Comparison of the Similarity measures used for the experiment.



	precision	recall	accuracy	fmeasure
Dice	0.873170732	0.577419355	0.863001745	0.695145631
Jaccard	0.321364452	0.92746114	0.716147719	0.477333333
FCM	0.380851064	0.511428571	0.681598897	0.436585366

Figure 7: Precison, Recall,Accuracy & F-Measure

The ROC space (Figure 8) shows that all the three methods have better performance than random prediction. However, FCM based method differs in the way that it finds the cognitive similarity between the abstract and the full article. The ROC space plot for FCM with other cognitive similarity measurement methods like Fuzzy Grey Cognitive Maps (FGCM) would provide a better understanding on the performance of FCM-based similarity detection.

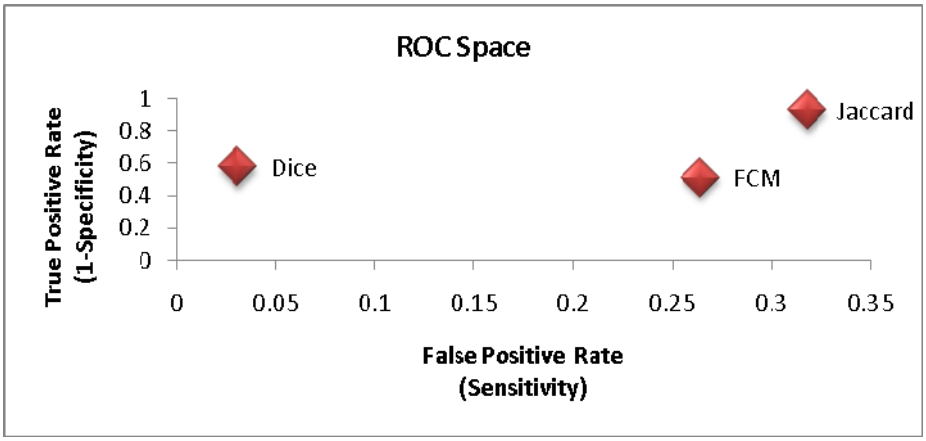


Figure 8: ROC space curve (Sensitivity vs 1-Specificity)

5.2 Impact of Abstract phrasing

To find the impact of abstract's phrasing on conceptual similarity, we attempted re-phrasing the abstract of the following article:

Performanc eevaluationof a new route optimization technique for mobile IP, Moheb R. Girgis, Tarek M. Mahmoud, Youssef S. Takroni and Hassan S. Hassan, International Journal of Network Security & Its Applications (IJNSA), Vol.1, No.3, October 09, pp 63 – 73(Refer Table 2).

The abstract was rephrased twice and FCM based similarity measure conveyed significant change in the similarity of the abstract with the article on rephrasing.

Original Abstract	Mobile ip (MIP) is an internet protocol that allows mobile nodes to have continuous network connectivityto the internet without changing their IP addresses while moving to other networks. The packets sent from correspondent node (CN) to a mobile node (MN) go first through the mobile node's home agent (HA), thenthe HA tunnels them to the MN's foreign network. One of the main problems in the original MIP is thetriangle routing problem. Triangle routing problem appears when the indirect path between CN and MN through the HA is longer than the direct path. This paper proposes a new technique to improve theperformance of the original MIP during the handoff. The proposed technique reduces the delay, thepacket loss and the registration time for all the packets transferred between the CN and the MN. In thistechnique, tunneling occurs at two levels above the HA in a hierarchical network. To show theeffectiveness of the proposed technique, it is compared with the original MIP and another technique forsolving the same problem in which tunneling occurs at one level above the HA. Simulation resultspresented in this paper are based on the ns2 mobility software on Linux platform. The simulations resultsshow that our proposed technique achieves better performance than the others, considering the packetdelay, the packet losses during handoffs and the registration time, in different scenarios for the locationof the MN with respect to the HA and FAs.
Rephrased Version 1	The mobile IP enables the user to have a seamless connectivity using the same IP address. The home agent (HA) tunnels the packets from the correspondent node (CN) to the mobile node (MN) in the foreign network .However, this original working of mobile IP created the triangle routing problem. Triangle routing problem appears when the indirect path between CN and MN through the HA is longer than the direct path. In this paper, we propose a new technique to improve the performance of the original MIP during the handoff. The proposed technique reduces the delay, the packet loss and the registration time for all the packets transferred between the CN and the MN. In this technique, tunneling occurs at two levels above the HA in a hierarchical network. To show the effectiveness of the proposed technique, it is compared with the original MIP and another technique for solving the same problem in which tunneling occurs at one level above the HA. Simulation results presented in this paper are based on the ns2 mobility software on Linux platform. The simulations results show that our proposed technique achieves better performance than the others, considering the packet delay, the packet losses during handoffs and the

	registration time, in different scenarios for the location of the MN with respect to the HA and FA.
Rephrased Version 2	The mobile IP enables the user to have a seamless connectivity using the same IP address. The home agent (HA) tunnels the packets from the correspondent node (CN) to the mobile node (MN) in the foreign network .However, this original working of mobile IP created the triangle routing problem. Triangle routing problem appears when the indirect path between CN and MN through the HA is longer than the direct path. In this paper, we propose a new technique to improve the performance of the original MIP during the handoff. The proposed technique reduces the delay, the packet loss and the registration time for all the packets transferred between the CN and the MN. In this technique, tunneling occurs at two levels above the HA in a hierarchical network. This method is compared against the original mobile IP and another optimization scheme which performs tunneling in the previous level. Simulations were done on the ns2 mobility software on Linux platform. The simulation results show that our proposed technique achieves better performance than the others, considering the packet delay, the packet losses during handoffs and the registration time, in different scenarios for the location of the MN with respect to the HA and FAs.

Table 2: Original abstracts and rephrased versions

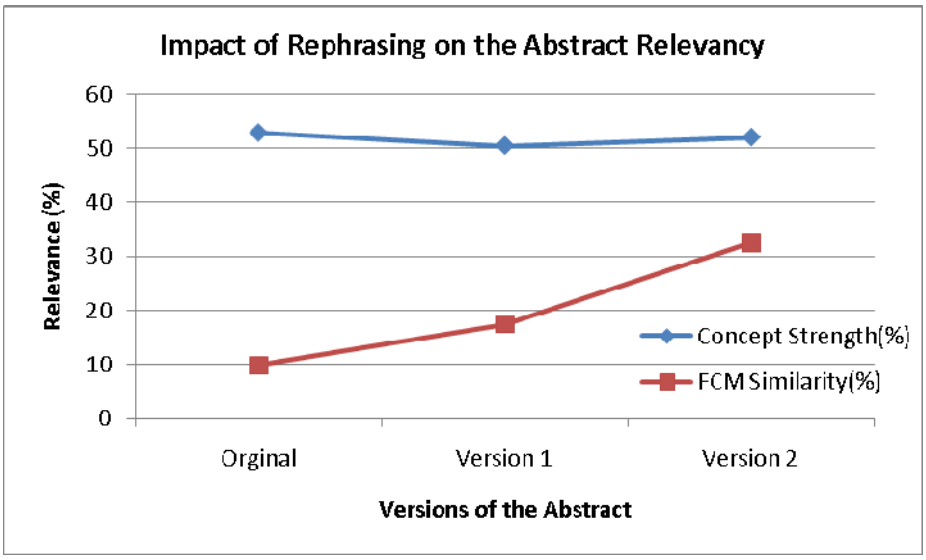


Figure 9: Impact of re-phrasing the abstract on the relevance of the abstract

Dice, Jaccard and Cosine similarity values (refer figure 9) did not change much with the re-phrasing of the abstract. FCM shows variation of the similarity with the structure of the text. A patient reviewing would agree that the level of version 2 sounds much better and clearer than the original and version 1 of research abstract.

Concept strength is calculated as the percentage of non-stop words in the abstract. From the figure 9, it is evident that the way the abstract is phrased has an impact on the similarity values and thus will directly impact the originality and the relevance values.

5.3 Empirical analysis of Research Abstract Originality

Abstracts are said to contain the essence of any research publication. Therefore, finding the originality of the research publication via abstracts is equally intelligent [Mahalakshmi et. al., 09]. Therefore, we have experimented with the abstracts of 00 research publications (sampling with replacement) and tabulated the similarity results in Figure 10a and originality results in Figure 10b. We have compared the FCM based abstract similarity measures with Dice's co-efficient and Jaccard co-efficient [Mihalcea et. al., 06; Nevin, 96]. FCM based similarity analysis yet outperforms every other approach in detecting the originality of scientific publication with respect to the abstract, except for a few surprises.

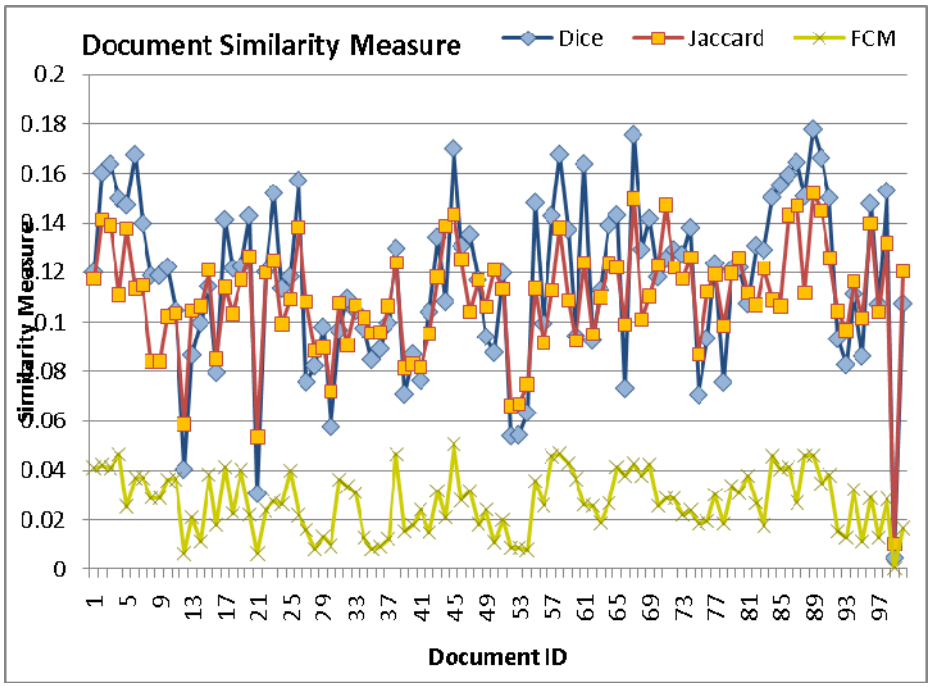


Figure10a: Comparison of statistical and ontological approach in detecting abstract based document similarity for 100 scientific abstracts across abstracts corpus of size = 2000. FCM outperforms by reducing the extreme fluctuations in the conducted experiments.

The reason where abstract based document originality goes for a failure would be for those abstracts which do not carry a detailed representation of the underlying idea, and, many times during our experiments, we found inconsistency of ideas as

expressed in the research publication across the abstracts, i.e. the abstracts were found to contain inspiring thoughts as well but the respective research description in the paper was not so convincing. Another reason may be that the abstracts are actually the short texts and therefore, to analyse the originality with the short text would be misleading.

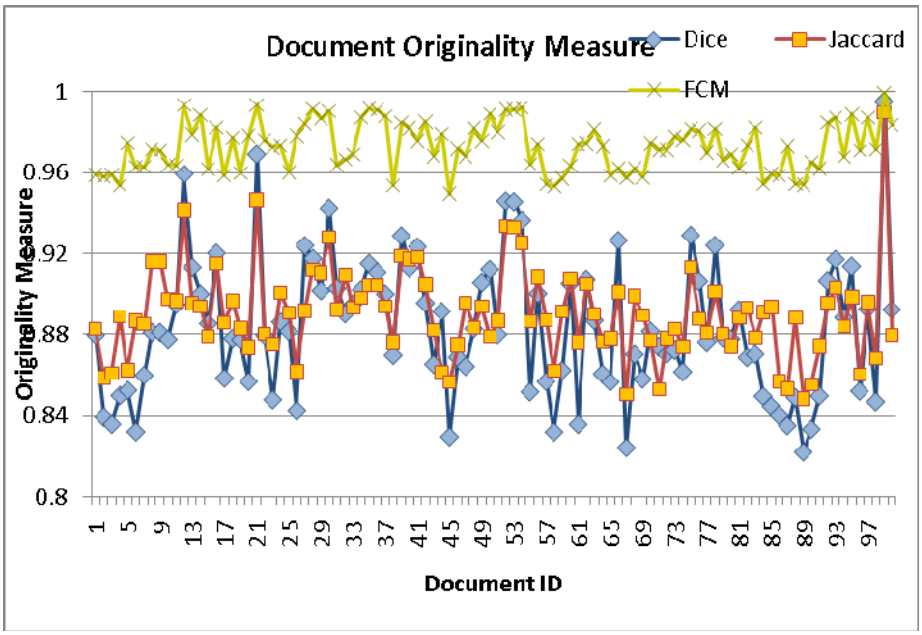


Figure10b: Comparison of statistical and ontological approach in detecting abstract based document originality for 100 scientific abstracts across abstracts corpus of size = 2000. FCM outperforms Jaccard in almost every experiment and yet surpasses Dice's approach in the conducted experiments.

5.4 Establishing Knowledge Networks via analysis of Research Articles vs. Research Abstracts

Problems of research abstracts are evident (as discussed in section 5.2) and therefore, analysing the entire research article to assess the research contribution of the author would be more appealing. From figure 11 the behaviour of similarity graph is the same as that for research abstracts since Jaccard measure of semantic similarity is purely syntactical. However, FCM method (figure 12) has less variation in similarity computation when comparing abstracts approach with full article approach.

From figure 13, it can be observed that the trend of the graph for both Jaccard and FCM based method is the same, however, variations within the results are greatly normalised in FCM based approach. Therefore, we can conclude that conceptual similarity has the following behaviours:

1. Supports the decision of syntactical similarity

2. Outperforms any other syntactical similarity by normalising the variations in similarity.

The reason may be that word level variations are transformed into conceptual variations and therefore the normalised results. Normalizing the obtained values, the FCM based originality calculations run on the corpus of abstracts alone suggested 49% of the publications to be original, whereas the calculations based on full-text corpus suggested only 41% of the publications to be original. This reduction in the number of original documents can be attributed to the presence of the related work section in the full-texts. The elaboration of the related work in the full-text documents brings down the originality measure of the documents owing to the increase in the similarity that exists through the related work section.

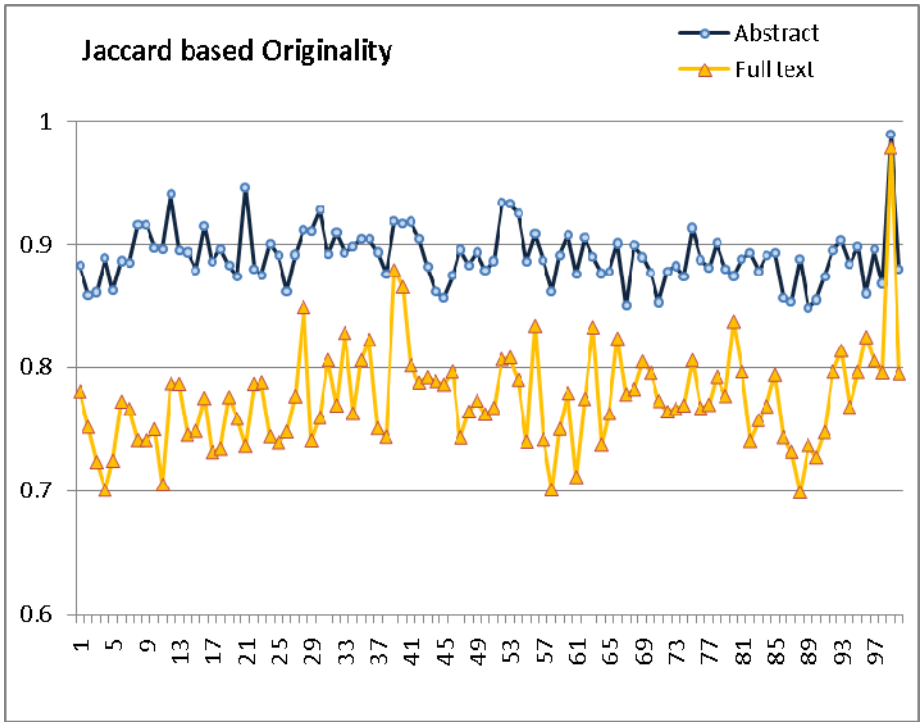


Figure 11: Comparison of statistical approach in detecting document similarity for Jaccard 1: 100 scientific abstracts across abstracts corpus of size = 2000. Jaccard 2: 100 scientific publications across scientific publication corpus of size =2000.

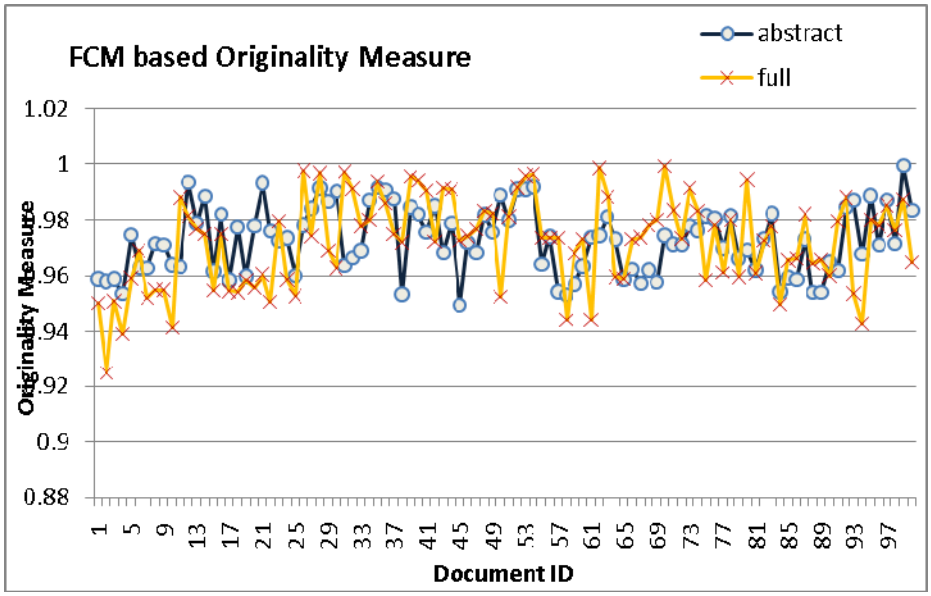


Figure12: Comparison of conceptual approach in detecting document similarity for (i) Full-Texts: 100 scientific abstracts across abstracts corpus of size = 2000. with (ii) Abstract: 100 scientific publications across scientific publication corpus of size =2000.

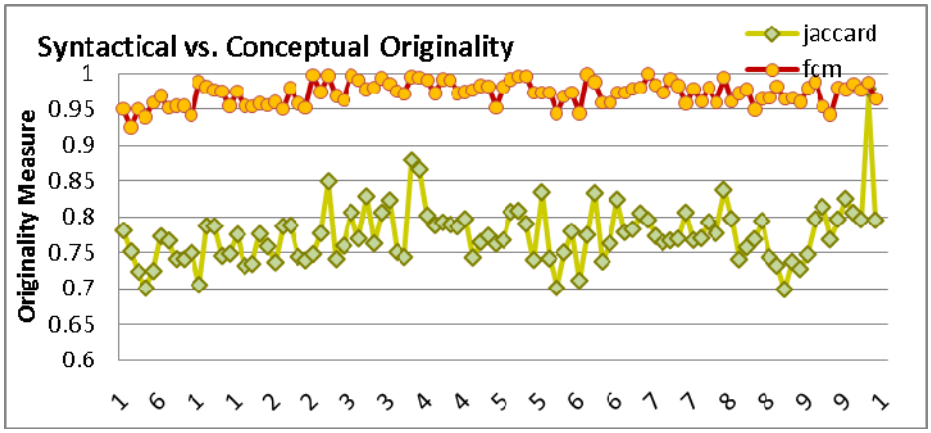


Figure13: Comparison of statistical and conceptual approach in detecting document similarity for Full-Texts: Jaccard Vs. FCM based approach

This could be overcome by not considering the related-work section of the full-texts, thereby providing a level ground for research articles with elaborate related work. Another method of overcoming this issue would be to consider the titles alone.

The articles which do not surpass the originality threshold in the corpus were not being considered for the knowledge network.

5.5 Cluster formation and Knowledge Networks

With the quality of the documents determined, the abstracts are clustered to form topic clusters. Clustering is done by using k-means algorithm with the Euclidean distance as a similarity metric. The number of clusters is determined by the Silhouette values of the data in the clusters. For the 49 documents which had better originality levels in the analysis of research abstracts, the process produced five clusters, the information of which is presented in Figure 14.

Once the clusters are formed, the clusters can be analysed to identify the author with the most contribution to the topic. The clusters are labelled with the most frequent bi-gram appearing in the cluster. However, this approach created ambiguity in the cluster topics as there was more than one cluster which had the same label. This problem can be overcome by considering tri-grams or using external knowledge labels to arrive on descriptive labels for the clusters.

Efficient Hybrid Multicast Routing Protocol for Ad-Hoc Wireless Networks
MMAC: A Mobility-Adaptive, Collision-Free MAC Protocol for Wireless Sensor Networks
Energy Efficient, Collision Free Medium Access Control for Wireless Sensor Networks .
Macro/micro-mobility fast handover in hierarchical mobile IPv6
LEAP: Efficient Security Mechanisms for Large-scale Distributed Sensor Networks

Table 3: Documents in Cluster 1 – Sensor Networks

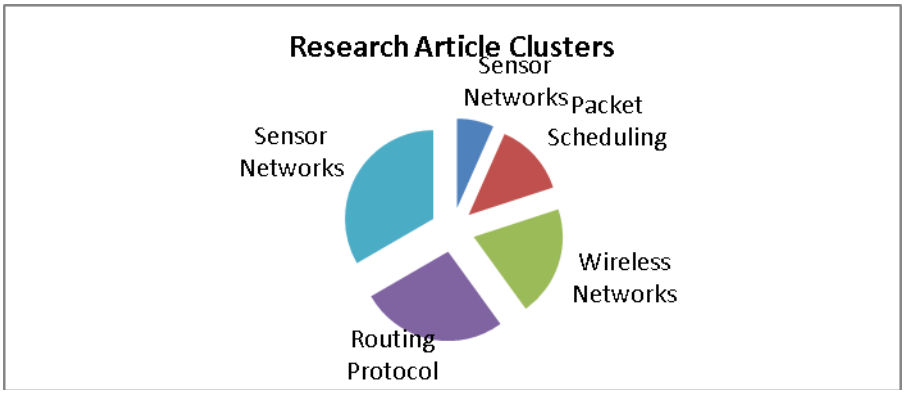


Figure14: Clusters of the research articles which had better originality values

The EigenTrust Algorithm for Reputation Management in P2P Networks
Wireless Sensor Networks and Applications: a Survey
On Computing Farthest Dominated Locations
WIRELESS SENSOR NETWORK SECURITY ANALYSIS
SECURITY IN WIRELESS SENSOR NETWORKS
Wireless Sensor Network Security: A Survey
Decentralized QoS-Aware Check pointing Arrangement in Mobile Grid Computing
PSFQ: A Reliable Transport Protocol for Wireless Sensor Networks
MAC Protocols for Wireless Sensor Networks: A Survey
QoS and energy aware routing for real-time traffic in wireless sensor networks

Table 4: Documents in Cluster 2- Packet Scheduling

A Survey on Energy-Efficient MAC Protocols for Wireless Sensor Networks
An Energy Efficient MAC in Wireless Sensor Networks to Provide Delay Guarantee
A Traffic-Aware Energy Efficient Routing Protocol for Wireless Sensor Networks
An Efficient Route Discovery Mechanism for Mobile Ad Hoc Networks
Evaluation of Energy-Aware QoS Routing Protocol for Ad Hoc Wireless Sensor Networks
Energy Aware Routing for Wireless Sensor Networks
Performance Analysis of Energy-Aware QoS Routing Protocol for Ad Hoc Wireless Sensor Networks
An Energy-Aware QoS Routing Protocol for Wireless Sensor Networks
An Assessment of Case-Based Reasoning for Spam Filtering

Table 5: Documents in Cluster 3- Wireless Networks

Improving TCP Performance over Mobile Wireless Environments using Cross Layer Feedback
Authenticated Multi-Step Nearest Neighbour Search
Minimizing Mobile IP Handoff Latency
A Comparison of Mechanisms for Improving Mobile IP Handoff Latency for End-to-End TCP
Measurement and Characterization of Network Traffic Utilization between Real Network and Simulation Modeling in Heterogeneous Environment
Evaluating the Performance and Accuracy of Network Traffic Management via Simulation Modelling in Heterogeneous Environment

Table 6: Documents in Cluster 4 – Routing Protocol

Fault-tolerant and Efficient Data Propagation in Wireless Sensor Networks using Local, Additional Network Information
Reputation-Based Trust Management System for P2P Networks
Trust and Reputation Model in Peer-to-Peer Networks
Voice over Wireless Local Area Network (WLAN) Performance Analysis
Performance Analysis of Wireless Controller Area Networks with Priority Scheme
Voice over Wireless Local Area Network (WLAN) Performance Analysis
Energy Efficient Resource Management in Virtualized Cloud Data Centers
Collision Free and Energy Efficient MAC protocol for Wireless Networks
ENERGY EFFICIENT ROUTING IN WIRELESS SENSOR NETWORKS
Mining Email Social Networks
Data Transport Reliability in Wireless Sensor Networks —A Survey of Issues and Solutions
HARP - HYBRID AD HOC ROUTING PROTOCOL
Routing Protocols for Mobile Ad-hoc Networks: Current Development and Evaluation
DAPR: A Protocol for Wireless Sensor Networks Utilizing an Application-based Routing Cost
A survey on routing protocols for wireless sensor networks
MAntS-Hoc: A Multi-agent Ant-based System for Routing in Mobile Ad Hoc Networks
A Survey on Congestion Control Mechanisms in High Speed Networks
A Global Contribution Approach to Maintain Fairness in P2P Networks
A Survey on Routing Protocols for Wireless Sensor Networks

Table 7: Documents in Cluster 5- Sensor Networks

Article ID (Cluster no., ID)	Position 1	Position 2	Position 3
1 1	JayantaBiswas	MuktiBarai	S. K. Nandy
1 2	Muneeb Ali	TashfeenSuleman	ZartashAfzalUzmi
1 3	VenkateshRajendran	Katia Obraczka	J.J. GarciaLunaAceves
1 4	M.H. Habaebi		
1 5	Sencun Zhu	SanjeevSetia	SushilJajodia

Table 8: Author Cluster 1

The document clusters thus formed are blindly transformed to author clusters (Table 8). In other words, the authors of documents finding a place in every document cluster are grouped and thus, the author cluster is formed. In table 9, the co-authorship information is indicated by A and the knowledge network information

is indicated by K. The empty cells depict 'no connection' within the knowledge network. Here, we have assumed that the first author is the responsible author for every publication and therefore, the knowledge network is linked from every first author to every other first author.

In future, other intelligent ways of finding author cluster from clusters of research abstracts may be attempted. The author clusters thus formed are verified for duplication i.e. an author may be part of more than one cluster. Such authors are preserved to form the author nucleus. There may be more than one nucleus for every cluster by following the blind transformation of abstract cluster to author cluster. This may be resolved by comparing the position of 'nuclei' authors in their respective publications and similar iterative analysis over other author attributes like citation count (if needed) would attempt to merge every cluster towards a single cluster with one unique nucleus for every cluster.

	111	112	113	121	122	123	131	132	133	141	151	152	153
111		C	C	K			K			K	K		
112	C		C										
113	C	C											
121	K				C	C	K			K	K		
122				C		C							
123				C	C								
131	K			K				C	C	K	K		
132							C		C				
133							C	C					
141	K			K			K				K		
151	K			K			K			K		C	C
152											C		C
153											C	C	

Table 9: Knowledge Network for Cluster 1

6 Conclusion & Future work

The paper discusses a method of evolving a knowledge network in which the quality of a research author is calculates in two dimensions of originality and relevancy within the perspective of a research area. A FCM-based approach is used to measure the originality and the relevancy if a research publication. This method used in this system is different from the other statistical approaches to measure the impact of an author like citation count etc. The system has its limitations in the way the quality measures are affected by the way an abstract is constructed and the dependency of the system on the offline ontology. The correctness and completeness of the ontology used for the construction of FCM will play a major role in the accuracy of the system. Future work will concentrate on the above mentioned limitations of the system. Also, addition of other dimensions to the author quality measure too will be considered. The knowledge network thus evolved will be a reliable recommendation system to find the authors who have had an impact on a specific research area.

References

[ACM, 98]ACM, http://www.acm.org/about/class/1998

[Alzahrani and Saliim, 10]Salha Alzahrani and Naomie Salim, "Fuzzy semantic-Based String Similarity for Extrinsic Plagiarism Detection", Lab Report for PAN at CLEF 10.

[Backstrom, 11] Lars Backstrom, Jure Leskovec: Supervised random walks: predicting and recommending links in social networks. WSDM 11: 635-644

[Banerjee, 07] Banerjee S, Ramanathan K, Gupta A, Clustering short text using Wikipedia. In: Proceedings of the 30th international ACM SIGIR conference on research and development in information retrieval (SIGIR) ,07, pp 787-788.

[Citeseer] http://citeseer.ist.psu.edu/

[Davis et. al., 04] John Davis, Alistair Duke and York Sure, OntoShare – An Ontology-based Knowledge Sharing system for virtual communities of Practice, Journal of Universal Computer Science, vol. 10, no. 3 (04), 262-283

[DBconnect] DBconnect: mining research community on DBLP data, in Proceedings of the 9th WebKDD and 1st SNA-KDD 07 workshop on Web mining and social network analysis, COPYRIGHT ACM, 07.

[DBLife] DBLife, http://dblife.cs.wisc.edu/

[DBLP] DBLP, DBLP Bibliography Homepage, *www.informatik.uni-trier.de/~ley/db/*

[Evelyn et. al., 10] Evelyn Perez Cervantes and Jesus P. Mena-Chalco. 2010. A New Approach to Detect Communities in Multi-weighted Co-authorship Networks. In Proceedings of the 2010 XXIX International Conference of the Chilean Computer Science Society (SCCC '10). IEEE Computer Society, Washington, DC, USA, 131-138. DOI=10.1109/SCCC.2010.31 http://dx.doi.org/10.1109/SCCC.2010.31

[Girvan and Newman, 02] Michelle Girvan and M. E. J. Newman. Community structure in social and biological networks. In Proceedings of the National Academy of Science USA, 99:8271-8276, 02.

[Guan & Wong, 99] Guan, T. & Wong, K.F.; KPS a Web information mining algorithm.. WWW8, 99.

[Gupta et. al, 11] Manish Gupta, Charu C. Aggarwal and Jiawei Han, Finding Top-k Shortest Path Distance Changes in an Evolutionary Network, Advances in Spatial and Temporal Databases, 2011, pp. 130- 148.

[Han, 09] Jiawei Han, Mining Heterogeneous Information Networks by Exploring the Power of Links, J. Gama et al. (Eds.): DS 09, LNAI 5808, pp. 13–30, Springer-Verlag Berlin Heidelberg, 09.

[Han et. al., 05] Han, H., Zha, H., & Giles, C. Name disambiguation in author citations using a k-way spectral clustering method, Proceedings of the 5th ACM/IEEECS Joint Conference on Digital Libraries, 05, pp. 334-343.

[Hastie, 09] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (09). "14.3.12 Hierarchical clustering" (PDF). The Elements of Statistical Learning (2nd ed.). New York: Springer. pp. 5–528

[Pothen et. al., 90] A. Pothen, H. Simon, and K. P. Liou. Partitioning sparse matrices with eigenvectorsof graphs. SIAM J. Matrix Anal. Appl., 11:430–452, 90.

[James Z. Wang et. al., 10] James Z. Wang, Farha Ali and Pradip K. Srimani, "An efficient method to measure the semantic similarity of ontologies", International Journal of Pervasive Computing and Communications,Vol. 6 No. 1, 10,pp. 88-103.

[Jung, 08] Jung, J.J.: Ontology-based Context Synchronization for Ad Hoc Social Collaborations, Knowledge-Based Systems, 21 (7), 2008, 573-580.

[Jung, 09] Jung, J.J.: Contextualized mobile recommendation service based on interactive social network discovered from mobile users, Expert Systems with Applications, 36 (9), 2009, 11950-11956.

[Jung, 10a] Jung, J.J.: Ontology Mapping Composition for Query Transformation on Distributed Environments, Expert Systems with Applications, 37 (12), 2010, 8401-8405.

[Jung, 10b] Jung, J.J.: Integrating Social Networks for Context Fusion in Mobile Service Platforms, Journal of Universal Computer Science, 16 (15), 2010, 2099-2110.

[Jung, 10c] Jung, J.J.: Reusing Ontology Mappings for Query Segmentation and Routing in Semantic Peer-to-Peer Environment, Information Sciences, 180 (17), 2010, 3248-3257.

[Jung, 11a] Jung, J.J.: Service Chain-based Business Alliance Formation in Service-oriented Architecture, Expert Systems with Applications, 38 (3), 2011, 2206-2211.

[Jung, 11b] Jung, J.J.: Boosting Social Collaborations Based on Contextual Synchronization: An Empirical Study, Expert Systems with Applications, 38 (5), 2011, 4809-4815.

[Jung, 12a] Jung, J.J.: Attribute selection-based recommendation framework for short-head user group: An empirical study by MovieLens and IMDB, Expert Systems with Applications, 39 (4), 2012, 4049-4054.

[Jung, 12b] Jung, J.J.: Evolutionary Approach for Semantic-based Query Sampling in Large-scale Information Sources, Information Sciences, 182 (1), 2012, 30-39.

[Jung, 12c] Jung, J.J.: Discovering Community of Lingual Practice for Matching Multilingual Tags from Folksonomies, Computer Journal, 55 (3), 2012, 337-346.

[Kanungo et. al., 02] Kanungo. T, Mount, D. M., Netanyahu, N. S,Piatko, C. D.; Silverman, R.; Wu, A. Y. "An efficient k-means clustering algorithm: Analysis and implementation". IEEE Trans. Pattern Analysis and Machine Intelligence 24: 881–892, 02.

[Kernighan, 70] B. W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. Bell System Technical Journal, 49:291–307, 70.

[Kosko, 86] Kosko B."Fuzzy Cognitive Maps", Int. Journal of Man-Machine Studies, Vol. 24, 86, pp. 65-75.

[Lee and Das, 11] Wei-Nchih Lee and Amar K. Das, "Evaluating Ontology Based Semantic Similarity Measures for Treatment Comparison",Proceedings of AMIA 11, pp 110

[Libra], http://libra.msra.cn/

[Mahalakshmi et. al., 09] G. S. Mahalakshmi, S. Sendhilkumar, AlaguIrulappan, PreethamMirinda (09), Ontology Based Relevance Analysis for Automatic Reference Tracking, International Journal of Computer Applications in Technology (IJCAT): Special Issue on Computer Applications in Knowledge-Based Systems, ISSN 0952-8091, Vol. 35, Nos. 2/3/4, pp. 165-173.

[Makoto et. al., 04] Makoto Takeya, Hitoshi Sasaki , KeizoNagaoka and Nobuyoshi Vonezawa,"A Performance scoring method based on quantitative comparison ofConcept maps by a teacher and students", Concept maps: theory, methodology, technology , Proc. Of the first int. Conference on concept mapping, A. J. Cañas, J. D. Novak, F. M. González, eds, Pamplona, Spain 04

[Manjula, 07] Manjula Dissanayake and SimaanM.AbouRizk, "Qualitative simulation of construction performance using fuzzy cognitive maps",IEEE, Proceedings of the 07 Winter Simulation Conference S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.,July 07,pp:2134 -2140.

[Martin and Carl, 08] Martin Rosvall and Carl T. Bergstrom, Maps of random walks on complex networks reveal community structure PNAS 08 105 (4) 1118-1123

[Mihalcea et. al., 06] R. Mihalcea, C. Corley, and C. Strapparava. 06. Corpus-based and knowledge-based measures of text semantic similarity. In Proceedings of AAAI-06.

[Moore, 01] Andrew W. Moore, "K-means and Hierarchical clustering",November 16th, 01.http://www.autonlab.org/ tutorials/kmeans09.pdf.

[Nascimento et. al., 03] Mario A. Nascimento, J¨org Sander, and Jeffrey Pound. Analysis of sigmod's co-authorship graph. SIGMOD Record, 32(2):57–58, 03.

[Nevin, 96] Nevin Heintze. Scalable Document Fingerprinting.Proceedings of the Second USENIX Workshop on Electronic Commerce, Oakland, California,November 18-21, 96.

[Newman, 03] M. E. J. Newman. The structure and function of complex networks. SIAM Review, 45(2):167–256, 03.

[Noshir and Peter, 02] Noshir S. Contractor, Peter R. Monge, Theories of Communication Networks, Oxford University Press, 02.

[Osmar et. al., 07] Osmar R. Za¨ıane, Jiyang Chen, and Randy Goebel, Mining Research Communities in Bibliographical Data; In Proceedings of the 9th WebKDD and 1st SNA-KDD 07 workshop on Web mining and social network analysis (WebKDD/SNA-KDD '07). ACM, New York, NY, USA,07, 74-81.

[Pena, 05] Pena A. Sossa H. and Gutiérrez, F," Knowledge and Reasoning Supported by Cognitive Maps", Proceedings of Mexican International Conference on Artificial Intelligence 05 (MICAI'05), Lecture Notes in Artificial Intelligence, Vol.3789. Springer, Monterrey, Mexico, November, (05)

[PROTÉGÉ, 11] http://protege.stanford.edu

[Rajesh et. al., 08] Rajesh Thiagarajan, Geetha Manjunath, and Markus Stumptner, "Computing Semantic Similarity Using Ontologies", International Semantic Web Conference (ISWC), Karlsruhe, Germany, 08.

[Rekha and Sasikumar, 10] Rekha Ramesh and Sasikumar M. Use of Ontology in Addressing the issues of Question Similarity in Distributed Question Bank. ICWET 10, Mumbai.

[Rosy et. al., 10] Rosy Madaan, Ashutosh Dixit, A.K. Sharma, Komal Kumar Bhatia; A Framework for Incremental Hidden Web Crawler, (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 03, 10, pp.753-758

[Roy et. al., 09] Roy; C. K.; Cordy, J. R.; Koschke,R. Comparison and Evaluation of Code Clone Detection Techniques and Tools: A Qualitative Approach. Science of Computer Programming, 09, v.74, n.7.

[Salmeron, 10] Jose Salmeron, Modeling grey uncertainty with Fuzzy Grey Cognitive Maps, Expert Systems with Applications 37 (10) 7581–7588.

[Salton and McGill, 83] Salton, G. and McGill, M. J., "Introduction to Modern Information Retrieval", McGraw-Hill, New York, NY, 83.

[Sanchez, 11] David Sanchez, Montserrat Batet and David Isern, "Ontology-based information content computation", ScienceDirect, Elsevier, Knowledge-Based Systems 24, pp 297–303, 11.

[Smeaton et. al., 02] A. F. Smeaton, G. Keogh, C. Gurrin, K. McDonald, and T. Sodring. 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):39–43, 02.

[Varlamis et. al., 10] Varlamis, I.; Eirinaki, M.; Louta, M.; A Study on Social Network Metrics and Their Application in Trust Networks, Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference. 2010, pp 168 - 175

[Vladimir and Asle, 03] Vladimir Oleshchuk and Asle Pedersen, "Ontology Based Semantic Similarity comparison of Documents", Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA'03), IEEE Computer society, pp 735-738, 03.

[Wasserman and Faust, 94] S. Wasserman and K. Faust. Social network analysis: Methods and applications, Cambridge University Press, 94.

[Zhen and Jiang, 11] Zhen, L., Jiang, Z.: Hy-SN: Hyper-graph based semantic network. Knowledge-Based Systems 23(8), 809–816 (10)