

## Tracing best semantic path using co-citation proximity analysis

**A. Balaji**

KCG College of Engineering, India

**S. Sendhilkumar**

**G. S. Mahalakshmi**

Anna University, India

ORIGINAL

### Abstract

**Objective.** The objective of this work is to find the best semantic path of research papers that matches the given research publication. This paper elaborates on the finding of continuous research progress analysis from semantic perspective.

**Methodology.** Previous work reported the work progress analysis and the integrated as well as optimized approaches to find the progressive research citations which had carried forward the essence of the base reference paper. In this work, we propose to identify the most useful research papers which are semantically closer to the research context, and lie in the citation path of the base paper.

**Result.** Our data set is generated for the popular paper of Hirsch published in 2005, in which, h-index is proposed. The paper has 5299 direct citations till date and the results of the proposed approach indicate very promising findings in view of measuring scientific research progress.

**Conclusion.** The inference reveals a couple of research papers connected as a path among the citation thread, which have significantly progressed the idea of base research paper into a more elaborate yet related context of research..

### Keywords

*Citation mining; Co-citation mining; Power graph; Research progress – Language models; Semantic analysis.*

## Ubicando el mejor camino semántico usando el análisis de proximidad de la cocitación

### Resumen

**Objetivo.** El objetivo de este trabajo es encontrar la mejor trayectoria semántica de los trabajos de investigación que coincida con la publicación de la investigación en cuestión. En este trabajo se profundiza en los hallazgos del análisis del progreso de la investigación continua desde una perspectiva semántica.

**Metodología.** En trabajos anteriores se informó sobre el análisis del progreso del trabajo y los enfoques integrados y optimizados para encontrar las citas de investigación progresivas que habían llevado adelante la esencia del documento de referencia básico. En este trabajo se propone identificar los trabajos de investigación más útiles y semánticamente más cercanos al contexto de la investigación, y que se encuentran en la trayectoria de citación del trabajo de base.

**Resultado.** Nuestro conjunto de datos se genera para el popular artículo de Hirsch publicado en 2005, en el que se propone el h-index. El documento tiene 5299 citas directas hasta la fecha y los resultados del enfoque propuesto indican hallazgos muy prometedores con vistas a medir el progreso de la investigación científica.

**Conclusión.** La inferencia revela un par de trabajos de investigación conectados como un camino entre el hilo de la cita, que han progresado significativamente la idea de un trabajo de investigación de base en un contexto de investigación más elaborado pero relacionado.

### Palabras clave

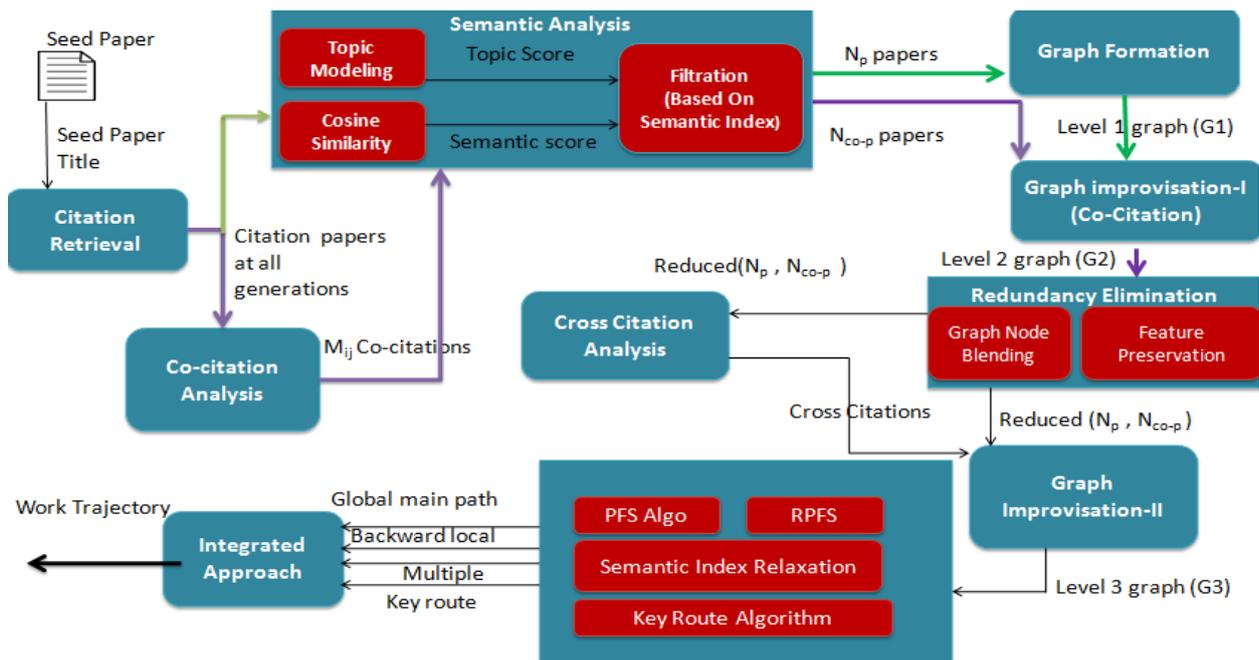
*Análisis semántico; Gráfico de potencia; Minería de citas; Minería de cocitación; Progreso de la investigación - Modelos de lenguaje.*

# 1 Introduction

Technological research progress trajectory determines the way a research idea has progressed so far. Earlier works on analyzing work progress (Hummon & Doreian, 1989; Hummon et. al., 1990; Hummon & Carley, 1993; Carley et al. 1993; Batagelj, 2003) tend to address solutions to measure the productivity of researcher. However, there lies the other side of the coin; the research idea would have progressed in multiple directions most of which need not be related to the semantic core of the research context addressed in the base paper. Therefore, the need for measuring the progressive research path with respect to semantic context (Moore et. al. 2006; Verspagen, 2007; Mina et al., 2007; Carlero-Medina & Noyons, 2008; Lucio-Arias & Leydesdorff, 2008; Harris et al. 2009; Lu et al. 2012) has arisen. This paper measures the most similar semantic path in the citation thread using optimized language models.

In previous work (Balaji et al, 2016), we employed cosine similarity to analyse and track the semantically significant paths of research progress which is further used to trace the developing trajectory of a research paper. The integrated optimal path analysis approach involved enables to view the true path in which the research idea said in the seed research paper has progressed. For this, we used semantic power graph representation of citation trajectories. The benefits of using semantic power graphs has already been discussed in (G. S. Mahalakshmi, & S. Sendhilkumar, 2013).

# 2 Analysing the research progress trajectories



- N - Number of citation papers at all generations (up to 2013)
- N<sub>p</sub> - Papers after citation semantic elimination
- N<sub>co-p</sub> - Papers after co-citation semantic elimination
- M<sub>ij</sub>- Number of co-citations in N papers
- Reduced (N<sub>p</sub> ,N<sub>co-p</sub>) - Number of papers after redundant elimination
- PFS - Priority First Search
- RPFS - Reverse Priority First Search

Figure 1 - Analysing the Research Progress Trajectory ( )

The idea (Balaji et al, 2016) is extended using co-citation proximity analysis (Gibbs & Joran, 2009) for finding semantic similarity. During this process, the text documents are subjected to 3 step process (refer figure 2) which include:

- Extraction of co-citation by parsing the document
- Applying CPA to group citations on similar discussions
- Extracting the documents by analyzing the reference section

The result of CPA is then taken as input with the seed document to the content analyser module. Pre-processing is performed on it in which TF-IDF vectors are generated together with Log frequency weighting. The vectors are then augmented with semantic information. The Wordnet corpus (<http://www.nltk.org/howto/wordnet.html>) in NLTK is used for extracting words having similar meaning (synonym and hypernym relationships). Term re-weighting is performed to augment the weights of the vectors of all the technical papers in the dataset. These scores are used to construct the ESA model stored in pickle file format. Finally, Cosine similarity between the seed document and the dataset is calculated to recommend the 'k' most similar documents from the dataset along with their similarity score sorted in the order of similarity.

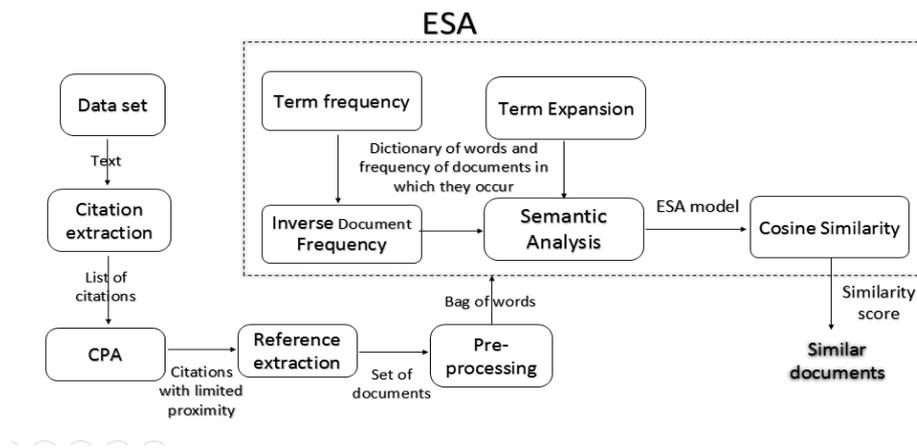


Figure 2 - Block diagram of proposed method for finding similar research papers using semantic and co-citation proximity analysis

Table 1 - Seed Publication

Paper Title
Hirsch, J.E. "An Index to Quantify an Individual's Scientific Research Output" 2005 PNAS

### 3 Co- citation analysis

If A, C are both cited by B, they may be said to be related to one another, even though they don't directly reference each other. Thus A and C are co-citations of B (refer Fig 3). Therefore, we proceed to find the co-citations of papers and trace all possible co-citations. The co-citations listed together in the same context are retained as such. Ex. [a,b,c]. Here a, b, c are co-citations grouped together and they need not be subjected to proximity analysis. The citations are assigned to their corresponding items in the bibliography. The other citations that occur at different positions within the document are subjected to proximity analysis where the proximity between co-citations are decided by the following heuristics.

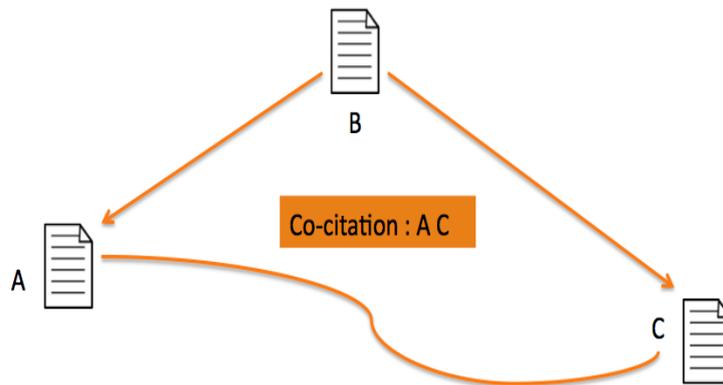


Figure 3 - Co-citation Illustration (Balaji et al, 2016)

The underlying assumption is that the closer the citations are to each other the more likely it is that they are related. Based on this proximity analysis, the CPI (Citation Proximity Index) is calculated (Gibbs & Joran, 2006). CPI is calculated as  $1/2^n$  where  $n$  is the number of levels between two citations. If for example two citations are given in the same sentence the probability that they are very similar is higher (CPI = 1), than if they are in the same paragraph (CPI = 1/2), and if they are in the same Section CPI = 1/4 and so on. Giving more weightage to citations within same page is found to degrade the results in some cases. Hence this level is eliminated. Thus we look at 4 levels:

1. Co-citations grouped together (ex. [a,b,c]) – This is handled differently as already discussed
2. Citations in same lines
3. Citations in same paragraphs
4. Citations in same sections.

The extracted citations are mapped to the corresponding items in bibliography. For this the “Reference” section is parsed and corresponding items are mapped and title of the paper is extracted. Since different journals follow different format for bibliography, this step requires framing proper rules to extract the title correctly. Here we also assume that all the documents cited in a paper lie within the dataset.

## 4 Integrated approach

The citation graphs are evolved with a systematic inclusion of co-citations and cross-citations (Balaji et al, 2016). The graph obtained is referred as G3 (figure 4 – 6).

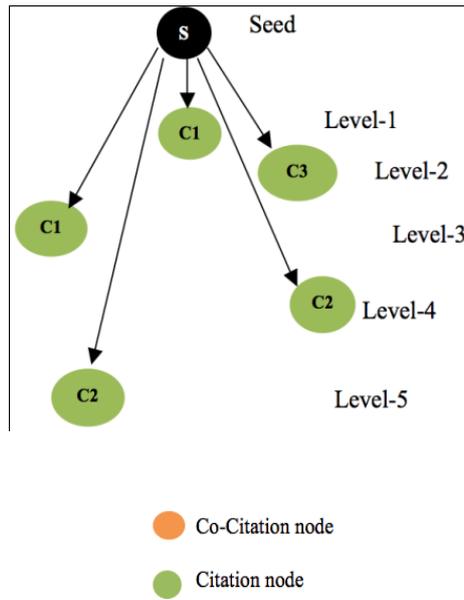


Figure 4 - DAG – G1 (Balaji et al, 2016)

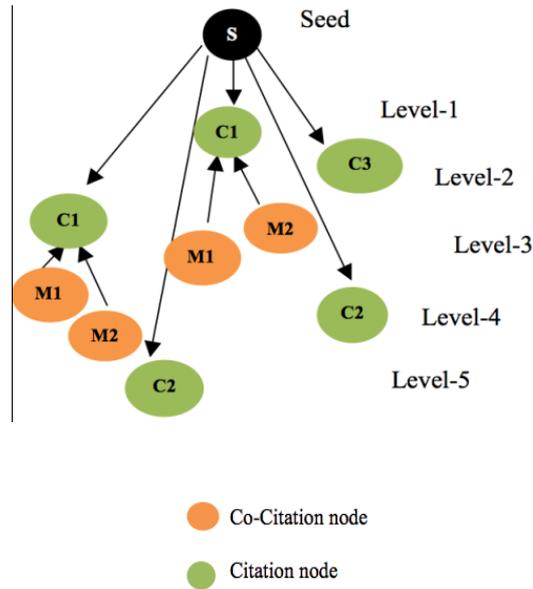


Figure 5 - G2 (G1 with co-citations) (Balaji et al, 2016)

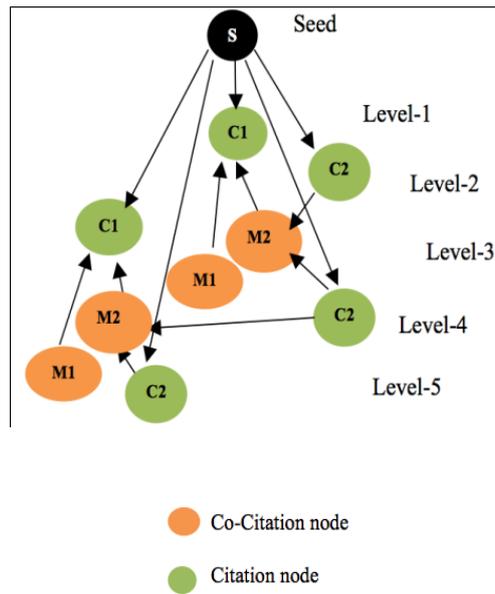


Figure 6 - G3 (G2 with cross-citations) (Balaji et al, 2016)

We use the following four approaches (G. S. Mahalakshmi, & S. Sendhilkumar, 2013) to track the work progress of a research article, namely, (i) Global Main Path, (ii) Backward Local Main Path, (iii) Multiple Main Path and (iv) Key Route Main Path. Using all these approaches the significant set of research progress paths are obtained. The resulting paths are ranked semantically after analyzing various features and finally, the best semantic path of research progress is obtained.

The global main path searches forward from the source to the sinks and is traced using Priority First Search (G. S. Mahalakshmi, & S. Sendhilkumar, 2013; Balaji et al, 2016). The Backward Local Main Path is found using Reverse Priority First Search Algorithm (G. S. Mahalakshmi, & S. Sendhilkumar, 2013; Balaji et al, 2016). In this paper, we propose approaches for main path analysis using semantic approaches.

#### 4.1 Multiple main path analysis

Multiple Main Path approach uses Semantic Index Relaxation (SRI). For a discipline that has many subfields, one may want to also discover the important paths at the next level. In SRI, one basically relaxes the search constraint. The constraints are relaxed by bringing in the next longest path. We make the average semantic score lesser so that more papers get included in the path. Thus we obtain various paths that are less important than the significant path in this approach.

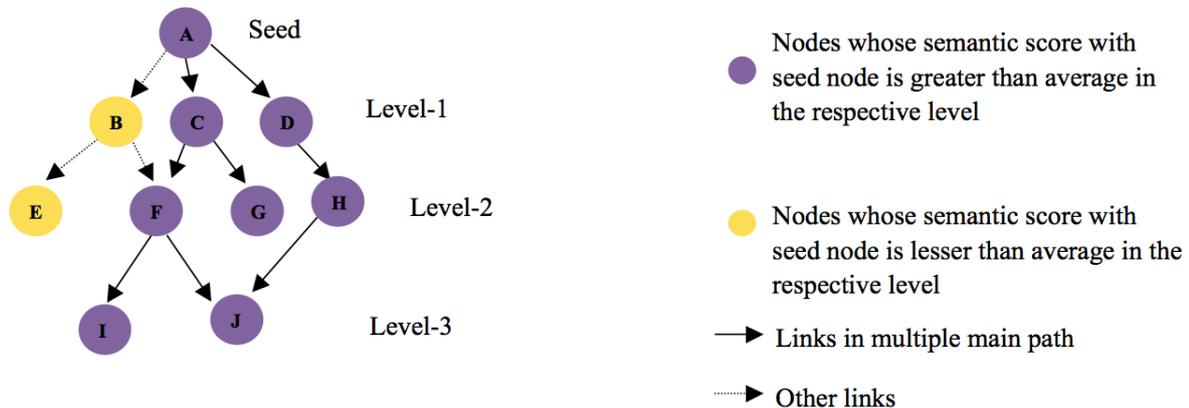


Figure 7 - Sample Graph using Multiple Main Path

Consider a sample graph (refer Fig 7) which contains node A as seed node, nodes B, C and D in level-1(depth=1), nodes E, F, G and H in level-2(depth=2), and nodes I and J in level-3(depth=3). At every level we find the average semantic score of all nodes. By relaxing the threshold value we consider all the nodes that are higher than the threshold value. In level-1, node C and D have higher semantic score than the average threshold value. In level-2, the nodes F, G and H and in level-3 nodes I and J have higher semantic scores than the threshold value. The multiple paths obtained are A-C-G, A-C-F-I, A-C-F-J, A-D-H-J.

### 4.2 Key route main path analysis

Using the Key Route Algorithm we take the most significant link and begin a search from the key rather than from the source or sink. This key is a research article with a very close semantic score compared to the seed. We call this the key-route search. It guarantees that this key route is included in the main path. The key here, traces the path through the next nodes that have a high semantic score greater than the average semantic score of that particular level and then to the next and so on until the leaf node is reached.

Consider a sample graph (refer Fig 8) which contains node A as seed node, nodes B, C and D in level-1(depth=1), nodes E, F, G and H in level-2(depth=2), and nodes I and J in level-3(depth=3). Take node C as the key. Traversing through node C to next nodes with high semantic scores greater than the average semantic score of every level we result in the path C-F-J.

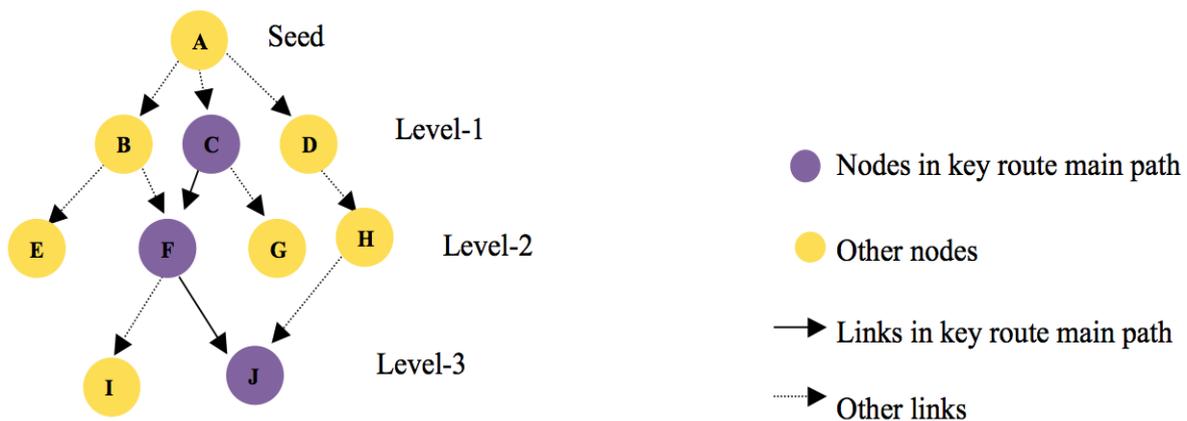


Figure 8 - Sample Graph using Key Route Main Path

### 4.3 Best semantic path

From the graph (G3), we arrange the nodes of the graph level-wise based on the depth value from the seed node. In this approach, the cosine similarity value between the seed and the respective nodes in all levels are determined. For all the levels in the graph, we find the maximum semantic score at each level separately. The node with the maximum semantic score at each level with respect to the seed is considered to be the best node at that particular level. From the paths already determined, we consider the paths that pass through these best nodes as the best semantic match path.

Consider a sample graph (refer Fig 9), which contains node A as seed node, nodes B, C and D in level-1(depth=1). The nodes E, F, G and H in level-2(depth=2), and nodes I and J in level-3(depth=3). Suppose the node C has the highest semantic score with respect to node A in level-1, then node C is said to be the best node in level-1. Similarly we consider the maximum semantic score at levels 2 and 3 and the best nodes are F and J respectively. Then the path A-C-F-J is said to be the best semantic match path that traverses through all best nodes.

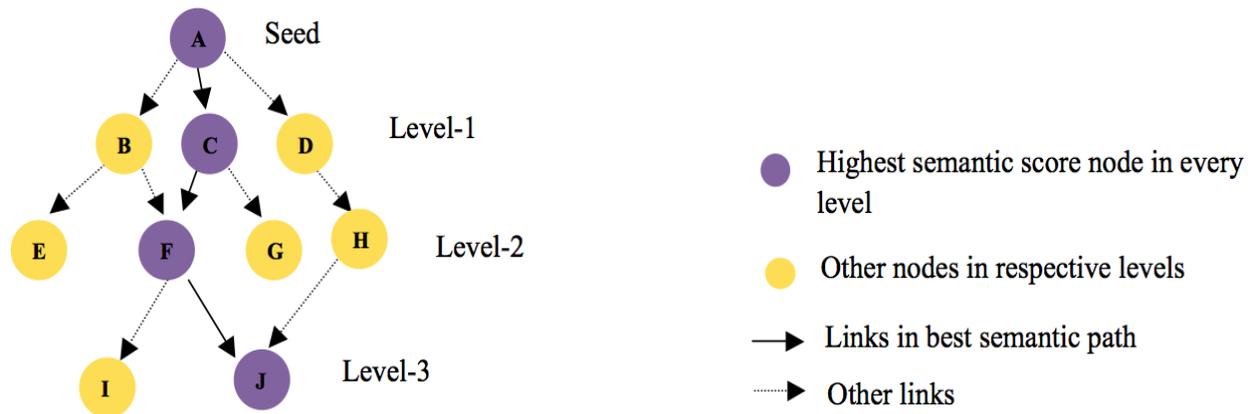


Figure 9 - Sample Graph using Best Semantic Path

## 5 Ranking and optimisation of semantic research paths

The paths obtained in integrated approach are ranked based on the following criteria:

- Length of Path: More length indicates more frequently the works are carried out.
- Popularity: Significant idea attracts more works, hence good work progress (through citation count). The citation counts for all the nodes in a path are added and the average is taken as popularity score.
- Relevancy Score: Average of similarity scores of nodes in a path with respect to the seed paper, gives relevancy score which indicates more relevant work progress.

Two versions of optimization approaches are already discussed in the literature (Balaji et al., 2016). In this paper, we extend the optimization further to incorporate CPA based semantic similarity for filtration.

### Optimization approach – III: Using average semantic score and popularity based filtration

From the graph (G3), we arrange the nodes of the graph level wise based on the depth value from the seed node. In this approach, the cosine similarity value between the seed node and the respective nodes in the first level is alone determined. Then we find the average of semantic scores of all nodes at first level alone and the first level's average semantic score value is taken as threshold and the nodes whose semantic score is greater than the average are only retained in first level. For the set of retained nodes alone in first level, we consider their corresponding second level nodes and for those nodes alone, popularity count is considered for all the nodes in second level and the nodes whose popularity count value is greater than the average popularity score of that level, are alone retained. For these retained nodes in second level we proceed for third level and the filtration is done for these third level nodes using popularity count. This process is continued for all levels with semantic filtration at first level and popularity filtration at all the other levels. From this approach we obtain paths that survive this kind of double attribute filtration (refer Fig 10 & algorithm 1).

**Algorithm 1:** Optimization Approach - III

**Input:** G3 paths **Output:** Optimized path III

- **Step 1:** The similarity between all papers of the first level is determined using cosine similarity algorithm.
- **Step 2:** The average semantic score for the first level alone is calculated.
- **Step 3:** Retain the nodes whose semantic score are above the first level's average.
- **Step 4:** Find the average of the popularity for the second level.
- **Step 5:** Retain the nodes whose semantic score are above the second level's average.
- **Step 6:** Repeat steps 4 and 5 for the remaining levels.
- **Step 7:** Trace the path for the retained nodes.

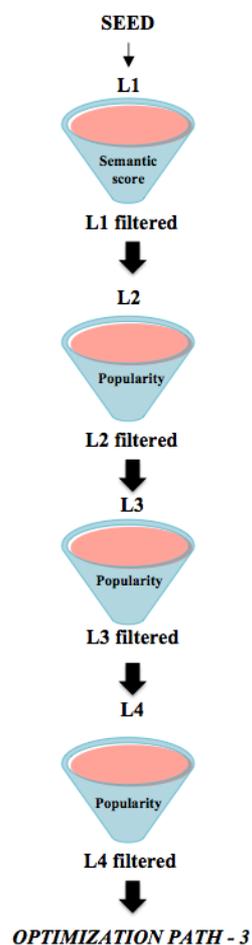


Figure 10 - Optimization Approach III

Consider a sample graph (refer Fig 11), which contains node A as seed node, nodes B and C in level-1(depth=1), nodes D, E, F in level-2 (depth=2), and nodes G and H in level-3(depth=3).In level-1, the average semantic score of B and C are taken. The node C has higher semantic score than level-1's average so node C is retained whereas the node B is filtered. In level-2 average popularity score of E and F are only considered and E, having higher value than threshold is retained. Node D is ignored since its parent node B got filtered in first level itself. In level-3, nodes G and H are considered since their parent node E alone is retained in that level. Between G and H, H passes the popularity filtration phase to give the optimum path as A-C-E-H.

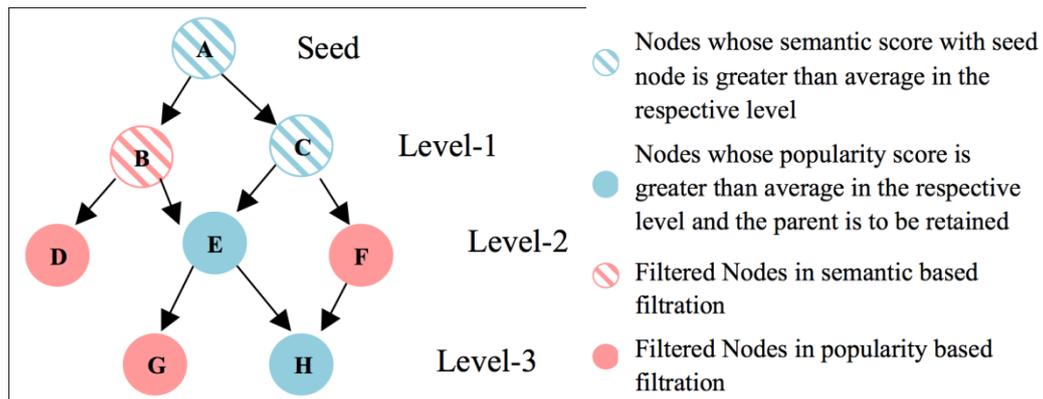


Figure 11 - Sample Graph using Optimization Approach III

## 6 Results and discussion

### 6.1 Data set

The seed paper for which we find the work progress trajectory is Hirsch, J.E. “An Index to Quantify an Individual’s Scientific Research Output”, 2005, PNAS which has 2706 citations at the first generation. Table 3 shows the total citations and the available citations for the first and second generations.

### 6.2 Path analysis

Refer Table 2 for the input and output path details for Global Main Path, Backward Local Main Path and Key Route Main Path Approach.

Table 2 - Path details for Global, Backward Local and Key Route Main Path Analysis

Paths	Global Main Path	Backward Local Main Path	Key Route Main Path
Input	Seed Paper, Leaf Node	Seed Paper, Leaf Node	Key Node, Any Node
Output	Paths from Seed Paper to Leaf Node	Paths from Leaf Node to Seed Paper	Paths from Key Node to Any Node

- Best Semantic Path: The best semantic path is traced through the nodes that have the maximum semantic score at every level (refer Fig 12). The best semantic path obtained is 000-209-102-141-168.
- Most Significant Path: The nodes that are common in all the three-optimization approaches are considered as the most significant works. The nodes that are the most significant papers in the trajectory are:  
011,013,016,019,024,025,027,048,050,053,054,055,056,080,087,095,096,097,098,099,100,102,103,105,107,111,118,122,125,126,127,135,136,137,140,143,146,147,148,150,153,174,194,198-206-209-215

### 6.3 Analyzing optimization approaches

We have already proposed optimization approaches I and II in (Balaji et al., 2016). In optimization II (Balaji et al., 2016) we evolved a path from 2005 to 2013 that is a narrow and ideal path as the features of the nodes such as co-citee and levels are ideally close. This infers that these nodes are good papers and many people are following the work. For optimization III, we consider two factors for evaluating the paths: one is the semantic score and the other is popularity. We get a path that is more of popularity based than semantic score based. The nodes in the path have very close semantic scores ranging from 0.3 to 0.4 and have good popularity scores. Consider node 131 in path S1 (Refer table 4). This node occurs in many levels but has not been co-cited and has high popularity. Hence node 131 is a good paper and many people are following the work although there is no parallel work being done. So this approach results in a path that combines nodes that are closely related to each other based on work progress. Refer figures 10-11 for additional information.

**Table 3 - Dimensionality Reduction and Evaluation of Optimization Results**

Approaches	No. Of Paths Before Optimization	Start,End	No. Of Paths After Optimization	Progressive Path
Optimization - I	17103	000,056	26 forward (P)	P2
	906	056,209	2 backward (Q)	Q1
Optimization - II	1439	000,167	1 (R)	R1
Optimization - III	2	000,180	1 (S)	S1

Table 3 shows the dimensionality reduction in the number of paths due to various optimization approaches used. Refer table 4 for the paths after optimization and ranking based on relevancy and popularity. For Optimization I, we have 26 forward and 2 backward paths. In the 26 forward paths, since the path P2 ranks second based on relevancy and popularity (Refer P in table 4), it is considered as the progressive path. Between the 2 backward paths, path Q1 ranks first in both popularity and relevancy (Refer Q in table 4) and is taken as the progressive path. For Optimization II and III there is only one path available namely R1 and S1 (Refer R and S in table 4), so those paths are taken as the progressive paths. Refer Table 5 that shows an analysis between the optimization approaches. We take meta-information such as the semantic score, popularity and year of publication of every node and also check if the node occurs at any level or is a co-citee or not. Based on this information we validate our results.

**Table 4 - Paths after using Optimization Approaches**

Reference	Path No.	Path Description	Ranking Based on Relevancy	Ranking Based on Popularity
P	1	000-198-209-102-056	1	17
	2	000-198-209-102-097-056	2	2
	3	000-209-102-056	3	10
	4	000-198-209-102-107-056	4	23
	5	000-198-209-102-107-097-056	5	26
	6	000-198-209-107-056	6	16
	7	000-209-102-097-056	7	20
	8	000-209-102-107-056	8	19
	9	000-198-209-107-097-056	9	22
	10	000-209-102-107-097-056	10	25
	11	000-056	11	1
	12	000-209-107-056	12	9
	13	000-095-056	13	4
	14	000-206-095-056	14	13
	15	000-097-056	15	3
	16	000-206-096-056	16	12
	17	000-209-107-097-056	17	18
	18	000-096-056	18	2
	19	000-206-096-097-056	19	21

Reference	Path No.	Path Description	Ranking Based on Relevancy	Ranking Based on Popularity
	20	000-096-097-056	20	11
	21	000-194-095-056	21	8
	22	000-194-097-056	22	7
	23	000-194-096-056	23	5
	24	000-194-107-056	24	6
	25	000-194-096-097-056	25	14
	26	000-194-107-097-056	26	15
Q	1	056-107-209	1	1
	2	056-097-107-209	2	2
R	1	000-174-167	1	1
S	1	000-118-131-180	1	1

Table 5 - Analysis of Path Results using Meta Information

Approach	Nodes	Year	Level	Co-Citee	Relevancy	Popularity
OPT - I Backward (Q2)	056	2011	Y	N	0.05020	0
	097	2011	Y	N	0.04289	7
	107	2010	Y	Y	0.04173	8
	209	2006	N	Y	0.05049	218
OPT - I Backward (Q1)	056	2011	Y	N	0.05020	0
	107	2010	Y	Y	0.04173	8
	209	2006	N	Y	0.05049	218
OPT - II (R1)	000	2005	Y	Y	0.04548	3001
	174	2010	Y	Y	0.03950	100
	167	2013	Y	Y	0.04928	0
OPT - III (S1)	000	2005	Y	Y	0.04548	3001
	118	2009	Y	Y	0.03873	43
	131	2009	Y	N	0.03374	54
	180	2011	Y	N	0.04689	7

## 9 Conclusion

In this paper, we have proposed to track the work progress of a research publication across timeline using semantic similarity as well as co-citation proximity based approaches. The semantically relevant citation network is mined for various graph analysis approaches to arrive at the significant research progress trajectory. In particular, we have tracked the path of work progress using the integrated approach methods which involved Multiple Main Path, Key Route Main Path and Best Semantic Path. In addition, we have also obtained deliverables related to most significant research progress trajectory.

However, the bibliographic corpus is tend to evolve over time, and therefore, applying evolutionary algorithms to pick the best similarity measure from multi-perspective based similarity measures would associate more meaning to the problem under context. In addition, besides using semantic score and popularity based features, citation context relevancy, citation classification and novelty of research articles shall also be taken into account. The research papers shall be classified into application and theoretic categories which shall provide additional insight into the semantic paths obtained.

## References

- Balaji A, S. Sendhilkumar, & G.S. Mahalakshmi (2016), Progressive Path Analysis using Optimized Discrete and Continuous Average Semantic Filters, *Aust. J. Basic & Appl. Sci.* (Vol.10(2) pp.224-233).
- Batagelj, V. (2003). Efficient algorithms for citation network analysis. University of Ljubljana, Institute of Mathematics, physics and Mechanics Department of Theoretical Computer Science, Preprint Series (Vol. 41, p. 897). doi: <https://doi.org/arXiv:cs/0309023>.
- Calero-Medina, C., & Noyons, E. (2008). Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field. *Journal of Informetrics*, (Vol. 2(4), pp. 272-279). doi: <https://doi.org/10.1016/j.joi.2008.09.005>.
- Carley, K. M., Hummon, N. P., & Harty, M. (1993). Scientific Influence An Analysis of the Main Path Structure in the Journal of Conflict Resolution, *Science Communication* (Vol. 14(4), pp. 417-447). doi: <https://doi.org/10.1177/107554709301400406>.
- Gipp, B., & Beel, J. (2009). Identifying related documents for research paper recommender by CPA and COA. In *International Conference on Education and Information Technology (ICEIT'09)*, Lecture Notes in Engineering and Computer Science (Vol. 1, pp. 636-639).
- Gipps, Bela, & Jöran Beel (2009). Citation Proximity Analysis (CPA)-A new approach for identifying related work based on Co-Citation Analysis. *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*. ( Vol. 2). Rio de Janeiro (Brazil): International Society for Scientometrics and Informetrics.
- Harris, J. K., Luke, D. A., Zuckerman, R. B., & Shelton, S. C. (2009). Forty years of secondhand smoke research: the gap between discovery and delivery. *American journal of preventive medicine* (Vol. 36(6), pp. 538-548). doi: <https://doi.org/10.1016/j.amepre.2009.01.039>.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America* (Vol. 102(46), pp. 16569-16572). doi: <https://doi.org/10.1073/pnas.0507655102>.
- Hummon, N.P., & Carley, K. (1993). Social networks as normal science *Social Networks* (Vol. 15(1) pp.71-106). doi: [https://doi.org/10.1016/0378-8733\(93\)90022-D](https://doi.org/10.1016/0378-8733(93)90022-D).
- Hummon, N.P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks* (Vol. 11(1), pp. 39-63). doi: [https://doi.org/10.1016/0378-8733\(89\)90017-8](https://doi.org/10.1016/0378-8733(89)90017-8).
- Hummon, N.P., Doreian, P., & Freeman, L.C. (1990). Analyzing the structure of the centrality-productivity literature created between 1948 and 1979. *Science Communication* (Vol. 11(4), pp.459-480). doi: <https://doi.org/10.1177/107554709001100405>.
- Lu, L. Y. Y., Lan, Y. L., & Liu, J. S. (2012, June). A novel approach for exploring technological development trajectories. In *Management of Innovation and Technology (ICMIT), 2012 IEEE International Conference on* (pp. 504-509). IEEE. doi: <https://doi.org/10.1109/ICMIT.2012.6225857>.
- Lucio-Arias, D., & Leydesdorff, L. (2008). Main-path analysis and path-dependent transitions in HistCite™-based historiograms. *Journal of the American Society for Information Science and Technology* (Vol. 59(12), pp.1948-1962). doi: <https://doi.org/10.1002/asi.20903>.
- Mahalakshmi G. S. & Sendhilkumar S. (2013) Optimizing Research Progress Trajectories with Semantic Power Graphs, Chapter in [Pattern Recognition and Machine Intelligence](#), Volume 8251 of the series [Lecture Notes in Computer Science](#) (pp 708-713).
- Mina, A., Ramlogan, R., Tampubolon, G., & Metcalfe, J. S. (2007). Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research policy* (Vol. 36(5), pp.789-806). doi: <https://doi.org/10.1016/j.respol.2006.12.007>.
- Moore, S., Haines, V., Hawe, P., & Shiell, A. (2006). Lost in translation: a genealogy of the "social capital" concept in public health. *Journal of epidemiology and community health* (Vol. 60(8), pp.729-734). doi: <https://doi.org/10.1136/jech.2005.041848>.
- Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems* (Vol. 10(01), pp.93-115). doi: <https://doi.org/10.1142/S0219525907000945>.

## Author data

### A. Balaji

Department of Computer Science and Engineering, KCG College of Engineering, Chennai 600097, India.

[bala465ji@gmail.com](mailto:bala465ji@gmail.com)

### S. Sendhilkumar

Department of Information Science and Technology, Anna University, Chennai 600025, India.

[thamaraikumar@annauniv.edu](mailto:thamaraikumar@annauniv.edu)

### G. S. Mahalakshmi

Department of Computer Science and Engineering, Anna University, Chennai 600025, India.

[thamizhini@gmail.com](mailto:thamizhini@gmail.com)

**Received - Recibido:** 2016-06-14

**Accepted - Aceptado:** 2016-11-17



This work is licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by the [University Library System](#) of the [University of Pittsburgh](#) as part of its [D-Scribe Digital Publishing Program](#) and is cosponsored by the [University of Pittsburgh Press](#).