DOI: http://dx.doi.org/10.15446/ing.investig.vxxnx.xxxx

SAP Algorithm for Citation Analysis: An improvement to Tree of Science

Algoritmo SAM para análisis de citaciones: Una mejora al Árbol de la Ciencia

[Daniel-Stiven Valencia-Hernández][1], [Sebastian Robledo][2], [Ricardo Pinilla][3], [Néstor Darío Duque-Méndez][4], [Gerard Olivar-Tost][5]

ABSTRACT

Tree of Science (ToS) is a web-based tool which uses the network structure of paper citation to identify relevant literature. ToS shows the information in the form of a tree, where the articles located in the roots are the classics, in the trunk are the structural publications, and leaves are the most current papers. It has been found that some results in the leaves can be separated from the tree. Therefore, an algorithm (SAP) is proposed, in order to improve outcomes in the leaves. Two improvements are presented: articles located in the leaves are from the last five years, and also, they are connected to root and trunk articles through their citations. This improvement facilitates construction of the current literature for researchers.

Keywords: [**T**ree of Science], [**S**AP], [**A**lgorithm], [**C**itation analysis].

RESUMEN

Tree of Science (ToS) es una herramienta web que usa la estructura de la red de citaciones para identificar literatura relevante. ToS muestra la información en forma de árbol, donde los artículos localizados en las raíces son los clásicos, en el tronco son las publicaciones que le dan estructura al tema y las hojas son los artículos más recientes. Se ha encontrado que algunos resultados de las hojas pueden ser separados del tema del árbol. Por lo tanto, el algoritmo SAP es propuesto para mejorar los resultados de las hojas. Dos mejoras son presentadas: los artículos localizados en las hojas son de los últimos 5 años, y también, estos están conectados a la raíz y al tronco a través de sus citaciones. Esta mejora facilita la construcción de la literatura actual a los investigadores.

Palabras clave: [**T**ree of Science], [**S**AP], [**A**lgoritmo], [Análisis de citaciones].

Received: January __th 20xx
Accepted: January __th 20xx

Introduction

Tree of Science (ToS) is a web-based tool that uses graph algorithms to optimize the search for and selection of published papers. ToS was created at the Universidad Nacional de Colombia (Robledo, Osorio-Zuluaga, and López-Espinosa 2014), and the algorithm is explained elsewhere (Zuluaga et al., 2016). ToS is a specialized tool for researchers interested in tracking the way in which a particular topic evolves over time. Firstly, users must download Web of Science (WoS) query results. Then, they upload the file to ToS (tos.manizales.unal.edu.co). With this data, ToS shows the results in the form of a tree: root, trunk, and leaves. Papers in the root are the classics, while papers in the trunk are considered structurals publications, and current papers are the leaves. In addition, ToS uses scientometric techniques to recommend relevant literature.

Scientometry refers to the study of science, technology, and innovation, from a quantitative perspective. Moreover, scientometry focuses on the measurement of the impact of articles, journals, and institutions, along with the mapping of scientific areas (Leydesdorff, 2013). Examples include citation analysis (Köseoğlu, Sehitoglu and Craft, 2015), co-author analysis (Ioannidis, 2015), and the impact of institutions (Singh, Uddin and Pinto, 2015). Thus, the importance of scientometry is based on the possibility of identifying high impact articles, principal researchers, and recognizing emerging areas of knowledge (Hood and Wilson, 2001).

[1] Systems information administrator, Universidad Nacional de Colombia. Colombia.Institution, Country. E-mail: dsvalenciah@unal.edu.co

[2] Industrial Engineer, Universidad Nacional de Colombia, Colombia. M.B.A. Universidad Nacional de Colombia, Colombia. Ph.D. Universidad Nacional de Colombia. Affiliation: Research-Professsor, Universidad Católica Luis Amigó, Manizales. Colombia. E-mail: sebastian.robledogi@amigo.edu.co

[3] Mathematician. Universidad Nacional de Colombia, Colombia. MSc. Applied Mathematics. Universidad Nacional de Colombia, Colombia. E-mail: rpinillae@unal.edu.co

[4] Mechanical engineer. Universidad Nacional de Colombia, Colombia. MSc. Informatics. Ph.D. engineer. Universidad Nacional de Colombia, Colombia. E-mail: ndduqueme@unal.edu.co

[5] Mathematician. Ph.D. Applied Mathematics. Universidad Nacional de Colombia, Colombia. E-mail: golivart@unal.edu.co

How to cite: Valencia-Hernández, D. S., Robledo, S., Pinilla, R., Duque-Méndez. N.D. and Olivar-Tost, G. (20xx). SAP Algorithm for Citation Analysis: An improvement to Tree of Science. *Ingeniería e Investigación*, xx(x), xx-xx.
DOI: http://dx.doi.org/10.15446/ing.investig.vxxnx.xxxx

Scientometrics emerged in the 1930s, with the analysis of the distribution frequency of productivity between chemicals and physics (Lotka, 1926). After analyzing a number of publications, he concluded that the proportion of researchers making small contributions was 60%. Later, Price (1963), known as the father of scientometrics, formulated Price's Law, which explains that 25% of scientific authors are responsible for 75% of published articles (preferential attraction model). Finally, another important initial contribution in the scientometry field was the h-index (Garfield, 1972; Hirsch, 2005) which measures the impact of papers, and is well known in the scientific community nowadays. Thus, these results showed patterns in the scientific world, which can be identified by mathematical and statistical analyses.

Currently, thanks to the advances in technology such as the Internet, it is possible to apply and develop sophisticated scientometric techniques in different fields. For example, a study in nanotechnology and nanoscience shows metrics such as the annual growth rate, authorship patterns, and an index of collaboration (Karpagam, Gopalakrishnan, Natarajan and Ramesh Babu, 2011). Another investigation in bioenergy from biomass explains the exponential growth and changes in this field (Konur, 2012). Therefore, scientometrics has been a useful tool in recent years to identify emerging areas of science.

Although scientometry has evolved in the last few years, one of the main challenges is to find accurate methods for the characterization of a scientific area (Köseoğlu et al., 2015). For this reason, various researchers have proposed other indexes to determine the impact of publications such as the CDS-index (Vinkler, 2011), multivariate analysis techniques, time series (Leydesdorff, 2013), and modeling techniques (Mutschke and Mayr, 2014). However, co-citation analysis has become a well-established topic in scientometrics to identify "sleeping beauty publications" (Fang, 2019). Examples of applications of this scientometric techniques are found in reviews about obesity (Landinez, Robledo and Montoya 2019), Corporate Social Responsibility (Duque and Cervantes-Cervantes, 2019), and in agriculture (Robledo-Buriticá, Aguirre-Alfonso and Castaño-Zapata, 2019).

During the last years, some graph algorithms have been implemented in co-citation analysis to select relevant literature. For instance, HITS algorithm (Kleinberg 1999) was applied to reduce ranking bias (Jiang et al. 2016) and Google's PageRank algorithm to find the most prestigious papers (Chen et al. 2007). Nevertheless, much uncertainty still exists about tracking global knowledge using co-citation analysis (Parolo, Kujala, Kaski and Kivelä, 2019). Hence, this study seeks to improve the ToS algorithm in order to streamline the research process on a specific topic, in order to fulfill the need of non-conventional literature review techniques (Alulema and Largo 2019) that other studies have proposed (Sepulveda and Cravero, 2015).

The paper is structured as follows. First, a few basic definitions about graph theory are presented. Secondly, the methodology is described, detailing the algorithm step by step. Next, the SAP algorithm is applied to create a graph of citation analysis about Word-of-Mouth Marketing, in order to compare it with the current ToS results. Finally, conclusions are addressed, and limitations and implications are discussed.

Some basic definitions

Some basic definitions about graph theory are explained below (Johnsonbaugh, 1999):

Definition 1 (Undirected Graph) A graph (or undirected graph) consists of a set of vertices *V* and a set of edges *E* such that each edge e ∈ E is associated with an unordered pair of vertices. If there is a unique edge e associated with the vertices v and w, the following is written e = (v, w) or e = (w, v). In this context, (v, w) denotes an edge between v and w in an undirected graph and not an ordered pair.

Definition 2 (Directed Graph) A directed graph (or digraph) G consists of a set of vertices V and a set of edges E such that each edge e ∈ E is associated with an ordered pair of vertices. If there is a unique edge e associated with the ordered pair (v, w) of vertices, the following is written e = (v, w), which denotes an edge from v to w, v is the initial vertex and w is the terminal vertex of the edge e.

Definition 3 (indegree and outdegree of vertex) Let v be a vertex of a directed graph G. The degree of entry of v, denoted by indegree (v) is the number of edges in G with terminal vertex v. The degree of output of v denoted by outdegree (v) is the number of edges in G whose initial vertex is v.

Definition 4 (Subgraph) Let G = (V, E) a graph. G' = (V', E') is a subgraph of G if

a) V' ⊆ V and E' ⊆ E.

b) For each edge e' ∈ E', if e' is incident on v' and w', then v', w' ∈ V'.

Definition 5 (Connected graph) A graph G is connected if there is a walk between every pair of distinct vertices in the graph.

Definition 6 (Connected Component) A connected component of a graph G is a connected subgraph S of G such that no other connected subgraph of G contains S.

Data

In order to test the algorithm, we use data from Web of Science. This dataset contains information about articles published by journals from different areas of knowledge. From this dataset, we can extract the citation relationships between papers, authors, publication dates, journals, volume, page, and the Digital Object Identifier (DOI). Similarly, we can create a citation graph with the papers and their references (Zuluaga et al., 2016).

SAP Algorithm

The SAP algorithm was implemented in Python with the graph package igraph. Next, SAP operation is explained.

Description

The SAP algorithm presented in six steps.

1. From a subset of papers V, which is obtained from WoS, a directed graph G = (V, E) with all the papers and references are generated, where each directed edge (i,j) of E is a citation from paper pi to pj.

$$p_i \to p_j$$

2. Graph G is filtered

2.1. The largest connected component is obtained.

2.2. From the graph obtained in (2.1), loops are eliminated.

2.3. From the graph obtained in (2.2), if there are parallel edges with the same address only one is selected and the remaining ones are eliminated.

2.4. From the graph obtained in (2.3), vertices with indegree 1 and outdegree 0 are eliminated along with their edges. The filtered graph obtained in (2.4) is noted by G' = (V', E').

Igraph Description

2.1. Graph.clusters() (which show the different components of the graph) and giant() function are used to select the largest component.

2.2. and 2.3. graph.simplify() is used to eliminate repeated loops and edges.

2.4. Graph.vs.select() is used to select the variables that do not have indegree 1 and outdegree 0 or from which information cannot be extracted.

3. Construction of set V_{root}

3.4. The edges of V' with outdegree 0 are selected.

3.5. The vertices obtained in (3.1.) are organized in descending order by their degree of entry.

3.6. From the vertices ordered in (3.2), the first 10 are selected.

3.7. V_{root} is defined as the set of all the vértices selected in (3.3). Those vertices are called roots.

3.8. If r is a root, the sap of the r is defined as its indegree.

Igraph description

3.1. Graph.vs.select() is used to choose the vertices with outdegree 0.

3.2. And also 3.3. and 3.4. In this part, sorted() (python native function) is used to sort the vertices in descending order with respect to their indegree. Next, the first 10 vertices on this list are saved.

3.5. Indegree() is used to determine the degree of input of one of the vertices obtained from steps (3.2), (3.3), and (3.4).

4. Construction of set V_{leaves}

4.1. V' vertices with indegree 0 are selected.

4.2. From the vertices in (4.1), those for which there are at least three paths between this and the set of roots are selected.

4.3. From the vertices in (4.2), those whose age (time since publication) do not exceed 5 years from the current year are selected.

4.4. V_{leaves} is defined as the set of all vertices selected in (4.3).

4.5. If h belongs to V_{leaves}, its SAP is denoted as the number of paths that exists between h and the roots.

4.6. The vertices of the set V' – ($V_{leaf1} \cup V_{root}$) that are in the paths between the elements of V_{root} and V_{leaf1} with which it has a connection by one or more paths.

4.7. The sap of the vertices obtained in (4.6) are defined as the sum of the saps of the elements of V_{root} and V_{leaf1} with it has a connection by one or more paths.

4.8. From the vertices in (4.6), those whose age does not exceed five years from the current year are selected.

4.9. Let V_{leaf2} be the set of all the vertices selected in (4-8).

4.10. If V $\in$ V_{leaf2} its sap is defined by (4.7).

4.11. V_{leaves} is defined as the union between V_{leaf1} with V_{leaf2}, and so, $V_{leaves} = V_{leaf1} \cup V_{leaf2}$.

Igraph Description

4.1. Graph.vs.select() is used to choose the vertices with indegree 0.

4.2. Graph.shortest_paths_dijkstra() is used to identify paths between the vertices obtained in (4.1) and the roots. Next, those vertices for which there are more than three paths are selected.

4.3. And 4.4. A cycle on the leaves is applied to select and save the current ones (less than five years old).

4.5. Indegree() is used to determine the outdegree of each vertex obtained from steps 4.2, 4.3, and 4.4.

4.6. The next function is used to select the vertices between V_{leaf1} and V_{root} to save its sap value.

```
Paths(graph_adj, source, target, path = [])
        for index in graph_adj[source]:
                If (index not in path):
#the path variable is a path into source and target
                Else:
                        path.append(index)
                        paths(graph_adj, index, target, path)
                        path.pop()
```

4.7. And 4.8. Cycle is applied on the vertices obtained in (4.6) to select and save the leaves (papers less than five years old).

4.9. See step 4.6 and 4.7

4.10. A union is made between sets V_{leaf1} and V_{leaf2} with .append().

5. Construction of set V_{root}

5.1. The vertices of set V' – ($V_{leaf1} \cup V_{root}$) that are in the paths between the elements of V_{leaf1} and the elements of V_{root} are selected.

5.2. The sap of the vertices obtained in 5.1 is defined as the sum of the saps of roots and leaves with which they have a connection by one or more paths.

5.3. From the vertices obtained in (5.2), those whose age exceeds five years from the current year are selected.

5.4. The vertices obtained in (5.3) are put in descending order with regard to their sap.

5.5. From the vertices ordered in (5.4), the first 10 are selected.

5.6. V_{root} is defined as the set of all selected vertices in (5.5).

5.7. If t belongs to V_{root}, its sap is defined by 5.2.

Vertices in V_{trunk} are called trunks

Igraph Description

5.1. And 5.2. The function from steps (4.6) and (4.7) in the igraph is used to select the vertices between V_{leaf1} and V_{root} and their sap is saved.

5.3. The vertices obtained in (5.2) are selected by their age (more than five years from the current year).

5.4. 5.5 and 5.6. Sorted() (native python function) is used to sort the vertices in descending order with regard to their sap. Next, the first 10 vertices of this sorted list are saved.

5.7. See step (5.1) and (5.2).

6. Tree construction.

Subgraph. G = (G, E) of G' = (V', E'), where V = $V_{root} \cup V_{trunk} \cup V_{leaves}$ and E is considered a subset of the edges of E' which only affects the vertices of V. Subgraph G = (G, E) is called "Tree of Science" (Robledo et al. 2014).

Application

To illustrate the functioning of the SAP algorithm, it is compared to the results from ToS. We present the similarities and differences between the two procedures.

The first step is to define the research topic, in order to obtain the data from WoS. In this case, Word-of-Mouth Marketing (WOMM) is used as the search equation for the time period from January 2001 to August 22, 2017.

Title = (marketing) AND Topic = (Word of Mouth) Indexes: SCI-EXPANDED, SSCI, A&HCI

Exactly 317 papers were extracted with these references. With this data, both algorithms were applied in order to identify the SAP improvements. Table 1 shows the difference between both algorithms. The results in the roots and trunk of both algorithms are similar, 90% and 70% respectively. However, ToS presents better results in terms of number of citations: in total 2,819 in the roots and 1,752 in the trunk. Despite this, SAP has an outstanding performance with more than three times more citations than results from ToS.

Table 1. Differences between ToS and SAP algorithm

		Root	Trunk	Leaves
Similarities		90%	70%	23%
Differences in citations	ToS	2 819	1 752	741
	SAP	666	1 001	2 442

Source: Authors

Another important result of SAP is the age of the papers on the leaves. According to Robledo et al. (2014), leaves are current papers, and a quality indicator of an investigation is the number of young references. Moreover, Price (1976) suggests that at least 50% of the references should be from the past five years. Similarly, SAP meets this requirement, selecting papers from the past five years for the leaves (see Figure 1).

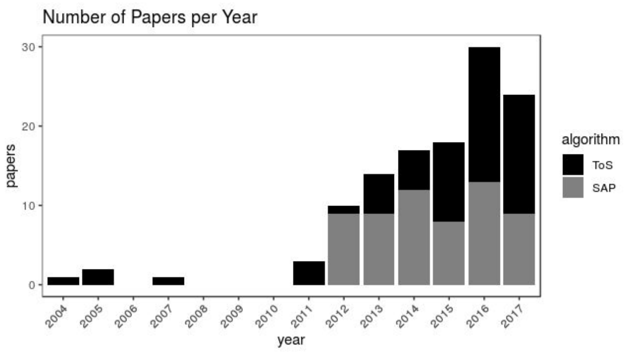

Figure 1. The number of paper per year per algorithm. Source: Authors

Conclusions

ToS is a scientometric tool which performs a citation analysis of a graph and shows the results in a form of a tree: root, trunk, and leaves. Most of the results presented by ToS are relevant and important (Robledo et al. 2014). However, there is a lack of precision in the leaves; sometimes publications are not connected to the roots and trunk, and additionally, the leaves are occasionally not current literature. Thus, the goal of this study is to propose a new algorithm called SAP, which improves the results in the leaves.

Results show that SAP is more accurate in terms of results in the leaves. SAP presents the most important current literature. However, this study is limited, and so must be further expanded, for example, to the evaluation both of different research topics and indicators, in order to understand the pros and cons of the new algorithm.

Acknowledgements

We thank the "Instituto Colombiano para Ciencia y Tecnología" (Colciencias) and its program "Jóvenes investigadores e innovadores". Also, this investigation was partially supported by Universidad Nacional de Colombia under project No. 32059 (Code Hermes) entitled "Consolidación de las líneas de investigación del Grupo de Investigación en Ambientes Inteligentes Adaptativos GAIA" within the "Convocatoria interna de investigación de la Facultad de Administración 2015, para la formulación y ejecución de proyectos de consolidación y/o fortalecimiento de los grupos de investigación. Modalidad I: Formulación y ejecución de proyectos de consolidación". Finally, we also would like to thank the Universidad Nacional - Sede Manizales library for its diffusion of ToS, and their support in this project.

References

Alulema, F. X. V. and Largo, F. L. (2019). Strategic portfolio of IT projects at universities: A systematic and non-conventional literature review. *Ingeniería e Investigación*, 39(2).

Chen, P., Xie, H., Maslov, S. and Redner, S. (2007). Finding scientific gems with Google's PageRank algorithm. *Journal of Informetrics*, *1*(1), 8-15.

Duque, P. and Cervantes-Cervantes, L.S. (2019). Responsabilidad Social Universitaria: una revisión sistemática y análisis bibliométrico. *Estudios Gerenciales*, 35(153), 451-464.

https://doi.org/10.18046/j.estger.2019.153.3389

Fang, H. (2019). A transition stage co-citation criterion for identifying the awakeners of sleeping beauty publications. *Scientometrics*, 121(1), 307-322. DOI: https://doi.org/10.1007/s11192-019-03195-9

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science (New York, N.Y.), 178(60)*, 471-479. DOI: https://doi.org/10.1126/science.178.4060.471

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, 102(46)*, 16569-72. DOI: https://doi.org/10.1073/pnas.0507655102

Hood, W. W. and Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics, 52(2)*, 291-314. DOI: https://doi.org/10.1023/A:1017919924342

Ioannidis, J. P. A. (2015). A generalized view of self-citation: Direct, co-author, collaborative, and coercive induced self-citation. *Journal of Psychosomatic Research, 78(1)*, 7-11. DOI: https://doi.org/10.1016/j.jpsychores.2014.11.008

Jiang, X., Sun, X., Yang, Z., Zhuge, H. and Yao, J. (2016). Exploiting heterogeneous scientific literature networks to combat ranking bias: Evidence from the computational linguistics area. *Journal of the Association for Information Science and Technology*, *67*(7), 1679-1702.

Johnsonbaugh R. (1999). Discrete Mathematics, New York: Macmillan Publishing Company.

Karpagam, R., Gopalakrishnan, S., Natarajan, M., and Ramesh Babu, B. (2011). Mapping of nanoscience and nanotechnology research in India: A scientometric analysis, 1990-2009. *Scientometrics, 89(2)*, 501-522. DOI: https://doi.org/10.1007/s11192-011-0477-8

Kleinberg, J. M. (1999). Hubs, authorities, and communities. ACM computing surveys (CSUR), 31(4es), 5.

Konur, O. (2012). The scientometric evaluation of the research on the production of bioenergy from biomass. *Biomass and Bioenergy, 47*, 504-515. DOI: https://doi.org/10.1016/j.biombioe.2012.09.047

Köseoğlu, M. A., Sehitoglu, Y. and Craft, J. (2015). Academic foundations of hospitality management research with an emerging country focus: A citation and co-citation analysis. *International Journal of Hospitality Management, 45(0)*, 130-144. DOI: https://doi.org/10.1016/j.ijhm.2014.12.004

Landinez, D., Robledo, S., and Montoya, D. (2019). Executive Function performance in patients with obesity: a systematic review. *Psychologia*, 15(2), 31-50. DOI: https://doi.org/10.1007/s11192-012-0917-0

Leydesdorff, L. (2013). Statistics for the dynamic analysis of scientometric data: The evolution of the sciences in terms of trajectories and regimes. Scientometrics, 96(3), 731-741. DOI: https://doi.org/10.1007/s11192-012-0917-0

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences, 16(2)*, 317-324. URL: https://www.jstor.org/stable/24529203

Mutschke, P. and Mayr, P. (2014). Science models for search: a study on combining scholarly information retrieval and scientometrics. *Scientometrics, 102(3)*, 2323-2345. DOI:

https://doi.org/ 10.1007/s11192-014-1485-2

Parolo, P. D. B., Kujala, R., Kaski, K. and Kivelä, M. (2019). Going beneath the shoulders of giants: tracking the cumulative knowledge spreading in a comprehensive citation network. *arXiv preprint arXiv:1908.11089*.

Price, D. J. de S. (1963). Little science, big science... and beyond. *Columbia University Press. New York*.

Price, D. D. de S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information Science*, *27*(5), 292-306. https://doi.org/10.1002/asi.4630270505

Robledo-Buriticá J., Aguirre-Alfonso, C.A. and Castaño-Zapata, J. (2019). Guía ilustrada de enfermedades en postcosecha de frutas y verduras y sus agentes causantes en Colombia. Bogotá, *Colombia: Academia Colombiana de Ciencias Exactas, Físicas y Naturales.*

Robledo, S., Osorio, G. and López, C. (2014). Networking en pequeña empresa: una revisión bibliográfica utilizando la teoría de grafos. *Revista vínculos*, 11(2), 6-16. URL: https://revistas.udistrital.edu.co/ojs/index.php/vinculos/article/view/9664

Sepúlveda, S. and Cravero, A. (2015). Protocol adaptations to conduct systematic literature reviews in software engineering: A chronological study. *Ingeniería e Investigación*, *35*(3), 84-91.

Singh, V. K., Uddin, A. and Pinto, D. (2015). Computer science research: the top 100 institutions in India and in the world. *Scientometrics*. DOI: https://doi.org/10.1007/s11192-015-1612-8

Vinkler, P. (2011). Application of the distribution of citations among publications in scientometric evaluations. *Journal of the American Society for Information Science and Technology, 62(10)*, 1963-1978. DOI: https://doi.org/10.1002/asi.21600

Zuluaga, M., Robledo, S., Osorio-Zuluaga, G. A., Yathe, L., Gonzalez, D. and Taborda, G. (2016). Metabolómica y Pesticidas: Revisión sistemática de literatura usando teoría de grafos para el análisis de referencias. *Nova, 13(25)*, 121–138. URL: http://hemeroteca.unad.edu.co/index.php/nova/article/view/1735

Please suggest possible peer reviewers / experts in the topic(s) addressed by this article. Editorial processing time could be reduced by doing so.

Peer reviewers must comply with the following requirements:

- Professionals holding an MSc or Ph.D.
- Active researchers in the topic of this article
- H-index higher than 2
- Authors' and reviewers' affiliation must be completely different
- Researchers working at Higher Education Institutions or Research Groups
- International peer reviewers are desirable (a different country from that of the authors)

Suggested peer reviewers:

Name	Active email addresses	Academic Degrees	Current Affiliation
Joaquín del Río Fernández	Joaquin.del.rio@upc.edu	Ph.D. Electronics Engineering	Universitat Politecnica de Catalunya, Barcelona SPAIN
Jesse Michael Fagan	jmfagan@unm.edu	Ph.D. in Business Administration	University of New Mexico

Table 1. Differences between ToS and SAP algorithm

		Root	Trunk	Leaves
Similarities		90%	70%	23%
Differences in citations	ToS	2 819	1 752	741
	SAP	666	1 001	2 442

Source: Authors