Scientometrics
https://doi.org/10.1007/s11192-018-2917-1

An integrated approach to path analysis for weighted citation networks

Hiran H. Lathabai[1] · Susan George[1] · Thara Prabhakaran[1] · Manoj Changat[1]

Received: 14 May 2018

Abstract

Profuse growth of scientometrics as a research field owes a discernible attribution to the introduction of citation networks and other scientograms. Centrality analysis, path analysis and cluster analysis are three major network analysis tools. Hummon and Doreian's introduction of path retrieval methods based on (1) traversal count as weight assignment (for arcs) method and (2) search methods such as local (forward) search and global search, marked the commencement of path analysis. Original Hummon–Doreian traversal count based weight assignment methods such as Search Path Link Count and Search Path Node Pair were computationally complex. Along with the computational improvement of these weights, Batagelj added another computationally efficient traversal count method to the path analysis literature known as Search Path Count. A major development in search methods was seen recently with the introduction of innovative search methods such as backward (local) search and key-route (local and global) search by Liu and Lu. They also powered the available and new local search methods with a parameter to control the search. Major advantage of Liu–Lu methods lies in the fact that these can reveal more paths or more papers that are usually missed out in classical methods. All these contributions considered unweighted citation networks as the object of analysis. Despite being a tool of tremendous potential, path analysis is much underexplored relative to other network analysis tools. Inspired by these, we generalise Liu–Lu integrated approach, the present state-of-art in path analysis to an integrated approach for weighted networks. We demonstrate a manifold improvement in analysis opportunities with the generalized integrated approach using FV gradient weights for weight assignment, on a case study of the field 'IT for engineering'. Integrated approach for weighted networks do not need additional implementation effort in PAJEK and this will be beneficial for a multitude of analysts and decision makers.

Keywords Citation networks · Path analysis · Weighted networks · Flow vergence effect · Flow vergence paths · Integrated approach

✉ Thara Prabhakaran
thara.dfs@gmail.com

Extended author information available on the last page of the article

Published online: 22 September 2018

Introduction

Network scientometrics is steadily gaining recognition among policymakers owing to the multitude of tools and analysis prospects it hoards. The notion of citation network or network of scientific papers was introduced by Garfield et al. (1964) and De Solla Price (1965) in early 1960's itself. Though scientograms Moya-Anegón et al. (2007) such as citation networks were present from the beginning of the modern age of scientometrics, other kinds of scientograms co-citation and bibliographic coupling (derived networks) were more admired till 1989. Path analysis had been added to the literature of citation network analysis in 1989 with the work of Hummon and Doreian (1989). Till then 'network scientometrics' was mainly vertex-centric and with the introduction of the main path and critical path, link centric approach also began to co-evolve in the network literature.

Identification of evolutionary trajectories of science and technology is of immense importance to a myriad of beneficiaries that include policymakers. Importance of methods to mine the literature and identify important evolutionary paths that serve as the backbone of the development of different fields lie there. Path analysis method introduced by Hummon–Doreian can be regarded as a breakthrough development in scientometrics in this context. Path analysis methods consist of two main steps-(1) computation of weights (traversal counts) of each arc (this converts unweighted citation networks to weighted citation networks) followed by (2) searching the weighted network to obtain important paths. Local search (greedy approach) on the traversal weighted network provides the *main path* and global search (critical path method) on the traversal weighted network provides the *critical path*. Hummon–Doreian method for traversal weight computation for each arc was based on search path link count (SPLC) and search path node pair (SPNP). Improvement in computation of these weights was realized by Batagelj (2003). Apart from computationally improving Hummon–Doreian traversal count weight assignment, introduction of computationally efficient weights, namely search path count (SPC) weights was another important development. As Hummon–Doreian weight assignment schemes (SPLC and SPNP methods) and Batagelj's SPC method are traversal count based measures, these form a genre that can be called as SPX methods. These were made available in the network visualization and analysis package PAJEK Batagelj and Mrvar (1998) in combination with the local and global search algorithms for the retrieval of main path and critical path. Important works in various disciplines that used path analysis for literature assessment is discussed in "Related literature" section. Recently owing to shortcomings of main path due to its dependency on local search, many variants of main paths such as global main path, the backward local main path, multiple main paths, key-route main paths, genetic knowledge persistence based main path, etc., were proposed. Genetic persistence based main path was an extension of genetic approach introduced in Martinelli and Nomaler (2014) for the identification of technologically important patents (determined as a path based vertex measure) to path analysis. Except genetic persistence based main path, all the others were introduced by Liu and Lu (2012) and suggested their use as an integrated approach to main path analysis, which embedded Hummon–Doreian's main path too. A parameter 'tolerance' was included to control the search process and is a determinant of the size of the path (subnetwork). The introduction of parameter even strengthened the usability of Hummon–Doreian's main path, which is termed as the 'forward local main path' by Liu and Lu. Another interesting thing is that global search on SPX weighted citation network will lead to the global main path, which is equivalent to Hummon–Doreian's critical path. Key-route main paths (local and global), multiple main paths were the novel additions in the integrated approach. More detailed discussion on these can be found in "Liu–Lu approach" section.

In network scientometrics, another important method for path analysis is genetic persistence based main paths. Introduced for retrieving important trajectories in patent citation networks, this method involves the computation of genetic persistence scores for each patent and then the search processes (forward as well as backward) on the resulting weighted network. Thus, when a citation network (unweighted) is converted into weighted one and subjected to search algorithms (the choice of weights, as well as search method, should be logical), important developmental trajectories can be revealed. SPC method and genetic persistence score computation method are two of such diligent methods for weight assignment. Search schemes like local search (forward and backward), global search (equivalent to CPM) and key-route search (local as well as global), etc. are the sensible search methods. Thus immense possibilities are there for the development of methods to retrieve more hidden trajectories either by designing methods to assign weights to arcs/ vertices or by finding novel search methods or the combination of both.

A scheme based on a neoteric metric is particularly investigated in this work for the purpose of weight assignment and generation of paths. An index (flow vergence index or FV index) was proposed in Prabhakaran et al. (2015) to identify important papers in a network based on their potential for flow vergence. The concept of flow vergence is discussed in "Flow vergence potential and the FV index" section. In Lathabai et al. (2015), a new metric-FV gradient, that could detect the pivot papers of paradigm shift based on the flow vergence effect is introduced. These are also briefly discussed in "Flow vergence gradient or FV gradient" section. An important thing to be noted here is that being an arc metric, a weighted network (arc weighted) can be created using FV gradient values. Paths retrieved using FV weighted citation network are referred in this work as flow vergence paths or FV paths and those that are created using traversal count methods can be termed as SPX paths. In this work, among SPX methods, SPC is used due to the reasons mentioned in "Liu–Lu approach" section. Both SPC and FV paths are created on the network of 'IT for engineering' and a comparative analysis of these are conducted to check the degree of uniqueness possessed them.

If an analyst comes with any weighted network (say FV weighted or citation relevance/ irrelevance score weighted or arcs weighted by any other means), he has at least three choices-(1) he can either go for search algorithms to generate important paths (direct search approach) or (2) simply ignore weights at the first place to generate SPX paths (ignore and proceed approach) or (3) without ignoring weights compute new weights using a new layer of weight assignment method so that new weights will be composite of both weights. Thus, in this work, treating citation network (unweighted) as a specific case of the weighted network, an integrated approach for path analysis is developed, which embeds the Liu and Lu's present integrated approach. As traversal count methods and search methods are available in PAJEK, first two options in the integrated approach for weighted networks (if networks are prior weighted) can be executed in PAJEK without the need for additional implementation efforts. However, PAJEK software is presently deficient when it comes to effective arc weight computation schemes. So the full potential of the third approach cannot be enjoyed now. Hence we recommend the implementation of more weight assignment schemes in PAJEK for the sake of a wide range of beneficiaries. As the methods discussed in this study are prone to the referencing behaviour of authors and other practices that may even result in misattribution of citations, these are not free from the inherent limitations of all such methods for citation network analysis.

Related literature

In this section, a brief discussion on important developments in citation network analysis and developments in path analysis literature are reported. Citation network analysis is almost as new (or old) as the modern scientometrics which commenced through the development of Science Citation Index by Garfield (1955) and the introduction of the scientogram-citation network or network of scientific paper De Solla Price (1965), Garfield et al. (1964). Major hindrances to the reliability of citation network analysis due to the citation practices were identified by Merton as *palimpsestic syndrome* (misattribution of citation to another work instead of the original) Merton (1965) and *obliteration by incorporation* (lack of proper attribution to the original source) Merton (1988). Structurally being (almost) acyclic directed graphs, a study of properties of the citation networks using graph theory soon commenced with Garner (1967). Some of the works in this direction are Bilke and Peterson (2001), Egghe and Rousseau (2002) and Valverde et al. (2007). Centrality analysis, path analysis and cluster analysis are three major possible analytic schemes for citation network analysis. Shibata et al. Shibata et al. (2011) used centrality and cluster analysis to detect emerging research fronts in regenerative medicine. Advantages of citation networks (direct) over indirect citation networks like co-citation and bibliographic coupling were identified in Shibata et al. (2009) and Leydesdorff et al. (2017). While the former concluded that direct citation networks are the most effective among the three for the detection of emerging research domains, the latter suggested that, at the journal level, diversity of citation networks is the best measure of interdisciplinarity compared to the diversity of co-citation and bibliographic coupling networks. A qualitative, as well as quantitative framework using citation network analysis for interdisciplinary research assessment, is proposed in Karunan (2017). These recent publications along with many others mark the present state of the art of citation network analysis. Citation network analysis which incorporates relevance/irrelevance of citations made is an emerging area of research. Evolution of path analysis, the prime topic of our focus is described next.

Hummon and Doreian's method for paths was based on search path link count (SPLC) and search path node pair (SPNP) weights as link measures Hummon and Doreian (1989). Key connective threads of knowledge flow in literature or the important pieces of knowledge evolution can be represented by these paths. This event paved the way for the development of path analysis as a tool for citation network analysis. Introduction of search path count (SPC) method Batagelj (2003), a relatively simpler weight assignment scheme increased its applicability in large citation network analysis. Implementation of SPC method in network analysis and visualization program called PAJEK Batagelj and Mrvar (1998) accelerated the use of path analysis. Study of evolutionary trajectories of medical knowledge can be found in Mina et al. (2007), which used main path analysis along with other network techniques. Another work Tampubolon and Ramlogan (2007) also explored the literature on medical treatment, especially the literature of 'coronary angioplasty' using the main path. Applications of path analysis for research on evolutionary trajectories and structural backbones of various research fields like human resource development and archaeology can be found in Jo et al. (2009) and Brughmans (2013) respectively. Combined usage of citation network analysis (including path analysis) along with bibliometric mapping for the exploration of the research field "Absorptive capacity" can be seen in Calero-Medina and Noyons (2008). The critical path is used for the analysis of literature for the investigation of 'Data Envelopment Analysis (DEA) and its application in financial services' Kaffash and Marra (2017).

An integrated approach for main path analysis was introduced by Liu and Lu recently. It embeds original SPX main path along with a facility to control the search using a parameter 'tolerance'. As forward search is used for the generation of the main path, the phrase 'forward local main path' is used to denote those paths. They also introduced a backward local main path which can be obtained by the use of SPX weights in combination with a reverse search (sink to source) scheme. Global main path method was there, but it is found equivalent to Batagelj's critical path method. Another important innovative method, the key-route main path method was also introduced and forms part of the integrated approach. Most of these methods are available in the latest version of PAJEK and are discussed in "Liu–Lu approach" section. Application of the paths in the integrated approach for the analysis of research theme 'data envelopment analysis' can be found in Lin et al. (2013). Three of the main paths in the integrated approach was used for path analysis of the "data quality" field in Xiao et al. (2014). Application of key-route main path can be found in Chuang et al. (2014) for the study of medical tourism developments. Study of knowledge diffusion paths of graphene for optoelectronics using key-route paths can be found in Chen et al. (2013). Receptivity of path analysis methods from various research fields indicates the immense potential hoarded by them. Despite being a technique in network scientometrics, which has already proven its worth, when it comes to the development of new methods, path analysis seems to be under-explored. Need for methods to retrieve paths that may be missed by the present state-of-the-art methods (i.e., methods in Liu–Lu integrated approach) is felt. This gap is attempted to be filled in this work by expanding Liu–Lu integrated approach to a generalized approach for weighted citation networks. After discussing briefly Liu–Lu integrated approach in "Liu–Lu approach" section, different methods that form the generalized approach is described in "Weighted citation networks" section.

Liu–Lu approach

In Liu and Lu (2012), an integrated approach to main path analysis was proposed by Liu and Lu. Most of the methods in Liu–Lu approach is implemented in PAJEK. The methods based on SPX weights are capable of providing forward local main path, backward local main path, global standard main path (which is same as critical path), key-route main path (local and global) and critical path. On a citation network (unweighted), methods for path retrieval is basically a two-step process-weight assignment followed by search scheme. Traversal weight assignment is an important method for weight assignment, introduced by Hummon and Doreian. Search path count (SPC) is a computationally simpler method for traversal weight assignment, developed by Batagelj. Batagelj observed that all the SPX methods produce almost same paths and suggested the usage of the SPC method as 'first choice' due to its relative simplicity and some 'nice properties' Batagelj (2003) once these are tested on real world large networks. These could probably be the reasons that made SPC method the most sought one among SPX methods. Before the introduction of Liu–Lu approach, path retrieval was based on SPX weight computation followed by search schemes such as local (forward) search and critical path method. While the former search scheme retrieves the main path, latter reveals the critical path. As SPC scheme best represents the SPX methods that serve as the base for Liu–Lu approach, it is briefly discussed with illustration as follows.

SPC method

SPC method begins by identification of all sources and sinks in a citation network. Once these are identified, for each arc in the network, the number of search paths passing through it is computed. SPC computation transforms the unweighted or equi-weighted citation networks to a weighted (SPC weighted) network. This is illustrated using the following sample network shown in Fig. 1 (top).

In the sample network shown in Fig. 1 (top), there are three sources (F, J and I) and two sinks (A and B). Various source-sink paths in the network are F-E-C-A, I-E-C-A, I-C-A, I-C-B, I-E-C-B, F-E-C-B, H-D-B, J-D-B, J-G-D-B and I-G-D-B. Now, each arc in the network need to be assigned weights according to their presence in search paths from source to sinks. i.e., each arc is weighted according to the number of search paths passing through them. For instance, let us consider arc E-C. From the above list, four source-sink paths namely F-E-C-A, I-E-C-A, I-E-C-B and F-E-C-B pass through this arc and hence weight of arc E-C will be **4**. Similarly, SPC weights for all the arcs can be computed. Weighted citation network (SPC weighted) thus obtained is shown in Fig. 1 (bottom).

Innovative notions in Liu–Lu approach is not on weight assignment part but on search schemes. Important search schemes in their integrated approach are discussed next.

Search schemes

Different search schemes in the Liu–Lu integrated approach, which are implemented in PAJEK are: (1) forward local search (2) backward local search (3) key-route search (local

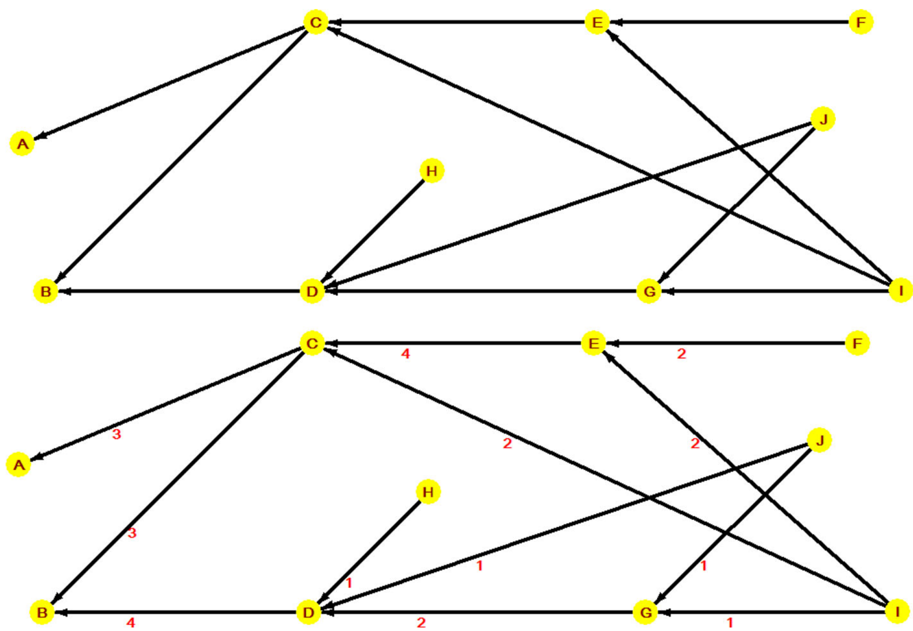


Fig. 1 (Top) sample (unweighted) citation network and (bottom) the weighted network obtained after SPC method

and global) and (4) critical path method or global (standard) search. All these are briefly discussed below.

Forward search

On a weighted network (like SPX weighted network), one of the simplest ways to trace a path of importance is local search. Firstly, from all the arcs that originate from the sources in the network, that arc with largest arc weight will be chosen. Then the target vertex of that arc is marked as next source and this greedy search continues till a sink is encountered. In case of a tie, during the search, target vertices of all such arcs are marked as new sources. In this way, a network can be partitioned into two subnetworks-one contains vertices and arcs that are identified by the local search and the other set contains the vertices that are missed by the local search. Upon extraction of the first subnetwork, main path or main paths can be obtained. This is termed as forward search since the search is initiated from source vertices and ends on sink vertices. Forward local main path search on SPC weighted network given in Fig. 1 (bottom) is shown in Fig. 2. This was the original search method implemented in early versions of PAJEK. As we can see, chances of missing important arcs that are not in the immediate vicinity of the first chosen arc are high.

Liu and Lu proposed a parameter namely tolerance to address this problem and it is available in recent versions of PAJEK. According to Liu and Lu, tolerance value in % “is an arbitrary number that controls the level of details we want to visualize” Liu and Lu (2012). For instance, a value of 20% (0.2 in PAJEK) relaxes the local search criteria from ‘search for largest arc’ to ‘search for 20% within largest’, which results in the more

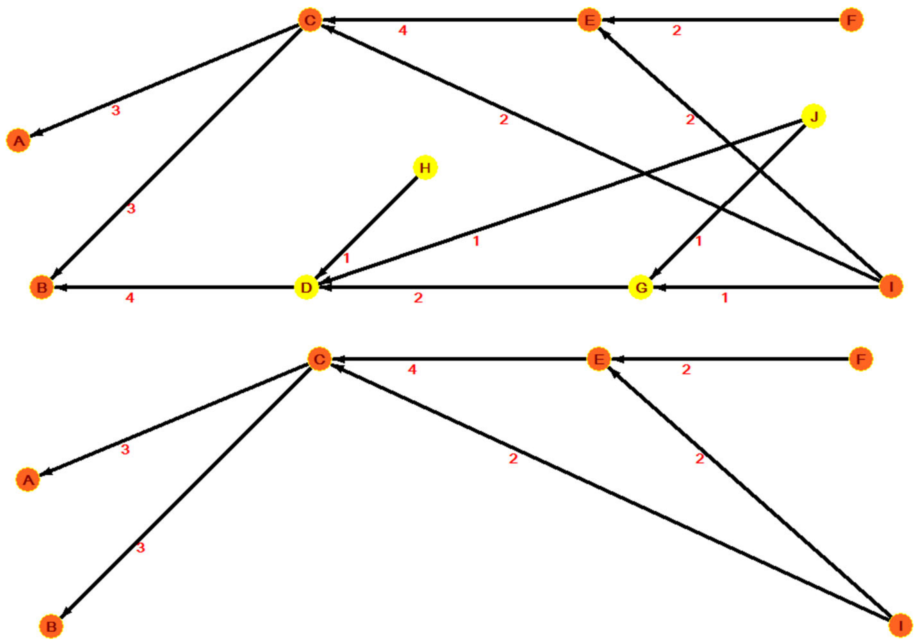


Fig. 2 (Top) forward local search on SPC weighted sample network where orange vertices represent the vertices highlighted in forward search and (bottom) extracted subnetwork as local forward main path

detailed inclusion of arcs and consequently more vertices or more paths itself. It is to be noted that if the parameter value is chosen as 0.2, if none of the arcs is found within 20% of the largest, only the largest will be chosen. In case if a large number of other arcs fall within 20% of the largest, the chance for revealing more vertices or paths is high. In the sample network, a tolerance of 0.1–0.4 does not reveal any other vertices because, at any level, none of the arcs is found to be within that 40% of the largest arc. But a .5 value for tolerance reveal more arcs in the first stage itself. With a 0.5 tolerance for forward search, along with arcs such as F-E, I-E and I-C which have SPC weight = 2 (largest weight), other arcs with weight = 1, i.e., I-G, J-G, J-D and H-D will be revealed and continuing like this we eventually end up with the whole network as main path.

Though the introduction of tolerance parameter is an enhancement of the original search method, the presence of a parameter may cause inconvenience to many of the users if they are not familiar with it or if they don't have much time to change parameter values and look for new details that appeared. Such users may prefer default parameter value 0% (0.00 in PAJEK) for their perusal. Therefore, in this work, the effect of tolerance will not be included or discussed during the analysis of case study network. In their integrated approach, Liu and Lu introduced not only a parameter to provide flexibility to search methods but also introduced some search methods too. One of such method is discussed next.

Backward search

The name forward search was given to the above search method because it searches from sources to sinks. Reverse searching, i.e., initiation of search from sink to source is also possible. This might result in a main path or paths that are more likely to be different from the ones obtainable in the forward search. The parameter tolerance can be used for this search too. With respect to the direction of citation (citing work to cited work or latest work to early work), backward search identifies the papers that could attract or inspire many other works with their intellectual merit and forward search identifies the papers that received knowledge or got inspired by the contribution of many other works. Liu and Lu used the opposite convention, i.e., they chose the direction of knowledge flow (cited work to citing or early to latest) as arc direction. So their interpretation of backward search actually applies to the forward search if the direction of citation is chosen as the direction of the arc like we did. According to the convention followed in this work, backward search in a citation network starts from a paper that initiated largest number of paths and ends upon one of the recent works that may be a distant successor of the first one, whereas forward search start from a recent one that has received knowledge through more knowledge flow paths and trace out its most significant origin(s).

Global search (standard) or critical path method

Local search methods (forward or backward) need not produce the path with the largest overall sum of traversal counts. The global main path was proposed in Liu and Lu (2012) as a path with largest overall traversal counts. Unlike the local paths that reveal a particular sequence of significant knowledge flows, the global search provides the path that has played an overall significant role in knowledge flow in the network. This method is same as that of the critical path method available in earlier versions of PAJEK, which is inspired by the critical path method in operations research. Implementation of both these methods in different names in recent versions of PAJEK is a redundancy that can be eliminated in future versions. Being a global search method, this one does not require the tolerance parameter.

Due to the structural peculiarities of this small sample network, the critical path is found to be same as that of the local forward main path. Despite the chance for having common papers, this exact similarity cannot be expected in large networks.

Key-route search

Source-sink search schemes (either local or global) do not guarantee the presence of the arc with the largest weight in the main path. This is a crucial drawback and to overcome this Liu and Lu introduced the innovative search method namely key-route search. The perceived view behind key-route search is that main path is the extension of the most significant link. Identification of key-route (the arc with largest arc weight) and execution of search from both the vertices of the arc ensures the presence of the key-route in the main path. In order to achieve this on a weighted citation, network search has to be initiated from both the vertices and search can be local or global. In case of local search, if the citation arcs in network are directed from citing vertex to cited vertex, a backward search has to be initiated from start vertex of the key-route arc and progressed till a source vertex is encountered and a forward search has to be done from end vertex of the key-route till a sink is found. In case of global key-route search, a global search is initiated from both the vertices of key-route instead of conducting a priority first search so that path with the highest sum of traversal weights that contain the key-route can be obtained. Liu and Lu also introduced the technique for ensuring or improving the chance of retrieving multiple paths by simultaneous search from top key-routes (say 1–10 arcs with large weights) than from a single largest key-route. The default value for this parameter is set as 1–10 in PAJEK, which means top ten key-routes will be selected for searching. In large networks, this may cause inconvenience to not-so-expert users as they might not change parameter value to a lesser range and suddenly ends up with large or multiple paths. It is recommended to keep the default value of the key-route parameter as 1. Additionally, the provision of tolerance while executing local search also controls the size of each path in the multiple paths by highlighting more arcs.

Weighted citation networks

Citation networks are essentially unweighted networks or equi-weighted networks where arcs (citation links) are of weight 1. Intrinsically all the citations are treated to be of equal importance or no distinction exists among arcs. However, if by any means we distinguish arcs, then these can be converted into weighted networks. One such possible distinction is based on the relevance of citations. If there is a means to assign weights to each citation arc by the assessment of the relevance of a citation from one paper to other, then a weighted citation network is created. For example, if the citation is highly relevant, then more weight is assigned and if the citation is less relevant (as if done as part of literature review), then less weight is assigned. Unweighted networks can be converted to weighted citation networks through many other rational methods. SPX methods are renowned methods in network scientometrics.

However, if an analyst has a weighted network at hand and is proceeding for path analysis, he has at least three choices for doing it. Three basic choices available to him is depicted in Fig. 3. Let x be the present weight of network (so network denoted as N_x) generated by method X.

1. *Ignore and proceed approach* He can ignore the weights and treat the network as unweighted citation networks and proceed to compute weights y using the method Y of his choice (for instance SPC) and use various search methods to generate main paths and critical path.
2. *Direct search approach* He can directly proceed to apply various search methods to generate (non-SPC) main paths or critical path.
3. *Layered approach* He can proceed (without ignoring the weights x) through another layer of weight assignment method Y to compute weights z, which is a composite of original weights x and weights y computed by method Y, i.e., $z = f(x, y)$.

As traversal weight assignment methods like SPX methods are coherent methods, it will be great to use such methods in 'Ignore and proceed approach' as an alternative weight computation method. SPX methods basically are algorithms that assign traversal count as weight. In PAJEK, if we run SPC weight computation, earlier weights will be automatically ignored and we will get the SPC weights of all the arcs. Thus, present version gives desired results for 'Ignore and proceed approach' on weighted networks if we choose to compute SPC weights. In the present version, if we do not opt for computation of SPC weights and directly proceed to search methods, new or different paths can be obtained. Thus, present version supports 'Direct search' approach too. Therefore, no additional implementation steps are required for path retrieval using first two approaches for weighted networks. If we are to transform unweighted citation networks to weighted ones with PAJEK, there are very few options. An important area the developers of PAJEK can invest their time and energy is the implementation of arc weight assignment methods. Apart from traversal count algorithms like SPC, weight based on citation relevance/irrelevance is one of the rational choices for weight assignment. Such a weight assignment scheme can work

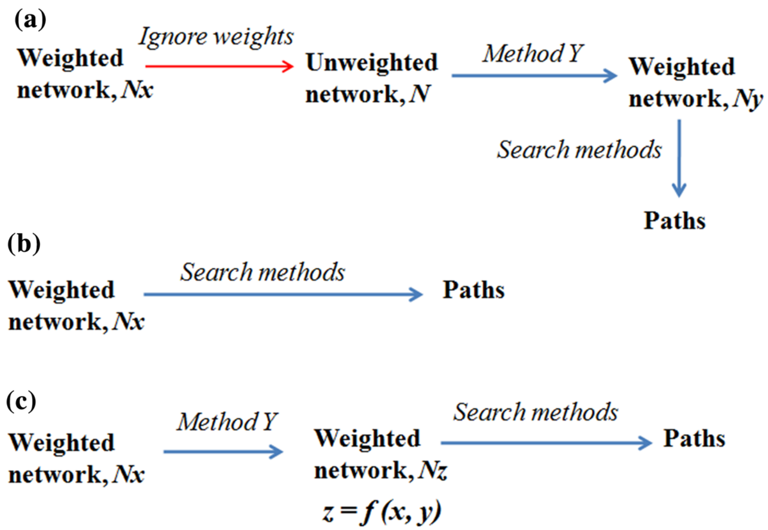


Fig. 3 **a** the ignore and proceed approach, where weights in weighted network Nx is ignored and a new method for weight assignment Y (preferably SPC) is sought, **b** the direct search approach in which weighted network Nx is used for search and generation of paths, **c** the layered approach in which method Y act as a layer that helps to convert original weights x to z

well with direct search approach as well as layered approach. For instance, the usage of relevancy of citations as a new layer in legal citations network Liu et al. (2014) after SPC computation (first layer). In that work, relevancy weights of citations were multiplied with SPC weights making $z = f(x, y) = x \times y$. However, in case of scientific citation networks, effective methods to assign relevancy of citations are not yet available. Thus, full potential of integrated approach that consist of three approaches cannot be realized now. In this work, we focus on first two approaches.

Now, a particular weight assignment scheme introduced recently in the literature is of great potential for scientometric applications. It is the 'flow vergence gradient' scheme. Before discussing this scheme, the origin of this scheme and important notions associated with this is revisited for the benefit of the readers.

FV gradient: a revisit

Citation network, a kind of information network is an interconnected structure of scholarly publications which are linked through arcs of citations. Like all the information networks, there is no physical flow among individual units (scholarly publications) through the citation links. However, the basic accord among information scientists or scientometric practitioners is that these links can be treated as a representation of the flow of information from the cited work to the citing work. Most of the vertex measures are incapable of reflecting one important property of a vertex (work) in a citation network that arises due to the flow of information through it. A paper in a network can be of the following types: (1) the ones that received knowledge from other works and is yet to transmit knowledge to any other works (2) the ones that have not received knowledge from other works, but transmitted knowledge to other works (3) the ones that received as well as transmitted knowledge (4) the ones neither received nor transmitted knowledge. Works of the fourth kind are isolated works and their merit cannot be determined by the means of the network approach. All other kinds are characterized by the flow of information which may be dominant in the cited side or the citing side. Thus a dominance in terms of *flow vergence*, i.e., either flow convergence or divergence is exhibited by a connected work in a citation network. A work can be said to be convergence driven or in flow convergence mode if its knowledge reception is greater than knowledge transmission. A work is said to be in flow divergence if its knowledge transmission is greater than knowledge reception. The property of a vertex in a citation network by the virtue of which it can be treated either in one of the flow vergence states-flow divergence or flow convergence modes is called the flow vergence property. In order to identify the flow vergence property, the basic network metrics like indegree and outdegree can be used together. These are briefly discussed as follows:

Indegree In a network, the degree of a vertex is the number of neighbours (i.e., number of vertices connected to that vertex). In a directed network, whose underlying structure is digraph, a vertex can have two types of neighbours according to the direction of links. If there are arcs directed towards the vertex of our interest from other vertices, then the total number of such neighbours is termed as the indegree. If the adjacency matrix of the network is $A = [a_{ij}]$, where a_{ij} can take value 1 only if there is an arc from i to j, then, indegree of vertex v is given by:

$$\mathrm{indeg}_v = \sum_{i=1}^{n} a_{iv} \tag{1}$$

Outdegree If there are arcs directed outwards from the vertex of our interest toward other vertices, then the total number of such neighbours is termed as the outdegree. For a network whose adjacency matrix is $A = [a_{ij}]$, the outdegree of vertex v is given by:

$$\text{outdeg}_v = \sum_{i=1}^{n} a_{vi} \tag{2}$$

In case of citation networks, as an arc represent the citation link, indegree of vertex v (i.e., paper v) gives the number of citations received by v and outdegree gives the number of citations made. As knowledge flow is regarded to be from the older or cited one towards the newer or citing one, indegree can be viewed as to reflect the outflow of knowledge and outdegree, the inflow of knowledge. As far as flow vergence property is concerned, if knowledge influx is greater than outflux, then that work can be treated as a work in flow convergence mode. Therefore, a paper with outdegree greater than indegree can be treated as the one in convergence driven mode. Similarly, a paper with indegree greater than outdegree can be treated as a divergence driven paper. Greater this difference, chance for this paper to reach newer papers through diverging paths is high. According to their connectivity to cited papers and citing ones, each paper can be placed at different levels to distinguish them. Higher the difference between knowledge outflux and influx, higher the level that paper can be placed. Now, we are in a position to attempt the definition of 'flow vergence potential' of a paper.

Flow vergence potential and the FV index

Flow Vergence potential is the ability of a paper to improve its flow vergence or remain in a high flow vergence level. Though this cannot be precisely measured, an indication of this potential is possible. The difference in flows, i.e., the difference between indegree and outdegree is the key to determine whether outflow is strong enough to overcome inflow. To express this in the scale $[-1,1]$ we divide this difference by the size of the immediate neighbourhood of the paper (degree of the paper gives the size of the immediate neighbourhood of a paper). This ratio is termed as degrees ratio of the flow vergence in Prabhakaran et al. (2015) and Lathabai et al. (2017). For a paper i it is expressed as:

$$W_{\text{dr}_{(\text{fv})_i}} = \frac{\text{indeg}_i - \text{outdeg}_i}{\text{deg}_i} = \frac{\text{indeg}_i - \text{outdeg}_i}{\text{indeg}_i + \text{outdeg}_i} \tag{3}$$

If $W_{\text{dr}_{(\text{fv})_i}} > 0$ indegree outnumbers outdegree and influx are clearly outweighed by outflux. If $W_{\text{dr}_{(\text{fv})_i}} < 0$, outdegree outnumbers indegree and outflux has not yet outweighed influx. This ratio indicates the direction of dominant flow of knowledge and it is the apparent flow vergence potential. Higher this value higher the chance for this paper to remain in its high flow vergence state or to improve its flow vergence to a higher level. Apparent flow vergence potential of the all the vertices in sample network shown in Fig. 4 (left) can be calculated as shown below and based on the scores obtained, the vertices can be distinguished according to their dominant mode of knowledge flow as convergent mode (red) and divergent mode (green), shown in 4 (right).

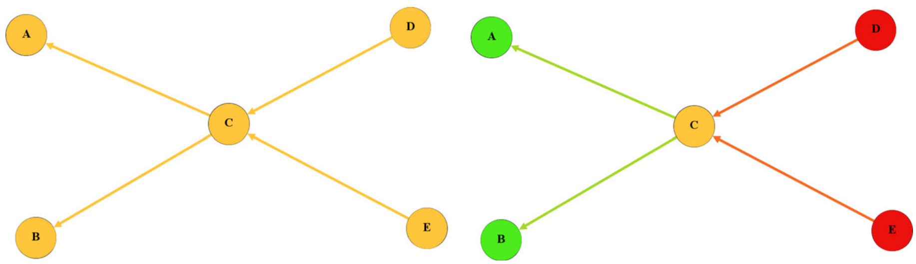


Fig. 4 (Left) sample network, (right) network with nodes distinguished using their apparent flow vergence potentials

$$\begin{aligned}
W_{\mathrm{dr}_{(\mathrm{fv})_A}} &= \frac{\mathrm{indeg}_A - \mathrm{outdeg}_A}{\mathrm{indeg}_A + \mathrm{outdeg}_A} = \frac{1-0}{1+0} = 1 \\
W_{\mathrm{dr}_{(\mathrm{fv})_B}} &= \frac{\mathrm{indeg}_B - \mathrm{outdeg}_B}{\mathrm{indeg}_B + \mathrm{outdeg}_B} = \frac{1-0}{1+0} = 1 \\
W_{\mathrm{dr}_{(\mathrm{fv})_C}} &= \frac{\mathrm{indeg}_C - \mathrm{outdeg}_C}{\mathrm{indeg}_C + \mathrm{outdeg}_C} = \frac{2-2}{2+2} = 0 \\
W_{\mathrm{dr}_{(\mathrm{fv})_D}} &= \frac{\mathrm{indeg}_D - \mathrm{outdeg}_D}{\mathrm{indeg}_D + \mathrm{outdeg}_D} = \frac{0-1}{0+1} = -1 \\
W_{\mathrm{dr}_{(\mathrm{fv})_E}} &= \frac{\mathrm{indeg}_E - \mathrm{outdeg}_E}{\mathrm{indeg}_E + \mathrm{outdeg}_E} = \frac{0-1}{0+1} = -1
\end{aligned} .$$

From the computations, vertex C is found to have zero value, which indicates a balanced flow at the moment (according to the apparent flow vergence potential score) and is not placed in divergence mode as shown in Fig. 4 (right). Also, the ratio that represents the apparent flow vergence potential does not indicate anything about the quality of citations, i.e., the quality of works that received information from this work. As the quality of citations do matter, the metric of flow vergence potential has to be designed with the incorporation of a quality-related term. Another important possibility in large networks is that for works with outdeg $= 0$, $W_{\mathrm{dr}_{(\mathrm{fv})}}$ will always be 1, irrespective of the indegree. Thus, two papers with different indegree, say indeg $= 1$ and indeg $= 30$ respectively, cannot be distinguished in terms of their contribution to the growth of the field represented by the network. In case of the first concern, a metric that could reflect the quality of connections Newman (2008) exist in the network literature-*eigenvector centrality*. Let us examine the usability of eigenvector centrality in addressing the above three drawbacks of degrees ratio of flow vergence as a measure of flow vergence potential, after going through a short description of the eigenvector centrality.

Eigenvector centrality

According to Bonacich (1972), Borgatti (2005), the eigenvector of a vertex i of a network is the corresponding principal eigenvector of the adjacency matrix of the underlying graph of the network. Principal eigenvector of a graph is the eigenvector corresponding to the principal eigenvalue, λ. For a paper i, it is mathematically expressed as:

$$\mathrm{eig}_i = \frac{1}{\lambda} \Sigma_{j \in M(j)} x_j \tag{4}$$

where, $M(j)$ is the set of neighbouring vertices, i.e., the ones connected to i and x_j represents the centrality score of neighbour j. Thus, from the equation, it is clear that eigenvector centrality of a vertex i is dependent on the centrality scores of its neighbours. Thus, for eigenvector to be high, either the concerned vertex has to be highly connected or neighbours of the vertex has to be highly connected, which implies that though more connections have an effect on eigenvector, moderate or less connected vertices can be of high eigenvector if its very few connections are highly connected. Thus a moderately or less cited work can be of high eigenvector centrality if these citations are made from works which are well cited. As the principal eigenvector is used for computation, values will always be positive and most of the program packages compute the score on a scale [0,1], reflecting the relative importance of the works. Computation of eigenvectors in large networks become affordable due to the power method of approximation, introduced long back by Hotelling (1933). This method starts by the choice of initial values for eigenvectors and upon many iterations of repeated multiplication with the adjacency matrix, these values will converge to the original eigenvector values.

Eigenvector centrality in directed acyclic networks

Adjacency matrix (with binary relations) of a directed acyclic network (DAN) is asymmetric and matrix equation based solution (analytic solution) often ends up with zero values for all the vertices. This mathematical problem in directed acyclic network is known as the zero-centrality problem. This is one of the situations where analytical solutions do not represent the nature of reality. Especially in case of citation networks, some of the well cited works or works that received citations from well cited works have to be more important in the network than others. Due to this problem with analytic approach for eigenvectors, many researchers regard eigenvector centrality itself as 'useless' for analysis of DANs. Usability of eigenvector centrality cannot be ruled against if any alternative ways of computation of the same exist that can *better reflect the nature of reality*. Approximation methods such as Hotelling's procedure can be used as an alternative in the following way.

In the iterative procedure, initially all the vertices are assigned a value 1, which creates an initial vector $\mathrm{Eig}(0) = [1, 1, \ldots 1]^T$, is a row vector of size N. That is, for every i, $\mathrm{eig}_i(0) = 1$. Now during the first iteration, eigenvector values are computed as normalized product of adjacency matrix and the initial vector of values 1. This can illustrated as:

$$\mathrm{Eig}(1) = \frac{A\mathrm{Eig}(0)}{\max\{A\mathrm{Eig}(0)\}}$$

With this step itself, a distinction between vertices is achieved and with each iteration the eigenvector values are improved for relative importance. Now in second iteration, adjacency matrix is again multiplied to the previous product and normalized again. This is same as multiplying A^2 to the initial vector Eig(0). Eigenvector values at step k is computed as:

$$\mathrm{Eig}(k) = \frac{A\mathrm{Eig}(k-1)}{\max\{A\mathrm{Eig}(k-1)\}}$$
$$= \frac{A^k\mathrm{Eig}(0)}{\max\{A^k\mathrm{Eig}(0)\}}$$

As the iterative procedure is supposed to converge to the values that can be found analytically, this procedure ultimately yield zero values for all vertices at some larger k. In that case the distinction of vertices is lost again. So the nature of reality (existence of distinction among vertices by virtue of their association with influential vertices) lies neither with small values of k nor with extremely large values of k. By setting a threshold for a suitable parameter as done in most of the iterative procedures, eigenvector centrality values that distinguish works in citation networks can be found out by stopping iterations before the procedure cease to distinguish them. In software packages like Gephi, the parameter sum change ($\triangle S$) is used to monitor the improvement in computation.

Prior to iterations, the sum of centrality values [in vector Eig(0)] will be N. After first iteration, sum of centrality values will drop to a considerable extent as many of the uncited works attain zero value for centrality and others will attain a score according to their relative importance. Then sum change is the difference between sum of centrality values prior to the first iteration and the sum of centrality values after the first iteration.

$$\triangle S_1 = \sum \mathrm{Eig}(0) - \sum \mathrm{Eig}(1) = \sum_{i=1}^{N} \mathrm{eig}_i(0) - \sum_{i=1}^{N} \mathrm{eig}_i(1)$$

After k iterations, the sum change will be:

$$\triangle S_k = \sum \mathrm{Eig}(k-1) - \sum \mathrm{Eig}(k) = \sum_{i=1}^{N} \mathrm{eig}_i(k-1) - \sum_{i=1}^{N} \mathrm{eig}_i(k)$$

Highest value of sum change can be witnessed after iteration 1 because of the drastic falls in centrality values. From second iterations onward, value of sum change decreases. Once sum change values drop below unity, pace of decrease of sum change also slows down. Change in sum change after iteration k can be expressed as:

$$\triangle(\triangle S_k) = \triangle S_{k-1} - \triangle S_k$$

A sample variation in these parameters are plotted against number of iterations and is shown in Fig. 5. The change in pace of the drop in sum change goes on decreasing and thus practically the convergence of eigenvector values for all vertices to zero cannot be witnessed even with millions or more iterations. However, once $\triangle S$ attain value below unity (note point A in Fig. 5), as the change in $\triangle S$ slows down, an impression of stability in centrality values can be felt. Even if that cannot be treated purely as stability, such region ($k > n$, for not so high values of k, shown in Fig. 5) can at least be treated as a *metastable* region. Stopping the procedure anywhere in metastable region might yield eigenvector values that help to distinguish most of the vertices (especially the ones that have same values for indegree) which takes us closer to reality. Usually, within 100 iterations, $\triangle S$ values will drop below unity and thus $k = 100$ can be regarded as a safe stopping point for iterations. One more evidence for the stability of eigenvector values in the region is

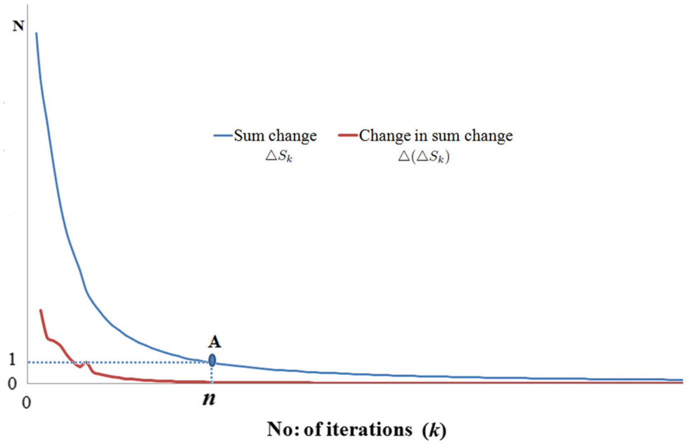


Fig. 5 Change in sum and change in pace of sum change of eigenvector centralities with number of iterations

discussed in "Metastability of eigenvector and FV index computation" section. Before that, the benefit of eigenvector incorporation in flow vergence index is discussed.

How eigenvector incorporation enables a better design of flow vergence index?

If a work is having $W_{\mathrm{dr}_{(\mathrm{fv})}} < 0$, apparently it is having less flow vergence potential and is in flow convergence mode. However, by the virtue of its quality of connections (if the citations are from much-cited works), if its eigenvector is high, can more confidence be placed over such work's ability to improve its flow vergence level?. If that is the case, then such works definitely have a greater flow vergence potential than the apparent one indicated by degrees ratio. How much higher value of eigenvector can indicate that papers which are apparently in low flow vergence are actually eligible to be in a higher level of flow vergence or in flow divergence mode?. Our observation on many datasets indicated that papers with $W_{\mathrm{dr}_{(\mathrm{fv})}} < 0$, improve their performance (either in terms of citations or quality of citations or both) if eig $> \| W_{\mathrm{dr}_{(\mathrm{fv})}} \|$. In other words, if addition of *eig* value to the $W_{\mathrm{dr}_{(\mathrm{fv})}}$ makes the resultant to take a positive value and if such papers are found to be improving their performance, then their apparent flow vergence potential clearly be an underestimated one and these papers may go unnoticed on assessment based on metrics like indegree or eigenvector on their standalone use. Thus, the addition of *eig* to $W_{\mathrm{dr}_{(\mathrm{fv})}}$ occurred to be a better option for the design of a metric for the flow vergence potential.

In case of two papers having outdeg $= 0$, $W_{\mathrm{dr}_{(\mathrm{fv})}} = 1$ and indegrees 1 and 30 respectively, their eigenvectors will be different. If all the 30 connections of the second paper are well or moderately cited, then that one may have a greater *eig* value than the first paper. If the eigenvector value of the first paper is too high so as to overcome the eigenvector of the second paper despite its high connectivity advantage, the first paper will be of extreme noteworthiness. In both cases, the addition of *eig* to $W_{\mathrm{dr}_{(\mathrm{fv})}}$ helps to distinguish two papers that were showing equal flow vergence potential apparently. Thus, the index for flow vergence potential of a paper or the FV index can be rightly expressed as given is Prabhakaran et al. (2015) as:

$$W_{\mathrm{FV}_i} = \frac{\mathrm{indeg}_i - \mathrm{outdeg}_i}{\mathrm{indeg}_i + \mathrm{outdeg}_i} + \mathrm{eig}_i \tag{5}$$

Now, a paper can be said to be in flow convergence mode if it's $W_{\mathrm{FV}} < 0$. That means, even by the addition of *eig* to the degrees ratio, if it does not take a positive value and thereby indicating no outweighing of the inflow by the outflow of knowledge, such a paper is to be treated as in flow convergence mode. If $W_{\mathrm{FV}} > 0$, such papers can be treated as in flow divergence mode, because either they have achieved sufficient level of flow vergence to be treated as a knowledge divergence paper or they are likely to improve their flow vergence to sufficient level or more. In the sample network shown in Fig. 4, let us examine how eigenvector helps to overcome the third problem (zero flow vergence problem). Using Eq. 5, the flow vergence index values of vertices can be computed as:

$$W_{\mathrm{FV}_A} = W_{\mathrm{dr}_{(\mathrm{fv})_A}} + \mathrm{eig}_A = 1 + 1 = 2$$

$$W_{\mathrm{FV}_B} = W_{\mathrm{dr}_{(\mathrm{fv})_B}} + \mathrm{eig}_B = 1 + 1 = 2$$

$$W_{\mathrm{FV}_C} = W_{\mathrm{dr}_{(\mathrm{fv})_C}} + \mathrm{eig}_C = 0 + .099 = .099$$

$$W_{\mathrm{FV}_D} = W_{\mathrm{dr}_{(\mathrm{fv})_D}} + \mathrm{eig}_D = -1 + 0 = -1$$

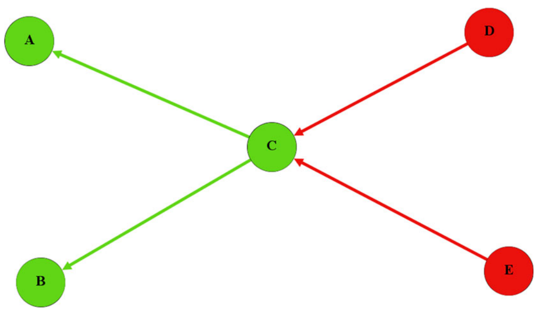
$$W_{\mathrm{FV}_E} = W_{\mathrm{dr}_{(\mathrm{fv})_E}} + \mathrm{eig}_E = -1 + 0 = -1$$

Now paper C can be treated as a work in flow divergence mode as its flow vergence potential is greater than zero and this is indicated in Fig. 6. Thus, flow vergence index design for the purpose of indication of flow vergence potential using eigenvector along with degrees ratio is a much better design over the one with degrees ratio alone because all the three problems with degrees ratio are overcome by this design. As mentioned in the previous section, stability (metastability) of the eigenvector is very important for incorporation of eigenvector in FV index. This is discussed next.

Metastability of eigenvector and FV index computation

As we have discussed earlier, in most of the citation networks there will be papers whose $W_{\mathrm{dr}_{(\mathrm{fv})}} < 0$ but whose eig $> ||W_{\mathrm{dr}_{(\mathrm{fv})}}||$ values. Such papers can be treated as underestimated papers (for their flow vergence potential) if apparent flow vergence (i.e., $W_{\mathrm{dr}_{(\mathrm{fv})}}$) was used as an indicator of flow vergence potential. Stability or metastability is therefore of utmost importance for the reliable computation of FV index. The choice of stopping point of

Fig. 6 Sample network with nodes distinguished using their flow vergence potentials computed through FV indices



iteration in the metastable region (i.e., when k crosses n but not too high) can be tested for its reliability using the number of underestimated papers. If generally in the metastable region and particularly prior to the chosen stopping point, a stability in the number of underestimated papers that are raised to the flow divergence mode by incorporation of eigenvector to the apparent flow vergence is visible, then the reliability of stopping the iteration in the metastable region can be confirmed. Prior to iteration as every vertex will have eig $= 1$, no papers exhibit convergence mode and out of N all those papers with $W_{\mathrm{dr}_{(\mathrm{fv})}} < 0$ can be treated as to be raised to divergence mode. Let that be N' where $N' < N$. Now, with first iteration itself, this value falls drastically and with further iterations, less drastic falls with occasional fluctuations can be witnessed. After some iterations, a stability is usually visible in the number of underestimated papers that can be raised to divergence mode. The starting point of such a stability, n' will be close to the beginning of metastable region as shown in Fig. 7, and hence n' won't be much higher. Thus, stopping the iteration procedure for eigenvector computation can be done for any $k > n'$, for not so high values of k. This n' is also usually attained well before 100 iterations and hence 100 will be a guaranteed safe stopping point.

Flow vergence gradient or FV gradient

Upon its publication, a paper cites one or more works in the network and after publication, it may receive one or more citations (sooner or later) from the newly published works. In citation networks, usually, knowledge flows are observed from papers of relatively high flow vergence potential to relatively low flow vergence potential. The difference in this flow vergence potential can be termed as the flow vergence gradient or FV gradient. For every pair of papers linked through an arc of citation, there exists a potential difference or flow vergence gradient. As the flow vergence potential is reflected by the FV index, the difference in FV indices of such pairs gives the FV gradient. For a citation arc l_{vu} from paper v to u, this gradient of flow vergences can be computed as the difference of FV index of the cited paper u and the FV index of the citing paper v and can be expressed as:

$$\triangle_{\mathrm{FV}_{uv}} = W_{\mathrm{FV}_u} - W_{\mathrm{FV}_v} \tag{6}$$

Normally, u will be having higher FV index than v, and hence $\triangle_{\mathrm{FV}_{uv}}$ for arc l_{vu} will be a positive value. However, if citing paper v earn much receptivity, its FV index builds up and

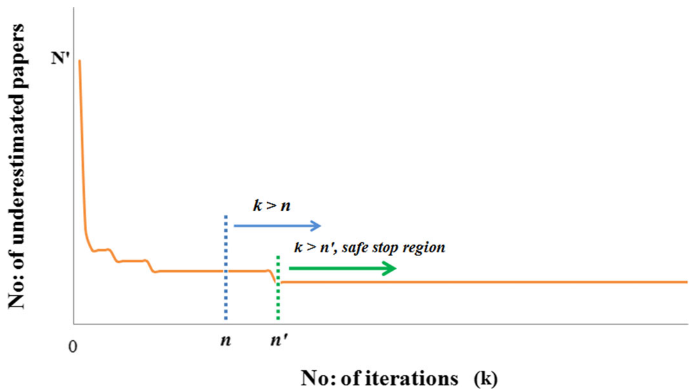


Fig. 7 Change in number of underestimated papers raised to flow divergence state (by eigenvector addition) with the number of iterations

overcome the flow vergence potential of cited paper u, then knowledge flow is visible as occurred from paper with relatively low FV index to a paper with relatively high FV index. This phenomenon was termed in Lathabai et al. (2015) as *flow vergence effect* (FV effect). This kind of reverse reset can happen if the citing paper in a FV effect pair has a strong 'implicit intellectual excellence'. This phenomenon can be used to detect the important knowledge flow associated with a paradigm shift. In the case of FV effect, the corresponding arc l_{vu}, takes a negative value for FV gradient, i.e., $\triangle_{\mathrm{FV}_{uv}} < 0$. Once the most crucial arc associated with a paradigm shift is identified, pivot paper can be identified as the source paper of that arc, as mentioned in Lathabai et al. (2015). Thus, as negative value for FV gradient can be used to identify FV effect, it can be used as a detector of pivot papers of paradigm shift.

In the network of 'Biotechnology for engineering' used in Lathabai et al. (2015), though all the arcs exhibiting FV effect can be identified using 6, only those that were present in main path and critical path were analysed for its content as part of that work. In that work, original SPC methods (using older version of PAJEK) were used for generating main path and critical path. In this work, the potential of FV gradient as a method for weight assignment or its ability to create a weighted network is explored. This network can be used for generation of different kind of main paths and critical path and such a path will reflect a continuum of crucial knowledge flows as some arcs are capable of reflecting the flow vergence effect. Another important thing to be noted is that negative or low value of FV gradients are better for reflecting such crucial flow of knowledge and hence such arcs should be given more priority. Thus a transformation of the following kind is necessary.

Let $\triangle_{\mathrm{FV}}$ represent the set of FV gradients of all arcs, then $\max.\triangle_{\mathrm{FV}} = \max\{\triangle_{\mathrm{FV}}\}$ and $\min.\triangle_{\mathrm{FV}} = \min\{\triangle_{\mathrm{FV}}\}$.

$$\triangle_{\mathrm{FV}_{ij}(\mathrm{norm.})} = 1 + \frac{\max.\triangle_{\mathrm{FV}} - \triangle_{\mathrm{FV}_{ij}}}{\max.\triangle_{\mathrm{FV}} - \min.\triangle_{\mathrm{FV}}} \tag{7}$$

This transformation helps to express the FV gradients of arcs in the range [1,2]. For the sample network shown in Fig. 4 (left), FV gradients of all the four arcs can be computed as:

$$\begin{aligned}
\triangle_{\mathrm{FV}_{AC}} &= W_{\mathrm{FV}_A} - W_{\mathrm{FV}_C} = 2 - .099 = 1.901\\
\triangle_{\mathrm{FV}_{BC}} &= W_{\mathrm{FV}_B} - W_{\mathrm{FV}_C} = 2 - .099 = 1.901\\
\triangle_{\mathrm{FV}_{CD}} &= W_{\mathrm{FV}_C} - W_{\mathrm{FV}_D} = .099 - (-1) = 1.099\\
\triangle_{\mathrm{FV}_{CE}} &= W_{\mathrm{FV}_C} - W_{\mathrm{FV}_E} = .099 - (-1) = 1.099
\end{aligned}$$

Maximum value among the FV gradients of all the four arcs, $\max.\triangle_{\mathrm{FV}}$= 1.901 and the minimum value among the FV gradients of four arcs, $\min.\triangle_{\mathrm{FV}}$= 1.099. Therefore, the after transformation expressed in Eq. 7, the values of all the four arcs can be computed as:

$$\begin{aligned}
\triangle_{\mathrm{FV}_{AC}(\mathrm{norm.})} &= 1 + \frac{\max.\triangle_{\mathrm{FV}} - \triangle_{\mathrm{FV}_{AC}}}{\max.\triangle_{\mathrm{FV}} - \min.\triangle_{\mathrm{FV}}} = 1 + \frac{1.901 - 1.901}{1.901 - 1.099} = 1\\
\triangle_{\mathrm{FV}_{BC}(\mathrm{norm.})} &= 1 + \frac{\max.\triangle_{\mathrm{FV}} - \triangle_{\mathrm{FV}_{BC}}}{\max.\triangle_{\mathrm{FV}} - \min.\triangle_{\mathrm{FV}}} = 1 + \frac{1.901 - 1.901}{1.901 - 1.099} = 1\\
\triangle_{\mathrm{FV}_{CD}(\mathrm{norm.})} &= 1 + \frac{\max.\triangle_{\mathrm{FV}} - \triangle_{\mathrm{FV}_{CD}}}{\max.\triangle_{\mathrm{FV}} - \min.\triangle_{\mathrm{FV}}} = 1 + \frac{1.901 - 1.099}{1.901 - 1.099} = 2\\
\triangle_{\mathrm{FV}_{CE}(\mathrm{norm.})} &= 1 + \frac{\max.\triangle_{\mathrm{FV}} - \triangle_{\mathrm{FV}_{CE}}}{\max.\triangle_{\mathrm{FV}} - \min.\triangle_{\mathrm{FV}}} = 1 + \frac{1.901 - 1.099}{1.901 - 1.099} = 2
\end{aligned}$$

The exercise of computation of flow vergence gradient and its normalized values will be more interesting if there are negative values and will be more useful if carried out on a network with much larger number of nodes as done in illustration and case study sections of Lathabai et al. (2015). However, as we are interested to explore how FV gradient weighted (normalised) citation networks can be used to retrieve more evolutionary trajectories, this exercise is done on the sample network shown in Fig. 1 and is discussed in the following section.

Path analysis on FV gradient weighted (normalised) network

Upon FV gradient computation, unweighted citation networks are converted to FV gradient weighted network and then weights are transformed to obtain FV gradient weighted (normalised) network using Eqs. 5, 6 and 7 respectively. Both these networks are shown in Fig. 8.

As discussed in "Weighted citation networks" section, two main approaches are possible for analysis of weighted networks-direct approach and Multi-layered approach. Firstly, the paths that can be retrieved if direct search methods are applied on FV gradient weighted (normalised) network is examined and compared against the ones that obtained using Liu–Lu approach or ignore and proceed approach, which are already discussed in "Search schemes" section. For finding the ability of new approach to retrieve unexplored paths or to highlight different papers over Liu–Lu approach, an indicator namely, *uniqueness index* is proposed. The idea behind this index is that two paths or subnetworks will be more distinct if more number of unique papers are present in the pair or if there are less number of common papers in it. Unique index or U index for a pair of paths P_i and P_j can be expressed as:

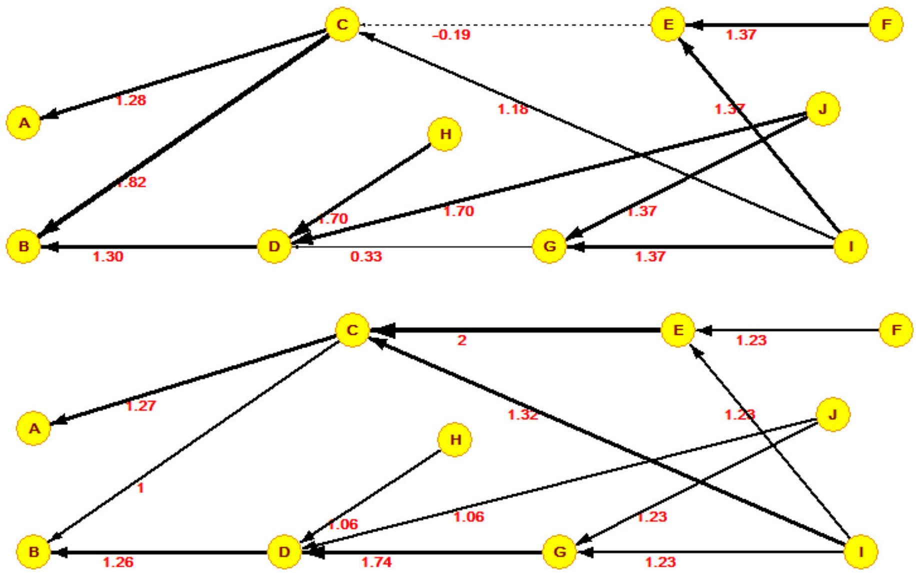


Fig. 8 (Top) FV gradient weighted sample network and (bottom) FV gradient weighted (normalised) sample network. Note that negative and low valued arcs (e.g., arc EC) obtained more importance after transformation

$$U_{P_iP_j} = \frac{n_U}{n_{P_i} + n_{P_j}} = \frac{n_{P_i} + n_{P_j} - n_C}{n_{P_i} + n_{P_j}} = 1 - \frac{n_C}{n_{P_i} + n_{P_j}} \tag{8}$$

where n_{P_i} and n_{P_j} are number of papers in subnetwork P_i and P_j respectively. n_U is the number of number of unique papers and n_C is the number of common papers.

U index can take value between 0.5 and 1. If both the subnetworks are entirely same, then $n_U = n_C = n_{p_i} = n_{P_j}$. Therefore, $U_{P_iP_j}$ will be 0.5. On the other hand, if both the subnetworks are entirely distinct, n_C=0 and $n_U = n_{P_i} + n_{P_j}$, then $U_{P_iP_j}$=1.

Interestingly, if one of the path is a subpath or subnetwork of the other, then $n_C = \min\{n_{P_i}, n_{P_j}\}$ and therefore,

$$U_{P_iP_j} = \frac{n_{P_i} + n_{P_j} - n_C}{n_{P_i} + n_{P_j}} = \frac{n_{P_i} + n_{P_j} - \min\{n_{P_i}, n_{P_j}\}}{n_{P_i} + n_{P_j}} = \frac{\max\{n_{P_i}, n_{P_j}\}}{n_{P_i} + n_{P_j}}$$

Direct method to path analysis using FV gradient

Paths obtained as a result of application of various search methods on FV gradient weighted (normalised) citation networks are shown in Fig. 9 and these are termed together as FV paths after the weight assignment method used. It has to be noted that for all the above path retrieval, tolerance is set to zero and in case of key-route, only the arc with highest FV gradient weight (normalised) is used. In case of multiple Key-route paths, FV gradient weight assignment scheme can provide a simple criterion for the choice of number of key-routes (top weighted arcs). Before transforming FV gradient weights to obtain positive values for shifting the prominence to lower weighted arcs, number of such 'zero-crossed' or negative arcs can be noted down. These will be the arcs that exhibited flow vergence effect. This number can be used as a choice for the parameter value which corresponds to the selection of number of key-routes to initiate local or global search. On a quick comparison to Liu–Lu paths based on SPC, FV paths are found to retrieved paths that are significantly distinct. More information on this can be obtained only after U index computation. Once path retrieval using weighted integrated approach is completed and computation of U indices, results of U indices for all the two pairs of paths (obtained for different search methods) can be compared. From Table 1, the newly introduced weight assignment scheme (FV gradient) are found to produce paths with much uniqueness to the SPC scheme, which is an important observation. Most similar paths are retrieved when search methods like critical path and local forward search. Being a global search method, critical path method can be expected to provide more similar results for different weight assignment schemes. However, the observation with local forward search may be specific to this network only. This cannot be always expected in large networks. For backward search and Key-route (local and global) search, U indices show high values, indicating a high level of distinction. Even from the analysis of the small network, the capability of a different weight assignment scheme to cause retrieval of unique paths or to draw our attention towards more unnoticed papers is revealed and more clarity can be achieved on the execution of these methods in large real-world networks.

Now, in the case study of the network of 'IT for engineering', which is a much large network with 3705 papers, after creating an FV gradient weighted network using two approaches (Ignore and proceed approach that gives SPC paths and Direct approach that provides FV paths), we compare SPC paths and FV paths. This comparison can be

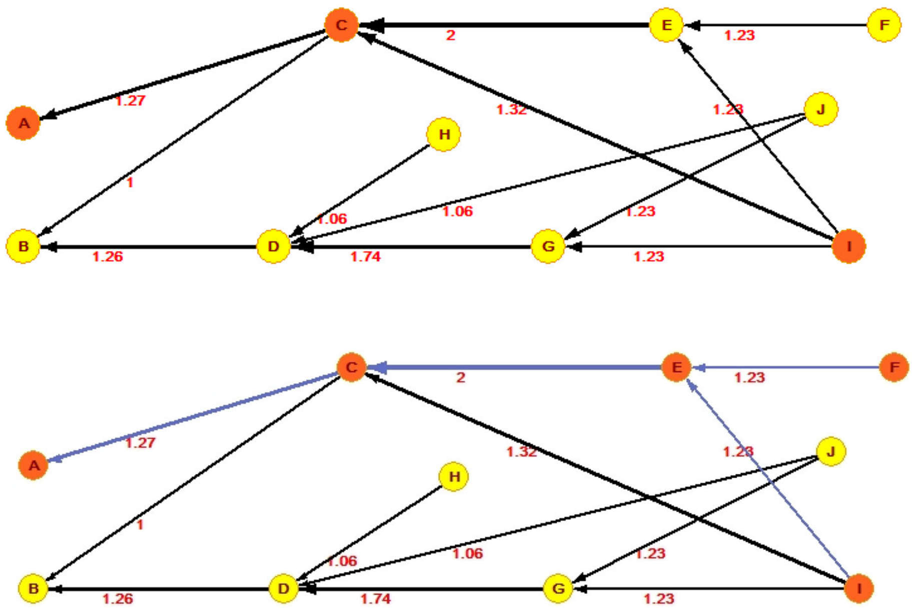


Fig. 9 (Top) forward local FV path obtained using forward search and (bottom) Same subnetwork obtained for Backward path search, Local and global key-route search and critical path search. Blue arcs and orange vertices highlight the retrieved paths. (Color figure online)

Table 1 *U* indices for the pair of paths obtained using two weight assignment methods and different search methods

Search method	*U* index for SPC and FV paths
Forward	0.66
Backward	0.9
Key-route (local)	0.9
Key-route (global)	0.9
Critical path method	0.55

generalized to SPX methods if SPC method and other SPX methods are found to retrieve almost same paths. If enough uniqueness is seen in paths retrieved using the direct approach, the integration of two approaches for path analysis is of great value to a multitude of beneficiaries and the strength of path analysis for scientometric applications can be improved manifold.

Case study of IT for engineering

Data and network

Citation network of 'Information technology for Engineering' analysed in Prabhakaran et al. (2015) is used here. Data was retrieved from WoS (Web of Science) for the period 1st Jan, 1999 to 28th Feb, 2013 using keyword 'Information Technology' and research area

'Engineering'. Original network consisted of 11645 papers and 4576 arcs. The minimal core Stephen (1983) (core with $k = 1$) of the citation network, which do not contain isolated papers is shown in Fig. 10. It consists of 3705 papers and 4576 arcs. Firstly, we identified all the main paths and critical path using SPC method in combination with various search schemes available in Liu–Lu approach. On our check (results are not shown), paths retrieved by SPC method is found to be almost same as that of other SPX methods, making SPC method a suitable representative of SPX methods. Then, the unweighted network is converted into weighted network after computation of FV gradients of all arcs and transforming those weights using Eq. 7. For that, eigenvector algorithm is run for 100 iterations (sum change value = 0.148951575 at $k = 100$). In Fig. 11a, b, the identification of values of n and n' (i.e., $n \approx 28$ and $n' = 40$) are depicted. Thus choice of $k = 100$ is justified. It is to be noted that as $N = 3705$, we obtained sum change after first iteration, $\triangle S_1$= 3547.345 and change in sum change after second iteration, $\triangle(\triangle S_2)$= 3539.76. Prior to the first iteration, number of underestimated papers that can be raised (i.e., papers with $w_{\mathrm{dr}_{(\mathrm{fv})}} < 0$) is found to be 2017. For legibility, we have excluded these huge values and the dynamics of $\triangle S_k$ from $k = 2$, $\triangle(\triangle S_k)$ from $k = 3$ is shown in Fig. 11a and dynamics of number of underestimated papers that can be raised is shown from $k = 1$ in Fig. 11b.

On the FV gradient (norm.) weighted network, different search schemes are applied and retrieved main paths and critical path, which are termed as FV paths. Content analysis of SPC paths and FV paths retrieved on various search schemes is conducted next and very brief description of key papers and trajectories are given.

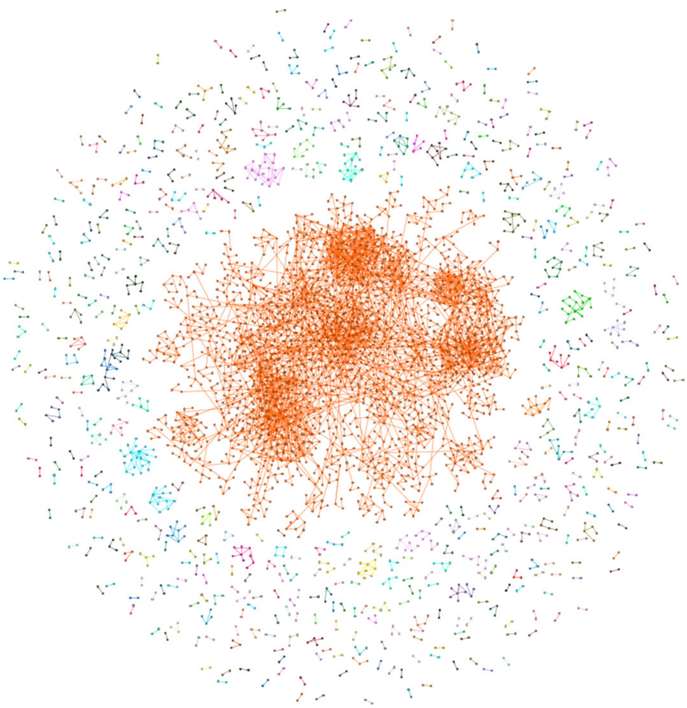


Fig. 10 Minimal core of network of 'IT for Engineering' used in Prabhakaran et al. (2015)

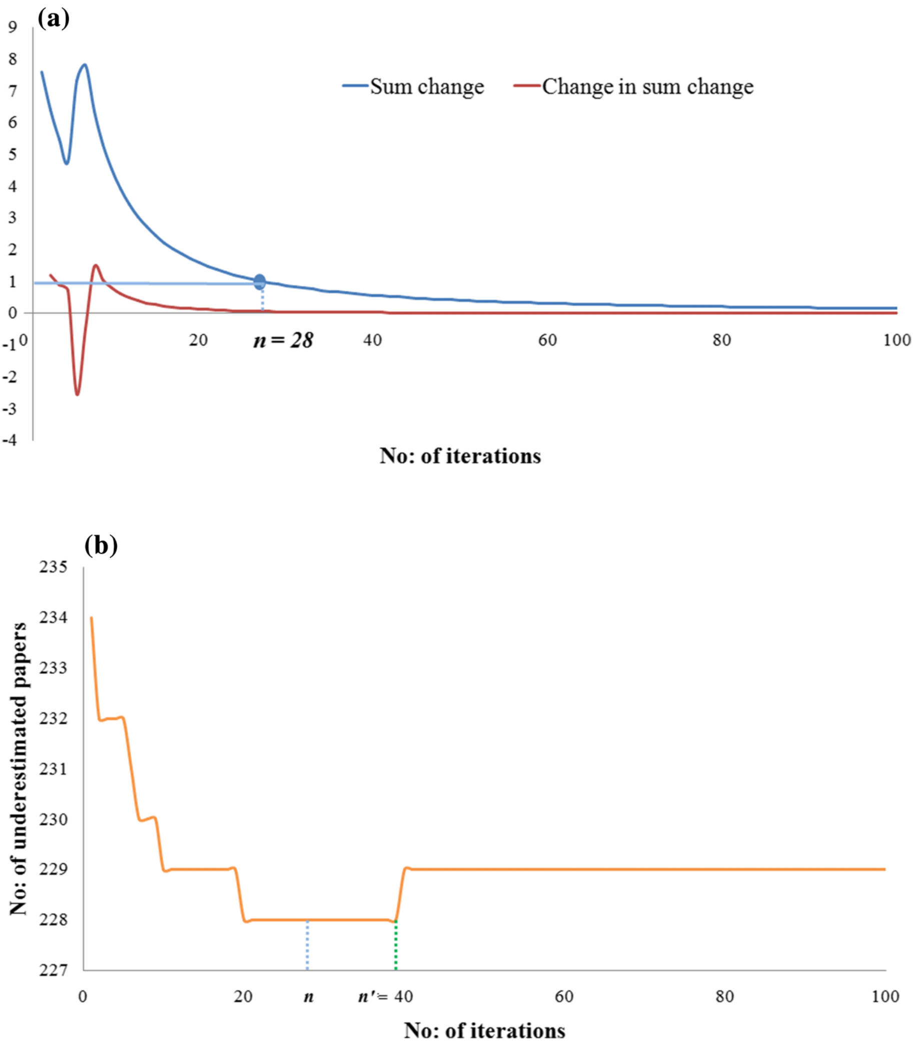


Fig. 11 Identification of safe stopping point of iterations for eigenvector centrality

Local forward FV and SPC path

When local forward search is applied on FV gradient weighted (normalised) and SPC weighted citation network of 'IT for engineering', paths shown in Fig. 12a, b are obtained.

Content analysis of SPC forward local main path (Fig. 12b) can be found in Prabhakaran et al. (2015). It clearly indicated a shift towards RFID technology for supply chain management and its capability to enhance product-service delivery. However, papers in FV forward local main paths (there are two disjoint main paths retrieved as shown in 12a) deal with different but specific aspects and benefits of RFID technology. The first path from 4881 Nuntasunti S, 2008 to 11214 Yang HJ, 2012 deals with how RFID technology enhances construction or civil engineering project management. Another one that extends from 1844 Albino V, 2002 to 10406 Huang GQ,2012 discussed the potential of RFID to enhance production processes itself and special emphasis was laid on its use in automotive part and accessory manufacturing sectors. Thus papers retrieved using FV gradient weight assignment scheme are also of good quality and highlighted previously unnoticed works related to RFID technology. Titles of all the papers in FV paths can be seen in Table 2.

Local backward FV and SPC path

Most of the papers in SPC backward path (Fig. 13 right) addressed the theme 'business and operational enhancement' by RFID technology and various IT innovations such as ERP (Enterprise Resource Planning). Several key papers missed in forward search is retrieved by backward search. Some of them such as 5756 Whitaker J, 2007, 5757 Delen D, 2007 and 8933 Sarac A, 2010 were identified in Prabhakaran et al. (2015) as key papers that described a paradigm shift that is powered by RFID technology. Some other themes addressed by papers in SPC backward path are RFID impact in healthcare, RFID and sensor technology based 'end-of-life management in closed loop supply chains'. On the other hand, FV backward path papers deal with an entirely different but important aspect of IT-'Human–Computer interaction'. Evolution of human–computer interface technologies through various information systems, web-based systems, other important communication systems, expert systems, etc., can be understood in this path (Fig. 13 left).

Key-route FV and SPC paths

For SPC weighted citation network, local key-route search and global key-route search (with parameter values for tolerance = 0 and number of key-routes = 1) retrieved more or

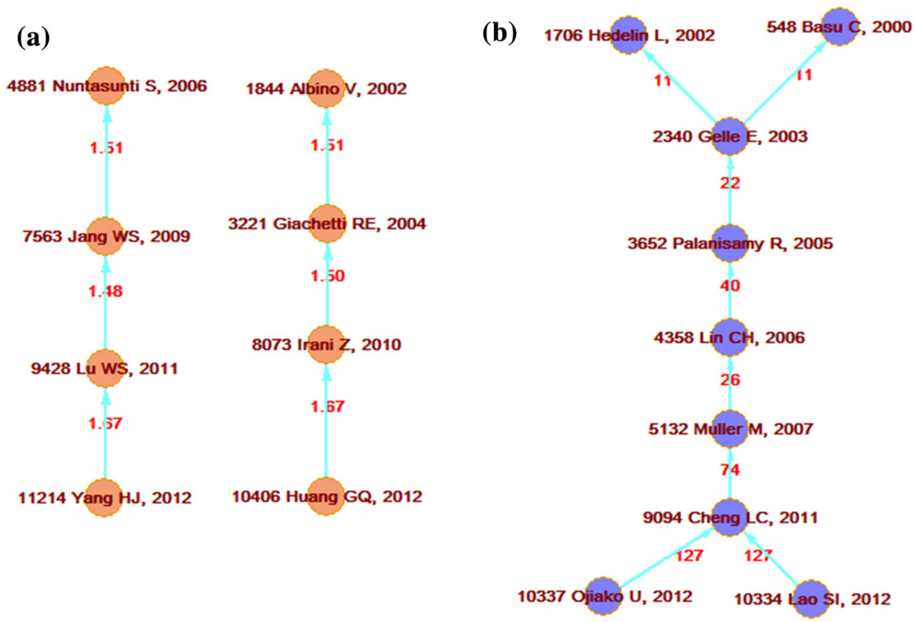


Fig. 12 **a** FV forward paths and **b** SPC forward path of 'IT for Engineering' (analysed in Prabhakaran et al. 2015)

Table 2 Unique papers found in various FV paths and their titles

Paper id.	FV path	Title
4881 Nuntasunti S, 2006	FV forward local	Experimental assessment of wireless construction technologies
7563 Jang WS, 2009	FV forward local	Embedded System for Construction Asset Tracking Combining Radio and Ultrasound Signals
9428 Lu WS, 2011	FV forward local	Scenarios for applying RFID technology in construction project management
11214 Yang HJ, 2012	FV forward local	Design and implementation of an identification system in construction site safety for proactive accident prevention
1844 Albino V, 2002	FV forward local	Analysis of information flows to enhance the coordination of production processes
3221 Giachetti RE, 2004	FV forward local	A framework to review the information integration of the enterprise
8073 Irani Z, 2010	FV forward local	Radio frequency identification (RFID): research trends and framework
10406 Huang GQ, 2012	FV forward local	RFID-enabled product-service system for automotive part and accessory manufacturing alliances
840 Venkatesh V, 2000	FV backward local	Creating an effective training environment for enhancing telework
2849 Yi MY, 2003	FV backward local	Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model
4871 Thong JYL, 2006	FV backward local	The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance
5553 Kim HW, 2007	FV backward local	A balanced thinking-feelings model of information systems continuance
7763 Cyr D, 2009	FV backward local	Perceived interactivity leading to e-loyalty: development of a model for cognitive-affective user responses
9854 Kim J, 2011	FV backward local	Antecedents of application service continuance: a synthesis of satisfaction and trust
10740 Lin JL, 2012	FV backward local	A Tale of Four Functions in a Multifunctional Device: Extending Implementation Intention Theory
269 Gunasekaran A, 1999	FV key-route	Agile,manufacturing: a framework for research and development
1850 Yoo SB, 2002	FV key-route	Web-based knowledge management for,sharing product data in virtual enterprises
2731 Liao SH, 2003	FV key-route	Knowledge management technologies and applications- literature review from 1995 to 2002

Table 2 continued

Paper id.	FV path	Title
3931 Liao SH, 2005	FV key-route	Technology management methodologies and applications-a literature review from 1995 to 2003
5706 Wang TY, 2007	FV key-route	The influences of technology development on economic performance-the example of ASEAN countries
7208 Badescu M, 2009	FV key-route	The impact of information technologies on firm productivity: empirical evidence from Spain
7495 Harbi S, 2009	FV key-route	Establishing high-tech industry: the Tunisian ICT experience
9680 Chien SC, 2011	FV key-route	Building the measurement framework of technology efficiency with technology development and management capability-evidence from the ASEAN countries

less the same papers as that of backward local search. The global key-route main path (Fig. 15 left) is found to be a subnetwork of the local key-route main path (Fig. 14 right). Papers that are not present in backward path dealt mainly with ‘challenges in operations management’ and other production related aspects. These are some among the many roots of the evolution of RFID technology for production and supply chain management. On the other hand, FV gradient weighted (normalised) network gave exactly same subnetwork on local and global key-route search (shown in Fig. 14 left). These papers mainly dealt with the use and practice of IT for knowledge management and technology management in industries and firms for business enhancement as well as overall performance enrichment. This again establishes the capability of FV gradient scheme to cause retrieval of unique

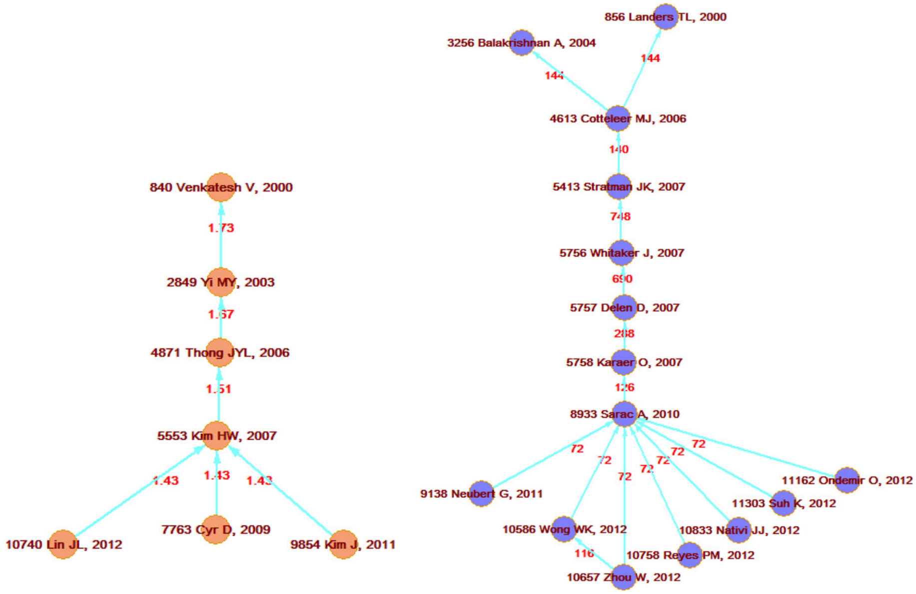


Fig. 13 (Left) FV backward path and (right) SPC backward path of ‘IT for Engineering’

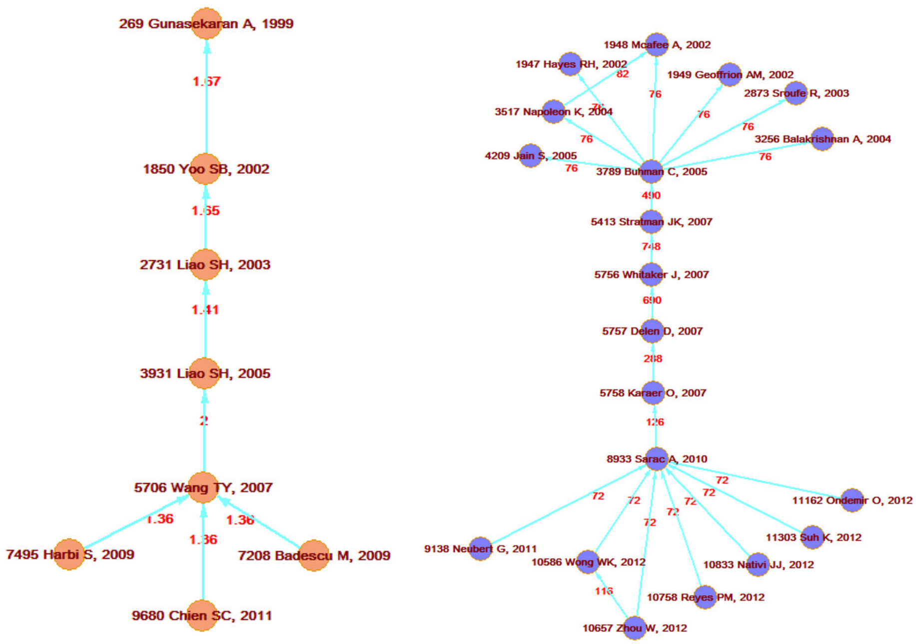


Fig. 14 (left) FV key-route main path (local and global), (right) SPC key-route (local) main path of 'IT for Engineering'

papers with different themes while SPC weighted networks retrieve paths with more or less the same theme on different search methods. Another important aspect of the FV gradient based scheme is that if the analysts are craving for multiple path retrievals using multiple key-routes, then the choice of parameter value can be made based on the number of arcs with negative values for FV gradient weight. As these arcs exhibit 'FV effect', most of the important developments in the network can be covered using the paths retrieved by that choice. In the case of our network, there are 63 arcs exhibiting 'FV effect'. Using these 63 key-routes (choice of parameter value 1–63), a subnetwork as shown in Fig. 15 right can be obtained. There are 27 connected components and this will be sufficient for a fair cluster analysis. With these many paths, it is hard to miss most important developments in the field. However, analyses of 301 papers in these 27 clusters are beyond the scope of this work and also not our present objective. This is included as a demonstration of the power of FV gradient method as a weight assignment scheme.

FV critical path and SPC critical path

On application of critical paths in networks weighted using both schemes, FV critical path and SPC critical path are obtained. Interestingly both of these are found to be exactly the same as shown in the depiction in Fig 16. Thus, the structure of this network is such that at the global level different weight assignment schemes seem to pick the same set of critically important papers. SPC critical path was analysed for its content in Prabhakaran et al. (2015). As most of the papers are found in SPC backward, forward and key-route main paths, the evolutionary trajectory shown by the critical path also revolves around the theme

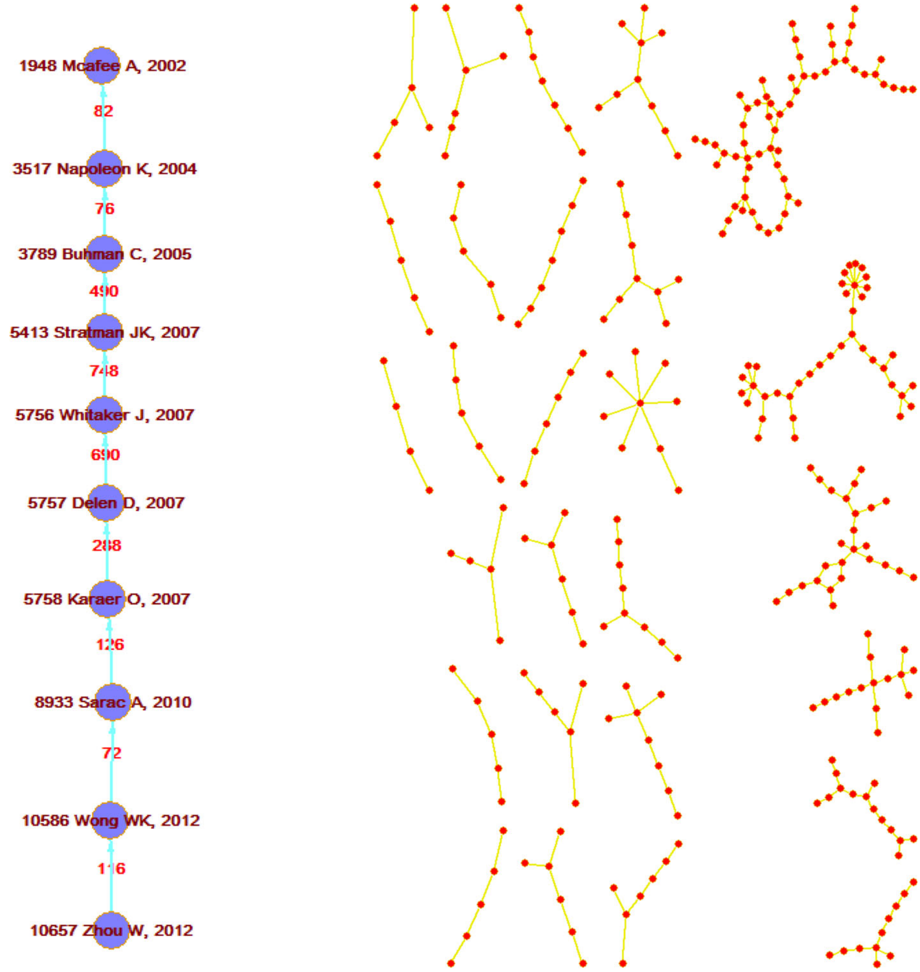


Fig. 15 (Left) SPC global key-route main path, (right) 27 clusters that can be obtained using top 63 FV gradient key-routes in 'IT for engineering'

RFID technology and its role in business, operations and overall quality enhancement of firms as well as industries.

U index values of every pair of paths obtained using direct and ignore and proceed (Liu–Lu) approach are shown in Table 3 as *U index matrix*. FV paths obtained during local forward, local backward, local key-route and global key-route etc., do not share any common paper with their SPC counterparts, thus *U* indices value of these 4 pairs are found to be 1. On the other hand, as an exact match is obtained for critical path method, *U* index for FV critical path and SPC critical path is 0.5. Another important finding is that the four unique paths or subnetworks obtained for FV gradient weight (normalised) assignment method is distinct among themselves and also seems to address different themes. Thus, the capability to highlight diverse evolutionary trajectories is greater for FV weight assignment scheme than SPC scheme. This can be due to the over influence of the overall structure of network on SPC weights of arcs. As search methods can be applied on weighted networks

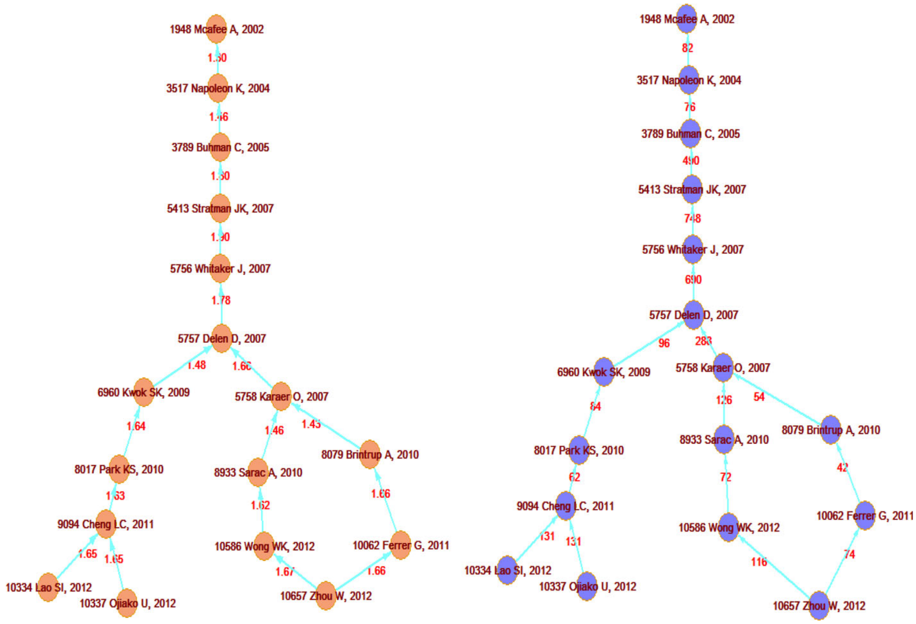


Fig. 16 (left) FV critical path , (right) SPC critical path for 'IT for engineering' (analysed in Prabhakaran et al. (2015)

Table 3 *U* index matrix between SPC paths and FV paths

	1	2	3	4	5	6	7	8	9	10
(1) SPC forward	–	1	1	1	0.89	1	1	1	1	0.89
(2) SPC backward	–	–	0.63	0.72	0.68	1	1	1	1	0.68
(3) SPC keyroute	–	–	–	0.67	0.73	1	1	1	1	0.73
(4) SPC keyroute (global)	–	–	–	–	0.63	1	1	1	1	0.63
(5) SPC critical path	–	–	–	–	–	1	1	1	1	0.5
(6) FV forward	–	–	–	–	–	–	1	1	1	1
(7) FV backward	–	–	–	–	–	–	–	1	1	1
(8) FV keyroute	–	–	–	–	–	–	–	–	0.5	1
(9) FV keyroute (global)	–	–	–	–	–	–	–	–	–	1
(10) FV critical path	–	–	–	–	–	–	–	–	–	–

in PAJEK, no additional implementation steps are needed for the integrated approach for weighted networks. However, more arc weight assignment schemes like FV gradient (norm.) are recommended to be made available in PAJEK.

Conclusion

The network approach to scientometrics or network scientometrics offers a lot of tools and techniques for analysts and for a wide range of applications that ranges from decision making at an organizational level to national policymaking related to science and technology, education and research, industry and commerce, etc. Network analysis can be broadly classified as Centrality analysis, Path analysis and Cluster analysis. In the last three decades, an explosive growth is witnessed in the research area network analysis and most of the tools were found suitable for scientometrics. Despite its huge potential, path analysis is the most underexplored and underdeveloped area among the network analysis trio. Path analysis was introduced by Hummon and Doreian when they introduced traversal based weight assignment schemes to citation networks along with search methods to trace important pieces of knowledge flows in the network. A much computationally efficient method for weight assignment was introduced by Batagelj namely search path count (SPC method) along with improved computation schemes for SPLC and SPNP weights. Recently, Liu and Lu identified the weaknesses of original search schemes used by Hummon–Doreian and introduced novel search schemes. Backward search, Key-route search (local and global) were their key contributions. An integrated approach was proposed by them with the suggestion that decision makers should conduct all these methods of path retrieval and their decision making should be based on insights from evolutionary knowledge flow reflected in all these paths. A genetic knowledge persistence based path analysis is also available in the literature, proposed for finding the technological evolution using patent citation networks. This is the state-of-the-art of path analysis in 'network scientometric literature'.

Availability of SPX methods in the software program PAJEK is a major factor that made it the most sought package for path analysis of citation networks. Presently the availability of Liu and Lu search methods made path analysis a much greater tool for literature mining for policy making applications. However, unless the analysts are ready to learn a bit about the working of the search schemes and the parameters introduced to control the search, they might encounter operational difficulties. Being built upon SPX methods as weight assignment scheme, the chance for finding distinct paths and unexplored themes are much low unless parameter values are properly modified. For decision makers without much expertise or having less time at their disposal, this will be a huge inconvenience. To address this problem, some other path retrieval methods that can identify distinct paths or unexplored themes and perform well without the need for parameter adjustments is required. In this work, our attempt is to fill this gap. We introduce a novel method for path retrieval, viewing the unweighted network as a special case of a weighted network. If the unweighted network can be converted to weighted one through SPX method and be used to retrieve paths through search methods, why can't weighted networks (weighted through other schemes) be directly used to retrieve different key paths? Direct approach to path analysis is possible and can be done in present or previous versions of PAJEK. This is one of the approaches possible for path analysis for weighted networks. This may unleash a lot of analytic opportunities and strengthen the applicability of path analysis to manifold. This approach forms the core contribution of this paper. The second approach named as 'Ignore and proceed' that give same results as Liu–Lu approach because Liu–Lu approach is embedded in the integrated approach for the weighted network we are proposing. We demonstrated the importance of a weight assignment scheme that can reflect an important phenomenon-'Flow Vergecne effect' on a sample network and compared it against SPC weight assignment. Encouraging results were obtained and then we proceeded to conduct the path analysis using flow vergence based weight assignment scheme on a real-world network.

Large network of ‘IT for engineering’ is chosen and firstly direct approach for path analysis is applied. The paths thus obtained are collectively termed as ‘FV paths’. Diversity of themes addressed in FV paths is commendable while SPC method was able to find more roots of a particular theme of evolution (RFID technology) or more contributory works that added value to it or that enhanced the evolutionary process to the level, we see today. This implies the potential of FV gradient method as an efficient weight assignment scheme when it comes to the ability to retrieve trajectories with diverse themes. Due to the lack of availability of citation relevancy assignment scheme, effective demonstration of various weight assignment schemes in combination with the relevancy method is not possible now. In PAJEK, even though layered approach is possible *in principle*, shortage of effective weight assignment schemes is a barrier to explore the full potential of layered approach. However, implementation of first two approaches of integrated approach for weighted networks (direct approach and ignore and proceed approach) is already available in PAJEK. If more weight assignment schemes, for instance, FV gradient method or citation relevance/irrelevance assignment are implemented, PAJEK will become better equipped for the trio approaches and we can anticipate a tremendous growth in the applicability of path analysis. Until then, our recommendation to various analysts is that along with SPX methods, methods such as FV gradient based methods or any other suitable methods can be used for retrieval of more paths using direct approach if weight assignment can be done using an external package. A less related methodical application we envisage for FV gradient weighted (normalised) network is the design of community detection method, which will be one of our further endeavors.

Acknowledgements Authors are grateful to Govt. of Kerala, for funding this work under Innovative programmmes/Research projects scheme, State Plan Grant (No. Pl.A1/1074/DFS/17). This work used the facility provided by ‘Scientometric lab’ (order No.Pl.A1/Annual plan 16-17/Imp.plan/16 dtd. 29/11/2016), Department of Futures Studies, University of Kerala.

References

Batagelj, V. (2003). Efficient algorithms for citation network analysis. arXiv preprint cs/0309023.

Batagelj, V., & Mrvar, A. (1998). Pajek-program for large network analysis. *Connections*, *21*(2), 47–57.

Bilke, S., & Peterson, C. (2001). Topological properties of citation and metabolic networks. *Physical Review E*, *64*(3), 036106.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*(1), 113–120.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71.

Brughmans, T. (2013). Networks of networks: A citation network analysis of the adoption, use, and adaptation of formal network techniques in archaeology. *Literary and Linguistic Computing*, *28*(4), 538–562.

Calero-Medina, C., & Noyons, E. C. M. (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*, *2*(4), 272–279.

Chen, Y.-B., Liu, J. S., & Lin, P. (2013). Recent trend in graphene for optoelectronics. *Journal of Nanoparticle Research*, *15*(2), 1454.

Chuang, T. C., Liu, J. S., Lu, L. Y. Y., & Lee, Y. (2014). The main paths of medical tourism: From transplantation to beautification. *Tourism Management*, *45*, 49–58.

De Solla Price, D. J. (1965). Networks of scientific papers. *Science*, *149*, 510–515.

Egghe, L., & Rousseau, R. (2002). Co-citation, bibliographic coupling and a characterization of lattice citation networks. *Scientometrics*, *55*(3), 349–361.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, *122*(3159), 108–111.

Garfield, E., Sher, I. H., & Torpie, R. J. (1964). The use of citation data in writing the history of science. Technical report, Institute for Scientific Information Inc., Philadelphia, PA.

Garner, R. (1967). A computer oriented, graph theoretic analysis of citation index structures. In In B. Flood (Ed.), *Three Drexel information science research studies* (pp. 3–46).

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417.

Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, *11*(1), 39–63.

Jo, S. J., Jeung, C.-W., Park, S., & Yoon, H. J. (2009). Who is citing whom: Citation network analysis among hrd publications from 1990 to 2007. *Human Resource Development Quarterly*, *20*(4), 503–537.

Kaffash, S., & Marra, M. (2017). Data envelopment analysis in financial services: A citations network analysis of banks, insurance companies and money market funds. *Annals of Operations Research*, *253*(1), 307–344.

Karunan, K., Lathabai, H. H., & Prabhakaran, T. (2017). Discovering interdisciplinary interactions between two research fields using citation networks. *Scientometrics*, *113*(1), 335–367.

Lathabai, H. H., Prabhakaran, T., & Changat, M. (2015). Centrality and flow vergence gradient based path analysis of scientific literature: A case study of biotechnology for engineering. *Physica A: Statistical Mechanics and its Applications*, *429*, 157–168.

Lathabai, H. H., Prabhakaran, T., & Changat, M. (2017). Contextual productivity assessment of authors and journals: A network scientometric approach. *Scientometrics*, *110*(2), 711–737.

Leydesdorff, L, Wagner, C. S., & Bornmann, L. (2017). Betweenness and diversity in journal citation networks as measures of interdisciplinarity: A tribute to eugene garfield. *Scientometrics*.

Liu, J. S., Chen, H.-H., Ho, M. H.-C., & Li, Y.-C. (2014). Citations with different levels of relevancy: Tracing the main paths of legal opinions. *Journal of the Association for Information Science and Technology*, *65*(12), 2479–2488.

Liu, J. S., & Lu, L. Y. Y. (2012). An integrated approach for main path analysis: Development of the hirsch index as an example. *Journal of the Association for Information Science and Technology*, *63*(3), 528–542.

Liu, J. S., Lu, L. Y. Y., Lu, W.-M., & Lin, B. J. Y. (2013). Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega*, *41*(1), 3–15.

Martinelli, A., & Nomaler, Ö. (2014). Measuring knowledge persistence: A genetic approach to patent citation networks. *Journal of Evolutionary Economics*, *24*(3), 623–652.

Merton, R. K. (1988). The matthew effect in science, ii: Cumulative advantage and the symbolism of intellectual property. *ISIS*, *79*(4), 606–623.

Merton, R. K. (1965). *On the shoulders of giants: A Shandean postscript*. San Diego: Harcourt.

Mina, A., Ramlogan, R., Tampubolon, G., & Metcalfe, J. S. (2007). Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research Policy*, *36*(5), 789–806.

Moya-Anegón, S. G. F. D., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Munoz-Fernández, F. J., & Herrero-Solana, V. (2007). Visualizing the marrow of science. *Journal of the Association for Information Science and Technology*, *58*(14), 2167–2179.

Newman, M. E. J. (2008). The mathematics of networks. *The New Palgrave Encyclopedia of Economics*, *2*(2008), 1–12.

Prabhakaran, T., Lathabai, H. H., & Changat, M. (2015). Detection of paradigm shifts and emerging fields using scientific network: A case study of information technology for engineering. *Technological Forecasting and Social Change*, *91*, 124–145.

Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks*, *5*(3), 269–287.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the Association for Information Science and Technology*, *60*(3), 571–580.

Shibata, N., Kajikawa, Y., Takeda, Y., Sakata, I., & Matsushima, K. (2011). Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications. *Technological Forecasting and Social Change*, *78*(2), 274–282.

Tampubolon, G., & Ramlogan, R. (2007). Networks and temporality in the development of a radical medical treatment. *Graduate Journal of Social Science*, *4*(1), 54–77.

Valverde, S., Solé, R. V., Bedau, M. A., & Packard, N. (2007). Topology and evolution of technology innovation networks. *Physical Review E*, *76*(5), 056118.
Xiao, Y., Lu, L. Y. Y., Liu, J. S., & Zhou, Z. (2014). Knowledge diffusion path analysis of data quality literature: A main path analysis. *Journal of Informetrics*, *8*(3), 594–605.

Affiliations

Hiran H. Lathabai[1] · Susan George[1] · Thara Prabhakaran[1] · Manoj Changat[1]

Hiran H. Lathabai
hiranhl007@gmail.com

Susan George
george.susan@gmail.com

Manoj Changat
mchangat@gmail.com

[1] Department of Futures Studies, University of Kerala, Thiruvananthapuram 695581, India