Journal of Informetrics 7 (2013) 651–664

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

# A layered framework to study collaboration as a form of knowledge sharing and diffusion

Yuxian Liu[a], Ronald Rousseau[b,c,d], Raf Guns[b,*]

[a] Library of Tongji University, Tongji University, Siping Street 1239, Shanghai 200092, China
[b] Institute for Education and Information Sciences, IBW, University of Antwerp, Venusstraat 35, Antwerp B-2000, Belgium
[c] Faculty of Engineering Technology, VIVES-KHBO (Association KU Leuven), Zeedijk 101, Oostende B-8400, Belgium
[d] KU Leuven, Leuven B-3000, Belgium

ARTICLE INFO

*Article history:*
Received 21 February 2013
Received in revised form 3 April 2013
Accepted 3 April 2013
Available online 12 May 2013

*Keywords:*
Collaboration
Diffusion
Layered systems
Networks
Centrality indicators

ABSTRACT

Collaboration can be described using layered systems such as the article–author–institute–country structure. These structures can be considered 'cascades' or 'chains' of bipartite networks. We introduce a framework for characterizing and studying the intensity of collaboration between entities at a given level (e.g., between institutions). Specifically, we define the notions of significant, essential and vital nodes, and significant, essential and vital sub paths to describe the spread of knowledge through collaboration in such systems. Based on these notions, we introduce relative and absolute proper essential node (PEN) centrality as indicators of a node's importance for diffusion of knowledge through collaboration.

We illustrate these concepts in an illustrative example and show how they can be applied using a small real-world example. Since collaboration implies knowledge sharing, it can be considered a special form of knowledge diffusion.



## 1. Introduction

In recent years the diffusion of scientific ideas over or through different units has received a growing share of academic attention. Rowlands (2002), for instance, introduced the notion of a journal diffusion factor as a measure of the transdisciplinary reception of a journal. This concept was further developed in (Frandsen, 2004; Frandsen, Rousseau, & Rowlands, 2006). Based on these works, Liu and Rousseau (2010) considered diffusion over ESI-fields. These examples illustrate the fact that diffusion of scientific ideas can be studied from different points of view. The units over which ideas are diffused need not be journals, but can also be ESI-fields, subject categories in the Web of Science, countries, institutes and so on. Publications, being carriers of their authors' scientific ideas, diffuse these ideas. The notion of diffusion itself is often operationalized through citations. In the context of the information sciences diffusion can be described as virtual movement through cognitive space (Liu, 2011).

In previous work (Liu & Rousseau, 2010) we studied mainly diffusion through citations. Yet, also in that work we included a form of diffusion through publications. This form originated from the fact that a group of researchers publishes in different units. This type of knowledge diffusion is an internal mechanism by which the group expands its own borders and diffuses what is known within the group to the outside world. The current paper focuses on a special aspect of knowledge diffusion, namely diffusion and sharing of knowledge through collaboration. Typically, when two or more researchers collaborate, they

* Corresponding author. Tel.: +32 32654510.
*E-mail addresses:* yxliu@tongji.edu.cn (Y. Liu), ronald.rousseau@ua.ac.be (R. Rousseau), raf.guns@ua.ac.be (R. Guns).

1751-1577/$ – see front matter 
http://dx.doi.org/10.1016/j.joi.2013.04.002

**Table 1**
A set of articles and details about collaborations.

| Articles | Authors | Institutes | Countries |
|---|---|---|---|
| A1 | Au1 | In1 | Co1 |
| | Au2 | In1 | Co1 |
| A2 | Au2 | In1 | Co1 |
| A3 | Au1 | In1 | Co1 |
| | Au3 | In2 | Co2 |
| | Au4 | In2 | Co2 |
| | Au5 | In3 | Co2 |
| | | In4 | Co3 |

directly share their knowledge and expertise. Hence, collaboration can be considered a strong form of knowledge diffusion. Moreover, knowledge diffusion through citations takes more time, – the time needed to carry out the research, write down the results, get the paper accepted and published, and finally citations are received. Knowledge diffusion through collaboration, however, is nearly instantaneous since during the collaboration process authors share knowledge and expertise, even before the paper has been written. The shared knowledge diffuses over the research groups, departments, institutes and countries to which the collaborating authors belong.

In earlier studies of diffusion (Liu & Rousseau, 2010; Liu, Rafols, & Rousseau, 2012) we investigated how a knowledge body expands its borders among units such as subject areas or countries. The point is that in these articles connections between different units were not taken into account.

Recently we introduced the notion of layered systems (Rousseau, Liu, & Ye, 2012) to study knowledge diffusion (see also Zhao, Tan, & Ye, 2012). An example of such a layered system is the 'citing articles – citing authors – citing universities – citing countries' system. Each layer pair, such as 'articles – authors', forms a bipartite network, where links can only occur *between* the layers (and never *within* one layer). Consequently, such layered systems can be considered 'cascades' or 'chains' of bipartite networks. Similar systems have previously been explored by Morris and Yen (2004, 2005). In Rousseau et al. (2012), we proposed a fractional counting system for the number of different units in a layer. In this way each layer gave rise to a unique fractionally counted number of units over which diffusion occurred. Layers themselves were compared by using the Gini evenness measure calculated over the diffusion values per layer. One criticism on this approach could be that there are many more authors (scientists) than institutes, and many more institutes than countries, creating a natural unbalance.

In this paper, we study the question how collaboration and the resulting diffusion of shared knowledge can be studied within the framework of layered systems. Layered systems will be considered as undirected networks; units in this network will be referred to as nodes. The connection between collaboration and knowledge diffusion is further explored in the discussion section.

This article is a thoroughly revised, extended and corrected version of (Liu & Rousseau, 2012), correcting mistakes in the original calculations of the illustrative example.

## 2. Collaboration in a layered system

As mentioned before, we focus on collaboration (as shown in the byline of published articles). Collaborating authors automatically lead to collaborating institutes (if authors belong to different institutes) and possibly to collaborating countries. We therefore consider collaboration a major factor in the spread of knowledge over research units, institutes and countries. An article or a set of articles on the same topic constitutes the first layer (layer one, denoted as $L_1$); their authors then form a second layer (layer two, denoted as $L_2$). Depending on the application other layers may consist of research groups, departments, faculties, schools, universities, regions or countries.

Our approach and methods are first explained by a fictional example, shown in Table 1 and illustrated in Fig. 1. We assume that we deal with three articles on the same topic. This set of articles, authors, institutions and countries connected as in Table 1 and Fig. 1 will be called the 'illustrative example'.

Note that there never is a direct link between nodes in the same layer, consistent with the definition of a bipartite network. If we assume that articles A1, A2 and A3 are on topic T, then we want to know how knowledge about topic T can spread by the collaborations shown in Table 1. Concretely: how is knowledge on topic T spread, by author collaboration, from one institute to another, and from one country to another? To study this we introduce the following terminology.

## 3. Definitions

We first recall that a *walk* is a sequence of nodes $n_1, n_2, \ldots, n_m$, such that each node pair $(n_i, n_{i+1})$ in the sequence is connected by a link. There are no further restrictions on walks. The length of a walk is $m - 1$, the number of links traversed. A cycle is a walk which begins and ends at the same node. A *path* is a walk without any cycles. In other words, in a path links and nodes can only be traversed once.

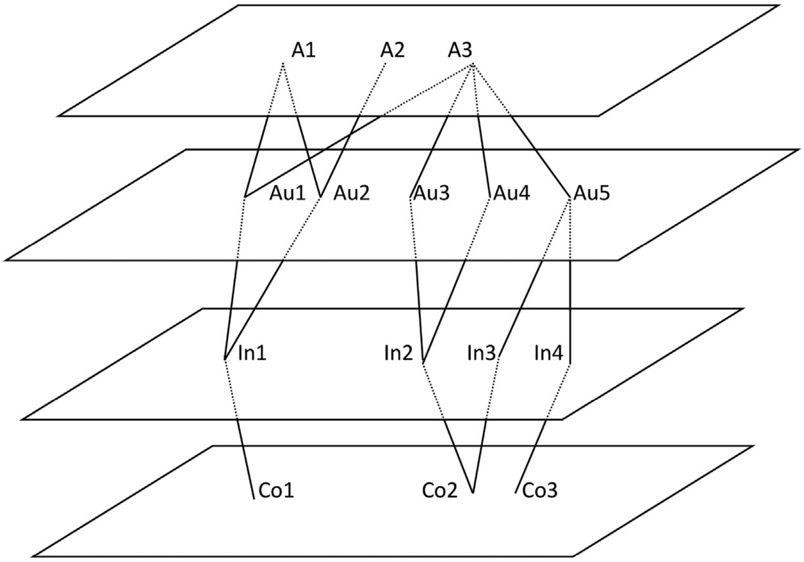


**Fig. 1.** Visual representation of the layered system given in Table 1; A stands for article, Au for author, In for institute and Co for country.

### 3.1. A-paths and collaboration paths

As we focus on collaboration as a means for spreading ideas we only consider paths that pass through the article layer. These paths imply the occurrence of co-authorship and hence collaboration. In order to distinguish them from paths in general they will be called *A-paths*.

If there is an A-path between two nodes in the same layer then this path will be called a *collaboration path* for these nodes. The length of the collaboration path is the number of traversed links. We propose the following important distinction.

### 3.2. Simple and generalized collaboration paths

We assume a layered system in which each node belongs to exactly one layer $L_i (i = 1, \ldots, r)$, where $L_1$ denotes the article layer. A *simple collaboration path* is a collaboration path between two nodes in the same layer $L_s (1 \leq s \leq r)$, such that the path passes through the layers as follows: $L_s, L_{s-1}, \ldots, L_1, \ldots, L_{s-1}, L_s$. The length of a simple collaboration path is $2(s-1)$. We will refer to non-simple collaboration paths as *generalized collaboration paths*. The length of a generalized collaboration path is by definition greater than 2(s-1).

Informally, we can say that simple collaboration paths first go ‘up’ to the article layer and then back ‘down’ to the layer where they started. Looking at the illustrative example, the following is a simple collaboration path of the institutional layer: In2–Au3–A5–Au5–In3, whereas Au1–A3–Au3–In2–Au4 is a generalized collaboration path of the author layer.

Next we introduce some terminology charactering the role of nodes and paths in this framework.

### 3.3. Significant, essential and vital nodes in a collaboration path

Given two nodes $n_1$ and $n_2$ in the same layer one may determine the node or nodes that occur the most in all collaboration paths connecting $n_1$ and $n_2$, excluding $n_1$ and $n_2$ themselves. Such nodes will be called *significant nodes* for $n_1$ and $n_2$. Clearly, if there is no collaboration path between $n_1$ and $n_2$ they have no significant nodes.

If a node occurs in all collaboration paths connecting $n_1$ and $n_2$ it is called an *essential node*. We will consider $n_1$ and $n_2$ to be essential nodes for themselves. The other essential nodes will be called proper essential nodes for $n_1$ and $n_2$. In this context $n_1$ and $n_2$ are trivial essential nodes. Clearly a proper essential node is always significant, but there may be node pairs without proper essential nodes. This happens, e.g., when two authors (AuX and AuY) collaborate on two articles A1 and A2. Then A1 and A2 are significant, but not essential nodes for the author pair (AuX, AuY).

When a node lies on all collaboration paths starting (or ending, which is the same) in node $n_1$ then this node is called a *vital node* for $n_1$. The node $n_1$ itself is a trivial vital node for itself, while the other ones, if they exist, are proper vital nodes for $n_1$. It is not required that there is always a collaboration path between $n_1$ and any other node in the same layer. Note that a proper vital node of $n_1$ is not necessarily a proper essential node for each pair $(n_1, n_2)$, where $n_2$ is another node at the same layer. Specifically, it may be the case that $(n_1, n_2)$ is a proper vital node for $n_1$ and a trivial essential node for $(n_1, n_2)$.

### 3.4. Significant, essential and vital sub paths on a collaboration path

Given two nodes $n_1$ and $n_2$ in the same layer one may determine the sub path or sub paths (such sub paths do not have to be A-paths) that occur the most in all collaboration paths connecting $n_1$ and $n_2$ and have the longest length among these. Such sub paths will be called *significant sub paths* for $n_1$ and $n_2$. Significant sub paths for $n_1$ and $n_2$ may include the nodes $n_1$ and $n_2$.

Sub paths that occur in all collaboration paths connecting $n_1$ and $n_2$ and, among these, have maximal length are called *essential sub paths*. Clearly an essential sub path is always significant, but there may be node pairs without essential sub paths. Essential sub paths for $n_1$ and $n_2$ may include the nodes $n_1$ and $n_2$.

If a sub path occurs in all collaboration paths starting in $n_1$ it is called a *vital sub path* for $n_1$. Again, a vital sub path for $n_1$ may include the node $n_1$.

Nodes on an essential sub path for $n_1$ and $n_2$ are essential nodes for $n_1$ and $n_2$. Yet, a path between two essential nodes for $n_1$ and $n_2$ is not always an essential sub path for $n_1$ and $n_2$. It is only an essential sub path if all intermediate nodes are also essential nodes. Every node on an essential sub path is an essential node, but an essential node is not necessarily part of an essential sub path. An example of the latter would be an article A1 that is written by authors Au1 and Au2 from institute In1 and authors Au3 and Au4 from institute In2. A1 is a proper essential node for In1 and In2 but the pair has no essential sub paths.

Nodes on a vital sub path are vital nodes.

### 3.5. Overview

In summary, we defined two special types of nodes and sub paths for node pairs, namely significant and essential nodes and sub paths. Furthermore, we defined one special type of nodes and sub paths for a single node, namely vital nodes and sub paths. The terminology is summarized in Table 2. Examples of each of these will be provided further on.

### 3.6. Finding all collaboration paths

It is practically impossible to find all collaboration paths in a real network by hand. Hence we need an algorithm to do this.

This algorithm is outlined in Fig. 2. It has been implemented and is used for all examples in this article. It performs a breadth-first search, beginning at start node *s*. We build and maintain a queue of paths and nodes to be explored. At each step we add a node to the current path. If this node is the destination node *e* and the current path contains a node from the article layer, we have a valid collaboration path and add it to the result set $P_s$ (if it is a simple collaboration path) or $P_g$ (if it is a generalized collaboration path). If not, we add all unvisited neighbours of the current node to the queue.

Once all collaboration paths have been found, the result sets can be mined for significant, essential and vital nodes and sub paths. We note that the problem of detecting significant and essential sub paths is equivalent to the problem of finding the longest common substring (Gusfield, 1997).

### 3.7. Collaboration paths in the illustrative example

#### 3.7.1. The author layer

Table 3 shows all simple collaboration paths connecting authors in the author layer, while Table 4 gives the generalized collaboration paths. All simple collaboration paths of the author layer have length 2. All nodes occurring in these tables, except possibly begin or end nodes, are significant nodes for some author pair. Table 5 shows proper essential nodes and essential sub paths in the author layer.

The author pairs (Au3, Au4), (Au3, Au5) and (Au4, Au5) do not have essential sub paths, but they do have significant sub paths. For (Au3, Au4) In2–Co2–In3–Au5–A3 is the significant sub path; note that the direction does not matter. For (Au3, Au5) In2–Au4–A3 is a significant sub path, while the sub path In2–Au3–A3 is significant for the pair (Au4, Au5). Au1 has no vital nodes, and neither has Au2. Nodes Au3, Au4 and Au5 have A3 as a vital node. There are no vital sub paths in this example.

Considering these examples, one can see that a generalized collaboration path such as Au3–In2–Co2–In3–Au5–A3–Au4 is of a special nature. As nodes In2 and In3 are situated in the same country there is only a weak link connecting them by Co2. Consequently, this collaboration path could be considered too weak to truly establish a noteworthy connection

**Table 2**
Overview of special types of nodes and sub paths.

| Defined for ... | Node | Sub path |
|---|---|---|
| Two nodes | Significant node | Significant sub path |
| | Essential node | Essential sub path |
| One node | Vital node | Vital sub path |

**Table 3**
Simple collaboration paths in the author layer.

| Author pairs | Collaboration paths |
|---|---|
| (Au1, Au2) | Au1–A1–Au2 |
| (Au1, Au3) | Au1–A3–Au3 |
| (Au1, Au4) | Au1–A3–Au4 |
| (Au1, Au5) | Au1–A3–Au5 |
| (Au2, Au3) | None |
| (Au2, Au4) | None |
| (Au2, Au5) | None |
| (Au3, Au4) | Au3–A3–Au4 |
| (Au3, Au5) | Au3–A3–Au5 |
| (Au4, Au5) | Au4–A3–Au5 |

**Table 4**
Generalized collaboration paths in the author layer.

| Author pairs | Collaboration paths | Length |
|---|---|---|
| (Au1, Au2) | None | |
| (Au1, Au3) | Au1–A3–Au4–In2–Au3 | 4 |
| | Au1–A3–Au5–In3–Co2–In2–Au3 | 6 |
| (Au1, Au4) | Au1–A3–Au3–In2–Au4 | 4 |
| | Au1–A3–Au5–In3–Co2–In2–Au4 | 6 |
| (Au1, Au5) | Au1–A3–Au3–In2–Co2–In3–Au5 | 6 |
| | Au1–A3–Au4–In2–Co2–In3–Au5 | 6 |
| (Au2, Au3) | Au2–A1–Au1–A3–Au3 | 4 |
| | Au2–In1–Au1–A3–Au3 | 4 |
| | Au2–A1–Au1–A3–Au4–In2–Au3 | 6 |
| | Au2–A1–Au1–A3–Au5–In3–Co2–In2–Au3 | 8 |
| | Au2–In1–Au1–A3–Au4–In2–Au3 | 6 |
| | Au2–In1–Au1–A3–Au5–In3–Co2–In2–Au3 | 8 |
| (Au2, Au4) | Au2–A1–Au1–A3–Au4 | 4 |
| | Au2–In1–Au1–A3–Au4 | 4 |
| | Au2–A1–Au1–A3–Au3–In2–Au4 | 6 |
| | Au2–In1–Au1–A3–Au3–In2–Au4 | 6 |
| | Au2–A1–Au1–A3–Au5–In3–Co2–In2–Au4 | 8 |
| | Au2–In1–Au1–A3–Au5–In3–Co2–In2–Au4 | 8 |
| (Au2, Au5) | Au2–A1–Au1–A3–Au5 | 4 |
| | Au2–In1–Au1–A3–Au5 | 4 |
| | Au2–A1–Au1–A3–Au3–In2–Co2–In3–Au5 | 8 |
| | Au2–A1–Au1–A3–Au4–In2–Co2–In3–Au5 | 8 |
| | Au2–In1–Au1–A3–Au3–In2–Co2–In3–Au5 | 8 |
| | Au2–In1–Au1–A3–Au4–In2–Co2–In3–Au5 | 8 |
| (Au3, Au4) | Au3–In2–Co2–In3–Au5–A3–Au4 | 6 |
| | Au3–A3–Au5–In3–Co2–In2–Au4 | 6 |
| (Au3, Au5) | Au3–In2–Au4–A3–Au5 | 4 |
| | Au3–A3–Au4–In2–Co2–In3–Au5 | 6 |
| (Au4, Au5) | Au4–In2–Au3–A3–Au5 | 4 |
| | Au4–A3–Au3–In2–Co2–In3–Au5 | 6 |

**Table 5**
Proper essential nodes and essential sub paths in the author layer.

| Author pairs | Proper essential nodes | Essential sub paths |
|---|---|---|
| (Au1, Au2) | A1 | Au1–A1–Au2 |
| (Au1, Au3) | A3 | Au1–A3 |
| (Au1, Au4) | A3 | Au1–A3 |
| (Au1, Au5) | A3 | Au1–A3 |
| (Au2, Au3) | Au1, A3 | Au1–A3 |
| (Au2, Au4) | Au1, A3 | Au1–A3 |
| (Au2, Au5) | Au1, A3 | Au1–A3 |
| (Au3, Au4) | A3 | None |
| (Au3, Au5) | A3 | None |
| (Au4, Au5) | A3 | None |

**Input**: network $G = (V, E)$; start node $s \in V$; end node $e \in V$

**Output**: the sets $P_s$ of simple collaboration paths and $P_g$ of generalized collaboration paths

**Data** (other than those mentioned above):

- $c$: current node
- $p$: a sequence of nodes (a path)
- $q$: a sequence of pairs (a queue)
- $R$: a set of nodes

```
begin
    p ← empty sequence
    q ← empty sequence
    Ps ← empty set Ø
    Pg ← empty set Ø

    append pair (s, p) to q
    while q is not empty do
            dequeue first pair from q → c, p
            append c to p
            if c = e then
                    if p contains at least one node from article layer then
                        if p is a simple collaboration path then
                                add p to Ps
                        else
                                add p to Pg
                        end if
                    end if
                    continue while
            end if
            R ← all neighbours of c \ p
            for each n ∈ R do
                    append pair (n, p) to q
            end for each
    end while
end
```

**Fig. 2.** Algorithm to find all collaboration paths.

between Au3 and Au4 (of course, these authors are already strongly related via their direct connection as co-authors of A3). While simple collaboration paths typically imply diffusion of knowledge between authors, institutions or countries, generalized collaboration paths represent a much weaker connection. For this reason we attach only a *potential* diffusion value to generalized collaboration paths. This potential diffusion is diluted the more layers these collaboration paths pass through. Moreover no or little diffusion potential can be associated to generalized collaboration paths passing through the country layer.

### 3.7.2. The institutional layer

Table 6 shows all simple collaboration paths connecting institutes in the institutional layer, while Table 7 shows all generalized collaboration paths. Simple collaboration paths between institutes have a length equal to four. All nodes occurring in these tables, except possibly begin or end nodes, are significant nodes for some institute pair. Table 8 shows proper essential nodes and essential sub paths in the institutional layer.

Here another special case occurs which has nothing to do with collaboration. As author 5 has two affiliations, everything author 5 knows, whether this knowledge is published or not, is knowledge present at these two institutes and diffused over the two from the moment Au5 has made his/her knowledge public, e.g. in a seminar or through informal talks (knowledge which is never made public is by definition not shared).

A3 is a vital node for all institutions. Au1 is a vital node for Institute 1. The sub path Au1–A3 is a vital sub path for Institute 1.

**Table 6**
Simple collaboration paths in the institutional layer.

| Institution pairs | Collaboration paths |
|---|---|
| (In1, In2) | In1–Au1–A3–Au3–In2 |
| | In1–Au1–A3–Au4–In2 |
| (In1, In3) | In1–Au1–A3–Au5–In3 |
| (In1, In4) | In1–Au1–A3–Au5–In4 |
| (In2, In3) | In2–Au3–A3–Au5–In3 |
| | In2–Au4–A3–Au5–In3 |
| (In2, In4) | In2–Au3–A3–Au5–In4 |
| | In2–Au4–A3–Au5–In4 |
| (In3, In4) | None |

**Table 7**
Generalized collaboration paths in the institutional layer.

| Institution pairs | Collaboration paths | Length |
|---|---|---|
| (In1, In2) | In1–Au2–A1–Au1–A3–Au3–In2 | 6 |
| | In1–Au2–A1–Au1–A3–Au4–In2 | 6 |
| | In1–Au1–A3–Au5–In3–Co2–In2 | 6 |
| | In1–Au2–A1–Au1–A3–Au5–In3–Co2–In2 | 8 |
| (In1, In3) | In1–Au2–A1–Au1–A3–Au5–In3 | 6 |
| | In1–Au1–A3–Au3–In2–Co2–In3 | 6 |
| | In1–Au1–A3–Au4–In2–Co2–In3 | 6 |
| | In1–Au2–A1–Au1–A3–Au3–In2–Co2–In3 | 8 |
| | In1–Au2–A1–Au1–A3–Au4–In2–Co2–In3 | 8 |
| (In1, In4) | In1–Au2–A1–Au1–A3–Au5–In4 | 6 |
| | In1–Au1–A3–Au3–In2–Co2–In3–Au5–In4 | 8 |
| | In1–Au1–A3–Au4–In2–Co2–In3–Au5–In4 | 8 |
| | In1–Au2–A1–Au1–A3–Au3–In2–Co2–In3–Au5–In4 | 10 |
| | In1–Au2–A1–Au1–A3–Au4–In2–Co2–In3–Au5–In4 | 10 |
| (In2, In3) | None | |
| (In2, In4) | None | |
| (In3, In4) | In3–Co2–In2–Au3–A3–Au5–In4 | 6 |
| | In3–Co2–In2–Au4–A3–Au5–In4 | 6 |

**Table 8**
Proper essential nodes and essential sub paths in the institutional layer.

| Institution pairs | Proper essential nodes | Essential sub paths |
|---|---|---|
| (In1, In2) | Au1, A3 | Au1–A3 |
| (In1, In3) | Au1, A3 | Au1–A3 |
| (In1, In4) | Au1, A3, Au5 | Au1–A3 |
| | | Au5–In4 |
| (In2, In3) | A3, Au5 | A3–Au5–In3 |
| (In2, In4) | A3, Au5 | A3–Au5–In4 |
| (In3, In4) | Co2, In2, A3, Au5 | In3–Co2–In2 |
| | | A3–Au5–In4 |

### *3.7.3. The country layer*

Tables 9 and 10 show all collaboration paths connecting countries in the country layer. Simple collaboration paths between countries have length six. Table 11 shows proper essential nodes and essential sub paths in the country layer.

Similar to what happened at the institutional layer we observe that a path such as Co2–In3–Au5–In4–Co3 is not a collaboration path as it does not pass through the article layer.

Node A3 is a vital node for each country in the country layer; moreover, In1 and Au1 are vital nodes for Co1; Au5 and In4 are vital nodes for Co3. Co1 and Co2 have no vital sub paths; Au5–In4–Co3 is a vital sub path for Co3.

Scientific ideas can be spread through collaboration paths to different units in each layer. Using an image we may compare layers to terraces in a hilly region, and collaboration paths to aqueducts irrigating the land on different terraces. Whether the land is fully irrigated or how long it will take to irrigate a terrace is determined by the way the aqueducts are connected to these terraces.

**Table 9**
Simple collaboration paths in the country layer.

| Country pairs | Collaboration paths |
|---|---|
| (Co1, Co2) | Co1–In1–Au1–A3–Au3–In2–Co2 |
| | Co1–In1–Au1–A3–Au4–In2–Co2 |
| | Co1–In1–Au1–A3–Au5–In3–Co2 |
| (Co1, Co3) | Co1–In1–Au1–A3–Au5–In4–Co3 |
| (Co2, Co3) | Co2–In2–Au3–A3–Au5–In4–Co3 |
| | Co2–In2–Au4–A3–Au5–In4–Co3 |

**Table 10**
Generalized collaboration paths in the country layer.

| Country pairs | Collaboration paths | Length |
|---|---|---|
| (Co1, Co2) | Co1–In1–Au2–A1–Au1–A3–Au3–In2–Co2 | 8 |
| | Co1–In1–Au2–A1–Au1–A3–Au4–In2–Co2 | 8 |
| | Co1–In1–Au2–A1–Au1–A3–Au5–In3–Co2 | 8 |
| (Co1, Co3) | Co1–In1–Au2–A1–Au1–A3–Au5–In4–Co3 | 8 |
| | Co1–In1–Au1–A3–Au3–In2–Co2–In3–Au5–In4–Co3 | 10 |
| | Co1–In1–Au1–A3–Au4–In2–Co2–In3–Au5–In4–Co3 | |
| | Co1–In1–Au2–A1–Au1–A3–Au3–In2–Co2–In3–Au5–In4–Co3 | 10 |
| | Co1–In1–Au2–A1–Au1–A3–Au4–In2–Co2–In3–Au5–In4–Co3 | 12 |
| | | 12 |
| (Co2, Co3) | None | |

**Table 11**
Proper essential nodes and essential sub paths in the country layer.

| Country pairs | Proper essential nodes | Essential sub paths |
|---|---|---|
| (Co1, Co2) | In1, Au1, A3 | Co1–In1 |
| | | Au1–A3 |
| | | In2–Co2 |
| (Co1, Co3) | In1, Au1, A3, Au5, In4 | Au5–In4–Co3 |
| (Co2, Co3) | In2, A3, Au5, In4 | A3–Au5–In4–Co3 |

### 3.8. *Essential node centrality: a new form of node centrality*

The illustrative example shows that some nodes are far more likely to occur as essential or vital nodes than others. For instance, A3 is a proper essential node for 18 node pairs, whereas A1 is a proper essential node for only one node pair and A2 for none. Clearly, A3 is far more important to establish knowledge diffusion through collaboration than the other two articles. We therefore introduce two new centrality indicators specific to layered collaboration systems. The first one is an absolute indicator, while the second one is a relative one.

*Absolute proper essential node centrality* (in short: absolute PEN centrality) is defined as the number of times a node is a proper essential node in a given layered system.

*Relative proper essential node centrality* (in short: relative PEN centrality) is defined as absolute PEN centrality divided by the maximum number of times a node can be a proper essential node, given the layer structure (but of course, not the link structure). This maximum number is determined as follows. Let $n$ be a node in the first layer; then $P_n = \{(x, y); x \neq y; x, y \in L_j; j > 1\}$. If $n \in L_k$, $k > 1$, then $P_n = \{(x, y); x \neq y; x \neq n; y \neq n; x, y \in L_j; j > 1\}$. The number of elements in $P_n$, denoted as $\#P_n$, is then equal to: $\#P_n = \sum_{j>1} \binom{\#L_j}{2} = \sum_{j>1} \left(\frac{(\#L_j).(\#L_j - 1)}{2}\right)$, if $n \in L_1$; and $\#P_n = \binom{\#L_k - 1}{2} + \sum_{\substack{j \neq k \\ j>1}} \binom{\#L_j}{2} = \frac{(\#L_k - 1).(\#L_k - 2)}{2} + \sum_{\substack{j \neq k \\ j>1}} \left(\frac{(\#L_j).(\#L_j - 1)}{2}\right)$, if $n \in L_k$, $k > 1$. Finally, the relative PEN centrality of node $n$ is equal to its absolute PEN centrality divided by $\#P_n$.

Relative PEN centrality can be seen as a variant of flow betweenness centrality (Freeman, Borgatti, & White, 1991), which considers the relative frequency with which a given node is part of paths between other node pairs. The main difference is that flow betweenness considers all paths, while we only count cases in which the node is part of all collaboration paths between two given nodes. Table 12 shows the absolute and the relative PEN centrality of the nodes in the illustrative example.

**Table 12**
PEN centrality values for the illustrative example.

| Node | Absolute PEN centrality | Relative PEN centrality |
|---|---|---|
| A1 | 1 | 1/19 |
| A2 | 0 | 0 |
| A3 | 18 | 18/19 |
| Au1 | 8 | 8/15 |
| Au2 | 0 | 0 |
| Au3 | 0 | 0 |
| Au4 | 0 | 0 |
| Au5 | 6 | 6/15 |
| In1 | 2 | 2/16 |
| In2 | 2 | 2/16 |
| In3 | 0 | 0 |
| In4 | 2 | 2/16 |
| Co1 | 0 | 0 |
| Co2 | 1 | 1/17 |
| Co3 | 0 | 0 |

**Table 13**
Real-world example overview.

| Articles | Authors | Institutes | Countries |
|---|---|---|---|
| A1. The Indian diaspora in cancer research: a bibliometric assessment for Canada and the USA, p. 110–120. | Aparna Basu (Au1)<br>Philip Roe (Au2)<br>Grant Lewison (Au3) | CSIR (In1)<br>Evaluametrics Ltd. (In2)<br>King's College London (In3) | India (Co1)<br>UK (Co2)<br>UK (Co2) |
| A2. Characteristics of bibliometrics articles in library and information studies (LIS) and other journals, p. 449–451 | Gemma Derrick (Au4)<br>Koen Jonkers (Au5)<br>Grant Lewison (Au3) | CCHS-CSIC (In4)<br><br>King's College London (In3) | Spain (Co3)<br><br>UK (Co2) |
| A3.Trends in the collaborative structure of the Spanish pharmacological scientific production and its influence over research impact, | Maria Bordons (Au6)<br>Javier Aparicio (Au7)<br>Rodrigo Costas (Au8) | CCHS-CSIC (In4)<br><br>CWTS (In5) | Spain (Co3)<br><br>the Netherlands (Co4) |
| A4. New indicators based on the 'Funding acknowledgement' information in the Web of Science: analysis of the effect of peer review over the impact of scientific journals, p. 193–205 | Rodrigo Costas (Au8)<br>Thed N. van Leeuwen (Au9) | CWTS (In5) | the Netherlands (Co4) |

## 4. Discussion

### 4.1. The relation between collaboration and diffusion

We have made a distinction between simple and generalized collaboration paths. While the former imply actual knowledge diffusion, the latter imply only potential diffusion (at best). Note that the two types are distinguished by the number of times collaboration paths change direction: simple collaboration paths change direction only once, whereas generalized ones have two or more changes. The number of direction changes in a collaboration path is an indicator of this potential: as the number increases, the potential decreases.

A further step would be to incorporate the temporal dimension and study how collaboration can, for instance, accelerate knowledge diffusion. We leave this as a suggestion for future research.

### 4.2. Fractional count versus integer count

When one wants to count to what extent an author contributes to an article, a natural approach (though certainly not the only one) is to use fractional counting. However, co-authorship is the strongest form to diffuse knowledge from one person to another. In theory, when two scientists collaborate they completely share all the new knowledge that they produce (and the old one they use in their research). This is one aspect of the well-known fact that if person A has some goods and shares them with person B then A has fewer goods than before, but when A shares knowledge with B he still has the full knowledge. Once the research is finished co-authors have the same potential to transfer the information included in the joint research to others. Hence, when the focus is on the diffusion of scientific ideas, one should give full credit to every co-author, mega-authorship articles being a reasonable exception (Cronin, 2001; Kretschmer & Rousseau, 2001).

The same principle can be applied to the collaboration among institutes and countries. If one person has two affiliations, then this person has a larger potential to transfer his/her knowledge from one affiliation to the other, than a person with only one affiliation.

**Table 14**
Overview of simple collaboration paths in the real-world example.

| Node pair | Simple collaboration paths |
|---|---|
| (Au1, Au2) | Au1–A1–Au2 |
| (Au1, Au3) | Au1–A1–Au3 |
| (Au2, Au3) | Au2–A1–Au3 |
| (Au3, Au4) | Au3–A2–Au4 |
| (Au3, Au5) | Au3–A2–Au5 |
| (Au4, Au5) | Au4–A2–Au5 |
| (Au6, Au7) | Au6–A3–Au7 |
| (Au6, Au8) | Au6–A3–Au8 |
| (Au7, Au8) | Au7–A3–Au8 |
| (Au8, Au9) | Au8–A4–Au9 |
| (In1, In2) | In1–Au1–A1–Au2–In2 |
| (In1, In3) | In1–Au1–A1–Au3–In3 |
| (In2, In3) | In2–Au2–A1–Au3–In3 |
| (In3, In4) | In3–Au3–A2–Au5–In4 |
| | In3–Au3–A2–Au4–In4 |
| (In4, In5) | In4–Au7–A3–Au8–In5 |
| | In4–Au6–A3–Au8–In5 |
| (Co1, Co2) | Co1–In1–Au1–A1–Au3–In3–Co2 |
| | Co1–In1–Au1–A1–Au2–In2–Co2 |
| (Co2, Co3) | Co2–In3–Au3–A2–Au5–In4–Co3 |
| | Co2–In3–Au3–A2–Au4–In4–Co3 |
| (Co3, Co4) | Co3–In4–Au7–A3–Au8–In5–Co4 |
| | Co3–In4–Au6–A3–Au8–In5–Co4 |

### 4.3. Collaborative articles versus single author articles

In Fig. 1, author 1 and author 2 are both active researchers. It is obvious though that author 1 makes a more significant contribution to the diffusing of scientific knowledge than author 2. A single-authored article does not diffuse ideas through co-authorship. When this single author moreover has only one institutional address, such as is the case for Au2, then this author's contribution to knowledge diffusion is minimal, at least in the specific framework of this article.

## 5. A real example

Besides the illustrative example we also provide a real-world case, which consists of four articles from the Proceedings of the STI 2012 Conference (Table 13). A graphical representation is provided in Fig. 3.

The entire system in Table 13 consists of 4 articles, 9 authors, 5 institutes and 4 countries. Tables 15–17 summarize the numbers of generalized collaboration paths per layer. The number of simple collaboration paths (Table 14) increases

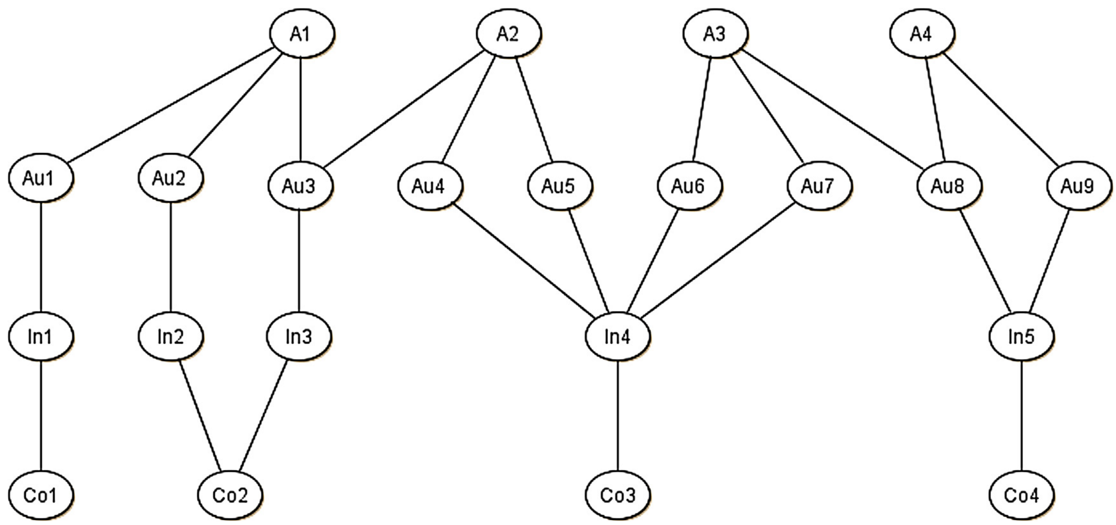


**Fig. 3.** Graphical representation of real-world example.

with the number of joint articles by two entities. The number of generalized collaboration paths, on the other hand, can be interpreted as the potential diffusion value, especially for node pairs that do not share any joint articles. Thus, we can see that simple collaboration paths and generalized collaboration paths are complementary concepts that capture the different characteristics of diffusion through collaboration.

Most essential sub paths are short (with length 1, see Table 18). The longest essential sub path is found between In2 and In3; here we have two institutes each of which has one author in the example with one joint article.

**Table 15**
Numbers of generalized collaboration paths in the author layer.

| | Au1 | Au2 | Au3 | Au4 | Au5 | Au6 | Au7 | Au8 | Au9 |
|---|---|---|---|---|---|---|---|---|---|
| Au1 | | 1 | 1 | 4 | 4 | 8 | 8 | 8 | 16 |
| Au2 | | | 0 | 4 | 4 | 8 | 8 | 8 | 16 |
| Au3 | | | | 1 | 1 | 4 | 4 | 4 | 8 |
| Au4 | | | | | 0 | 3 | 3 | 4 | 8 |
| Au5 | | | | | | 3 | 3 | 4 | 8 |
| Au6 | | | | | | | 0 | 1 | 4 |
| Au7 | | | | | | | | 1 | 4 |
| Au8 | | | | | | | | | 0 |
| Au9 | | | | | | | | | |

**Table 16**
Numbers of generalized collaboration paths in the institute layer.

| | In1 | In2 | In3 | In4 | In5 |
|---|---|---|---|---|---|
| In1 | | 1 | 1 | 4 | 16 |
| In2 | | | 0 | 4 | 16 |
| In3 | | | | 2 | 16 |
| In4 | | | | | 2 |
| In5 | | | | | |

**Table 17**
Numbers of generalized collaboration paths in the country layer.

| | Co1 | Co2 | Co3 | Co4 |
|---|---|---|---|---|
| Co1 | | 0 | 4 | 16 |
| Co2 | | | 2 | 16 |
| Co3 | | | | 2 |
| Co4 | | | | |

**Table 18**
Proper essential nodes and essential sub paths in the real-world example.

| Node pair | Proper essential nodes | Essential sub paths |
|---|---|---|
| (Au1, Au2) | A1 | Au1–A1 |
| (Au1, Au3) | A1 | Au1–A1 |
| (Au1, Au4) | A1, A2, Au3 | Au1–A1<br>Au3–A2 |
| (Au1, Au5) | A1, A2, Au3 | Au1–A1<br>Au3–A2 |
| (Au1, Au6) | A1, A2, Au3, In4 | Au1–A1<br>Au3–A2 |
| (Au1, Au7) | A1, A2, Au3, In4 | Au1–A1<br>Au3–A2 |
| (Au1, Au8) | A1, A2, A3, Au3, In4 | Au1–A1<br>Au3–A2<br>A3–Au8 |
| (Au1, Au9) | A1, A2, A3, Au3, Au8, In4 | Au1–A1<br>Au3–A2<br>A3–Au8 |
| (Au2, Au3) | A1 | Au2–A1–Au3 |
| (Au2, Au4) | A2, Au3 | Au3–A2 |
| (Au2, Au5) | A2, Au3 | Au3–A2 |
| (Au2, Au6) | A2, Au3, In4 | Au3–A2 |
| (Au2, Au7) | A2, Au3, In4 | Au3–A2 |

Table 18 (*Continued*)

| Node pair | Proper essential nodes | Essential sub paths |
|---|---|---|
| (Au2, Au8) | A2, A3, Au3, In4 | Au3–A2<br>A3–Au8 |
| (Au2, Au9) | A2, A3, Au3, Au8, In4 | Au3–A2<br>A3–Au8 |
| (Au3, Au4) | A2 | Au3–A2 |
| (Au3, Au5) | A2 | Au3–A2 |
| (Au3, Au6) | A2, In4 | Au3–A2 |
| (Au3, Au7) | A2, In4 | Au3–A2 |
| (Au3, Au8) | A2, A3, In4 | Au3–A2<br>A3–Au8 |
| (Au3, Au9) | A2, A3, Au8, In4 | Au3–A2<br>A3–Au8 |
| (Au4, Au5) | A2 | Au4–A2–Au5 |
| (Au4, Au6) | In4 | None |
| (Au4, Au7) | In4 | None |
| (Au4, Au8) | A3, In4 | A3–Au8 |
| (Au4, Au9) | A3, Au8, In4 | A3–Au8 |
| (Au5, Au6) | In4 | None |
| (Au5, Au7) | In4 | None |
| (Au5, Au8) | A3, In4 | A3–Au8 |
| (Au5, Au9) | A3, Au8, In4 | A3–Au8 |
| (Au6, Au7) | A3 | Au6–A3–Au7 |
| (Au6, Au8) | A3 | A3–Au8 |
| (Au6, Au9) | A3, Au8 | A3–Au8 |
| (Au7, Au8) | A3 | A3–Au8 |
| (Au7, Au9) | A3, Au8 | A3–Au8 |
| (Au8, Au9) | A4 | Au8–A4–Au9 |
| (In1, In2) | A1, Au1 | In1–Au1–A1 |
| (In1, In3) | A1, Au1 | In1–Au1–A1 |
| (In1, In4) | A1, A2, Au1, Au3 | In1–Au1–A1 |
| (In1, In5) | A1, A2, A3, Au1, Au3, Au8, In4 | In1–Au1–A1 |
| (In2, In3) | A1, Au2, Au3 | In2–Au2–A1–Au3–In3 |
| (In2, In4) | A2, Au3 | Au3–A2 |
| (In2, In5) | A2, A3, Au3, Au8, In4 | Au3–A2<br>A3–Au8 |
| (In3, In4) | A2, Au3 | Au3–A2 |
| (In3, In5) | A2, A3, Au3, Au8, In4 | Au3–A2<br>A3–Au8 |
| (In4, In5) | A3, Au8 | A3–Au8 |
| (Co1, Co2) | A1, Au1, In1 | Co1–In1–Au1–A1 |
| (Co1, Co3) | A1, A2, Au1, Au3, In1, In4 | Co1–In1–Au1–A1 |
| (Co1, Co4) | A1, A2, A3, Au1, Au3, Au8, In1, In4, In5 | Co1–In1–Au1–A1 |
| (Co2, Co3) | A2, Au3, In4 | Au3–A2<br>In4–Co3 |
| (Co2, Co4) | A2, A3, Au3, Au8, In4, In5 | Au3–A2<br>A3–Au8<br>In5–Co4 |
| (Co3, Co4) | A3, Au8, In4, In5 | Co3–In4<br>A3–Au8<br>In5–Co4 |

Intuitively, Au3 (Lewison) and Au8 (Costas) are important nodes in that their removal would result in the network to fall apart into three separate components. These two nodes enable knowledge diffusion and sharing through collaboration (within the boundaries of the current example). Indeed, these two nodes are proper essential nodes for several node pairs (Table 18). Au8 is also a vital node for three other nodes. Their importance is also reflected in their high PEN centrality (Table 19).

Table 19 summarizes all non-zero PEN centralities. As expected, nodes in higher layers tend to have higher essential node centrality values. A notable exception is In4 (CCHS-CSIC), whose unique position leads it to obtain the highest essential node centrality in the system.

From Table 20, we can see that each layer contains nodes that are dependent on one or more other nodes to diffuse and share their knowledge throughout the system. This heavy reliance on other nodes does not automatically render these nodes peripheral or unimportant. For instance, Au1 depends on A1 but is in its turn a vital node for In1.

**Table 19**
Absolute and relative PEN centrality in the real-world example (only non-zero values).

| Node | Absolute PEN centrality | Relative PEN centrality | Node | Absolute PEN centrality | Relative PEN centrality |
|---|---|---|---|---|---|
| A1 | 17 | 0.327 | Au3 | 23 | 0.523 |
| A2 | 29 | 0.558 | Au8 | 14 | 0.318 |
| A3 | 22 | 0.423 | In1 | 3 | 0.063 |
| A4 | 1 | 0.019 | In4 | 28 | 0.583 |
| Au1 | 7 | 0.159 | In5 | 3 | 0.063 |
| Au2 | 1 | 0.023 | | | |

**Table 20**
Vital nodes and sub paths in the real-world example.

| Node | Vital nodes | Vital sub paths |
|---|---|---|
| Au1 | A1 | |
| Au9 | Au8 | |
| In1 | Au1, A1 | In1–Au1–A1 |
| In5 | In4, A3, Au8 | |
| Co1 | A1, Au1, In1 | Co1–In1–Au1–A1 |
| Co3 | In4 | |
| Co4 | In4, In5, A3, Au8 | |

We omitted determining significant nodes and sub paths, since it is still feasible to determine essential nodes and sub paths for this small example. In very large systems essential nodes or sub paths may become quite rare; in those cases, one would have to turn to significant nodes and sub paths to describe the system.

## 6. Conclusion

In this article we introduced the notions of collaboration paths (simple ones and generalized ones), and significant, essential and vital nodes and sub paths in a layered collaboration network. Based on these notions relative and absolute PEN centrality are proposed as indicators of a node's importance for diffusion of knowledge through collaboration. All concepts are illustrated in an example. Scientific ideas are not simply diffused to different countries or different fields, they are diffused through a layered system. Diffusion through the units in one layer depends on the structure of units in other layers as well as the collaboration structure between these units. In this sense we link the diffusion of scientific ideas to social and geographical structures. This framework leads to new mechanisms to study the diffusion of scientific ideas through units in different layers. Several hypotheses can be studied within this framework, such as "Is it true that: The more collaboration paths between two nodes, the easier a scientific idea is diffused? Is it more difficult to diffuse scientific ideas between two institutes in one country than that between two institutes in two countries?". Of course it is impossible to answer these questions based on just a few examples: a study of large numbers of cases is necessary. Hence, these and other derived hypotheses are proposed for future research projects.

## Acknowledgements

This research was supported by NFSC grant no. 71173154. R. Rousseau thanks Elias Sanz-Casado for his hospitality during his research stay at Carlos III University at Madrid (Spain).

## References

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, *52*(7), 558–569.
Frandsen, T. F. (2004). Journal diffusion factors: A measure of diffusion? *Aslib Proceedings*, *56*(1), 5–11.
Frandsen, T. F., Rousseau, R., & Rowlands, I. (2006). Diffusion factors. *Journal of Documentation*, *62*(1), 58–72.
Freeman, L. C., Borgatti, S. P., & White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks*, *13*(2), 141–154.
Gusfield, D. (1997). *Algorithms on strings, trees and sequences: Computer science and computational biology*. USA: Cambridge University Press.
Kretschmer, H., & Rousseau, R. (2001). Author inflation leads to a breakdown of Lotka's law. *Journal of the American Society for Information Science and Technology*, *52*(8), 610–614.
Liu, Y. (2011). *The diffusion of scientific ideas in time and indicators for the description of this process* (Doctoral thesis). Antwerp University.
Liu, Y., Rafols, I., & Rousseau, R. (2012). A framework for knowledge integration and diffusion. *Journal of Documentation*, *68*(1), 31–44.
Liu, Y., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. *Journal of the American Society for Information Science and Technology*, *61*(2), 340–351.

Liu, Y., & Rousseau, R. (2012). Diffusion through collaboration in a layered system. In E. Archambault, Y. Gingras, & V. Larivière (Eds.), *Proceedings of STI 2012 Montréal* (pp. 552–560). Montréal (Québec): Science-Metrix, Observatoire des Sciences et des Technologies.

Morris, S. A., & Yen, G. G. (2004). Crossmaps: Visualization of overlapping relationships in collections of journal papers. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(Suppl. 1), 5291–5296.

Morris, S. A., & Yen, G. G. (2005). *Construction of bipartite and unipartite weighted networks from collections of journal papers*. Retrieved from http://arxiv.org/abs/physics/0503061

Rousseau, R., Liu, Y., & Ye, F. Y. (2012). A preliminary investigation on diffusion through a layered system. *Journal of Informetrics*, *6*(2), 177–191.

Rowlands, I. (2002). Journal diffusion factors: A new approach to measuring research influence. *Aslib Proceedings*, *54*(2), 77–84.

Zhao, S. X., Tan, A., & Ye, F. Y. (2012). Distributive h-indices for measuring multi-level impact. *Journal of the American Society for Information Science and Technology*, *63*(10), 2074–2086.