Decision Support Systems 52 (2011) 217–231

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

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

An approach to identify influential building blocks and linkages in an information resource network

Sudip Bhattacharjee [a], James R. Marsden [a,c,*], Harpreet Singh [b]

[a] Department of Operations and Information Management, 2100 Hillside Road, U-1041, School of Business, University of Connecticut, Storrs, CT 06269-1041, United States
[b] Information Systems and Operations Management (ISOM), 800 West Campbell Road, SM33, School of Management, University of Texas at Dallas, Richardson, TX 75080-3021, United States
[c] KU-Leuven, Leuven, Belgium

ARTICLE INFO

Article history:
Received 18 February 2010
Received in revised form 1 July 2011
Accepted 28 July 2011
Available online 12 August 2011

Keywords:
Information resource network
IRN metrics
Directed graph
Digital goods
Digital good markets
Time ordered information network

ABSTRACT

An information resource network (IRN) is a time-ordered and potentially interrelated set of information elements. Examples include papers within a research domain, blog postings dealing with a certain topic, and information records within a company. We present a structured analysis to identify influential building blocks and linkages in a general IRN and show that our approach can be used for large networks of information nodes. Our method compensates for biases that can emerge at the edges of such time-dependent networks. Importantly, our focus is on the information elements and not on the authors of such information. We illustrate this process using one example of a resource network — research papers in a given domain. Our method can be implemented in any domain that can be represented as time-ordered, interrelated components of information sets.

1. Introduction

In the digital world we live in, massive amounts of information are continuously generated, captured and stored relating to the actions of individuals and organizations. Given the vast amounts of information, an individual or even an organization can be hard pressed to identify the important pieces of information relevant for their purpose, a task often akin to finding the proverbial “needle in a haystack”. In a variety of settings, however, there are important components and connections that relate elements of information. We employ the term *information resource networks (IRN)* to refer to such structures. In these settings, an individual investigator's interest is in the information elements and not in the individual actors or participants in the network, the latter being the usual focus of social network analysis. In our usage, an IRN is a time-ordered and potentially interrelated set of information elements. Examples include papers within a research domain, blog postings containing information elements in the form of specific word strings or related to a specific topic, and time-ordered information records within a company such as email conversations. The exponentially growing information space is reflected in the expanding sizes for IRN's. In this setting, individuals may be unable to analyze each information element as to whether the element is a key player or building block in an IRN. Our purpose is to detail the concept of an IRN and to present a structured method to identify influential building blocks and linkages in an IRN even where the candidate set of information elements might be quite large. We illustrate this process using one example of a resource network — a network of research papers in a given domain. As discussed below, our method can be implemented in any domain that can be represented as time-ordered, interrelated components of information sets.

Consider the following examples of the use of an IRN. An individual student or researcher might be interested in identifying the main linkages of research findings that develop a field, rather than in a complete enumeration of all work in the domain. Such a subset of the “key building block” articles for the specified research domain would enable the individual to spend more time being productive and less time searching. In analyzing a company's information consistency and controls, an auditor might be interested in the flow of information patterns and linkages across time, looking for critically important banned (IT control restricted) patterns or inconsistent patterns. A national health organization may identify inaccurate, or possibly dangerous misinformation, floating in the blogosphere. In order to counteract the incorrect medical information, the organization may find it useful to first understand how such information elements are disseminated across blogs. Identifying key “main flow” blogs (blogosphere information building blocks, if you will) enables the health

* Corresponding author.
E-mail addresses: sbhattacharjee@business.uconn.edu (S. Bhattacharjee), jmarsden@business.uconn.edu (J.R. Marsden), harpreet@utdallas.edu (H. Singh).

0167-9236/$ – see front matter
doi:10.1016/j.dss.2011.07.006

organization to effectively disseminate correct information by tapping the main flows rather than trying to identify and send out correct information to all possible information outlets — a virtually impossible task.

The explosion in information availability (see [1,12]) including the "deluge" of information in emails, blogs, instant messages and other online resources, can result in a complex set of potential linkages in an IRN. Having a structured method to analyze the time-ordered information in an IRN and identify the key information elements reduces complexity and enables focused action. Our purpose is to present such a structured method and illustrate its use in one application area. The focus here is on the information elements themselves and not on the authors or individual developers of the information resource. This differs from existing studies of information linkages including various approaches in citation analysis. The latter can be generally grouped in two major categories: i) Theoretical/Behavioral studies, and ii) Quantitative studies relying on network properties. The main focus of (i) is to identify individual behavioral patterns behind citations, that is, identify factors determining why authors cite the work of their peers (see [4] for a comprehensive review). This is a form of social network analysis, and is not a concern of our research. In (ii), studies utilize citations or references to study the evolution of a field and identify important authors/articles/journals in the evolution of the field. Here, three types of networks have been analyzed: a) co-citation networks, b) co-authorship networks, and c) inter-citation networks. Co-citation networks can be constructed using either authors or papers as nodes with a link formed when the two authors or two papers cite the same third author or paper. These networks, while structural, are non-directional in nature [6]. Co-authorship networks are the most widely studied networks in citation analysis, where the nodes are authors and a link is formed when two authors collaborate on the same article. These are also non-directional networks [2,8–10,13–16,19,20]. Inter-citation networks have received limited attention, where nodes might be authors or papers and a link is formed when one cites the other one (paper or author). These are directional in nature. Popular metrics used in inter-citation networks are detailed in [18]. However the bibliometric indicators are temporally biased toward nodes that appear early in the evolution process. A popular metric, PageRank, also suffers from the same bias [11].

Significant research has focused on identifying influential authors and journals, with a plethora of measurement indices. For example, [3] compare nine variants of the *h* index using data from biomedicine. They concluded that a better prediction of assessments occurs while using the impact, rather than the quantity, of the "productive core" of a scientist's research output. While impact can be measured in multiple ways, the most popular existing metrics show a temporal bias [17]. In Table 1, we contrast the uniqueness of our research approach with existing research. It should be noted that inter-citation networks and social networks could be formulated to possess time-stamped and context-specific information, but this has rarely been the case. We use the term IRN to denote a specific network formulation and provide appropriate non-biased metrics. While it may be possible to transform an inter-citation network into an IRN, the IRN concept (as explained in more detail below) is broader and applicable in a wider set of scenarios than just citation analysis. Thus, we focus our presentation on IRNs as a separate network formulation.

In the next section, we introduce a set of new metrics appropriate for analyzing a time directed IRN. These metrics certainly draw from a variety of previous network analyses, but they are specifically tailored for the IRN setting. Section 3 introduces a detailed example of an IRN and details the steps involved in identifying candidate nodes (research papers) for the example. This is followed in Section 4 by a formal presentation of the example IRN and accompanying analysis. Concluding remarks and potential application areas are then presented.

2. IRN metrics

In an IRN, nodes are information elements and arcs are time-ordered relationships between the nodes. Hence an IRN is a directed, acyclic graph, and the temporal nature of the elements is an integral part of the design of the network and associated metrics. Along with the time-ordered relationship of nodes and links, unbiased measurements of the relationships (i.e., metrics) are critical to a reliable IRN structure. As discussed below and illustrated in our IRN examples, the exact nature of these arcs or relationships is application specific. If, for example, information element (node) A uses information element B as a building block, then there is a directed arc from node A to node B. We employ the following notation and definitions:

N_i: Total number of information elements (nodes) in a given network i;
$indegree_j$: number of directed arcs to node j from other nodes in network i;
$outdegree_j$: number of directed arcs from node j to other nodes in network i.

Raw measures of indegree may bias analysis results toward an information resource that was produced toward the beginning of the domain phenomenon under study. A similar type of bias, but in the reverse direction, may occur for outdegree measures. Our review of

Table 1
Comparison of information resource networks (IRN) with typical formulations of social networks and citation analysis.

Link properties	Social network	Citation analysis	Information Resource Network (IRN)
Directional	No. Most ties in social networks are bi-directional. In most cases if person X is linked to Y, it is assumed that Y is also linked to X. For example — Friend's network.	Maybe. Co-authorship and co-citation networks are non-directional, but inter-citation networks are directional.	Yes. The nodes are information elements and links are information flows between these elements. Since information flows from an element originating first to ones appearing later, these networks will be directional. For example — links between blog posts, research papers, etc.
Time-stamped	No. It is difficult to determine the exact time of the formation of link between nodes.	Maybe. Although it is easier to determine the exact time of link formation, most studies ignore it, and most bibliometric indicators are temporally biased.	Yes. The exact time of the link formation can be determined. In addition the time of link formation is considered while incorporating any link in calculating node level measures.
Context-specific	Maybe. The boundary of the network may be based on "snow ball" (pick a few nodes randomly and trace their links) or "whole network" (all nodes in the context) approach.	Maybe. Most times all links are considered, whether the citations are from same context (domain) or not.	Yes. In IRN, links or information flow between nodes in the same context is considered.
Focus on information or author	Author	Mostly author	Information elements

the literature related to network applications in social connections, and citation networks (see, for example, [7]) do not reveal a method to reduce biases introduced at the edges of such inherently time-ordered information elements. Hence we introduce normalized indegree and outdegree metrics below.

Information elements in an IRN research database can be time indexed, say, by 1, 2, …, t, ..,T. Let n_t represent the number of information elements indexed in period t. In general, an information element, A, can only have a time-ordered relationship to information element B as a building block if B is indexed in period t or earlier. Thus we utilize the following:

for an information element indexed in period t,

$$\textit{max indegree} = (n_t - 1) + \sum_{k \in t+1}^{T} n_k$$

for an information element indexed in period t,

$$\textit{max outdegree} = (n_t - 1) + \sum_{k \in 1}^{t-1} n_k.$$

Hence:

norm-indegree$_j$ (normalized indegree): *indegree*$_j$/*max indegree*$_j$; and,
norm-outdegree$_j$ (normalized outdegree): *outdegree*$_j$/*max outdegree*$_j$.

Our normalized indegree and outdegree metrics differ from the corresponding metrics utilized in social networks. This reflects the time dependent nature of IRN. Similarly the concepts of power, centrality, and density are often used in network analysis (see the interesting discussion in [5]), but we suggest modified measures that are appropriate and relevant for the IRN setting. We introduce each in turn.

2.1. Information innovator or fundamental building block in an IRN

In an IRN, it is possible that certain information elements or nodes appear to bring new information into the network, or “start the conversation”. Such a node plays a “fundamental building block” role in the IRN. That is, suppose a node has a time-ordered relationship directed to few, if any, earlier nodes in the IRN, but the same node has a large number of time-ordered information relationships directed to it, giving the node a very low out-degree relative to a very high in-degree. Hence we introduce the concept of “building block value” (bbv_j) defined as follows:

$$bbv_j = indegree_j \Big/ \left(outdegree_j + 1\right).$$

The “1” is added to avoid division by zero in the case of a node having no outward arcs (a new information node, if you will).

To overcome the bias toward nodes with earlier time stamps and thus possibly gaining on the *bbv* by virtue of being in the network longer, we utilize a normalized form for *bbv* as follows:

$$norm{-}bbv_j = norm{-}indegree_j \, / \left(norm{-}outdegree_j + 1\right).$$

As with the normalized indegree and normalized outdegree measures presented earlier, *norm-bbv*$_j$ help avoid early entrant bias. As relative measures, the normalized values focus on how many links actually exist for an information node in an IRN relative to how many such links could possibly exist.

2.2. Seminal node

What does it mean to say that a node makes a seminal information contribution to an IRN? Suppose that a node is a “must link” node in a network in that the node continues to have directed arcs from subsequent nodes that have directed links to intermediate time nodes which themselves already have directed arcs to the early node. In a sense, a seminal node is a type of exceptional building block node, where the network's late time-indexed nodes continue to have directed arcs to the seminal node. We define the following measures where sv_j measures the seminal value of node j's information element:

$link_{kj} = 1$ if node k is linked to node j, 0 otherwise
$sv_j = \sum_{kr} \{(link_{kj})(link_{rk})(link_{rj})\}$.

An example may help to clarify this measure. Consider an IRN where the nodes (information elements) are research papers in a specific research domain. Suppose that John C's paper in 1990 cites Joe B's 1985 paper and Joe B's 1985 paper cites Jim A's 1980 paper. Then $link_{CB} = 1$, $link_{BA} = 1$ and, if C also cites A, then $link_{CA} = 1$ which adds a value of 1 to the value of sc_A. Thus the highest possible value for sv_A is $indegree_A$, the number of papers citing A. This maximum value is reached only when all papers citing other papers that cite A also cite paper A itself.

Our measure of seminal value emphasizes the continued importance of directed arcs across multiple time periods. That is, the seminal value of node A increases if subsequent nodes that link to nodes with connecting directed arcs to A also have direct connecting arcs to A itself. Thus when a node has a high sv_j this suggests that the node maintains sufficient information importance across time to attain the status of a seminal node within the specified IRN. Our measure is designed to overcome bias toward articles that appear early in the lifecycle of a domain's contributions.

2.3. IRN network density

One final metric that we put forth is the density of an IRN, which we define as:

$$\textit{IRN density} = \frac{\text{Total number of directed arcs}}{\sum_j \textit{max indegree}_j}$$

where *max indegree*$_j$ is as defined above. We note here that our density calculation differs from the standard social network density because of the time-directed nature of the network. We also note that

$$\sum_j \textit{max indegree}_j = \sum_j \textit{max outdegree}_j.$$

Given that information is contained in the nodes, a denser network suggests a greater inter-dependence of information within such a network. In addition, since adding a single node to a dense network may not significantly impact the metrics of the network, we might say that dense networks have high stability. On the other hand, adding a node to a sparse network may result in significant change of the network metrics. Hence we posit the above measure of *IRN density* as a network density that has an impact on the relative importance of linkages within the network.

Appendix A provides a summary table of definitions and formulae for all IRN metrics introduced above.

3. IRN demonstration: domain of analysis and data

Research in information systems (IS), a broad area encompassing multiple facets of the interaction of technology, people and organizations, has shown tremendous growth in the last few decades. To illustrate our approach, we focus on research related to two digital good products, music and software, and their associated market dynamics. Significant research exists at the intersection of these products and their market characteristics, including pricing, piracy issues, digital rights management (DRM), legal and economic frameworks of revenue

Table 2
Set of journals used in analysis.

Journal ID	Journal	Journal ID	Journal
AE	Applied Economics	JEP	Journal of Economic Perspectives
AEL	Applied Economics Letters	JIE	Journal of Industrial Economics
AMAPSS	Annals of the American Academy of Political and Social Science	JLE	Journal of Law and Economics
BIT	Behavior & Information Technology	JM	Journal of Marketing
CACM	Association for Computing Machinery. Communications of the ACM	JMIS	Journal of Management Information Systems
CJE	Canadian Journal of Economics	JMR	Journal of Marketing Research
DSS	Decision Support Systems	JOCEC	Journal of Organizational Computing and Electronic Commerce.
EIT	Ethics and Information Technology	JORS	Journal of the Operational Research Society
EL	Economics Letters	JPBM	Journal of Product and Brand Management.
HR	Human Relations	JPE	Journal of Political Economy
ICC	Industrial and Corporate Change	JPPM	Journal of Public Policy & Marketing
IEEEEM	IEEE Transactions on Engineering Management	JRM	Journal of Research in Marketing
IEEESE	IEEE Transactions on Software Engineering	LRP	Long Range Planning
IEP	Information Economics and Policy	MCS	Media Culture & Society
IJEC	International Journal of Electronic Commerce	MISQ	MIS Quarterly
IJF	International Journal of Forecasting	MS	Management Science
IJIO	International Journal of Industrial Organization	NMS	New Media & Society
IJMM	International Journal on Media Management.	OME	Omega
IM	Information & Management	OS	Organization Science
INT	Interfaces	RANDJE	The RAND Journal of Economics
IPTLJ	Intellectual Property & Technology Law Journal	RERCI	Review of Economic Research on Copyright Issues
ISR	Information Systems Research	RES	Review of Economics and Statistics
ITM	Information Technology and Management	RIO	Review of Industrial Organization
JASA	Journal of the American Statistical Association	RP	Research Policy
JB	Journal of Business	SCMIJ	Supply Chain Management-an International Journal
JBE	Journal of Business Ethics	SDR	System Dynamics Review
JCE	Journal of Cultural Economics	SLR	Stanford Law Review
JEB	Journal of Education for Business	TFSC	Technological Forecasting and Social Change
JEE	Journal of Evolutionary Economics	UCLALR	UCLA Law Review
JEI	Journal of Economic Issues		

management, and marketing of bundled goods and services. In the US, such digital products are a multi-billion dollar industry, a mainstay of intellectual property and copyright regime, and a significant component of the US economy. This domain of digital products and markets is in a state of rapid advancement and has a rich, evolving set of research “information elements” making the two areas appropriate selections for demonstrating the IRN approach.

Further, while these digital goods domains have their own distinct research lines, they are related in some aspects of the underlying theory. Hence the level of commonalities can also be analyzed as a joint IRN. In particular, we sought IRN nodes (research papers) in these two domains that focused on topics such as the interface of digital goods and markets, piracy concerns, revenue and revenue sharing structures, consumer behavior and attitudes, legal issues, or the impact of technology in these two domains. We include research articles in these digital goods domains that incorporate technical elements to reduce piracy and enforce digital rights; legal approaches to digital rights and subsequent consumer behavior; social attitudes toward digital rights, encryption, usability, etc.; and economic mechanisms such as new models of pricing structure to counter free-riding, increase revenues and profitability.

For the selected digital goods, we focus on market related issues. Thus we are not considering research on software production and development, music acoustics, and the impacts of music on human psychology, and other research areas which typically do not emphasize market analysis. In subsequent discussion, we use the terms “software” and “music” as shorthand to identify the domains of the IRNs under discussion.

The first steps in structuring the relevant IRNs involved performing a set of keyword searches for each of the selected research domains on the following databases: i) ABI Inform, ii) Web of Science, iii) SSRN; and, iv) Google Scholar. The searches utilized a broad set of keywords to include relevant research, yielding an initial set of 389 and 140 research papers in the software and music research domains respectively. Sources included working papers, proceedings, book chapters, and journals. The first filtering step was to remove working papers, book chapters, and conference proceedings from the article sets. While certainly an arbitrary choice, we made this decision based on two factors: 1) information systems conference proceedings papers most often represent work-in-progress and are typically an intermediate step to successful development of a research journal article, and 2) important/seminal papers on theory building and analysis are found in archival journals. We note that structuring an IRN in any domain will almost certainly involve selection filtering based upon relevance criteria determined by the user's interest and intended use of the IRN.

After reducing the candidate information node set to journal articles, we further reduced the set by removing articles in journals failing to reach minimum standards on “importance”, using filters based on the following journal quality measures: 1) ISI Web of Knowledge impact ratings and half life ratings, and 2) perceptions of journal quality using various published works on journal rankings. Journals rated on the ISI Web of Science with an impact factor of less than 0.5 and half life less than five years were candidates to be dropped. However, recognizing that it takes time to be accepted into the ISI Web of Science ratings, we included relatively new and rising journals whose mission and scope fit the research domains to be analyzed. Table 2 provides the list of journals containing the information node elements (papers) used in our analysis.[1] Finally, we also checked the citation list of included papers to identify relevant articles that may have been missed in the initial keyword search. In this step, only a small number of additional papers were found and included in the research paper sets. The small number identified in this step validated that the keyword search process was fairly robust.

[1] The set of journals cover subject areas including economics, law, supply chain, technology, MIS, culture, media, behavior, computing, marketing, education, business, statistics, system dynamics, and social change. Journals that indicated no focus on information systems, information technology, digital goods, or markets were excluded from analysis, given the domains of the IRN being developed.

Table 3
Size of initial and final data sets for music and software domains.

Research domain	Number of papers in initial candidate set	Number of papers in reduced set (quality screened journals)	Number of papers added from citation list of reduced set	Number of papers in final set
Music	140	55	10	65
Software	389	59	8	67

It also suggests another important criterion — the low prevalence of "extraneous" citations among the chosen set of quality journals. The risk of such citations diluting the research network under study would be a concern if simple citation counts were utilized. After completing the steps detailed above, the final set of information resource elements (research articles) included 67 software papers and 65 music papers (a complete list of the final paper set and abbreviated paper ID for each is included in sections of the References). Table 3 traces the sizes of the data sets at each step of the filtering process.

4. Demonstration of information resource network approach

Figs. 1, 2, and 3 present the representations of our IRNs for the software, music, and combined software and music research paper sets. Comparing Figs. 1 and 2 visually, we see a more densely connected IRN for software than for music. In addition, contemporaneous research links are quite evident in the software research network, but very limited in the music research network. This suggests that published research in software more closely follows concurrent and past research activity and findings. There are far more connected nodes in the software IRN than in the music IRN. In fact, there are a substantial number of isolated nodes in the music IRN, while there are only a few in the software IRN. This is potentially a reflection of the relative maturity of the software IRN together with a focus on research topics that are related. In contrast, the music IRN shows the diversity of research ideas generated in that research domain.

We applied our IRN metrics, (detailed in Section 2), to the two IRN domains using the various research paper sets. As detailed in Table 4, the density of the software research network is almost three times than that of the music research network. Tables 5 and 6 provide measures for the top papers (based on our IRN metrics) in the music and software IRNs, respectively. Figs. 4 and 5 present the data in chart form for a visualization that compares the different metrics.

As detailed in Table 4, the density of the software research network is almost three times than that of the music research network. Tables 5 (music) and 6 (software) provide comparative rank measures for the top (higher ranked) papers using each of our metrics: indegree rank, norm-indegree rank, seminal node rank, innovator node rank, norm-innovator rank. Figs. 4 and 5 present the data in chart form for a visualization that compares the different metrics.

In the combined network illustrated in Fig. 3, there are many references to software research from music research papers, but none in the other direction. Again, that may be a reflection of the maturity of the software research domain compared to the music research domain, given that software appeared in the digital good marketplace before music. We note that digitized software faced significant piracy, pricing, and related issues in the market. Digitized music, which followed later, faced similar market forces. In addition, some research in the music domain references the software domain but only to point out the inherent differences in the two areas and to call for greater research to understand the unique dimensions of the music domain. In short, while both music and software are digital goods, these domains involve different market forces and issues with some commonality. It is the related yet distinct nature of the two research domains which prompted us to use them as examples in this demonstration. We summarize our measures for the top papers in the combined network in Table 7.

We note that the software IRN is characterized by greater consistency in rankings across metrics than is the music IRN. In fact, using the normalized *bbv* ranking and ignoring the non-normalized *bbv* ranking yields almost identical ranking across the four measures

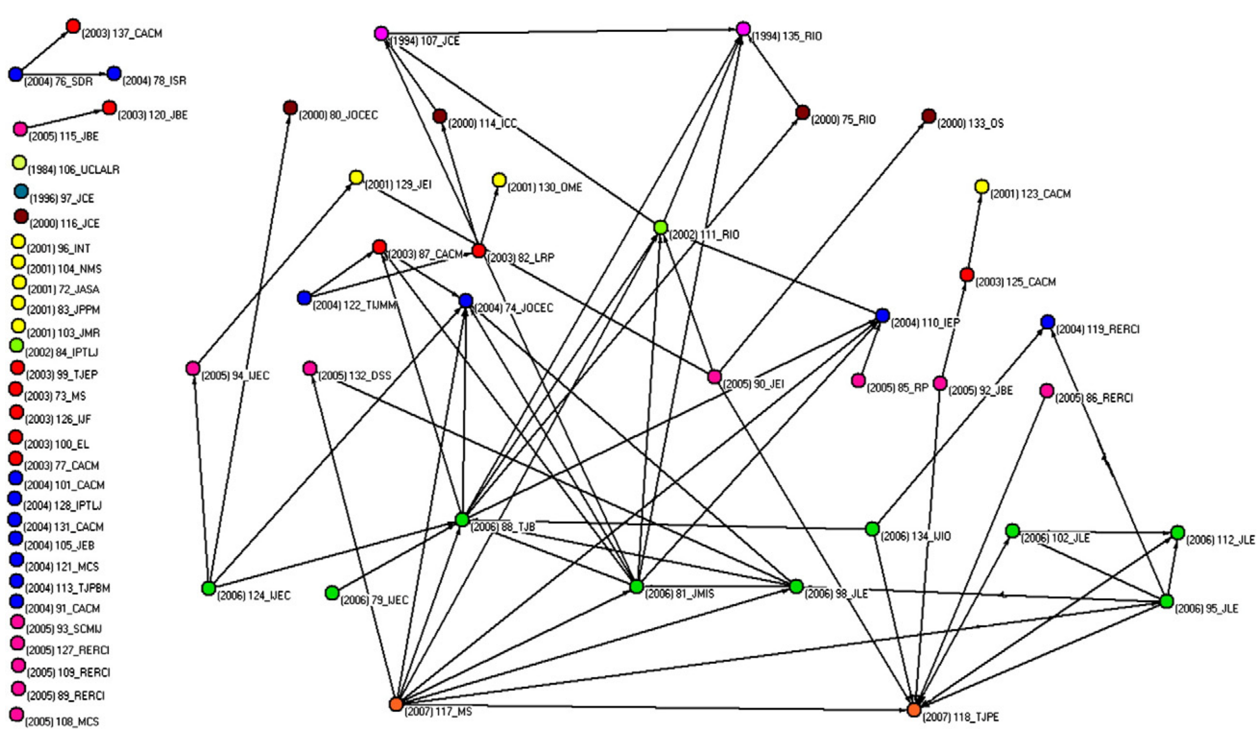

Fig. 1. IRN — music domain.

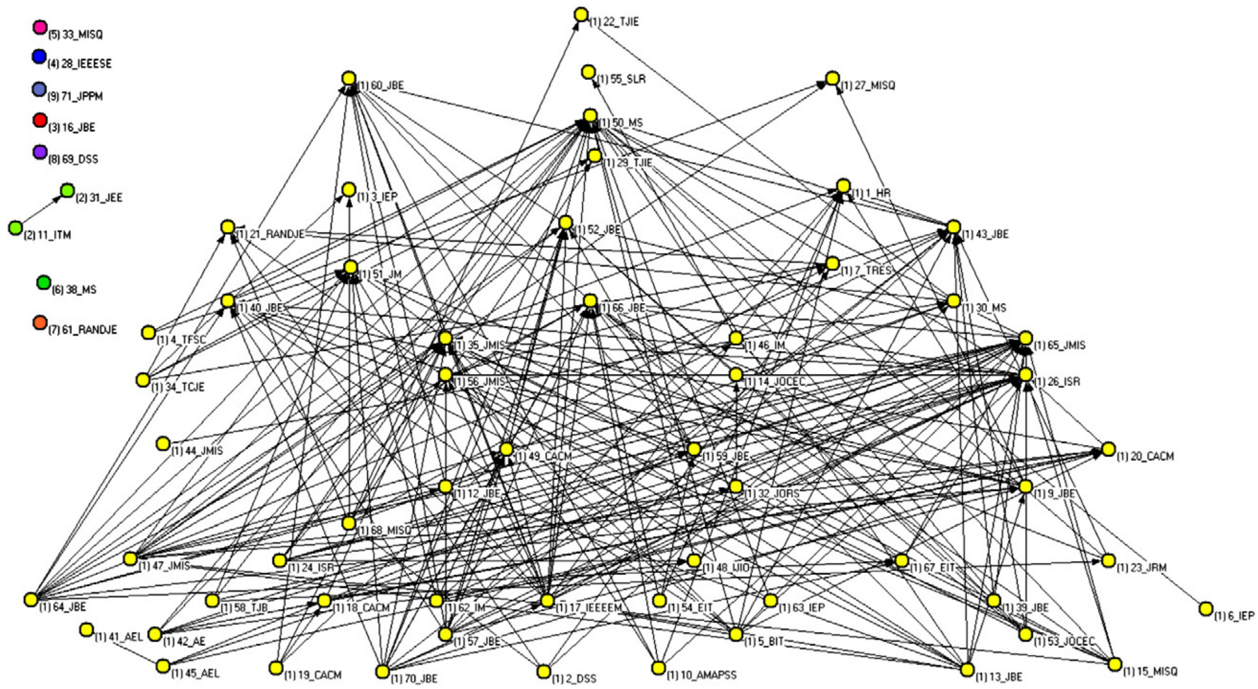

Fig. 2. IRN — software domain.

in the software IRN. We conjecture that this may reflect that the software IRN has reached a more stable setting while the music IRN is, comparatively, still in an emerging state. The relative higher density of the software network may play a stabilizing role here. However, our purpose here is illustrate the IRN approach rather than to explain the level of similarity in the rankings.

4.1. Path analysis of the IRN

We now identify major interconnected "paths of knowledge" in the example IRNs. Scientific knowledge increases over time and is cumulative. New research articles build on previous articles and previous articles are often cited until new results modify or contradict

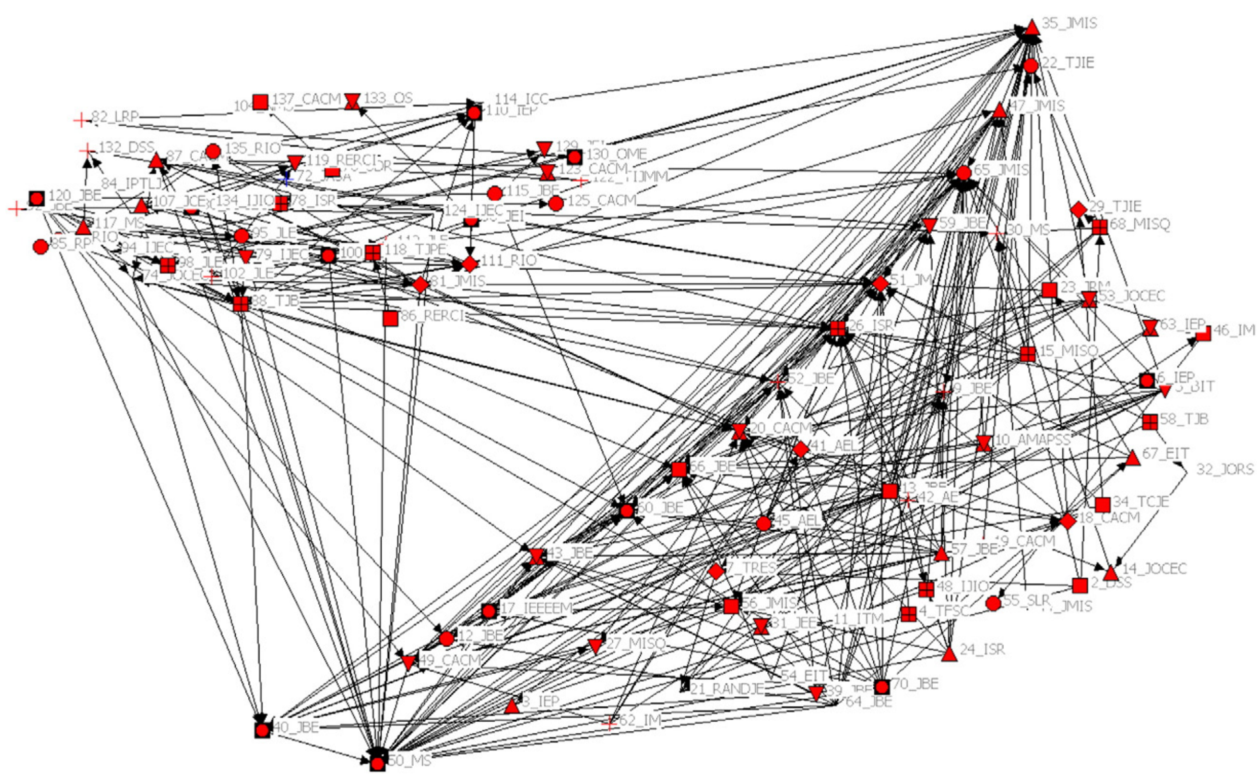

Fig. 3. Combined IRN — music and software.

Table 4
Research network density.

IRN	Number of papers	Total number of links	Maximum possible number of links	IRN density
Music	65	69	2346	0.0294
Software	67	224	2369	0.094
Combined	132	341	9483	0.036

them. While the metrics developed above help to identify critical IRN nodes, the metrics do not directly delineate the historical development or the path of discovery in the research domain. To identify IRN nodes that were vital to the development of the IRN domain over time, we used a technique called main path analysis of a network. As discussed earlier, [7] demonstrated such a path for a relatively small research network, but there was no compensation for the bias of link weights at the edges of the network. Further, the method employed exhaustive search on a network to determine link weights, a process that can be prohibitively expensive for a large IRN. Here we can think of a directed arc in an IRN as an information conduit and the IRN as a system of conduits through which knowledge or information flows. If an IRN node builds on existing knowledge in the field and makes substantial new contribution by modifying or contradicting existing results, that node will be both a starting point and target for many directed arcs and will thus act as an important hub node through which knowledge flows. In an IRN, a node that falls on the path between many nodes is more important than an isolated node. The directed, acyclic graph of most important IRN nodes over a specified time will constitute one or more main paths.

We define a *source node* as a node with zero out-degree and a *sink node* as a node with zero in-degree. In our IRN examples, the earliest nodes (papers) are likely members of the set of sources and the most

Table 5
Ranking of articles in music network.

Information node (paper) ID (full citations in references)	Indegree rank	Norm-indegree rank	Seminal node rank	Innovator node rank	Norm-innovator node rank
118_TJPE	1	1	2	3	1
88_TJB	2	2	6	19	2
112_JLE	8	3	2	10	3
102_JLE	10	4	9	20	4
98_JLE	10	4	9	25	5
81_JMIS	10	4	9	27	6
74_JOCEC	2	7	1	2	7
110_IEP	6	8	6	4	8
111_RIO	4	9	2	9	9
132_DSS	10	11	9	4	10
87_CACM	6	11	6	4	12
135_RIO	4	13	2	1	13
119_RERCI	10	14	14	4	14
129_JEI	10	17	14	4	16

Table 6
Ranking of articles in software network.

Information node (paper) ID	Indegree rank	Norm-indegree rank	Seminal node rank	innovator node rank	Norm-innovator node rank
50_MS	1	1	1	1	1
65_JMIS	2	2	4	8	2
35_JMIS	3	3	3	10	3
26_ISR	4	5	2	3	3
66_JBE	5	4	5	15	5
56_JMIS	6	12	8	2	7
20_CACM	9	11	5	7	10
49_CACM	11	14	12	3	14
30_MS	15	17	13	5	15

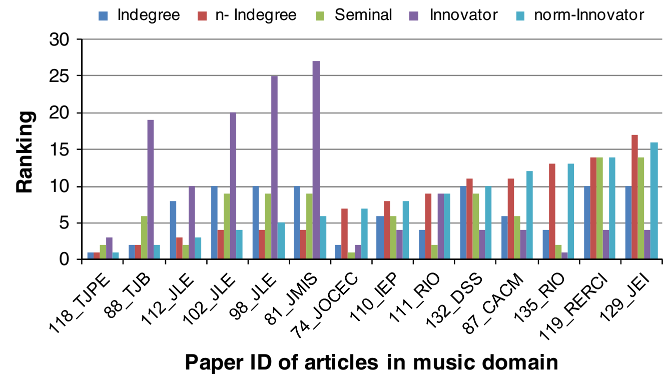

Fig. 4. Article rankings: music network.

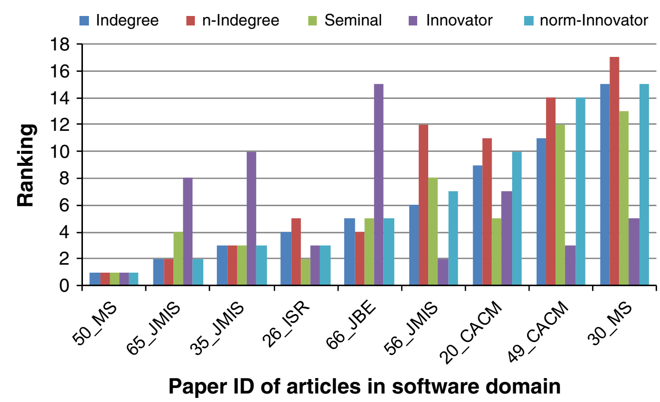

Fig. 5. Article rankings: software network.

recent nodes (latest papers) are likely members of the set of sinks. (N.B. the exceptions would be if there were any pre-time citations due to pre-publication circulation of the work). In our main path analysis, we first find all paths in the IRN that run between any source and any sink. We then assign each arc on any such path a "path weight" based on how many of these paths the cited work falls on relative to the total number of paths. That is, for a given IRN node (research paper), we divided the number of paths of which the node appears by the total number of paths. Once the weights are determined, we extract main paths based on these weights. We then remove all the arcs with value less than a certain *critical value* and extract the maximal connected graph from the remaining arcs. The critical value choice is arbitrary, so we complete the analysis for several networks resulting from a variety of critical values. For example, using a critical value of 0.10 yields nodes that that appear on at least 10% of the paths while using a

Table 7
Ranking of articles in combined network.

IRN node (paper) ID	Indegree rank	Norm-indegree rank	Seminal node rank	bbv rank	Norm-bby rank
50_MS	1	3	1	1	3
35_JMIS	2	4	2	10	4
66_JBE	5	8	3	4	7
65_JMIS	3	5	4	8	5
26_ISR	4	6	5	18	8
51_JM	6	9	8	5	9
60_JBE	7	12	8	2	12
1_HR	12	21	12	3	19
118_TJPE	13	1	15	17	1
88_TJB	16	2	20	54	2

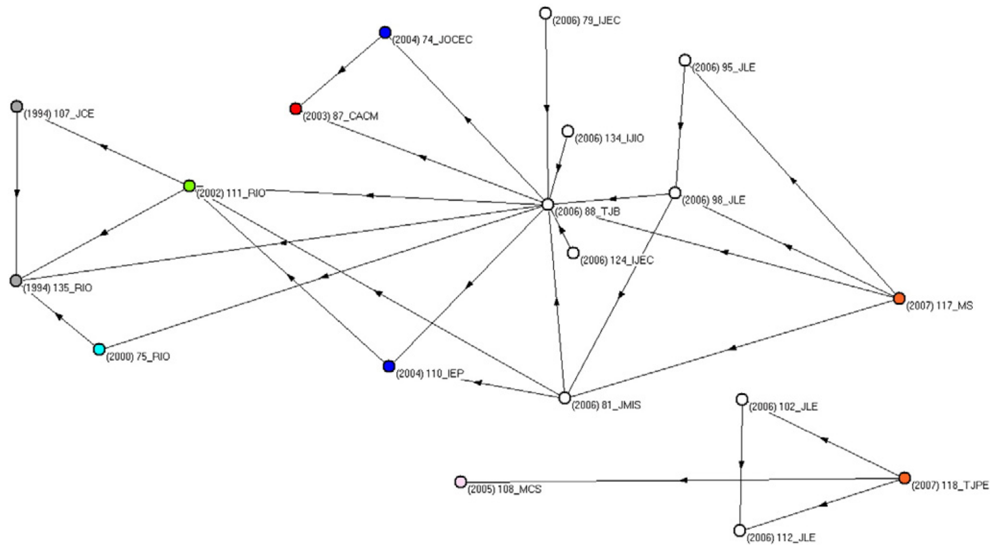

Fig. 6. Music IRN main paths — critical value of 0.025.

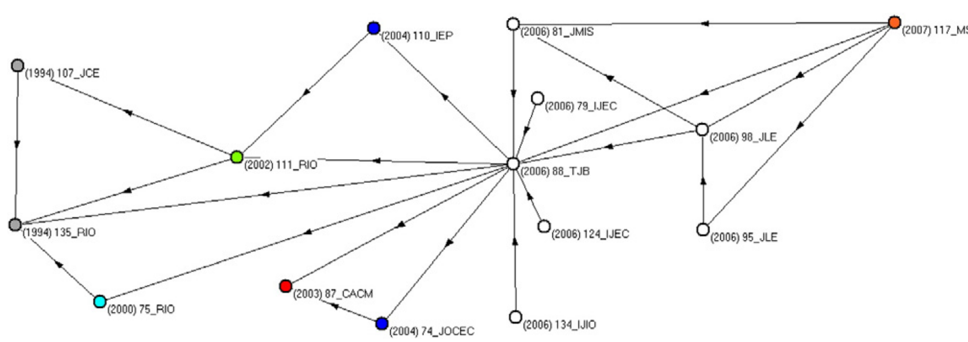

Fig. 7. Music IRN main paths — critical value of 0.05.

critical value of 0.15 yields the set of nodes appearing on at least 15% of the paths. As one would anticipate, the 0.15 critical value maximal connected graph is "smaller" than the 0.10 maximal connected graph. We extracted networks based on cut-off values in steps of 0.025, starting at 0.025 and ending when there were no connected components.

The path analysis provides a set of nodes that are measurably inter-linked within the IRN. These nodes are identified as lying on at least a specified percentage of all paths from sources to sinks. Take, for example, a professor structuring a new graduate seminar that includes analysis of music as a digital good. If the professor was considering including up to 10 such papers, then she might consider the critical value outcome for 0.1 which yields 11 such papers. Or perhaps a Ph.D. or faculty researcher is investigating the software digital goods research area. Using a critical value of 0.05 helps the researcher to reduce the space from 140 candidate papers to 17 that appear central to the research domain. Using a critical value of 0.025 would increase the number of papers to 28 while using a critical value of 0.075 would reduce the number to 14. Path analysis can be a helpful first step in reducing the space to a manageable initial set of key papers, a structured way to identify a "starting point" of obtaining knowledge that appears central to the research arena. Conversely, it can also identify areas for future research.

Figs. 6–9 present the main paths for our music IRN using cutoff values of 0.025, 0.05, 0.075, and 0.1, respectively. Figs. 10–13 present the main paths for our software IRN, again using cutoff values of 0.025, 0.05, 0.075, and 0.1.

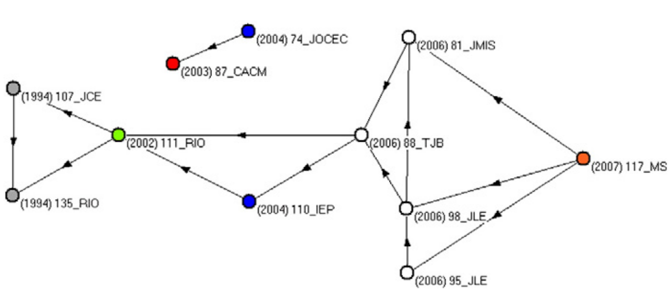

Fig. 8. Music IRN main paths — critical value of 0.075.

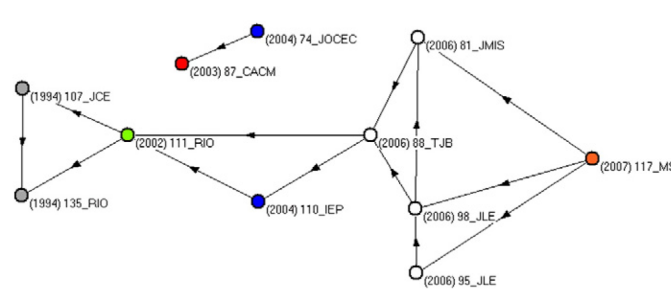

Fig. 9. Music IRN main paths — critical value of 0.1.

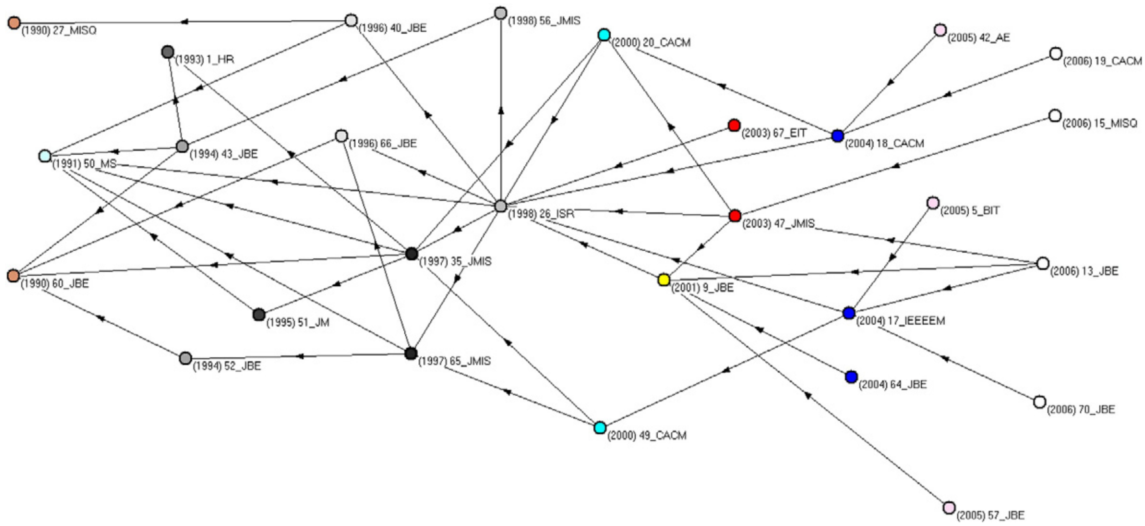

Fig. 10. Software IRN main paths — critical value of 0.025.

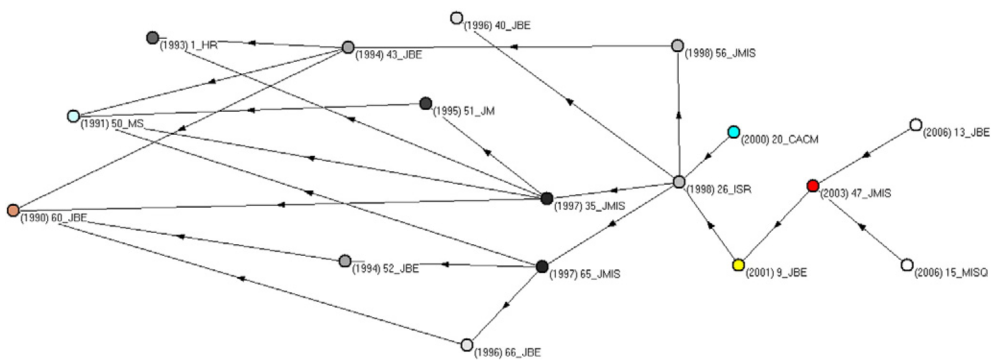

Fig. 11. Software IRN main paths — critical value of 0.05.

There is no rule set for pre-selecting the "most appropriate" cutoff value. Rather, this selection is dependent upon the purpose of the investigation. For example, one purpose we suggested earlier involved a faculty member selecting a set of papers on specific topics. If the goal was to select 10–15 papers in each of the two domains we consider here, then a cutoff value in the range of 0.075 would appear appropriate. A researcher seeking to "tool up" in a field might select a lower cutoff value, hence including additional articles.

Tables 5 and 6 provide information on the top ranked nodes in the music and software IRNs. In comparing the set of top ranked nodes in these tables to nodes included in our main paths, we find that the top seven ranked papers are contained in the main path for software with a cutoff of 0.075. In the music research domain, the main path with a cutoff value of 0.075 contains only two of the top ranked papers, but the main path with a 0.025 cutoff contains each of the top nine ranked papers from Table 5.

By restricting path analysis to subsets of an IRN, path analysis can also be used to further parse the domain and be of additional value to investigators. For example, by "time-restricting" the network, we can use path analysis to identify nodes that are key in the most recent segments of the network, that is, nodes that are leading the information trend. In the same way, we may identify nodes that might be important early in an IRN, but at some point fell off in importance and have little significance in more recent times of the network.

4.2. Lags in information dissemination

The metrics computed above are impacted by the timing of information resource dissemination. In our examples, the "delay in disseminating an information resource" can negatively impact the presence of in-directed arcs (citations, if you will) in the time period immediately following the node's emergence. The focus of the field of discussion may have shifted during this lag, or multiple other

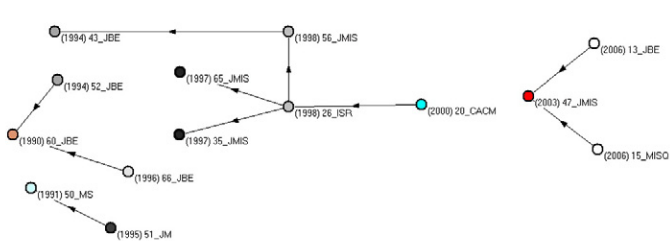

Fig. 12. Software IRN main paths — critical value of 0.075.

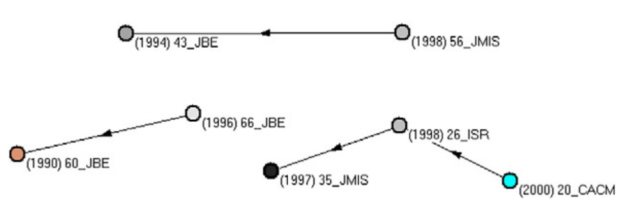

Fig. 13. Software IRN main paths — critical value of 0.1.

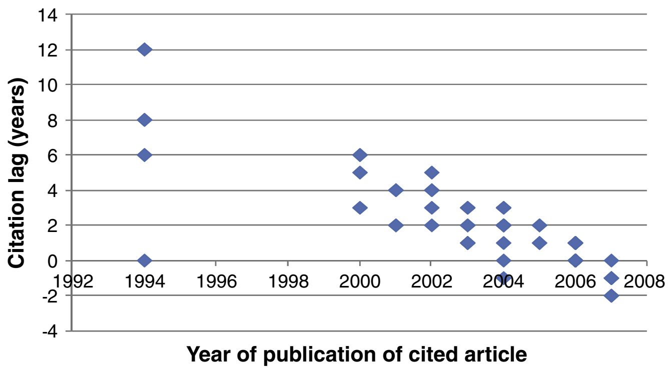

Fig. 14. Citation lag for music IRN.

information resource nodes may have espoused the core concept in different ways during the dissemination delay. That is, if the publication of a paper (in our continuing example) is delayed significantly, there may be a diminished metric values of the paper (node). Thus we consider lags in information dissemination (citation lags) for each of our research domains.

Figs. 14 and 15 provide scatter plots of the citations to papers appearing in each of the years for the time duration of the music and software research domains, respectively. In each case, the scatter plots suggest that citation lag has been decreasing, a not surprising result in our age of faster information availability among researchers through electronic posting (pre-publication) on various journal websites, electronic library and individual subscription access, and research communities such as SSRN where researchers share their latest research. In fact, there are several citations with a zero year lag (and one specific paper that was widely promoted and circulated prior to publication having a negative lag) suggesting a great awareness of current on-going research. We have included this specific paper in the analysis based on its wide circulation and pre-publication citations. In Fig. 14, there are a few negative citation lags because of pre-publication citation. In most corporate communication and public Internet discourse, such negative lags would be highly unlikely.

Researchers seem to understand the importance of time in research arenas that can change so rapidly. In addition, the continuing lags suggest the importance of researchers becoming active in "pre-publication networks" to facilitate access to research without incurring the significant time delay inherent in academic journal publication environments.

In summary, Section 4 demonstrates how an information resource network can be structured and analyzed to identify influential information resources within a domain under consideration. As discussed earlier, the number of influential resources extracted from such a network is dependent on the context or use of the information, and, in our approach, it is the investigator that makes this determination. In the next section, we illustrate a related result which focuses on the sources of the information nodes that we used in this section — rather than the information nodes themselves.

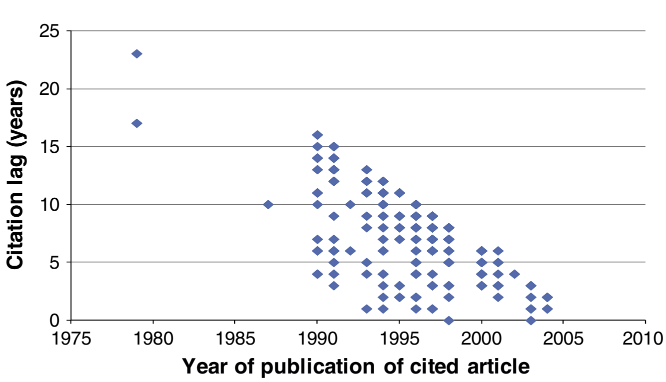

Fig. 15. Citation lag for software IRN.

5. Classifying the sources of information nodes

We now illustrate the flexibility of our approach and metrics by applying the processes in a related context. The nodes in an IRN can be sourced in many different ways, including from blogs. Blogs may be hosted by different providers, and the question of interest may relate to the prominence of certain providers for specific information context in the blogs. Yet another investigation (on instant messages, say) may focus on the provider network where most influential messages originated or from where significant messages were transmitted. This has significant application in resource planning related to managing important segments of the network, including controlling segments of cyberspace or mobile networks for law enforcement purposes. In an IRN within a company, a corresponding information source would be the individual or group that initially developed or introduced the information (data) into the company IRN. Hence an IRN can also be a useful asset in analyzing the prominence of various sources of information nodes in a specific domain.

Consider our running example. The nodes of a related IRN might represent the information source (journal) for the information elements in our earlier IRNs. In this IRN, the arcs between nodes are links between information sources (research journals) enabling the computation of IRN properties at the information source (journal) level. For example, a directed arc from an article appearing in journal A to an article appearing in journal B indicates a citation flow from journal A to journal B. We now apply our method and metrics to this new IRN with the focus now on the information sources rather than the information itself.

Table 8
Journal ranking in music IRN.

IRN node source (Journal ID*)	*norm-j-indegree rank*	*norm-Innovator rank*	Number of papers
(a) (Including intra-journal citations)			
JPE	1	1	1
JB	2	2	1
JLE	2	2	4
JMIS	2	2	1
DSS	5	5	1
IJEC	5	6	3
ISR	7	7	1
RERCI	7	8	5
IEP	7	9	1
JOCEC	10	10	2
JBE	11	11	3
LRP	11	12	1
CACM	13	13	8
OME	14	14	1
JEI	14	15	2
(b) (Excluding intra-journal citations)			
JPE	1	1	1
JB	2	2	1
JMIS	2	2	1
JLE	4	4	4
DSS	5	5	1
RERCI	6	6	5
ISR	7	7	1
IEP	7	8	1
JOCEC	9	9	2
LRP	10	10	1
CACM	11	11	8
OME	12	12	1
OS	13	13	1
ICC	13	14	1
JCE	15	16	3
RIO	16	15	3

* See Table 2 for full journal names.

Using the journal identifications noted earlier, we formed journal IRNs for the music domain, for the software domain, and for the music and software domains combined. Tables 8a–b, 9a–b, and 10a–b summarize two key measures (normalized indegree ranking and normalized journal *bbv* ranking — defined below) and indicate the number of relevant papers appearing in the top fifteen ranked journals in each category in music, software, and combined, respectively. The "a" version tables include intra-journal citations, while the "b" version tables exclude the intra-journal citations. We provide both sets of tables as a means to view how integrated a journal's articles in these subject areas are with the broad set of articles in the areas. Our metrics are defined as follows:

j-indegree$_k$: number of incoming arcs to information nodes appearing in source k from other nodes appearing in other sources
j-outdegree$_k$: number of outgoing arcs from information nodes appearing in source k to other nodes appearing in other sources
norm-j-indegree$_k$ (normalized indegree): *j-indegree*$_k$/*j-max indegree*$_k$ (*j-max indegree*$_k$ defined as the sum of max indegrees for all nodes appearing in source k)
norm-j-outdegree$_j$ (normalized outdegree): *j-outdegree*$_k$/*j-max outdegree*$_k$
(*j-max outdegree*$_k$ defined as the sum of max outdegrees for all nodes in source k)

Similar to Section 2, an "innovator" journal node is defined by:

$$j{-}bbv_k = j{-}indegree_k \,/\, (j{-}outdegree_k + 1).$$

The normalized "innovator" measure is given by:

$$norm{-}j{-}bbv_k = norm{-}j{-}indegree_k \,/\, (norm{-}j{-}outdegree_k + 1).$$

Table 9
Journal ranking in software IRN.

Journal ID	norm-j-indegree rank	norm-Innovator rank	Number of papers
(a) (Including intra-journal citations)			
AEL	1	1	2
IEEEEM	1	2	1
EIT	3	3	2
IJIO	3	4	2
JRM	3	4	1
JEE	6	6	1
JORS	6	7	1
CACM	8	8	4
JOCEC	9	9	2
ISR	9	10	2
JMIS	11	11	5
IM	12	12	2
JBE	13	13	14
JM	14	16	1
RES	14	16	1
MISQ	16	14	4
RANDJE	17	15	2
(b) (Excluding intra-journal citations)			
IEEEEM	1	1	1
EIT	2	2	2
JRM	3	3	1
IJIO	3	4	2
JEE	5	5	1
JORS	5	6	1
JOCEC	7	7	2
CACM	8	8	4
JMIS	9	9	5
ISR	10	10	2
MISQ	11	11	4
JM	12	14	1
RES	12	14	1
RANDJE	14	12	2
IEP	15	13	3

Table 10
Journal ranking in combined IRN.

Journal ID	norm-j-indegree rank	norm-Innovator rank	Number of papers
(a) (Including intra-journal citations)			
JPE	1	1	1
JLE	2	2	4
JB	2	3	2
DSS	4	4	3
IJEC	4	5	3
AEL	6	6	2
RERCI	6	7	5
IEEEEM	6	8	1
JOCEC	9	9	4
IJIO	10	10	2
EIT	10	11	2
LRP	10	12	1
JRM	10	12	1
IEP	14	14	4
JEE	15	15	1
OME	15	15	1
JEI	15	17	2
JORS	15	18	1
(b) (Excluding intra-journal citations)			
JPE	1	1	1
JB	2	2	2
JLE	3	3	4
DSS	4	4	3
RERCI	5	5	5
IEEEEM	6	6	1
IJIO	7	7	2
EIT	7	8	2
LRP	9	9	1
JRM	9	9	1
IEP	11	11	4
JOCEC	12	12	4
JEE	13	13	1
OME	13	13	1
JORS	13	15	1

The structure of these metrics closely follows those defined in Section 2, and we do not detail them again here. As before, we focus on capturing the time-ordered nature of such linkages, and our approach minimizes the temporal bias that is inherent in existing metrics.

First we note that several of the journals listed have only one article in an area, yet many of them are ranked higher than other journals with several articles in an area. The metrics we set forward are based on impact in terms of normalized in-degrees (citations to an article) and building block (*bbv*) values (citations to an article relative to possible citations to that article from papers in the original IRN). We reiterate that normalized building block value (*bbv*) of a journal utilizes the number of citations relative to the number of possible citations (i.e. time subsequent research papers) that could have referred to the particular paper. Our measures should not be construed as indicators of the overall quality of a journal. Rather, we illustrate the use of IRNs and report how sets of papers (possibly only a single paper) in a specific information domain that have a specific journal as their source perform on specific IRN metrics. Including intra-journal citations did not result in much difference in the ordering of journals in either the music or software IRNs.

6. Concluding remarks

We introduced the concept of information resource networks (IRNs) and suggested their use in a variety of settings that involve time-stamped and directed introduction of information resources. IRNs provide a representation of potentially interrelated information nodes along with the specific linkages that occur. We developed and presented a set of IRN metrics and illustrated the IRN concept and metric computations using two information domains — the intersection of market forces with two digital goods, software and music. We also illustrated how the IRN approach can be utilized with the nodes indicating the information sources (here academic journals).

Our specific domain choices were used for illustration (and proof of concept) purposes. The structured IRN approach can be applied in any number of information resource settings (examples provided along with the exposition above) as long as the linkages or arcs between nodes are directed and nodes are time-stamped. To expand the validation of the IRN network formulation, future research will focus on additional domain-specific applications. Our goal in this work will be to investigate the consistency of usefulness or "goodness" of each of the individual metrics developed and presented here. We conjecture that the IRN approach will meet these challenges but we must provide sufficient empirical support.

Time-directed and interconnected information streams have proliferated inside corporations and in the public domain. E-mail conversations, document management system logs, supply chain transactions, blogs and the generation of "buzz" online, patent litigation, and numerous other applications abound for structured analysis of time-directed interconnections. The IRN approach provides a set of tools to discover patterns, linkages, and underlying relationships in such environments and to help investigators to better understand specific domains.

Appendix A

Table A1
IRN metrics: definitions and formulae.

	Definition	Formulae
IRN (Information Resource Network)	It is a time ordered directed graph which has (1) set I of information elements as its vertices and (2) set A of time ordered directed edges (arcs) between information elements.	$G=(I,A)$
N_i	Total number of information elements (nodes) in a given network i	$N_i = \sum_{k=1}^{T} n_k$, where n_k is total number of elements in IRN indexed in time k
$indegree_j$ (indegree)	Number of directed arcs to node j from other nodes in network i	
$outdegree_j$ (outdegree)	Number of directed arcs from node j to other nodes in network i	
$max\ indegree_j$ (maximum indegree)	Maximum outdegree for an information element j indexed in period t, $max\ indegree_j$ is the maximum possible directed arcs to node j from other nodes	$max\ indegree_j = (n_t - 1) + \sum_{k=t+1}^{T} n_k$
$max\ outdegree_j$ (maximum indegree)	Maximum outdegree for an information element j indexed in period t, $max\ outdegree_j$,is the maximum possible directed arcs from node j to other nodes	$max\ outdegree_j = (n_t - 1) + \sum_{k=1}^{t-1} n_k$
$norm-indegree_j$ (normalized indegree)	Normalized indegree is the ratio of indegree to maximum indegree for node j	$norm-indegree_j = \frac{indegree_j}{max\ indegree_j}$
$norm-outdegree_j$ (normalized outdegree)	Normalized outdegree is the ratio of outdegree to maximum outdegree for node j	$norm-outdegree_j = \frac{outdegree_j}{max\ outdegree_j}$
bbv_j *(building block value)*	Building block value of node j represents the new information that a node j brings in to network and is defined as the ratio of indegree to out degree for node j.	$bbv_j = \frac{indegree_j}{outdegree_j + 1}$
$norm-bbv_j$ *(normalized building block value)*	Normalized building block value of node j is the ratio of normalized indegree to normalized outdegree for node j.	$bbv_j = \frac{norm-indegree_j}{norm-outdegree_j + 1}$
sv_j (seminal node)	A seminal node is a type of exceptional building block node, where the network's late time-indexed nodes continue to have directed arcs to the seminal node.	$sv_j = \sum_{kr} \{(link_{kj})(link_{rk})(link_{rj})\}$ where $link_{kj}=1$ *if node k is linked to node j*, 0*otherwise*
D_i (IRN density)	IRN density is defined as ratio of total number of directed arcs in the network to maximum possible directed arcs in the network	$D_i = \frac{\sum_j outdegree_j + indgeree_j}{\sum_j max\ outdegree_j} = \frac{\sum_j outdegree_j + indgeree_j}{\sum_j max\ indegree_j}$

References

[1] R. Bapna, P. Goes, R. Gopal, J.R. Marsden, Moving from data-constrained to data-enabled research: Experiences and challenges in collecting, validating and analyzing large-scale e-commerce data, Statistical Science 21 (2) (2006) 116–130.
[2] A.L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, Evolution of the social network of scientific collaborations, Physica A: Statistical Mechanics and its Applications 311 (3–4) (2002) 590–614.
[3] L. Bornmann, R. Mutz, H.D. Daniel, Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine, Journal of the American Society for Information Science and Technology 59 (5) (2008) 830–837.
[4] M.M. Camacho-Minano, M. Nunez-Nickel, The multilayered nature of reference selection, Journal of the American Society for Information Science and Technology 60 (4) (2009).
[5] K. Faust, S. Wasserman, Social network analysis: Methods and applications, Cambridge University Press, New York, 1994.
[6] M. Gmür, Co-citation analysis and the search for invisible colleges: A methodological evaluation, Scientometrics 57 (1) (2003) 27–57.
[7] N.P. Hummon, P. Dereian, Connectivity in a citation network: The development of DNA theory, Social Networks 11 (1) (1989) 39–63.
[8] H. Kretschmer, Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the Web, Scientometrics 60 (3) (2004) 409–420.
[9] J.G. Liu, Z.G. Xuan, Y.Z. Dang, Q. Guo, Z.T. Wang, Weighted network properties of Chinese nature science basic research, Physica A: Statistical Mechanics and its Applications 377 (1) (2007) 302–314.
[10] X. Liu, J. Bollen, M.L. Nelson, H. Van de Sompel, Co-authorship networks in the digital library research community, Information Processing and Management 41 (6) (2005) 1462–1480.
[11] N. Ma, J. Guan, Y. Zhao, Bringing PageRank to the citation analysis, Information Processing and Management 44 (2) (2008) 800–810.
[12] J.R. Marsden, The Internet and DSS: massive, real-time data availability is changing the DSS landscape, Information Systems and E-Business Management 6 (2) (2008) 193–203.
[13] M.A. Nascimento, J. Sander, J. Pound, Analysis of SIGMOD's co-authorship graph, ACM SIGMOD Record 32 (3) (2003) 8–10.

[14] M.E.J. Newman, Scientific collaboration networks. I. Network construction and fundamental results, Physical Review E 64 (1) (2001) 016131.
[15] M.E.J. Newman, The structure of scientific collaboration networks, Proceedings of the National Academy of Sciences of the United States of America 98 (2) (2001) 404.
[16] M.A. Rodriguez, A. Pepe, On the relationship between the structural and socioacademic communities of a coauthorship network, Journal of Informetrics 2 (3) (2008) 195–201.
[17] D.F. Thompson, E.C. Callen, M.C. Nahata, New indices in scholarship assessment, American Journal of Pharmaceutical Education 73 (6) (2009).
[18] A. Verbeek, K. Debackere, M. Luwel, E. Zimmermann, Measuring progress and evolution in science and technology–I: The multiple uses of bibliometric indicators, International Journal of Management Reviews 4 (2) (2002) 179–211.
[19] R. Vidgen, S. Henneberg, P. Naude, What sort of community is the European Conference on Information Systems? A social network analysis 1993–2005, European Journal of Information Systems 16 (1) (2007) 5–19.
[20] L. Yin, H. Kretschmer, R.A. Hanneman, Z. Liu, Connection and stratification in research collaboration: an analysis of the COLLNET network, Information Processing and Management 42 (6) (2006) 1599–1613.

Citations for the Research Network Nodes
Music Research Network Papers

[21] (135_RIO). P.J. Alexander, Entry barriers, release behavior and multi-product firms in music recording industry, Review of Industrial Organization 9 (1) (1994) 85–98.
[22] (107_JCE). P.J. Alexander, New technology and market structure: evidence from the music recording industry, Journal of Cultural Economics 18 (2) (1994) 113–123.
[23] (80_JOCEC). K. Altinkemer, S. Bandyopadhyay, Bundling and distribution of digitized music over the internet, Journal of Organizational Computing and Electronic Commerce 10 (3) (2000) 209–224.
[24] (133_OS). N. Anand, R.A. Peterson, When market information constitutes fields: sensemaking of markets in the commercial music industry, Organization Science 11 (3) (2000) 270–284.
[25] (97_JCE). E.B. Andrew, How effective are international copyright conventions in the music industry? Journal of Cultural Economics 20 (1) (1996) 51.
[26] (78_ISR). A. Asvanund, K. Clay, R. Krishnan, M.D. Smith, An empirical analysis of network externalities in peer-to-peer music-sharing networks, Information Systems Research 15 (2) (2004) 155–175.
[27] (126_IJF). R. Bewley, W.E. Griffiths, The penetration of CDs in the sound recording market: issues in specification, model selection and forecasting, International Journal of Forecasting 19 (1) (2003) 111.
[28] (81_JMIS). S. Bhattacharjee, R.D. Gopal, K. Lertwachara, J.R. Marsden, Consumer search and retailer strategies in the presence of online music sharing, Journal of Management Information Systems 23 (1) (2006) 129–159.
[29] (87_CACM). S. Bhattacharjee, R.D. Gopal, G.L. Sanders, Digital music and online sharing: software piracy 2.0? Association for Computing Machinery, Communications of the ACM 46 (7) (2003) 107–111.
[30] (98_JLE). S. Bhattacharjee, R.D. Gopal, K. Lertwachara, J.R. Marsden, Impact of legal threats on online music sharing activity: an analysis of music industry legal actions, Journal of Law and Economics 49 (1) (2006) 91–114.
[31] (132_DSS). S. Bhattacharjee, R.D. Gopal, K. Lertwachara, J.R. Marsden, Whatever happened to payola? An empirical analysis of online music sharing, Decision Support Systems 42 (1) (2005) 104–120.
[32] (117_MS). S. Bhattacharjee, R.D. Gopal, K. Lertwachara, J.R. Marsden, R. Telang, The effect of digital sharing technologies on music markets: a survival analysis of albums on ranking charts, Management Science 53 (9) (2007) 1359–1374.
[33] (124_IJEC). J.C. Bockstedt, R.J. Kauffman, F.J. Riggins, The move to artist-led on-line music distribution: a theory-based assessment and prospects for structural changes in the digital music market, International Journal of Electronic Commerce 10 (3) (2006) 7–38.
[34] (72_JASA). E.T. Bradlow, P.S. Fader, A bayesian lifetime model for the "hot 100" billboard songs, Journal of the American Statistical Association 96 (454) (2001) 368–381.
[35] (115_JBE). J.S. Chiou, C.-Y. Huang, H.H. Lee, The antecedents of music piracy attitudes and intentions, Journal of Business Ethics 57 (2) (2005) 161–174.
[36] (128_IPTLJ). S.C. Christopher, The twisted path of the music file-sharing litigation: the cases that have shaped the litigation and the RIAA's litigation strategy, Intellectual Property & Technology Law Journal 16 (10) (2004) 6–12.
[37] (110_IEP). B.M. Cunningham, P.J. Alexander, N. Adilov, Peer-to-peer file sharing communities, Information Economics and Policy 16 (2) (2004) 197–213.
[38] (92_JBE). R.F. Easley, Ethical issues in the music industry response to innovation and piracy, Journal of Business Ethics 62 (2) (2005) 163–168.
[39] (125_CACM). R.F. Easley, J.G. Michel, S. Devaraj, The mp3 open standard and the music industry's response to internet piracy. Association for Computing Machinery, Communications of the ACM 46 (11) (2003) 90–96.
[40] (96_INT). P.S. Fader, B.G.S. Hardie, Forecasting repeat sales at cdnow: a case study 2001. 31, Interfaces 3 (2) (2001) 94–107.
[41] (89_RERCI). J. Farchy, H. Ranaivoson, DRMS: a new strategic stake for contents industries: the case of the online music market, Review of Economic Research on Copyright Issues 2 (2) (2005) 53–67.
[42] (129_JEI). T. Gallaway, D. Kinnear, Unchained melody: a price discrimination-based policy proposal for addressing the mp3 revolution, Journal of Economic Issues 35 (2) (2001) 279–287.
[43] (100_EL). A. Gayer, O. Shy, Internet and peer-to-peer distributions in markets for digital products, Economics Letters 81 (2) (2003) 197–203.
[44] (88_TJB). R.D. Gopal, S. Bhattacharjee, G.L. Sanders, Do artists benefit from online music sharing?*, Journal of Business 79 (3) (2006) 1503–1534.
[45] (74_JOCEC). R.D. Gopal, G.L. Sanders, S. Bhattacharjee, M. Agrawal, S.C. Wagner, A behavioral model of digital music piracy, Journal of Organizational Computing and Electronic Commerce 14 (4) (2004) 89–105.
[46] (84_IPTLJ). S.M. Heidmiller, Digital copying and file sharing on trial, Intellectual Property & Technology Law Journal 14 (4) (2002) 1–8.
[47] (94_IJEC). C.Y. Huang, File sharing as a form of music consumption, International Journal of Electronic Commerce 9 (4) (2005) 37–55.
[48] (101_CACM). E. Jason, G.F.T. Bailes, Managing p2p security. Association for Computing Machinery, Communications of the ACM 47 (9) (2004) 95–98.
[49] (127_RERCI). B. Keintz, The recording industry's digital dilemma: challenges and opportunities in high-piracy markets, Review of Economic Research on Copyright Issues 2 (2) (2005) 83–94.
[50] (104_NMS). M. Kretschmer, G.M. Klimis, R. Wallace, Music in electronic markets - an empirical study, New Media & Society 3 (4) (2001) 417–441.
[51] (120_JBE). K.K. Kwong, O.H.M. Yau, J.S.Y. Lee, L.Y.M. Sin, A.C.B. Tse, The effects of attitudinal and demographic factors on intention to buy pirated cds: the case of chinese consumers, Journal of Business Ethics 47 (3) (2003) 223–235.
[52] (123_CACM). C.K.M. Lam, B.C.Y. Tan, The internet is changing the music industry. Association for Computing Machinery, Communications of the ACM 44 (8) (2001) 62–69.
[53] (99_TJEP). W. Landes, D. Lichtman, Indirect liability for copyright infringement: napster and beyond, Journal of Economic Perspectives 17 (2) (2003) 113–124.
[54] (83_JPPM). J. Langenderfer, D.L. Cook, Copyright policies and issues raised by a&m records v. Napster: "the shot heard 'round the world" or "not with a bang but a whimper?", Journal of Public Policy & Marketing 20 (2) (2001) 280–288.
[55] (137_CACM). J. Lee, An end-user perspective on file-sharing systems. Association for Computing Machinery, Communications of the ACM 46 (2) (2003) 49–53.
[56] (73_MS). J. Lee, P. Boatwright, W.A. Kamakura, A bayesian model for prelaunch sales forecasting of recorded music, Management Science 49 (2) (2003) 179–196.
[57] (93_SCMIJ). G.J. Lewis, G. Graham, G. Hardaker, Evaluating the impact of the internet on barriers to entry in the music industry, Supply Chain Management-an International Journal 10 (5) (2005) 349–356.
[58] (108_MCS). A. Leyshon, P. Webb, S. French, N. Thrift, L. Crewe, On the reproduction of the musical economy after the internet, Media Culture & Society 27 (2) (2005) 177–209.
[59] (95_JLE). S.J. Liebowitz, File-sharing: creative destruction or just plain destruction? Journal of Law and Economics 49 (1) (2006) 1–28.
[60] (131_CACM). F.V. Lohmann, Voluntary collective licensing for music file sharing. Association for Computing Machinery, Communications of the ACM 47 (10) (2004) 21–24.
[61] (121_MCS). L. Marshall, The effects of piracy upon the music industry: a case study of bootlegging, Media Culture & Society 26 (2) (2004) 163–181.
[62] (106_UCLALR). R.G. Martin, Music video copyright protection: implications for the music industry, UCLA Law Review 32 (2) (1984) 396–429.
[63] (86_RERCI). N.J. Michel, Digital file sharing and the music industry: was there a substitution effect? Review of Economic Research on Copyright Issues 2 (2) (2005) 41–52.
[64] (75_RIO). F.G. Mixon Jr., R.W. Ressler, A note on elasticity and price dispersions in the music recording industry, Review of Industrial Organization 17 (4) (2000) 465–470.
[65] (103_JMR). W.W. Moe, P.S. Fader, Modeling hedonic portfolio products: a joint segmentation analysis of music compact disc sales, Journal of Marketing Research 38 (3) (2001) 376–385.
[66] (82_LRP). L. Molteni, A. Ordanini, Consumption patterns, digital technology and music downloading, Long Range Planning 36 (4) (2003) 389–406.
[67] (130_OME). O. Muammer, User segmentation of online music services using fuzzy clustering, Omega 29 (2) (2001) 193–206.
[68] (118_TJPE). F. Oberholzer-Gee, K. Strumpf, The effect of file sharing on record sales: an empirical analysis, The Journal of Political Economy 115 (1) (2007) 1–42.
[69] (113_TJPBM). P. Papadopoulos, Pricing and pirate product market formation, The Journal of Product and Brand Management 13 (1) (2004) 56–63.
[70] (90_JEI). O.V. Pavlov, Dynamic analysis of an institutional conflict: copyright owners against online file sharing, Journal of Economic Issues 39 (3) (2005) 633–663.
[71] (76_SDR). O.V. Pavlov, K. Saeed, A resource-based analysis of peer-to-peer technology, System Dynamics Review 20 (3) (2004) 237–262.
[72] (119_RERCI). M. Peitz, P. Waelbroeck, The effect of internet piracy on music sales: cross-section evidence, Review of Economic Research on Copyright Issues 1 (2) (2004) 71–79.
[73] (134_IJIO). M. Peitz, P. Waelbroeck, Why the music industry may gain from free downloading – the role of sampling, International Journal of Industrial Organization 24 (5) (2006) 907–913.
[74] (111_RIO). J.A. Peter, Peer-to-Peer File Sharing: the case of the music recording industry, Review of Industrial Organization 20 (2) (2002) 151–161.
[75] (77_CACM). G.P. Premkumar, Alternate distribution strategies for digital music. Association for computing machinery, Communications of the ACM 46 (9) (2003) 89–95.
[76] (85_RP). T. Puay, Digital copyright and the "new" controversy: is the law molding technology and innovation? Research Policy 34 (6) (2005) 852–871.
[77] (112_JLE). R. Rob, J. Waldfogel, Piracy on the high c's: music downloading, sales displacement, and social welfare in a sample of college students, Journal of Law and Economics 49 (1) (2006) 29–62.

[78] (109_RERCI). F. Rochelandet, F.L. Guel, P2p music sharing networks: why the legal fight against copiers may be inefficient, Review of Economic Research on Copyright Issues 2 (2) (2005) 69–82.
[79] (114_ICC). F. Silva, G.B. Ramello, Sound recording market: the ambiguous case of copyright and piracy, Industrial and Corporate Change 9 (3) (2000) 415–442.
[80] (116_JCE). E.A. Strobl, C. Tucker, The dynamics of chart success in the U.K. Pre-recorded popular music industry, Journal of Cultural Economics 24 (2) (2000) 113–134.
[81] (105_JEB). S.L. Taylor, Music piracy-differences in the ethical perceptions of business majors and music business majors, The Journal of Education for Business 79 (5) (2004) 306–310.
[82] (79_IJEC). Y. Tu, M. Lu, An experimental and analytical study of on-line digital music sampling strategies, International Journal of Electronic Commerce 10 (3) (2006) 39–70.
[83] (122_TIJMM). V.L. Vaccaro, D.Y. Cohn, The evolution of business models and marketing strategies in the music industry, The International Journal on Media Management 6 (1) (2004) 46–58.
[84] (102_JLE). A. Zentner, Measuring the effect of file sharing on music purchases, Journal of Law & Economics 49 (1) (2006) 63–90.
[85] (91_CACM). K. Zhu, B. MacQuarrie, Economics of digital bundling: the impacts of digitization on the music industry, Communications of the Association for Computing Machinery (CACM) 46 (9) (2004) 264–270.

Software Research Network Papers

[86] (13_JBE). S. Al-Rafee, T.P. Cronan, Digital piracy: factors that influence attitude toward behavior, Journal of Business Ethics 63 (3) (2006) 237–259.
[87] (45_AEL). A.R. Andres, Software piracy and income inequality, Applied Economics Letters 13 (2) (2006) 101–105.
[88] (19_CACM). B. Bagchi, P. Kirs, R. Cerveny, Global software piracy: can economic factors alone explain the trend? Association for Computing Machinery, Communications of the ACM 49 (6) (2006) 70–76.
[89] (42_AE). D.S. Banerjee, A.M. Khalid, J.E. Sturm, Socio-economic development and software piracy: an empirical assessment, Applied Economics 37 (18) (2005) 2091–2097.
[90] (48_IJIO). D.S. Banerjee, Software piracy: a strategic analysis and policy instruments, International Journal of Industrial Organization 21 (1) (2003) 97–127.
[91] (59_JBE). W.H. Bryan, The impact of national culture on software piracy, Journal of Business Ethics 26 (3) (2000) 197–211.
[92] (30_MS). E. Brynjolfsson, C.F. Kemerer, Network externalities in microcomputer software: an econometric analysis of the spreadsheet market, Management Science 42 (12) (1996) 1627–1647.
[93] (16_JBE). J.V. Calluzzo, J.C. Cante, Ethics in information technology and software use, Journal of Business Ethics 51 (3) (2004) 301–312.
[94] (2_DSS). S. Chakravarty, K. Dogan, N. Tomlinson, A hedonic study of network effects in the market for word processing software, Decision Support Systems 41 (4) (2006) 747–763.
[95] (24_ISR). Y.N. Chen, I. Png, Information goods pricing and copyright enforcement: welfare analysis, Information Systems Research 14 (1) (2003) 107–123.
[96] (65_JMIS). H.K. Cheng, R.R. Sims, H. Teegen, To purchase or to pirate software: an empirical study, Journal of Management Information Systems 13 (4) (1997) 49–60.
[97] (14_JOCEC). V. Choudhary, K. Tomak, A. Chaturvedi, Economic benefits of renting software, Journal of Organizational Computing and Electronic Commerce 8 (4) (1998) 277.
[98] (29_TJIE). J. Church, N. Gandal, Network effects, software provision, and standardization, The Journal of Industrial Economics 40 (1) (1992) 85–103.
[99] (50_MS). K.R. Conner, R.P. Rumelt, Software piracy: an analysis of protection strategies, Management Science 37 (2) (1991) 125–140.
[100] (41_AEL). C.A. Depken, L.C. Simmons, Social construct and the propensity for software piracy, Applied Economics Letters 11 (2) (2004) 97–100.
[101] (61_RANDJE). G. Ellison, D. Fudenberg, The neo-luddite's lament: excessive upgrades in the software industry, The Rand Journal of Economics 31 (2) (2000) 253–272.
[102] (11_ITM). C. Faugère, G.K. Tayi, Management Information Technology, Designing free software samples: a game theoretic approach, Information Technology and Management 8 (4) (2007) 263–278.
[103] (68_MISQ). J.M. Gallaugher, Y.M. Wang, Understanding network effects in software markets: evidence from web server pricing, Management Information Systems Quarterly 26 (4) (2002) 303–327.
[104] (7_TRES). N. Gandal, Competing compatibility standards and network externalities in the pc software market, The Review of Economics and Statistics 77 (4) (1995) 599–608.
[105] (21_RANDJE). N. Gandal, N. Gandal, Hedonic price indexes for spreadsheets and an empirical test for network externalities, The Rand Journal of Economics 25 (1) (1994) 160–170.
[106] (4_TFSC). M. Givon, V. Mahajan, E. Muller, Assessing the relationship between the user-based market share and unit sales-based market share for pirated software brands in competitive markets, Technological Forecasting and Social Change 55 (2) (1997) 131–144.
[107] (51_JM). M. Givon, V. Mahajan, E. Muller, Software piracy: estimation of lost sales and the impact on software diffusion, Journal of Marketing 59 (1995) 29–37.
[108] (40_JBE). R.S. Glass, W.A. Wood, Situational determinants of software piracy: an equity theory perspective, Journal of Business Ethics 15 (11) (1996) 1189–1198.
[109] (70_JBE). S. Goode, S. Cruise, What motivates software crackers? Journal of Business Ethics 65 (2) (2006) 173–201.
[110] (20_CACM). R.D. Gopal, G.L. Sanders, Global software piracy: you can't get blood out of a turnip. Association for Computing Machinery, Communications of the ACM 43 (9) (2000) 82–89.
[111] (26_ISR). R.D. Gopal, G.L. Sanders, International software piracy: analysis of key issues and impacts, Information Systems Research 9 (4) (1998) 380–397.
[112] (35_JMIS). R.D. Gopal, G.L. Sanders, Preventive and deterrent controls for software piracy, Journal of Management Information Systems 13 (4) (1997) 29–47.
[113] (64_JBE). P.B. Gupta, S.J. Gould, B. Pola, To pirate or not to pirate: a comparative study of the ethical versus other influences on the consumer's software acquisition-mode decision, Journal of Business Ethics 55 (3) (2004) 255–274.
[114] (32_JORS). H. Gurnani, K. Karlapalem, Optimal pricing strategies for internet-based software dissemination, The Journal of the Operational Research Society 52 (1) (2001) 64–70.
[115] (63_IEP). J.H. Hahn, The welfare effect of quality degradation in the presence of network externalities, Information Economics and Policy 16 (4) (2004) 535–552.
[116] (31_JEE). E. Haruvy, A. Prasad, Optimal freeware quality in the presence of network externalities: an evolutionary game theoretical approach, Journal of Evolutionary Economics 11 (2) (2001) 231–248.
[117] (58_TJB). E. Haruvy, V. Mahajan, A. Prasad, The effect of piracy on the market penetration of subscription software, Journal of Business 77 (2) (2004) S81–S108.
[118] (67_EIT). S. Hinduja, Trends and patterns among online software pirates, Ethics and Information Technology 5 (1) (2003) 49–61.
[119] (38_MS). S. Jain, P.K. Kannan, Pricing of information products on online servers: issues, models, and analysis, Management Science 48 (9) (2002) 1123–1142.
[120] (39_JBE). R.B. Kini, H.V. Ramakrishna, B.S. Vijayaraman, Shaping of moral intensity regarding software piracy: a comparison between thailand and u.s. Students, Journal of Business Ethics 49 (1) (2004) 91–104.
[121] (6_IEP). H. Kinokuni, Compensation for copying and bargaining, Information Economics and Policy 17 (3) (2005) 349–364.
[122] (28_IEEESE). B. Kitchenham, L.M. Pickard, S. Linkman, P.W. Jones, Modeling software bidding risks, IEEE Transactions on Software Engineering 29 (6) (2003) 542–554.
[123] (46_IM). C.M. Koen, J.H. Im, Software piracy and its legal implications, Information Management 31 (5) (1997) 265–272.
[124] (12_JBE). F.Y. Kuo, M.H. Hsu, Development and validation of ethical computer self-efficacy measure: the case of softlifting, Journal of Business Ethics 32 (4) (2001) 299–315.
[125] (17_IEEEEM). M. Limayem, M. Khalifa, W.W. Chin, Factors motivating software piracy: a longitudinal study, IEEE Transactions on Engineering Management 51 (4) (2004) 414–425.
[126] (52_JBE). J.M. Logsdon, J.K. Thompson, R.A. Richard, Software piracy: is it related to level of moral judgment? Journal of Business Ethics 13 (11) (1994) 849–857.
[127] (3_IEP). S. Mark, An investigation into sources of network externalities in the packaged pc software market, Information Economics and Policy 5 (3) (1993) 231–251.
[128] (55_SLR). P.S. Menell, Tailoring legal protection for computer software, Stanford Law Review 39 (6) (1987) 1329–1372.
[129] (22_TJIE). R. Michaels, Hedonic prices and the structure of the digital computer industry, The Journal of Industrial Economics 27 (3) (1979) 263–275.
[130] (53_JOCEC). B.K. Mishra, T.S. Raghu, A. Prasad, Strategic analysis of corporate software piracy prevention and detection, Journal of Organizational Computing and Electronic Commerce 15 (3) (2005) 223–252.
[131] (15_MISQ). T.T. Moores, J.C.J. Chang, Ethical decision making in software piracy: initial development and test of a four-component model, MIS Quarterly 30 (1) (2006) 167–180.
[132] (62_IM). T.T. Moores, J. Dhaliwal, The reversed context analysis of software piracy issues in singapore, Information Management 41 (8) (2004) 1037–1042.
[133] (49_CACM). T. Moores, G. Dhillon, Software piracy: a view from hong kong. Association for Computing Machinery, Communications of the ACM 43 (12) (2000) 88–93.
[134] (71_JPPM). J.C. Nunes, C.K. Hsee, E.U. Weber, Why are people so prone to steal software? The effect of cost structure on consumer purchase and payment intentions, Journal of Public Policy & Marketing 23 (1) (2004) 43–53.
[135] (47_JMIS). A.G. Peace, Dennis D.F. Galletta, J.Y.L. Thong, Software piracy in the workplace: a model and empirical test, Journal of Management Information Systems 20 (1) (2003) 153–177.
[136] (10_AMAPSS). N.L. Piquero, A.R. Piquero, Democracy and intellectual property: examining trajectories of software piracy, The Annals of the American Academy of Political and Social Science 605 (1) (2006) 104–127.
[137] (23_JRM). A. Prasad, V. Mahajan, How many pirates should a software firm tolerate? An analysis of piracy protection on the diffusion of software, International Journal of Research in Marketing 20 (4) (2003) 337–353.
[138] (44_JMIS). S. Raghunathan, Software editions: an application of segmentation theory to the packaged software market, Journal of Management Information Systems 17 (1) (2000) 87–113.
[139] (18_CACM). S.K. Shin, R.D. Gopal, G.L. Sanders, A.B. Whinston, Global software piracy revisited. Association for Computing Machinery, Communications of the ACM 47 (1) (2004) 103–107.
[140] (54_EIT). R.M. Siegfried, Student attitudes on software piracy and related issues of computer ethics, Ethics and Information Technology 6 (4) (2004) 215–222.
[141] (43_JBE). P.M. Simpson, D. Banerjee, C.L. Simpson Jr., Softlifting: a model of motivating factors, Journal of Business Ethics 13 (6) (1994) 431–438.
[142] (66_JBE). R.R. Sims, H.K. Cheng, H. Teegen, Toward a profile of student software pirates, Journal of Business Ethics 15 (8) (1996) 839–849.

[143] (5_BIT). M.T. Siponen, T. Vartiainen, Attitudes to and factors affecting unauthorized copying of computer software in Finland, Behaviour & Information Technology 24 (4) (2005) 249–257.
[144] (34_TCJE). J. Slive, D. Bernhardt, Pirated for profit, The Canadian Journal of Economics 31 (4) (1998) 886–899.
[145] (27_MISQ). D.W. Straub Jr., R.W. Collins, Key information liability issues facing managers: software piracy, proprietary databases, and individual rights to piracy, MIS Quarterly 14 (2) (1990) 143–156.
[146] (60_JBE). W.R. Swinyard, H. Rinne, A.K. Kau, The morality of software piracy: a cross-cultural analysis, Journal of Business Ethics 9 (8) (1990) 655–664.
[147] (57_JBE). J.H. Tang, C.K. Farn, The effect of interpersonal influence on softlifting intention and behavior, Journal of Business Ethics 56 (2) (2005) 149–161.
[148] (33_MISQ). A. Taudes, M. Feurstein, A. Mild, Options analysis of software platform decisions: a case study, MIS Quarterly 24 (2) (2000) 227–243.
[149] (1_HR). G.S. Taylor, J.P. Shim, A comparative examination of attitudes toward software piracy among business professors and executives, Human Relations 46 (4) (1993) 419–433.
[150] (69_DSS). M.E. Thatcher, T. Kim, D.E. Pingry, Welfare analysis of alternative patent policies for software innovations, Decision Support Systems 41 (4) (2006) 803–823.
[151] (56_JMIS). J.Y.L. Thong, C.S. Yap, Testing an ethical decision-making theory: the case of softlifting, Journal of Management Information Systems 15 (1) (1998) 213–237.
[152] (9_JBE). S.C. Wagner, G.L. Sanders, Considerations in ethical decision-making and software piracy, Journal of Business Ethics 29 (1–2) (2001) 161–167.

Sudip Bhattacharjee is an Associate Professor and Ackerman Scholar in the School of Business, University of Connecticut. He also serves as theAssistant Dept. Head of Operations and Information Management. His research interests include information systems economics, intellectual property rights, digital goods and markets, operations management, supply chains and distributed computing systems. He has extensive research consulting experience with several Fortune 100 firms on data-driven decision making in IT and operations. His research has appeared in premier journals such as Management Science, INFORMS Journal on Computing, Journal of Business, Journal of Law and Economics, ACM Transactions, Journal of Management Information Systems, IEEE Transactions, and other leading peer-reviewed journals and conference proceedings. His research has been highlighted in various media outlets such as Business Week, Washington Post, San Francisco Chronicle, Der Spiegel, Christian Science Monitor, slashdot.org, Business 2.0 Web Guide, and others. He serves on the editorial boards of Information Systems Research, Management Information Systems Quarterly, Decision Sciences, Information Systems Frontiers, and various other conference and workshops.

Dr. James R. Marsden, is the Treibick Family Endowed Chair in e-Business and Board of Trustees Distinguished Professor, at the Department of Operations and Information Management (OPIM) at the University of Connecticut. He has been at UConn since 1993 as Professor, serving fifteen years (1993-2008) as Head of OPIM. He helped develop both the Connecticut Information Technology Institute and the Treibick Electronic commerce Initiative and currently serves as Executive Director of both. Jim also severs as the UConn Director of ***edgelab***, the unique ongoing research partnership between GE and UConn now in its ninth eighth year of operation (see www.edgelab.com). Dr. Marsden has a lengthy publication record in market innovation and analyses, economics of information, artificial intelligence, and production theory. His research work has appeared or is forthcoming in ***Management Science; Journal of Law and Economics; American Economic Review; Journal of Economic Theory; Journal of Political Economy; IEEE Transactions on Systems, Man, and Cybernetics; Computer Integrated Manufacturing Systems; Decision Support Systems; Journal of Management Information Systems***, and numerous other academic journals. He received his A.B. from the University of Illinois and his M.S. and Ph.D. from Purdue University. Also holding a J.D, Jim has been admitted to both the Kentucky and Connecticut Bar.

Harpreet Singh is an Assistant Professor of Information Systems at the School of Management at University of Texas at Dallas. He received his PhD in Business Administration (Information System) from University of Connecticut. He also holds a Bachelors degree in Chemical Engineering from Punjab University, Chandigarh, India. His research has been published in refereed conference proceedings and is currently under review in various journals such as Academy of Management Journal, Information Systems Research and Management Information Systems Quarterly (MISQ).