Journal of Arid Environments 73 (2009) 862–871

Contents lists available at ScienceDirect

Journal of Arid Environments

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

Research impact within the international arid literature: An Australian perspective based on network theory

R.R.J. McAllister [a,b,*], L.R. Izquierdo [c], M.A. Janssen [d], D.M. Stafford Smith [a,e]

[a] Desert Knowledge Cooperative Research Centre, Alice Springs, NT 0871, Australia
[b] CSIRO Sustainable Ecosystems, 306 Carmody Road, St Lucia, QLD 4067, Australia
[c] Universidad de Burgos, Edificio la Milanera C/Villadiego s/n,09001, Burgos, Spain
[d] School of Human Evolution and Social Change, Arizona State University, P.O. Box 872402, Tempe, AZ 85287-2402, USA
[e] CSIRO Sustainable Ecosystems, P.O. Box 284, Canberra, ACT 2601, Australia

ARTICLE INFO

Article history:
Received 22 September 2008
Received in revised form
14 January 2009
Accepted 26 March 2009
Available online 9 May 2009

Keywords:
Bibliometrics
Dryland
Graph theory
Rangeland
Semiarid

ABSTRACT

While research-article impact is routinely judged by citation counts, there is recognition that a much broader view is needed to better judge the true value of citations. This paper applies a developing framework based on the application of network theory, where the network consists of journal articles on arid-systems research which are listed on ISI Web-of-Science. Keywords were used to identify articles related to arid-systems research. Linkages between articles were defined by citations, and we bound our analysis by focusing on how the Australian subsample contributes to the international arid-systems literature. The analysis showed that impact based on how articles contribute structurally to the flow of knowledge within the literature offers an alternative metric to citation counts. The analysis also presented a partitioned view of the Australian arid literature. This showed that there exists some citation-based structure within the literature, and we showed this structure better describes the literature than a partition based on which journal articles are published in.

1. Introduction

New tools for doing science are being developed in response to an increasing focus on questions which are interdisciplinary and involve interactions across temporally longer and spatially broader scales. In this new world where the dominant scale at which research is conducted demonstrates complex behaviour, there are also challenges in demonstrating impact from our research. Bibliometrics, such as citation counts, have traditionally been used to measure the impact of science publications. However, citation counts ignore the relative importance of some citations in terms of their role in integrating and spreading science. Here network theory is used to view the arid literature from a fresh perspective. What constitutes impact is redefined around the role articles play in developing science, and the literature is partitioned with a view to contributing to a discussion on what type of impact could be most relevant regarding science publications, and how this is achieved.

Network theory is applied to 43 103 articles listed on ISI Web-of-Science, a major journal-only database. Here we concentrate on analysing the structure of citations. In the network constructed, nodes are individual articles, and there is a link from one article to another if the former cites the latter. This is used to explore the diffusion of knowledge and theory within a scientific community (Chen and Hicks, 2004; Crane, 1972; Janssen, 2007; Janssen et al., 2006), and allows the exploration of the structure of the literature. We focus on a subsample to explore how parts of the literature contribute to a broader research effort. Our subsample is representative of the Australian arid literature.

Even though we present empirical analysis, there are methodology and data limitations which prohibit the literal interpretation of article specific results. Our dataset is limited to what articles were available on ISI Web-of-Science on 2nd August 2007. Many important articles are missing from this dataset. While we do identify important articles, the aim is not to judge which articles are best, but rather demonstrate characteristics of good articles.

Further, we use a rule-based definition of what constitutes an Australian article. This provides a consistently defined dataset, but because of differing article writing styles, there are inconsistencies

* Corresponding author at: CSIRO Sustainable Ecosystems, 306 Carmody Road, St Lucia, QLD 4067, Australia. Tel.: +61 7 3214 2359.
E-mail addresses: ryan.mcallister@csiro.au (R.R.J. McAllister), lrizquierdo@ubu.es (L.R. Izquierdo), marco.janssen@asu.edu (M.A. Janssen), mark.staffordsmith@csiro.au (D.M. Stafford Smith).

0140-1963/$ – see front matter Crown Copyright © 2009 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.jaridenv.2009.03.014

in the degree to which some articles in this dataset focus on Australia. What the paper presents, therefore, is an analysis of a sample of articles which we use to infer the characteristics how a subset of arid literature fits within its broader landscape.

We focus on just two ideas. One, within the Australian arid literature, we use new methods to measure and discuss an alternative view of what impact means. Two, we search for a natural structure within the arid literature which is determined by the structure of citations. This facilitates a discussion on new approaches to analysing literature datasets with the aim of introducing new readers to network theory and its potential for its application in the analysis of literature. Ultimately this contributes to the broad aim of strategic thinking about where to target research.

2. Bibliometrics and network theory

Since the early 1900s, bibliometrics have been used to measure the impact of publications, scientists and journals (Godin, 2006). Bibliometrics are the measurement of text and information in documents and document related processes (Borgman and Furner, 2002). Originally, psychologists started counting publications systematically. Since the 1950s bibliometrics have become a broader scholarly activity to measure scientific output (Garfield, 1955; Price, 1965).

The cumulative body of publications led to the systematic study of citation processes (Price, 1965). Generally only a small portion of publications are cited frequently (Price, 1976; Redner, 1998), and most are not cited at all. Highly cited publications have a higher probability of future citations (Price, 1976), and citation rates typically drop with the age of the publication.

There are a number of problems with citation analysis (e.g. MacRoberts and MacRoberts, 1989), such as the bias of the datasets used (often dominated by the English language) and the culture of publication and citation within a discipline. Authors frequently cite their own work, and journal editors sometimes coerce citations of other articles in their journal. Further, citations may be copied, and not necessarily read by those making the citation (Simkin and Roychowdhury, 2003). As an example, 7% of citations in ecology were not supportive of the statements they were purporting to support (Todd et al., 2007).

Since the early days of bibliometrics, literature has been conceptualised as networks. Representations include: treating articles as nodes that are linked by citations (Garfield, 1963; Kessler, 1963); using co-citation graphs (Small, 1973); treating authors as nodes that are linked by co-authorship (Logan and Shaw, 1987; Newman, 2001a,b,c, 2004a,b); and using co-word graphs (Leydesdorff, 1989). Such networks have been used to visualize and analyse the dynamics of specialities, the structure within scientific communities and the diffusion of information in science. Such analyses have provided useful insights to identify frontiers of research, and quantify the impact of scientific output.

An advantage of network theory is that it can be used to algorithmically analyse the structure of large datasets. Nonetheless, the analysis of large datasets comes with its own practical limitations: the data must be available in a format that allows automatic processing and, naturally, the resulting analysis cannot be conducted with the same depth as when dealing with small datasets. It is not practical for example, to study every single node individually, and this somewhat limits the nature of the questions which can be practically asked.

In this paper, we conduct three different types of analyses. One, we provide an overview of the network by showing some basic statistics such as the top journals in terms of published articles. Two, we measure the structural importance of individual articles. Here we identify the type of articles that are most important, using new sophisticated measures of importance which go beyond citation counts alone. Finally, we explore the broad structure of the arid literature by partitioning articles into citation groups or communities. These communities contain articles that cite each other more than they cite articles in other groups.

3. The arid literature as a citation network

There are two main types of network that have been used to study scientific literature: citation networks of articles (Hopcroft et al., 2004; Janssen, 2007; Janssen et al., 2006; Rosvallt and Bergstrom, 2008) and co-authorship networks of scientists (Logan and Shaw, 1987; Newman, 2001a,b,c, 2004a,b). Recent studies have also considered the co-evolution of these two networks (Börner et al., 2004). Our primary interest lies in the flow of ideas and concepts and in the identification of knowledge sub-domains within the literature; thus, the analysis is mainly focused on the citation network of articles where nodes represent individual articles, and directed links denote citations between articles.

3.1. Data

Our citation network was constructed using arid literature contained in the ISI database (Web-of-Science, http://portal.isiknowledge.com/) on the 2nd August 2007.

To obtain ISI Web-of-Science data, we searched for the literature that contained, in the abstract, title or keywords, any of the following terms: rangeland*, arid*, semiarid*, desert*, hyperarid*, dryland* (where * refers to a wildcard). The trend is for an increasing number of articles to be ISI listed per year (Fig. 1). The articles were then exported, including their cited reference list, into EndNote (software, version X1). From Endnote, the data were exported to a text file using a reference list (with 43 103 articles, below) with a custom defined EndNote reference style. The reference list was designed to be read by the program Matlab (version R2008a) for text parsing. During parsing, a unique identifier was created for each article, which included first author, year of publication, abbreviated journal title, volume and starting page number, for example,

ABDEL-WAHAB A 1998 J AFR EARTH SCI V27 P277

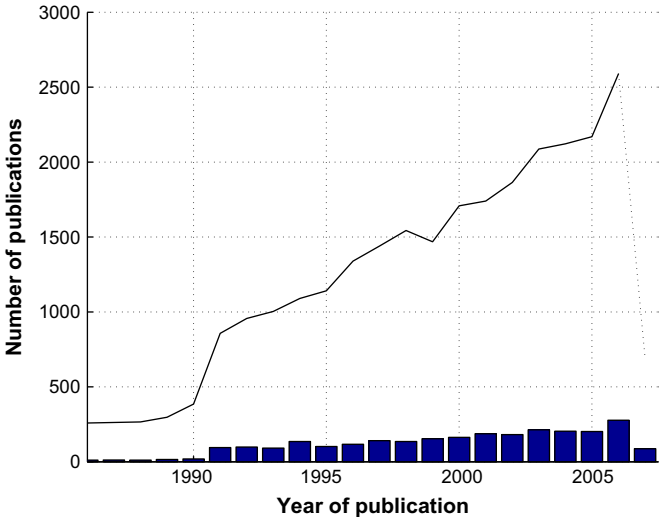

Fig. 1. Number and year of arid literature articles in the dataset; international (line) and Australian (bar). Note that 2007 does not include the full year's data.

Note that the unique identifier is the same as that format used by ISI to show citations for each piece of literature. The journal abbreviation style differed between the citing article and its reference list. We standardised all abbreviations to the ISI journal title abbreviation style before parsing.

After creating a unique key for each article, the dataset was re-parsed (also in Matlab). This time for each article parsed, its list of cited articles was also parsed, recording incidences where the article cited other articles within the scope of our study (i.e. international arid literature). At the same time quality checks were run. The unique key was broken into five parts (first author's last name, journal, year, volume, starting page number). If the unique keys were not equal, but three or more parts of the unique key matched, the citing literature and cited literature were printed to the screen in order to consider if they perhaps should match. This was a semi-manual process, so there are possibilities of error – however, no errors were found, indicating that where literature from within ISI data cites other data within ISI data, article data were accurate; though there are likely to be some errors which relate to how the data are stored on the ISI database which we do not account for. Hence analyses may be affected by some data errors, but we assert our broad conclusions hold.

The data were downloaded from ISI Web-of-Science 2nd August 2007, with the Web-of-Science data recorded as being updated on 29th July 2007.

- This yielded 43 103 articles.
- Duplicates were removed, leaving 43 007 articles (these occurred where articles were recorded as the original, and also again with corrected details).
- Articles with no author were removed, as were those which were not journal articles. Some 42 713 articles remained.
- Literature that related to Australia was tagged, defined as articles with either, *Australia*, Queensland, Northern Territory, New South Wales, in the abstract, title or keywords (i.e. excludes wetter states of Victoria and Tasmania, and note '*Australia*' accounts for Western Australia). This search was conducted in EndNote, searching only the data retrieved into the ISI search defined above. Some 3247 articles were defined as Australian. Different definitions of 'Australian' are feasible. Hence this subset is best interpreted as a representative sample from which general characteristics can be inferred.
- Articles which did not cite, nor were cited by, any other literature in the database were removed; leaving 30 233 articles.
- 796 components in the network were found. The largest component linked 27 313 articles, while the second largest linked only 59 articles. The largest component was assumed representative of the international arid literature.
- Hence, the dataset representative of the arid literature contained 27 313 interlinked articles (108 656 citations across this literature sample).
- Of this total of 27 313 articles, 2651 of these retained were tagged as Australian.

Our data are available for distribution (contact corresponding author).

3.2. Terminology

We define three key terms.

- **'Australian articles'** (or Australian arid articles) is used to mean the 2651 ISI Web-of-Science listed articles on arid systems tagged as Australian (see above). In this paper we only analyse these articles, but we do this in two different contexts.
- **'Domestic context'** is used to denote an analysis which, in analysing the Australian articles, only considers the citations to and from other Australian articles in the dataset.
- **'International context'** is used to denote an analysis which considers the citations to and from the full 27 313 articles in the dataset (see Section 3.1 above), including citations to and from non-Australian articles.

Unless otherwise stated, when we refer to data or analysis, we refer to articles from the full time frame of the data (i.e. $\leq$2nd August 2007).

4. Analysis

4.1. Basic overview of the arid literature

Within the 2651 articles in our dataset tagged as Australian, Tongway and Ludwig's (1990) article received the most citations from other Australian articles (Table 1). This was followed by Ludwig and Tongway (1995) and then Stafford Smith and Morton (1990). Some 748 of all Australian articles received no citations from other Australian articles (all un-cited articles did cite other articles within the database, otherwise they would have been excluded, see above). Within our timeframe, most Australian articles were published in the Journal of Arid Environments (5.3%, Table 2), followed by Wildlife Research and the Australian Journal of Experimental Agriculture. The journal which received the greatest share of citations was the Journal of Arid Environments (8.0%, Table 2).

4.2. Structural importance of individual articles: authorities and hubs

The structural importance of a node in a network (for example, how important a person is within a social network, or how important a road is within a traffic network) is usually measured using centrality metrics. Given the subjectivity of the term "importance", it is not surprising that there are various types of centralities in network theory (Freeman, 1978). All of them aim at quantifying the prominence of an individual node embedded in a network, but the criteria used to do this differ (Hanneman and Riddle, 2005).

The simplest type of centrality is based on the idea that having more links (i.e. citations) means being more important. This is the approach taken in the previous section, where the number of received citations (i.e. incoming links) was considered to be a useful measure of an article's importance.

A more sophisticated approach, proposed by Kleinberg (1999), acknowledges the fact that not all citations contribute to an article's importance in the same way. Citations received from important articles should weigh comparatively more than those received from less important ones. (This approach is used by many web search engines to rank pages.) This somewhat recursive definition is solved by simultaneously considering two interrelated types of centrality for each article: the authority score and the hub score.

The hub score relates to the citations made by the article (i.e. the node's outgoing links). A prominent hub is an article that cites many important authorities – e.g. a good critical review article that cites the most influential articles in the field. The authority score relates to the citations received by the article (i.e. the node's incoming links). An article has a high authority score if it is cited by the most important hubs. Thus, authoritative articles are likely to be research articles that were a major breakthrough in a specific field.

Table 1
Most cited Australian articles (see Section 3.2) (a) within the domestic context, (b) within the international context, and (c) the total citations as recorded on ISI Web-of-Science.

Top articles	Times cited		
	(a) Domestic context	(b) International context	(c) Total on ISI
TONGWAY DJ 1990 AUST J ECOL V15 P23	68	84	111
LUDWIG JA 1995 LANDSCAPE ECOL V10 P51	53	101	127
STAFFORD SMITH DM 1990 J ARID ENVIRON V18 P255	53	59	72
MABBUTT JA 1987 J ARID ENVIRON V12 P41	41	51	58
ELDRIDGE DJ 1994 AUST J SOIL RES V32 P389	38	85	116
MORTON SR 1988 AM NAT V132 P237	38	44	79
NANSON GC 1992 GEOLOGY V20 P791	37	43	108
KINGSFORD RT 2000 AUSTRAL ECOL V25 P109	31	33	91
DICKMAN CR 1995 J ARID ENVIRON V31 P441	30	30	44
GEORGE RJ 1999 AGR WATER MANAGE V39 P91	30	30	46
GREENE RSB 1992 AUST J SOIL RES V30 P55	30	34	49
NANSON GC 1995 PALAEOGEOGR PALAEOCL V113 P87	29	37	50
TONGWAY DJ 1989 AUST J ECOL V14 P263	29	33	44
ANDERSEN AN 1995 J BIOGEOGR V22 P15	27	31	103
DUNKERLEY DL 1995 J ARID ENVIRON V30 P41	27	36	46
KINGSFORD RT 1999 BIOL CONSERV V88 P231	26	26	42
KNIGHTON AD 1994 HYDROL PROCESS V8 P137	26	28	42
KINGSFORD RT 1995 ENVIRON MANAGE V19 P867	25	25	44
MONTAÑA C 1992 J ECOL V80 P315	25	54	61
MORTON SR 1995 J ENVIRON MANAGE V43 P195	25	25	40

Of particular interest would be an article that synthesises and builds on the most authoritative articles at the time of publishing (i.e. it is a foremost hub), and which later becomes a prominent authoritative article itself (i.e. it is cited by the leading hubs of the future).

Following Leicht et al. (2007), we calculated hub and authority scores for different time periods by altering the temporal range of the dataset. In this paper we used a rolling temporal range when calculating scores, starting with publication year $\leq$1990. Hub and authority scores were then repeatedly calculated adding one additional year's data – extending the time period year by year (i.e. next $\leq$1991, then $\leq$1992 etc.). In other words, we take the network as it existed in each year from 1990 to 2007 and compute the authority and hub scores for every article in each temporal range.

Articles were not strictly ranked. Instead, a simple rule was used to select examples of articles which can be regarded as either prominent hubs or authorities (or both). An Australian article was defined as a hub if its score was in the highest three hub scores of all Australian articles, for any given temporal range. An Australian article was defined as an authority if its authority score was in the highest three authority scores of all Australian articles for any temporal range. For both hub and authority scores, some articles appear in the top three in multiple temporal ranges: these articles are not distinguished from those which appear in a single temporal range.

Hub and authority scores were calculated using both the international and domestic contexts (see Section 3.2). The domestic context shows Australian articles which are important in the flow of ideas through the Australian literature (Tables 3 and 4). Graphically, authority and hub scores are related to the number of citations to and from articles, but this relationship is not perfect (Table 5). This points to the underlying differences between defining impact by citation counts compared to using concepts from network theory. Further, the inconsistencies between articles ranked highly using citation counts (Table 1) compared to hub and authority scores demonstrate the differences between impact metrics (Tables 3 and 4).

When calculating the scores using the international context (see Section 3.2), only Australian articles were considered when selecting the top three scores in a given temporal range. This showed the most important Australian articles in linking the international arid literature (Tables 3 and 4). Some articles were both internationally and domestically important (Tables 3 and 4). The relationship between an article's international and domestic scores is significant, but this relationship is not perfect (Table 5).

Intuitively, articles that are both authorities and hubs at some point in time (either domestically or internationally), have extended influence in the literature (Table 4). These 'authoritative-hubs', have an initial period with a high hub score, indicating they

Table 2
Top journals for Australian arid articles (domestic context, see Section 3.2) (a) percentage representation, (b) share of citations, and (c) 2007 impact factor.

Top journals	(a) % of Articles in domestic context	(b) % of Citations in domestic context	(c) 2007 ISI impact factor
Journal of Arid Environments	5.3%	8.0%	1.349
Wildlife Research	4.5%	5.1%	0.943
Australian Journal of Experimental Agriculture	3.5%	3.0%	0.948
Australian Journal of Agricultural Research	3.2%	3.6%	1.352
Australian Journal of Soil Research	3.1%	4.4%	1.310
Austral Ecology	3.1%	3.2%	1.674
The Rangeland Journal	2.8%	2.2%	0.545
Australian Journal of Botany	2.5%	1.7%	0.987
Palaeogeography, Palaeoclimatology, Palaeoecology	1.8%	3.3%	2.162
Australian Journal of Zoology	1.7%	1.6%	1.084

Table 3
Australian-article hubs in the international and domestic contexts.

	Article
Domestically important	ABENSPERGTRAUN M 1991 AUST J ECOL V16 P331
	ABENSPERGTRAUN M 1994 AUST J ECOL V19 P65
	ANDERSON VJ 1997 AUST J BOT V45 P331
	BERG SS 2004 J ARID ENVIRON V59 P313
	BUCKLEY R 1987 J ARID ENVIRON V13 P211
	COVENTRY RJ 1988 AUST J SOIL RES V26 P375
	ELDRIDGE DJ 1999 ACTA OECOL V20 P159
	ELDRIDGE DJ 1994 AUST J SOIL RES V32 P389
	ELDRIDGE DJ 1997 ARID SOIL RES REHAB V11 P113
	HENLE K 1989 OECOLOGIA V78 P521
	JAMES CD 1991 HERPETOLOGICA V47 P194
	JAMES CD 1991 J HERPETOL V25 P284
	LUDWIG JA 1996 RESTOR ECOL V4 P398
	LUDWIG JA 1999 CATENA V37 P257
	MILEWSKI AV 1994 J BIOGEOGR V21 P529
	RIDPATH MG 1987 EMU V87 P143
	TONGWAY DJ 1996 RESTOR ECOL V4 P388
	STAFFORD SMITH DM 1990 J ARID ENVIRON V18 P255
	VALENTIN C 1999 CATENA V37 P1
	WIENS JA 1991 ECOLOGY V72 P479
Internationally important	BARNET YM 1991 PLANT SOIL V135 P109
	CRAWFORD CS 1991 J ARID ENVIRON V21 P245
	DANSO SKA 1992 PLANT SOIL V141 P177
	GRAETZ RD 1987 REMOTE SENS ENVIRON V23 P313
	GRAETZ RD 1987 J ARID ENVIRON V12 P269
	HESHMATTI GA 1997 J ARID ENVIRON V35 P395
	HOLM AM 2002 J ARID ENVIRON V50 P1
	JONES TA 1998 J RANGE MANAGE V51 P594
	KELT DA 1996 ECOLOGY V77 P746
	MAUCHAMP A 1993 OIKOS V68 P107
	MURRAY BR 1994 OECOLOGIA V99 P216
	NOBLE JC 1989 ACTA OECOL-OEC GEN V10 P355
	PERRINGS C 1997 ECOL ECON V22 P73
	PREDAVEC M 1994 WILDLIFE RES V21 P569
	PUIGDEFABREGAS J 1998 LAND DEGRAD DEV V9 P393
	RANDALL JA 1994 AUST J ZOOL V42 P405
	REID KD 1999 SOIL SCI SOC AM J V63 P1869
	SHACHAK M 1998 ECOL APPL V8 P455
	STAFFORD SMITH DM 1990 J ARID ENVIRON V18 P255
	WATSON IW 1997 J ECOL V85 P815
	WHITFORD WG 1992 J ARID ENVIRON V22 P87

Table 4
Australian-article authorities in the international and domestic contexts.

	Article
Domestically and internationally important	BUCKLEY R 1987 J ARID ENVIRON V13 P211
	LUDWIG JA 1995 LANDSCAPE ECOL V10 P51
	MABBUTT JA 1987 J ARID ENVIRON V12 P41
	MONTAÑA C 1992 J ECOL V80 P315
	MORTON SR 1988 AM NAT V132 P237
	PECH RP 1986 INT J REMOTE SENS V7 P389
	TONGWAY DJ 1990 AUST J ECOL V15 P23
	TONGWAY DJ 1989 AUST J ECOL V14 P263
Only domestically important	JAMES CD 1991 OECOLOGIA V85 P553
	MORTON SR 1988 ECOL MONOGR V58 P19
	RIDPATH MG 1986 AUST WILDLIFE RES V13 P245
	RIDPATH MG 1986 IBIS V128 P177
	STAFFORD SMITH DM 1990 J ARID ENVIRON V18 P255
Only internationally important	AGUIAR MR 1999 TRENDS ECOL EVOL V14 P273
	FISHER DO 1993 J ARID ENVIRON V25 P397
	FRIEND GR 1989 AUST WILDLIFE RES V16 P1
	GRAETZ RD 1987 J ARID ENVIRON V12 P269
	HARRINGTON GN 1991 ECOLOGY V72 P1138
	JURADO E 1992 AUST J ECOL V17 P341
	KERLEY GIH 1991 J ARID ENVIRON V20 P63
	MCTAINSH GH 1989 J ARID ENVIRON V16 P11
	MORTON SR 1994 AUST J ZOOL V42 P501
	NOBLE JC 1989 ACTA OECOL-OEC GEN V10 P355
	PECH RP 1992 OECOLOGIA V89 P102
	PENRIDGE LK 1986 J ECOL V74 P925
	WALKER BH 1993 AMBIO V22 P80

build on existing authoritative articles, followed by a period where the articles themselves become authoritative (Fig. 2).

4.3. Communities within the arid literature network

In this section articles in the arid literature network are partitioned into separate groups or communities. Communities in a network are groups of nodes within which connections are dense, but between which connections are sparser (Newman, 2004a). Given our purposes we ignore the directed nature of the citation network and study the underlying undirected network instead. Fig. 3 shows an example of a network with three communities.

In recent years it has been observed that many real-world networks in various fields (e.g. sociology, biology, ecology, information, computer science) divide naturally into communities. It has also been shown that this feature (i.e. having a strong community structure) may be critical to understanding the system that the network represents (Newman, 2006a,b). Communities in a citation network may indicate different disciplines, different areas of research, different methodologies or different institutions. In any case, the presence of a strong community structure in a scientific citation network is likely to have important implications for the dynamic process of creating and disseminating knowledge. Once communities are identified, many insights can be gained by studying what intrinsic features articles within each community share and by identifying the key articles that bridge different communities. This type of analysis can reveal interesting insights about the underlying factors that may have caused the emergence of such communities.

Several methods have been developed to detect and quantify community structure in networks (for reviews see Danon et al., 2005, 2007; Newman, 2004a). A particularly intuitive method was proposed by Girvan and Newman (2002). The algorithm works by repeatedly identifying and removing links that are thought to lie between communities. These links are the ones with the highest (shortest path) betweenness centrality. The underlying idea is that if a link is joining two different communities, then there will be many shortest paths (those joining any node in one community with any node in the other community) that will run through it. As an example, consider the three communities sketched in Fig. 3. Any shortest path between two nodes belonging to different communities must run along the (only) link that joins those two communities. Thus, this method applied to the network in Fig. 3 would successfully identify the shaded communities after removing the three links with the highest betweenness centrality.

The main problem with the approach outlined above (henceforth GN) is that one does not know when to stop removing links. If run until the end, the algorithm will place each single node in a different community. To overcome this drawback (and also to provide a first-principles justification for the partition), Newman and Girvan (2004) proposed a new metric called modularity. Modularity is a measure of how good a community partition is. The larger the modularity, the better the partition is. Modularity is never greater than one, and equals zero if all the nodes are put in the same community. It can also be negative, suggesting that the network lacks community structure.

Nowadays maximising modularity seems to be widely accepted as the "definite current method of community detection"

Table 5
Correlation matrix. For Australian articles, percentage correlation between authority and hub scores, times cited and citations of others, for both international and domestic contexts (see Section 3.2). Only relationships significant with 95% confidence shown (non-significant results shown as n.s.). Bolding draws attention to key results.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Authority (domestic)	100.0	–	–	–	–	–	–	–
(2) Hub (domestic)	32.4	100.0	–	–	–	–	–	–
(3) Authority (international)	47.2	30.1	100.0	–	–	–	–	–
(4) Hub (international)	8.5	29.2	20.0	100.0	–	–	–	–
(5) Times cites (domestic)	**50.3**	28.9	49.4	13.7	100.0	–	–	–
(6) Citations to others (dom.)	n.s.	**20.3**	4.5	42.2	7.7	100.0	–	–
(7) Times cites (international)	43.3	29.2	**72.0**	23.0	79.0	4.3	100.0	–
(8) Citations to others (int.)	n.s.	16.4	9.7	**55.7**	n.s.	73.4	9.8	100.0

(Newman, 2006b, p. 8578), and much research in the field is devoted to developing algorithms that can be practically used to analyse large networks (see e.g. Clauset, 2005; Danon et al., 2007; Duch and Arenas, 2005; Newman, 2006a) and at different scales (Arenas et al., 2008).

Here we combine the intuitive approach proposed by Girvan and Newman (2002) with the maximisation of modularity by stopping the iterative removal of links when no further partitioning increases modularity. This method is henceforth referred to as GN-Mod. To check that the partition obtained with GN-Mod provides good modularity values, we also use two of the best modularity-maximising algorithms: Newman's (Newman, 2006a,b) eigenvector method (including refinement) and Clauset's (2005) fast algorithm based on local modularity.

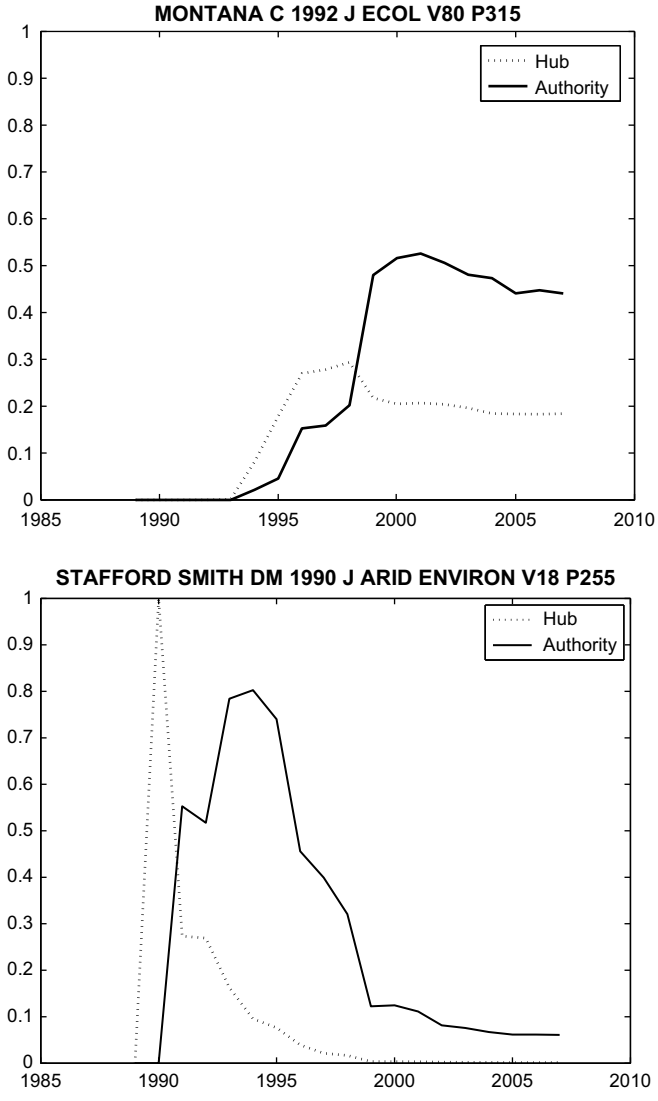

Fig. 2. Ideal profile of an article is one that shows an initial period with a high hub score, indicating it builds on existing authoritative articles, followed by a period where the article itself becomes authoritative (based on domestic context, see Section 3.2).

4.4. Evidence of structure in the arid literature landscape

Given the profile of an authoritative-hub, where in the structure of arid literature do these authoritative-hubs lie? The first requirement is to establish some type of structure in the literature. The Australia articles (domestic context) were partitioned using three alternative methods. Importantly, it was actually a subsample of the Australian literature, 2202 articles in its largest component, that was partitioned (out of a total of 2651, see Section 3.1). This was used because the partitioning methods require that the network be one component.

There is a perception that articles tend to cite the journals in which they are published (either because politics or genuine housing of research silos). For partitioning based purely on which of the 365 journals represented in the dataset articles were published in, modularity was 14% (Table 7). These results support the notion that while journals seem to represent research based partitions,

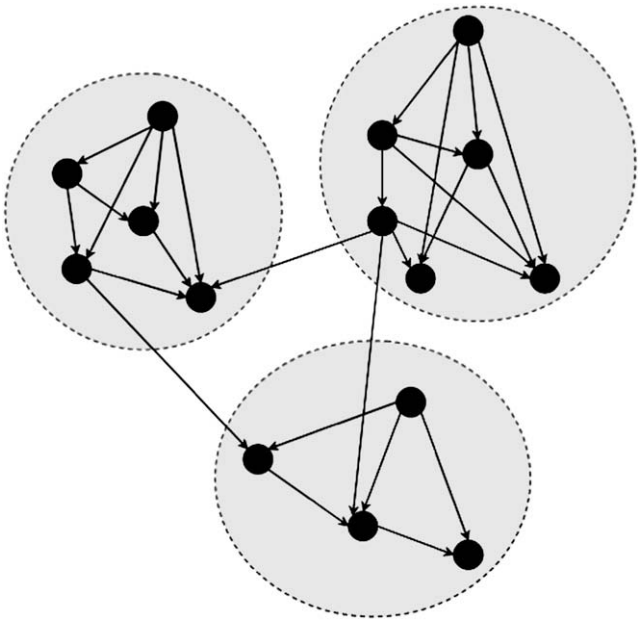

Fig. 3. An example network with community structure. The three communities are shaded.

Table 6
Australian articles that have been both an authority and hub (in either the international or domestic contexts).

	Article
Authoritative-hub	BUCKLEY R 1987 J ARID ENVIRON V13 P211 LUDWIG JA 1995 LANDSCAPE ECOL V10 P51 MABBUTT JA 1987 J ARID ENVIRON V12 P41 MONTAÑA C 1992 J ECOL V80 P315 (see Fig. 5, top) MORTON SR 1988 AM NAT V132 P237 PECH RP 1986 INT J REMOTE SENS V7 P389 STAFFORD SMITH DM 1990 J ARID ENVIRON V18 P255 (see Fig. 5, bottom) TONGWAY DJ 1990 AUST J ECOL V15 P23 TONGWAY DJ 1989 AUST J ECOL V14 P263

Table 7
Modularity for different Australian-article partitions.

Partitioning	Number of partitions	Modularity
Algorithm		
GN-Mod (Girvan and Newman, 2002)	25	76%
Newman's (Newman, 2006a,b) eigenvector method	34	76%
Clauset's (Clauset, 2005) fast algorithm based	38	74%
Index		
Journal title	365	14%

there are more important non-journal based partitions, as shown by the algorithm based partitioning (Table 7).

GN-Mod partitioning performs best and hence is used to present a structure of the Australian arid literature. The approach divided the 2202 strong literature into 25 partitions. For each of these partitions, the citations to and from the articles across partitions were summed. The summated citations represent the weights of the links (i.e. arcs) between partitions. For graphical purposes links with fewer than two cross-partition citations were removed. This isolated (disconnected) four partitions from the main group. Isolated partitions were removed leaving 21 partitions, representing 98% of the 2202 in the Australian literature (i.e. only small partitions were removed.) This structure is graphically depicted in Fig. 4 using a style based on Rosvallt and Bergstrom (2008). Partition names were assigned by the authors based on an interpretation of the dominant themes of the articles within each partition.

There are two features of the literature partitioning which guide a discussion on where in this landscape to target publications. First, most of the authoritative-hubs from Table 4 reside in the largest partition, identified as "PLANT ECOLOGY" literature (Table 8, Fig. 4). Second, within each partition, the average hub and authority scores were calculated (based on individual article scores at their temporal peak). While the partitioning was based only on the Australian literature (the largest component within this literature, i.e. 2202 articles), the hub and authority scores presented were calculated using both the Australian and international contexts. The results show that the "PLANT ECOLOGY" partition was the most authoritative, as well as being the best hub in the international context (Table 6). The "ANIMAL ECOLOGY" partition had the best average hub score based on the international context. There was wide variation across most the other partitions.

5. Discussion

The conjecture of this paper is that no idea is developed in isolation, but rather ideas co-develop through time and always embody preceding learning. This paper explores the arid literature as a case study of this process, based only on ISI listed journal articles (see Section 3.1). This case study highlights some important issues about how ideas are transferred through the literature. In particular, defining impact merely by the number of citations is too narrow. Rather article impact should be judged also on its role in the propagation of ideas throughout the literature landscape.

Network theory can be used to present this broader view of article impact. Before this is discussed further a few caveats of our approach are acknowledged. A dataset for a formal bibliometric analysis is derived by retrieving data in a systematic way from a standard dataset, which will make it possible to replicate the analysis. The ISI Web-of-Science database was used, one of the major databases on publications. This dataset includes most relevant ecological journals, but books, book chapters, and some journals, especially non-English language journals, are not included. Whatever choice is made in selecting a dataset, there will be incompleteness. Implicitly the assumption is that diffusion of information and knowledge can be captured by citations. There are many other forms in which knowledge diffuses, which are difficult to trace in a systematic manner.

To retrieve relevant publications a set of keywords were used to identify publications, based on discussion among the authors: rangeland*; arid*; semiarid*; desert*; hyperarid*; dryland*. Arid system literature that did not include any of these keywords was excluded from the study. And of course publications not on the ISI database on 2nd August 2007 were also excluded. Our study also focuses on Australian literature within this broader landscape, and Australian is defined as any article having a keyword of Australia*, Queensland, Northern Territory, or New South Wales. This definition is in some cases too inclusive and in other cases too exclusive. In summary, our dataset is a sample. And as such it is the general patterns which are useful rather than analysis of individual articles.

While our data represent a subset of the arid literature, they allow us to explore many useful general questions. It allows a researcher to think strategically about where to target future research, assuming the goal of publishing is to contribute to knowledge accumulation. In particular, from an Australian perspective, how can small isolated research communities make a contribution to the overall global knowledgebase?

If impact is defined by the structural role an article plays in the development of scientific knowledge, then how do you target impact? Our results present no surprise – the strategy for impact is to cite the 'best' articles and be cited by the 'best' articles. Digging deeper, the assertion embodied in our analysis is that 'best' is defined by the uniqueness of the linkages created by an article. If ten articles cite the same literature, and are cited by the same group of articles, then the metrics employed here discount the importance of the associated links. On the other hand, if only one article links the cited and citing groups of articles, then its role is greater. Further still, if that one article links literature from two otherwise unconnected groups of literature, then those links are particularly noteworthy.

Temporal dynamics are also critical. Building on the above example, if an article's citations create an important integration of knowledge, then it is probable that in following years countless other articles will make the same links. Critically the linkages at some point in time were important – and the fact that future contributions diminish the uniqueness of the linkages does not diminish the original article's role in developing a body of literature.

For an article to build unique links (at some point in time) it requires that it both cites broadly, and is cited broadly. The former strategy is directly controlled by the author(s), with the implication

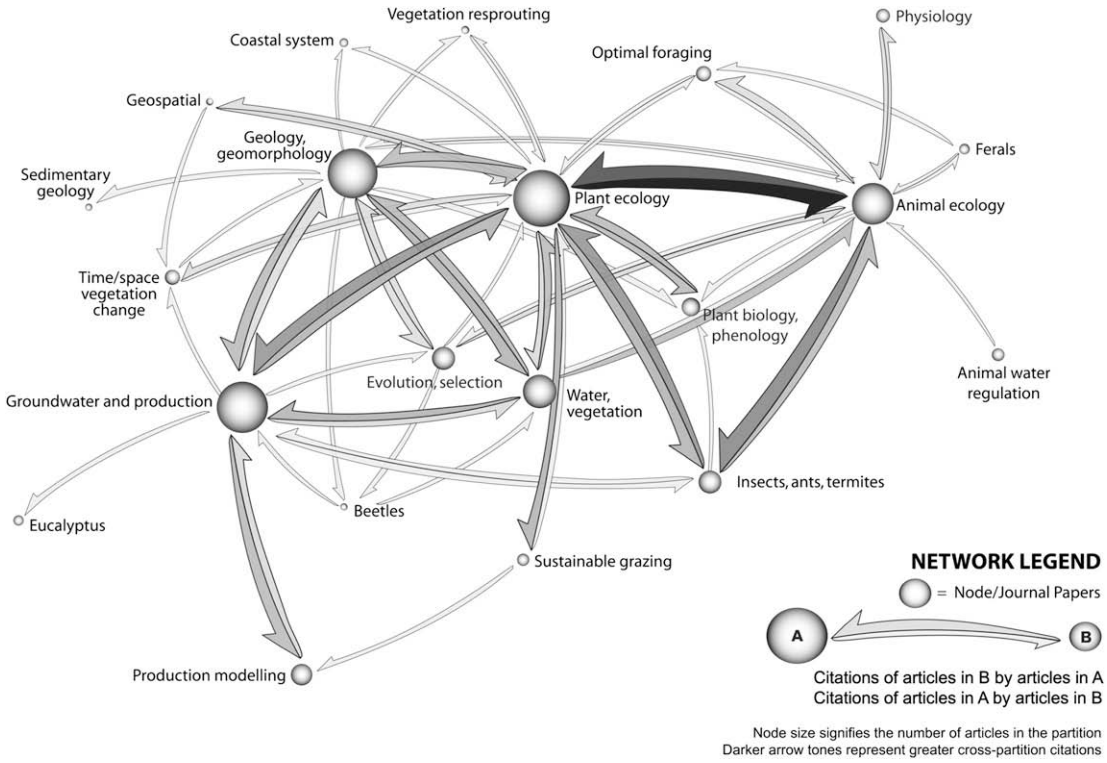

Fig. 4. Linkages between 21 partitions of the Australian arid literature (based on GN-Mod – see Table 7). Location of authoritative-hub articles (from Table 3): "Animal ecology" contains Buckley et al. (1987), Morton and James (1988), and Stafford Smith and Morton (1990); "Plant ecology" contains Ludwig and Tongway (1995), Mabbutt and Fanning (1987), Montaña (1992), Tongway and Ludwig (1990), and Tongway et al. (1989); and "Geospatial" contains Pech et al. (1986).

that an article should not be too confined within a discipline. It should build on a broad set of ideas from across the literature. Such advice could be perverted such that meaningless citations are included (as can occur with citation counts, a risk with citation-based journal impact factors).

The success of any strategy aimed at having an article cited broadly and by good articles is difficult to predict or plan. It may be hard to have an article cited widely across disciplines. And having an article cited by an article which itself has little impact (as defined here) does little to boost the importance of the article. To have an

Table 8
Average authority and hub scores within partitions, with each article's scores taken as its highest over all time periods.

Label (size)	Average maximum authority		Average maximum hub score	
	Domestic context	International context	Domestic context	International context
Animal ecology (249)	0.3680	0.0179	0.2359	0.9505
Groundwater and production (391): salinity and water use efficiency	0.0097	0.0077	0.0084	0.7958
Geology, geomorphology (364)	0.0398	0.0044	0.0116	0.9089
Insects, ants, termites (85)	0.0630	0.0051	0.0234	0.8271
Plant ecology (470)	0.3924	0.0487	0.2133	0.9778
Production modelling (61)	0.0011	0.0001	0.0002	0.6092
Water, vegetation (171)	0.0090	0.0018	0.0458	0.7802
Plant biology, phenology: (47)	0.0125	0.0056	0.0284	0.7774
Evolution, selection (80)	0.0001	0.0001	0.0014	0.3696
Coastal systems (12)	0	0	0	0.4947
Physiology (35)	0.0004	0.0001	0.0007	0.6347
Animal water regulation (22): frogs and lizards	0	0	0	0.2721
Eucalyptus (13): forest systems and genetics	0	0.0002	0.0001	0.7579
Sustainable grazing (35)	0.0056	0.0040	0.0073	0.8986
Beetles (9): incl. focus of genetics	0.0001	0.0001	0	0.2775
Optimal foraging (31): incl. animal diet selection, movement	0.0086	0.0011	0.0500	0.7614
Sedimentary geology (9)	0	0	0	0.5261
Vegetation resprouting (9)	0.0003	0.0004	0.0027	0.4834
Ferals (31): mainly pigs	0	0.0001	0.0001	0.4655
Time/space vegetation change (31)	0.0038	0.0197	0.0013	0.8406
Geospatial (11): remote sensing, but has some GIS	0.0063	0.0058	0.0029	0.85821

article cited widely (not just often) requires an article with strong inputs from other literature and a strong message. And the arguments developed here are that the ultimate *ex post* judgement of an article follows an analysis of the direct and indirect permutations of ideas through a complex structure of citations.

With the above in mind, two strategies may deliver real impact from academic research in different ways. In general, effort must be made to appropriately link an article with the quality and broader findings which have preceded it. If a multidisciplinary article is written, is must equally have broad applicability to future studies. If an article is written firmly to appeal within a discipline, it will need to be strong enough such that its broader impact is boosted by other broader articles to follow (either directly or indirectly linked).

Within the Australian literature, good strategising could likewise focus on either broad appeal or being a very strong article within a discipline with the hope that its links ultimately integrate research more broadly. Linking internationally requires the same broad principles at a different scale. A good arid Australian article must be either very strong within the Australian context (parallel to being strong within a discipline), in the hope that other literature provides the international links. Or alternatively, an article may seek to mix Australian and international contexts to broaden its links to the international community (parallel to being multidisciplinary).

An article's profile is one aspect of strategic publishing. The other relates to where to publish. In network terminology, the question is in which partition of the literature should one 'place' their research. There are methods which use the structure of a network to identify logical partitioning, but journal title is another way to partition the literature. Refreshingly, the partitions within the Australian arid literature based on the structure of the citations provide better partitioning than merely using the journal title. This means two things. One, partitioning based on domains of interest better reflects literature patterns than do journals. Two, where articles are published structurally is probably more important than in which journal articles are published. Where you publish structurally is important because some partitions within the literature seem to be associated with articles which are more authoritative and better hubs.

In moving towards new methods for analysis research impact, there are related questions which could be explored in time. The Australian literature was used as our case study, but there may be other subsets for which proving impact is even more complex. For example, what is the role of theory within the broader literature landscape? If the network is broken into finer-scale partitions, what binds the partitions together – are there partitions that seem to be dominated by issues, place or methods? And if there is a partition for 'theory', how does this (these) link beyond the realm of theory?

Acknowledgements

Much of the work reported here was supported by funding from the Australian Government Cooperative Research Centres Programme through the Desert Knowledge CRC; the views expressed herein do not necessarily represent the views of Desert Knowledge CRC or its participants. This is Publication No. 8 in the development of a Science of Desert Living. We wish to acknowledge comments from an anonymous reviewer, Kerry Collins, Oswald Marinoni, Anthony Richardson, and Amy Thams, and Jo Boniface for producing Fig. 4.

References

Arenas, A., Fernández, A., Gómez, S., 2008. Analysis of the structure of complex networks at different resolution levels. New Journal of Physics 10, 053039.

Borgman, C.L., Furner, J., 2002. Scholarly communication and bibliometrics. Annual Review of Information Science and Technology 36, 3–72.

Börner, K., Maru, J.T., Goldstone, R.L., 2004. The simultaneous evolution of author and paper networks. Proceedings of the National Academy of Sciences of the United States of America 101, 5266–5273.

Buckley, R., Wasson, R., Gubb, A., 1987. Phosphorus and potassium status of arid dunefield soils in central Australia and southern-Africa, and biogeographic implications. Journal of Arid Environments 13, 211–216.

Chen, C.M., Hicks, D., 2004. Tracing knowledge diffusion. Scientometrics 59, 199–211.

Clauset, A., 2005. Finding local community structure in networks. Physical Review E 72, 026132.

Crane, D., 1972. Invisible Colleges: Diffusion of Knowledge in Scientific Communities. University of Chicago Press, Chicago.

Danon, L., Diaz-Guilera, A., Duch, J., Arenas, A., 2005. Comparing community structure identification. Journal of Statistical Mechanics: Theory and Experiment.

Danon, L., Duch, J., Arenas, A., Diaz-Guilera, A., 2007. Community structure identification. In: Caldarelli, G., Vespignani, A. (Eds.), Large Scale Structure and Dynamics of Complex Networks From Information Technology to Finance and Natural Science. World Scientific, pp. 93–113.

Duch, J., Arenas, A., 2005. Community detection in complex networks using extremal optimization. Physical Review E 72, 027104.

Freeman, L.C., 1978. Centrality in social networks: conceptual clarification. Social Networks 79, 215–239.

Garfield, E., 1955. Citation indexes for science: a new dimension in documentation through association of ideas. Science 122, 108–111.

Garfield, E., 1963. Citation indexes in sociological and historical research. American Documentation 14, 289–291.

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America 99.

Godin, B., 2006. On the origins of bibliometrics 1. Scientometrics 68, 109–133.

Hanneman, R.A., Riddle, M., 2005. Introduction to Social Network Methods. University of California, Riverside, CA. Published in digital form at: http://www.faculty.ucr.edu/~hanneman/nettext/.

Hopcroft, J., Khan, O., Kulis, B., Selman, B., 2004. Tracking evolving communities in large linked networks. Proceedings of the National Academy of Sciences of the United States of America 101, 5249–5253.

Janssen, M.A., 2007. An update on the scholarly networks on resilience, vulnerability, and adaptation within the human dimensions of global environmental change. Ecology and Society 12.

Janssen, M.A., Schoon, M.L., Ke, W.M., Borner, K., 2006. Scholarly networks on resilience, vulnerability and adaptation within the human dimensions of global environmental change. Global Environmental Change: Human and Policy Dimensions 16, 240–252.

Kessler, M.M., 1963. Bibliographic coupling between scientific papers. American Documentation 14, 10–25.

Kleinberg, J.M., 1999. Authoritative sources in a hyperlinked environment. Journal of the ACM 46, 604–632.

Leicht, E.A., Clarkson, G., Shedden, K., Newman, M.E.J., 2007. Large-scale structure of time evolving citation networks. The European Physical Journal B 59, 75–83.

Leydesdorff, L., 1989. Words and co-words as indicators of intellectual organization. Research Policy 18, 209–223.

Logan, E.L., Shaw, W.M., 1987. An investigation of the co-author graph. Journal of the American Society for Information Science 38, 262–268.

Ludwig, J.A., Tongway, D.J., 1995. Spatial organisation of landscapes and its function in semi-arid woodlands, Australia. Landscape Ecology 10, 51–63.

Mabbutt, J.A., Fanning, P.C., 1987. Vegetation banding in arid western-Australia. Journal of Arid Environments 12, 41–59.

MacRoberts, M.H., MacRoberts, B.R., 1989. Problems of citation analysis – a critical review. Journal of the American Society for Information Science 40, 342–349.

Montaña, C., 1992. The colonization of bare areas in 2-phase mosaics of an arid ecosystem. Journal of Ecology 80, 315–327.

Morton, S.R., James, C.D., 1988. The diversity and abundance of lizards in arid Australia: a new hypothesis. American Naturalist 132, 237–256.

Newman, M.E.J., 2001a. Scientific collaboration networks. I. Network construction and fundamental results. Physical Review E 6401, 016131.

Newman, M.E.J., 2001b. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E 6401, 016132.

Newman, M.E.J., 2001c. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences of the United States of America 98, 404–409.

Newman, M.E.J., 2004a. Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences of the United States of America 101, 5200–5205.

Newman, M.E.J., 2004b. Who is the best connected scientist? A study of scientific coauthorship networks. In: Ben-Naim, E., Frauenfelder, H., Toroczkai, Z. (Eds.), Complex Networks, pp. 337–370.

Newman, M.E.J., 2006a. Finding community structure in networks using the eigenvectors of matrices. Physical Review E 74.

Newman, M.E.J., 2006b. Modularity and community structure in networks. Proceedings of the National Academy of Sciences of the United States of America 103, 8577–8582.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical Review E 69, 026113.
Pech, R.P., Graetz, R.D., Davis, A.W., 1986. Reflectance modeling and the derivation of vegetation indexes for an Australian semiarid shurbland. International Journal of Remote Sensing 7, 389–403.
Price, D.J.D., 1965. Networks of scientific papers. Science 149, 510.
Price, D.J.D., 1976. General theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science 27, 292–306.
Redner, S., 1998. How popular is your paper? An empirical study of the citation distribution. European Physics Journal B 4, 131–134.
Rosvallt, M., Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences of the United States of America 105, 1118–1123.
Simkin, M.V., Roychowdhury, V.P., 2003. Read before you cite!. Complex Systems 14, 269–274.
Small, H., 1973. Co-citation in the scientific literature: a new measure of the relationship between two documents. Journal of the American Society for Information Science 24, 265–269.
Stafford Smith, D.M., Morton, S.R., 1990. A framework for the ecology of arid Australia. Journal of Arid Environments 18, 255–278.
Todd, P.A., Yeo, D.C.J., Li, D.Q., Ladle, R.J., 2007. Citing practices in ecology: can we believe our own words? Oikos 116, 1599–1601.
Tongway, D.J., Ludwig, J.A., 1990. Vegetation and soil patterning in semiarid mulga lands of eastern Australia. Australian Journal of Ecology 15, 23–34.
Tongway, D.J., Ludwig, J.A., Whitford, W.G., 1989. Mulga log mounds – fertile patched in the semi-arid woodlands of eastern Australia. Australian Journal of Ecology 14, 263–268.