

Assessing Research Network and Disciplinary Engagement Changes Induced by an NSF Program

Jon Garner, Alan L. Porter, Nils C. Newman, and Todd A. Crowl^{1 2}

Abstract

To assess the success of the National Science Foundation's Research Coordination Network (RCN) program, a set of publication measures and visualization tools were used to determine how effective the program is in enhancing interdisciplinary publication and information sharing. The publication patterns of a set of researchers were compared before and after receiving RCN awards. These analyses show significant increases in basic collaboration measures – authors per paper and institutions per paper – following RCN support. Various indications suggest increased linkage among the RCN researchers in terms of extent of co-authoring and of cross-citing each other's work. RCN support appears to foster more interdisciplinary research. Diffusion scores (a new metric) showed that the diversity of articles citing RCN-related articles was similar to those citing control group research. Finally, the articles generated by the project activities showed as especially influential; they appear in high impact journals and are more highly cited.

Introduction

The National Science Foundation (“NSF”) instigated the Research Coordination Network (“RCN”) Program in 2000. This Biology Directorate Program's central goal is to:

“...foster communication and promote collaboration among scientists with common interests from a broad background across disciplinary, geographical, and organizational boundaries.”

RCN differs from most NSF core research programs in that funding does not support research directly, but rather supports networking of scientists who share a common research theme but who are not already collaborating. It seeks to catalyze development of research areas that grow at the shared margins of more conventional fields of study. The program announcement called for researchers to propose networking activities. That networking should focus on a broad research question or particular approach [www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf11001].

Bibliometrics have been used in various forms to evaluate research support, usually focusing on research outputs (especially publications) and one form of outcome (citations) (c.f., van Raan, 1988; 2000; De Bellis, 2009). De Bellis explores various issues, including: dealing with skewed data distributions, use of journal impact factors in lieu of direct citation comparisons to sidestep issues in comparing unequal citation time periods; different indicators' suitability for different units of analysis; dealing with

¹ Contact author: Alan Porter is Director, R&D, Search Technology, Inc. He is also co-director, Technology Policy and Assessment Center, Georgia Tech, and Professor Emeritus, Industrial & Systems Engineering, and Public Policy – aporter@searchtech.com.

Jon Garner [jon.garner@searchtech.com] and Nils Newman [newman@searchtech.com] are with Search Technology, Inc.

Todd Crowl is Professor in the College of Natural Resources, Utah State University -- todd.crowl@usu.edu.

² This research was supported by NSF award: Interdisciplinary Networking Impact of the Research Coordination Network (RCN) program (Award No. 0939622), with the help and guidance of Richard Inouye and Todd Crowl as Program Managers.

We expressly thank the reviewers who went into great detail to help improve this paper. They pointed us toward clarifying research questions, specifying which data explicitly pertain to each research question, checking data normality, and conducting appropriate statistical testing.

systematic and random noise in database records; cross-field comparison issues; and gauging research performance.

Martin and Irvine (1983) made a strong case for use of multiple bibliometric (and other, especially peer review, with its own issues) indicators in assessing research units. No absolute appraisal is warranted, but rather relative or comparative ones. As De Bellis (2009) points out, no bibliometric measure is conclusive; all face multiple sources of error. For instance, publications are not of equal importance and individual contributions to a publication are not equivalent; furthermore, publications are not the only important research output. Likewise, citations are variously motivated, so that crediting them as markers of influence is certainly fallible. Cross-field differences can be large in publication and citation norms.

With availability of suitable data and software for text analyses, one can investigate additional facets beyond publication activity per se. Balakrishnan et al. (2011) used bibliometrics to assess program awards in terms of the degree of interdisciplinarity, compared to that of the principal investigators' prior research. Rogers (2011) used bibliometrics in combination with interviews to situate a major nanotechnology research centers program in the field and assess its impact. Porter et al. (2003) examined authorship of prior, directly derived, and later publications by EPA-funded researchers to investigate whether teaming continued. Researchers collaborating on a project were the key unit of analysis – i.e., tracking their publications over time to provide Before vs. After comparisons.¹ They also compared topical emphases of publications to ascertain if the funding reflected in publications relating to program themes (not so much). They then benchmarked citation intensity to EPA STAR (Science to Achieve Results) supported papers to papers published at large in the same fields. Porter et al. (2010) profiled research proposal references, in addition to published papers, to gain perspective on the research communities engaged in ROLE (Research on Learning and Education, Education Directorate) projects. They used Integration scores and science overlay maps to locate the resulting research activity among the disciplines (see next section).²

Garner and Porter (2011) reported on assessment of the NSF Human and Social Dynamics (HSD) program, keying on cross-disciplinarity. They analyzed 2004 awards (22 full scale, plus 6 smaller awards treated separately) using two comparison groups. Sixteen similar size awards by the same NSF Directorate—Social, Behavioral and Economic Sciences (“SBE”) – proved to be non-equivalent in that they emphasized a particular area (cognitive-neurosciences) with high equipment costs. So, a second comparison group was generated of 20 SBE awards with 3 or more Principal Investigators, but these tended to be considerably lesser funding levels. As is the case here, they grappled with skewed citation distributions (a few papers are heavily cited) – they used auxiliary comparisons such as “papers cited 10 or more times per year since publication.” Their design is simpler in the sense that comparisons focused on publications deriving from NSF support, for which, one-tail t tests were run between the HSD papers and each of the comparison groups. They did examine Before vs. After changes in HSD co-authoring and cross-citing patterns (but not for the comparison groups).

The main goal of the present analysis is to gauge the efficacy of RCN support in building *interdisciplinary* teams and publications. Shifts in *research networking* among the researchers involved are a keen interest. The NSF program managers also expressed interest in locating RCN researchers' work among the disciplines, measuring how interdisciplinary their articles are, and exploring diffusion of their results across disciplines.

One could approach this research assessment through peer review mechanisms, bibliometrics, or a combination. This study relies upon bibliometrics – counting and analyzing research publication and citation activity – to present evidence about the effects of this research support. Limitations of such an approach are notable and are considered in later sections.

This paper shares selected results of this research assessment. The four key research questions investigated here are:

1. *Does RCN support lead to increased collaboration?*
We compare the numbers of co-authors, collaborating institutions, collaborating countries, and cross-citation intensity for the RCN researchers, “Before vs. After” the period of support, and compare to similar data for a Comparison Group (“CG”) of researchers.
2. *Does RCN support lead to more interdisciplinary research outputs?*
To get at interdisciplinarity, an “Integration score” is calculated – this reflects the intellectual diversity of the sources referenced in each paper. Scores are again compared Before vs. After for RCN researchers, and for the CG projects as well. We also benchmark against external samples of papers published in the same Subject Categories.
3. *How influential is RCN-supported research?*
Citation rates and journal impact factors for “**RCN-derived**” papers (i.e., those expressly associated with project support) are analyzed; and compared with Before & After publications, and to CG Before and After publications.
4. *How broadly does this research knowledge diffuse?*
We examine the breadth of Subject Categories that cite these research publications.

Analyses to address Questions 1 and 2 represent hypothesis testing. Questions 3 and 4 were exploratory (no explicit hypotheses of differences). Selected descriptive measures (e.g., of disciplinary participation) are also provided. Evidence to address these research questions was sought by the RCN program managers to help gauge how well the program is achieving its aims. Results could also inform the Foundation’s deliberations about possibly extending such network building support to other fields.

Data and Methods

Study Design

This study’s research design especially seeks to identify changes in the outputs (publications) and outcomes (citations) of groups of NSF-funded researchers attributable to RCN funding. It includes a Comparison Group (“CG”), that applied for, but did not receive, RCN funding, to check whether observed changes are likely attributable to underlying environmental or maturational changes, or not. NSF program officers, in consultation with us, determined that a set of RCN proposals from the same time period that rated well, but were not funded, was most attractive. By proposing, these researchers indicate comparable interests in instigating new research networks, within the scope of this Biology Directorate program announcement. Informal assessment by the program officers indicates that they involve highly successful researchers in terms of receipt of other NSF funding, before and since. However, the number of suitable, declined RCN proposals was more limited than the set of RCN awards under study. And, obviously, these researchers’ initiatives were judged less worthy; there are no data on the extent to which they pursued such network building. [Also, strong confidentiality must be maintained on which researchers are included.] All these factors mean that comparisons are suggestive, not definitive. Schematically, the design can be represented as per Figure 1.

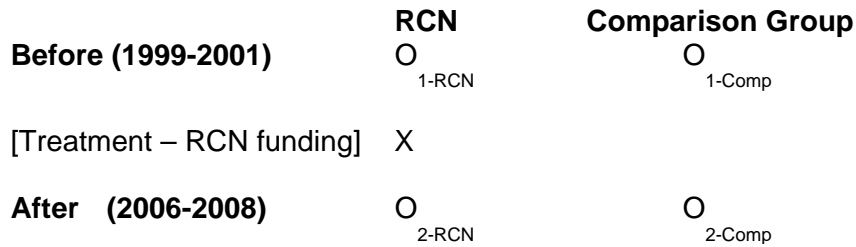


Figure 1. Experimental Research Design

[Where “O’s” represent observed outputs, such as research publications.]

This represents a nonequivalent control group (“NECG”) design – the archetypal quasi-experimental design (Campbell and Stanley, 1963; Cook and Campbell, 1979). Such designs can offer appealing internal validity, but pose a variety of issues in interpretation beyond those of a truly randomized experiment. In the random experimental design, most sources of bias are well-counteracted by the random assignment of subjects to experimental or control groups, leaving analysts to emphasize statistical testing to rule out noise effects in interpreting observed differences. For the NECG, valid interpretation depends on considering and ruling “down” (if not out) alternative biasing factors – as well as addressing concerns about signal vs. noise (via statistical testing). Judd and Kenny (1981) pose construct validity considerations for the dependent variable(s), and consideration of whether variables are consistently measured “Before and After,” and for the treatment and control groups. They note many NECG analytical options, including regression adjustment, change score analyses, transformations to promote stationarity (standardized change score analyses), and many complications.

Achen (1986) makes a strong case that statistically controlling variables in NECGs does not justify analysis as if the control (comparison) group were truly random. He explores statistical techniques for coping with non-equivalencies, such as two-stage least squares, but finds no complete remedies. Achen considers “censored samples” – nonrandom draws from the population of interest. This fits the HSD assessment discussed in the Introduction and also the present case for RCN. He finds that controlling for variables that contribute to important differences (e.g., if comparison group researchers are less prolific) is not generally sufficient.

This paper presents the raw data for comparisons of interest, augmented with some statistical testing. Many key comparisons are at the level of the research groups (i.e., colleagues involved in a given proposal or project). The numbers of such groups involved are small (13 RCN groups and 5 CG projects for many of the analyses) – heightening the possibilities of random noise effects distorting the comparisons. Statistical test results can be affected by these low and unequal “N’s.” We also examine normality of the data involved in particular comparisons. The reader should consider alternative interpretations. *Even were statistical tests between the RCN and CG to be significant, the non-equivalence suggests caution in interpreting differences.*

For research question #1 (Does RCN support lead to increased collaboration?), we include the comparison group because scientific research is increasing over time in the extent of research collaboration (Porter and Rafols, 2009). As emphasized, this gives a non-equivalent, control group design. Note that research questions 3 & 4 largely concern attributes of *the publications directly associated with the treatment* (“RCN-derived”). The comparison group does not pertain directly, so the pertinent analytical design reduces essentially to a case study. However, where informative, comparison outputs are provided as rough benchmarks in addressing these questions too. Because the comparisons pertinent to the four research questions differ, no single data format, comparison, or statistical test fits all.

Measuring Disciplines and Interdisciplinarity

The “Integration score” offers a measure of interdisciplinarity, developed in support of continuing evaluation of the National Academies Keck Futures Initiative [see www.keckfutures.org] (Porter et al., 2007; 2008). The Integration score keys on the notion of *intellectual integration* as definitive of interdisciplinarity (National Academies Committee on Facilitating Interdisciplinary Research, 2005, p. 188). This can be distinguished from other, complementary perspectives that emphasize *social* (collaborative) aspects of interdisciplinarity (c.f., the “Team Science” initiative – Stokols et al., 2008; Falk-Krzesinski et al., 2010).

Integration score is based on the diversity of the Web of Science (“WOS”) Subject Categories (“SCs”) cited by a given article or body of research (e.g., the works by one author during a given time period). It nicely fits the 3-dimensional conceptualization of diversity by Stirling (2007) – that distinguishes variety, balance, and disparity. Integration score is calculated as follows:

$$I = 1 - \sum_{i,j} p_i p_j s_{ij}$$

where p_i and p_j are the proportion of cited references corresponding to SC_i and SC_j in a given paper. The summation is taken over the cells of the paper’s cited SC x cited SC matrix. s_{ij} is the cosine measure of similarity between SCs i and j for cross-citation in the full WOS for Year 2007.

This corresponds to $1 -$ Stirling’s diversity measure (Rafols and Meyer, 2010). “Specialization score” (pertaining to the diversity of the publications themselves) is similar, with p_i and p_j reflecting the publication journals’ SCs; it is meaningful for sets of papers, not for individual ones. “Diffusion score” (for the diversity of SCs citing a given paper or set of papers) is analogous, with p_i and p_j pertaining to the citing papers’ SCs (Carley and Porter, 2011). Conceptually, it is a measure of the breadth of influence of a given research output.

Each of these measures ranges from 0 (e.g., for Integration, all of the references cited by a given research output fall in a single SC) to 1 (maximal diversity of those cited references). Note that, due to the incorporation of the cosine similarity measure, diversity of SCs takes into account not just how many different SCs are engaged, but how distant they are from each other (based on cross-citation propensities).

Loet Leydesdorff transforms the WOS journal-to-journal cross-citation matrix to obtain the 2007 WOS SC x SC matrix [www.leydesdorff.net].³ For instance, for five benchmark SC sample sets (2005 publications), the Integration scores average 0.43 using this full-year similarity matrix. [Our sixth benchmark SC, Mathematics, stands apart, behaving more like one’s image of a discipline unto itself. For 2005, its average Integration score is 0.20.]

Science overlay maps have been described in detail (Leydesdorff and Rafols, 2009; Rafols et al., 2010; Rafols and Meyer, 2010; Porter and Rafols, 2009). This methodology is also evolving. Porter et al. (2010) presented such an overlay to aid research evaluation by locating articles deriving from NSF Research on Learning and Education (ROLE) and on Research and Evaluation on Education in Science and Engineering (REESE). The base science map used therein had 244 SCs, including Arts and Humanities Citation Index (AHCI) coverage. The behavior of AHCI publications and citations is sufficiently distinct from that of Science Citation Index (SCI) and Social Science Citation Index (SSCI) ones that it seems preferable to exclude the AHCI fields from the maps.

The science overlay maps presented here draw upon the same full year (2007) WOS sample noted in conjunction with calculating Integration scores. Ismael Rafols factor analyzes the SC-to-SC cosine

matrix from Leydesdorff to extract 18 factors (Rafols et al., 2010).⁴ Those “Macro-Disciplines” describe the main disciplinary concentration of each group of SCs, based on our judgment. The SCs can be further aggregated into 4 “Meta-Disciplines.” These are exclusive assignments – a given SC is located in a single Macro-Discipline or Meta-Discipline for mapping clarity.

The methodology described here and in the next section relies on the WOS Subject Categories. Publications in other than WOS-indexed journals are not readily included. Use of SCs as the unit of classification locates an article based on the journal in which it was published, not upon the content of the article itself. Assignment of journals to SCs is based on a combination of citation patterns and editorial judgment at the Institute for Scientific Information (ISI) – it is not unambiguous. Recent analyses suggest that researcher views of where a given article would best be situated vary on the order of 50% from the observed SC assignment. However, those deviations tend to be “nearby” – i.e., the preferred SC is usually very close to the WOS-assigned one, so collective analyses, such as presented in science overlay maps, are quite robust (Rafols and Leydesdorff, 2009; Klavans and Boyack, 2009). However, fine distinctions should not be emphasized. This research assessment uses science overlay maps to gain perspective on disciplinary participation and uses other research network depictions (c.f., Figure 2 presented shortly) to evaluate local patterns.

Two recent articles (Porter et al., in press; Garner and Porter, 2011) presented Integration scores and science overlay mapping to help assess outputs of two NSF programs’ research funding.

Data

The 13 RCN projects that constitute the study sample were initiated between January, 2001, and March, 2003. They represent four NSF programs within the Division of Environment Biology -- ecosystem science, evolutionary processes, population and community ecology, and systematics and biodiversity science -- <http://www.nsf.gov/div/index.jsp?org=DEB>.

The NSF program managers selected 6 RCN proposals, that were not funded, as a Comparison Group (“CG”). [However, most results show only 5 because one lacked sufficient “Before and After” publications.] The principal investigators of these unsuccessful projects received high ratings and have a high rate of success with NSF funding and publication productivity. Sampling was done within NSF to protect privacy.

For each RCN proposal and CG proposal, all researchers named in the proposal or, for RCNs, in project reports to NSF, were identified. We then searched WOS to identify journal articles published by them. This required a series of steps to check that only articles by our RCN researchers were included. Until 2008, WOS only provided author last names and initials, so various checks were run:

- Against the full name field where available
- Using information on author affiliations
- Common co-authors
- Topical relevance

Abstract records were gathered for any publication year, then reduced for most analyses to the periods indicated in Figure 1 – **1999-2001** as a 3-year “**Before**” period and **2006-2008** as a 3-year “**After**” period. Given that RCN awards are for 5 years and cover various network-building activities (e.g., holding meetings, establishing websites), this seems like an appropriate time-frame. Publications by these researchers during the 2006-08 period should largely reflect RCN-related activities. “After less Before,” for the RCNs and CGs, should reflect RCN-instigated research activities as a differential. For example, taking one of the 13 projects – “Coordinated research on amphibian population declines in the neotropics” – RCN aimed for researchers from various disciplines to come together to address this topic, likely in

addition to their continuing research in their own fields. The hypothesis underlying Research Question 2 is that RCN support would result in “After” articles being, on average, more interdisciplinary than “Before” papers.

Of special interest, 235 **RCN-derived** publications are attributed to this support based on final reports (R. Inouye, NSF Program Officer, Pers. Comm.) – being published anytime from 2002 on.

Not all possible research outputs appear in WOS. Articles in journals not included in WOS plus books, reports, conference participation, etc., are not assessed. Coverage in the SCOPUS database was investigated, and determined that this would enrich the publication set, but not substantially. To discern differences attributable to RCN support, the WOS datasets suffice. These analyses further restrict to peer-reviewed journal articles, excluding a small number of other publications included in WOS (e.g., reviews, editorial materials).

We also compiled WOS abstract records for the publications that later cited RCN and CG researchers’ articles. These were tagged to associate them with the particular projects. Numbers were extensive. For instance, one RCN project included 38 authors and 799 articles located in WOS. On average, each RCN project researcher set received over 4500 citations. Table 1 gives some basic descriptive statistics; the Appendix provides more details on more measures. It is also not surprising that the RCN project groups are larger; the RCN goal of networking building encourages recruiting additional colleagues over the course of the award. CG data only have the respective proposals on which to draw while the RCN data also use annual and final reports to identify involved researchers. The CG reflects the best comparison set available, but obviously differs in composition from the RCNs.

Table 1. Researchers and Research Publication Outputs Compared

	RCN (N=13)			Comparison Group (CG)(N=5)		
	Range	Mean	Median	Range	Mean	Median
BEFORE (1999-2001)						
# of Articles/project group	37 to 218	140	138	20 to 68	37	31
# of Researchers publishing/project group	4 to 36	18	18	3 to 9	6	5
AFTER (2006-2008)						
# of Articles/project group	43 to 440	179	145	24 to 96	47	40
# of Researchers publishing/project group	6 to 36	20	19	3 to 10	6	5

Note to Table 1 – Researcher counts come from Proposals plus Final Reports and are those that published during the indicated time periods. E.g. 36 researchers published out of a group of 38 from the proposal set for the high end proposal on # of researchers “Before,” and also “After” (although not exactly the same 36).

A Shapiro-Wilks test gauged the normality of the data distributions. No test results would reject that the samples come from a normally distributed population; these 8 samples range from probabilities of 0.96 to 0.08 (close to rejection). Nonetheless, non-parametric testing appears more suitable.

How does group productivity compare? RCN-Before is more prolific than CG-Before (t test, $p = .001$); likewise for RCN-After vs. CG-After. Paired 2-tailed t tests show a statistically significant increase in # of articles/project group from RCN Before to After ($p = 0.013$). While the CG means & medians also increase from Before to After, this is not statistically significant ($p = 0.16$).⁵

VantagePoint software [www.theVantagePoint.com] was used for data cleaning and many analyses. A major challenge arises in name disambiguation (e.g., to verify that a given paper was authored by one of the target researchers, not someone else with similar last name and initials) (Tang and Walsh, 2010). This desktop text mining software is especially aimed at field-structured documents, such as WOS abstract records. It helped to:

- Extract information from multi-entry fields [e.g., author, Subject Categories (“SCs”)]
- Extract source information from the cryptic WOS “Cited Reference” format -- e.g., from “ODEGAARD F, 2000, BIOL J LINN SOC, V71, P583” obtain “BIOL J LINN SOC”
- Identify Cited SCs by applying a thesaurus to associate particular sources (journals) to WOS Subject Categories – e.g., “BIOL J LINN SOC” belongs to the “Evolutionary Biology” SC.

Summing up, these analyses focus on article abstract records retrieved from the Web of Science, as tallied in Table 2. Note that “After” could well include additional RCN-related work appearing after submission of project final reports. Notably, there are fewer citing publications for After than for Before articles because these have had fewer years in which to be cited.

Table 2. Research Publications and Citations Compared

	RCN (N=13)		Comparison Group (CG) (N=5)	
	Articles	Citing Publications	Articles	Citing Publications
BEFORE (1999-2001)	1657	48587	184	4362
AFTER (2006-2008)	2096	16916	236	1580
DERIVED (2002-2009)	235	6038		

Note to Table 2: Tabulation based on searches in the Web of Science.

Results

Disciplinary Participation: This was of interest to NSF to locate programmatic activities and to see their reach via publications (and citations to those).

Table 3 lists the leading subject categories in which the research has been published or cited. The prominent SCs listed include any that appeared in the “Top 15” for any of these six research sets. They are ordered based on the frequency of RCN-After articles. The top SC in all cases is Ecology; Environmental Sciences and Plant Sciences are either the second or third most prominent SC for each group. For instance, for “RCN After,” Plant Sciences has the third most of the 2096 articles with 17.1%; whereas for the “Citing RCN-After” articles, Plant Sciences holds second place with 16.5% of these 16,916 papers. In general, the six sets address similar broad cross-sections of the biosciences. The Comparison groups are more concentrated in agri-science. RCN-After shows increased publication intensity in “Geosciences, Multidisciplinary,” and less in Mycology.

Table 3. Distribution of Publications and Citations among Subject Categories

	RCN Before		RCN After		Citing RCN- After		CG Before		CG After		Citing CG- After	
Subject Category	Rank	% of 1657	Rank	% of 2096	Rank	% of 16916	Rank	% of 184	Rank	% of 236	Rank	% of 1580
Ecology	1	25.9%	1	29.7%	1	25.0%	1	20.7%	1	27.1%	1	23.7%
Environmental Sciences	3	14.2%	2	17.8%	3	13.5%	3	10.9%	2	16.5%	2	18.1%
Plant Sciences	2	20.9%	3	17.1%	2	16.5%	2	13.6%	3	12.3%	3	12.8%
Evolutionary Biology	4	8.6%	4	8.4%	4	7.7%						
Geosciences, Multidisciplinary	14	4.2%	5	7.2%	8	5.8%	7	6.5%	10	5.9%	4	8.1%
Meteorology & Atmospheric Sciences	7	6.8%	6	7.0%	7	6.7%	10	4.9%	11	5.9%	5	8.0%
Genetics & Heredity	6	7.2%	7	6.9%	5	7.3%						
Biodiversity Conservation	10	4.9%	8	6.3%	11	4.9%	8	4.9%	8	6.4%	8	7.1%
Forestry	11	4.6%	9	6.3%	12	4.8%	9	4.9%	13	3.0%	14	3.7%
Biochemistry & Molecular Biology	9	5.7%	10	5.5%	6	7.2%	4	7.6%	5	8.1%	11	5.4%
Multidisciplinary Sciences	8	5.9%	11	5.2%	10	5.0%			17	2.5%	12	4.3%
Mycology	5	7.3%	12	4.7%								
Marine & Freshwater Biology	13	4.3%	13	3.8%	13	4.2%	5	7.6%	7	8.1%	7	7.4%
Soil Science	12	4.5%	14	3.3%					18	2.5%		
Microbiology			15	3.0%	9	5.6%	11	4.9%	4	9.3%	6	7.9%
Agronomy							6	6.5%	9	5.9%	10	5.7%
Agriculture, Multidisciplinary							12	4.3%	12	3.4%		
Agricultural Engineering							15	3.3%	15	2.5%		
Biology	15	3.0%			15	4.0%					15	2.9%
Biophysics							16	3.3%				
Biotechnology & Applied Microbiology					14	4.2%	13	4.3%	6	8.1%	9	6.0%
Geochemistry & Geophysics							14	4.3%	16	2.5%		
Limnology											16	2.9%
Water Resources									14	3.0%	13	4.2%

Comparing the spread of publications over Subject Categories, using a Pearson's Chi-Square test, these don't differ significantly:

- RCN-Before vs. CG-Before – p value = 0.30
- RCN-After vs. CG-After – p value = 0.12
- Citing RCN-After vs. Citing CG-After – p = 0.24.

Table 3 shows that the research publications are concentrated in the biosciences. Figure 2 provides a visual perspective -- presenting the science overlay map for the RCN "After" articles. Overlay maps for RCN "Before," and for the articles citing RCN papers are generally similar. The CG research also maps quite similarly, suggesting that it is a reasonable comparison in terms of the research fields engaged. Figure 3 contrasts the relative publication activity by SC. Research fields (SCs) in which publications are more focused "After" than "Before" appear as Black nodes; fields in which publications were more focused "Before" appear shaded lighter. The pattern appears mixed, but shows more outreach to less-central SCs "After."

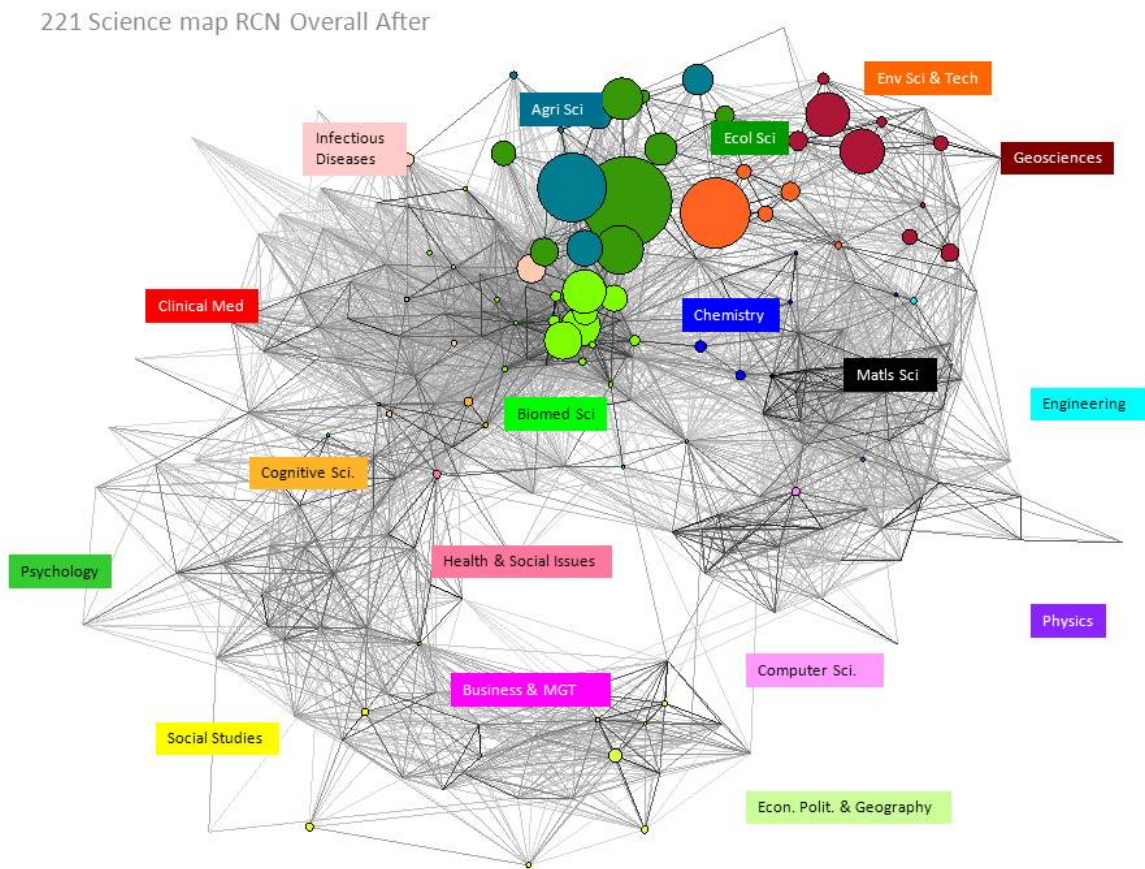


Figure 2. RCN Researcher "After" Publications

Note: In Figures 2 and 3, the gray background intersections are the 221 Subject Categories. They are located based on cross-citation relationships among all WOS journals in 2007. The 18 labeled "macro-disciplines" are based on factor analyses of that cross-citation matrix as well. Nearness in the map indicates closer relationship. The circular node sizes reflect the relative number of RCN researcher publications.

RCN Before vs After SC changes

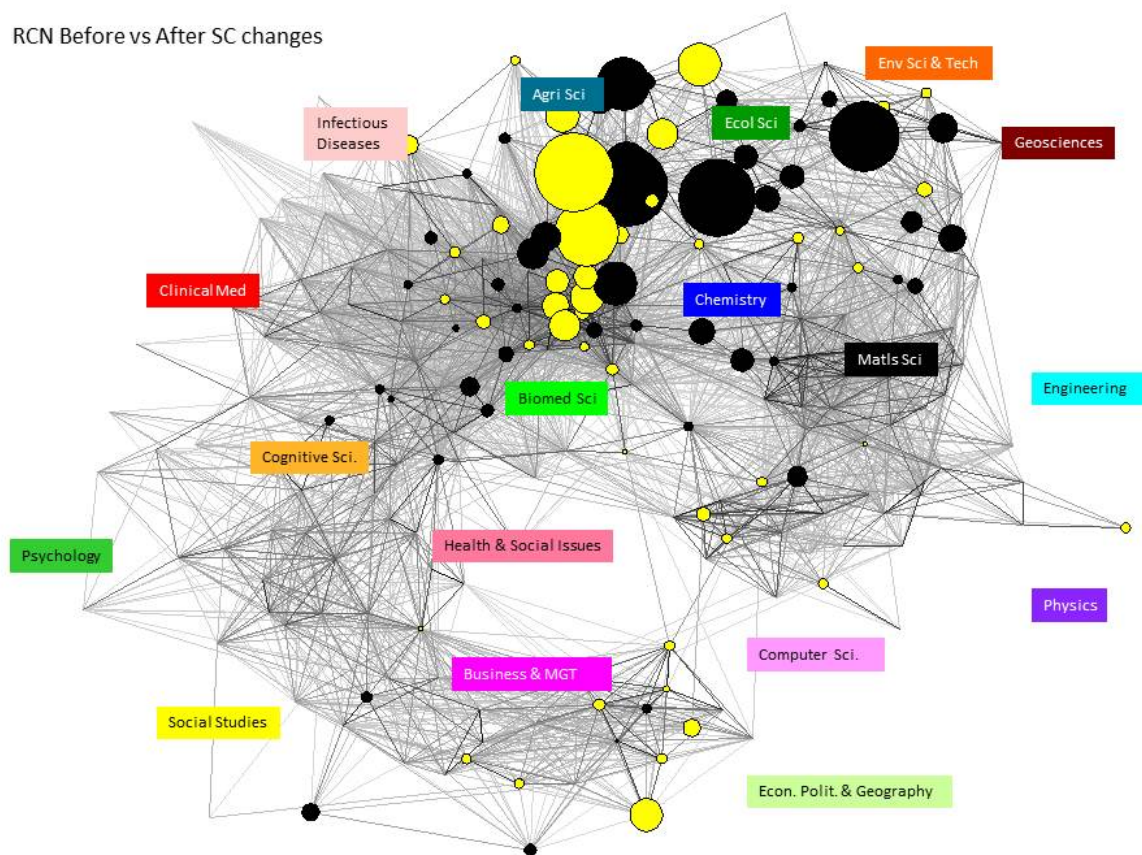


Figure 3. RCN Researcher Publication Emphases – After vs. Before

Note: Subject Categories with a larger share of After publications than Before publications are Black; those with relatively more Before appear shaded more lightly. Larger node sizes reflect a greater absolute difference between the percentages of each subject category.

In sum, RCN addresses a wide swath of biosciences; “After” RCN leans toward “macro” (environmental) domains over “micro” (biomedical ones). And, while CG publication concentrations show strong commonality with RCN interests, they also differ. As an insightful reviewer noted, evolutionary biology and genetics appear only in the RCN publication set; agriculture only in the CG.

Question 1: Does RCN support boost collaboration? One strong indicator of effective research networking is active collaboration in publications. Table 4 summarizes the overall extent of collaboration, showing a notable increase in this for the research outputs of the RCN researchers “After” the period of NSF-RCN support. However, the CG also appears to increase collaboration (next paragraph).

Table 4. Collaboration Measures Compared

[See Table 2 for numbers of articles in each category]

	RCN – Awardees (N=13)		Comparison Group (CG)(N=5)	
	Before	After	Before	After
Authors/Paper	4.46	5.66	3.96	5.04

Institutions/Paper	2.43	3.17	2.03	2.66
Countries/Paper	1.51	1.85	1.29	1.54

A 2-tailed Wilcoxon test shows that the increase in authors, institutions, and countries per paper is significant for the 13 RCN groups ($p < 0.001$), but not for the CGs (p 's of 0.31, 0.12, and 0.19). The Comparison groups are at a disadvantage in this regard in that there are only 5 to be analyzed and the number of publications is much smaller (Table 2). Auxiliary data show a similar pattern -- 85% of RCN groups increase on authors/paper; only 60% of CGs increase.⁶

Another perspective on team building can be gleaned by considering how many of the possible links among researchers in a RCN group take place. Squaring the number of researchers/project, subtracting N , and dividing by 2 gives the number of possible co-authorships. Doing the same without dividing by 2 gives the number of possible citation links (i.e., separately counting whether A cites B and whether B cites A).

Results range widely for each project/researcher team. Figure 4 shows the most dramatic increase in collaboration among one RCN project's researchers from the Before to the After period. As depicted, larger node size indicates more publications; connecting lines indicate co-authoring.

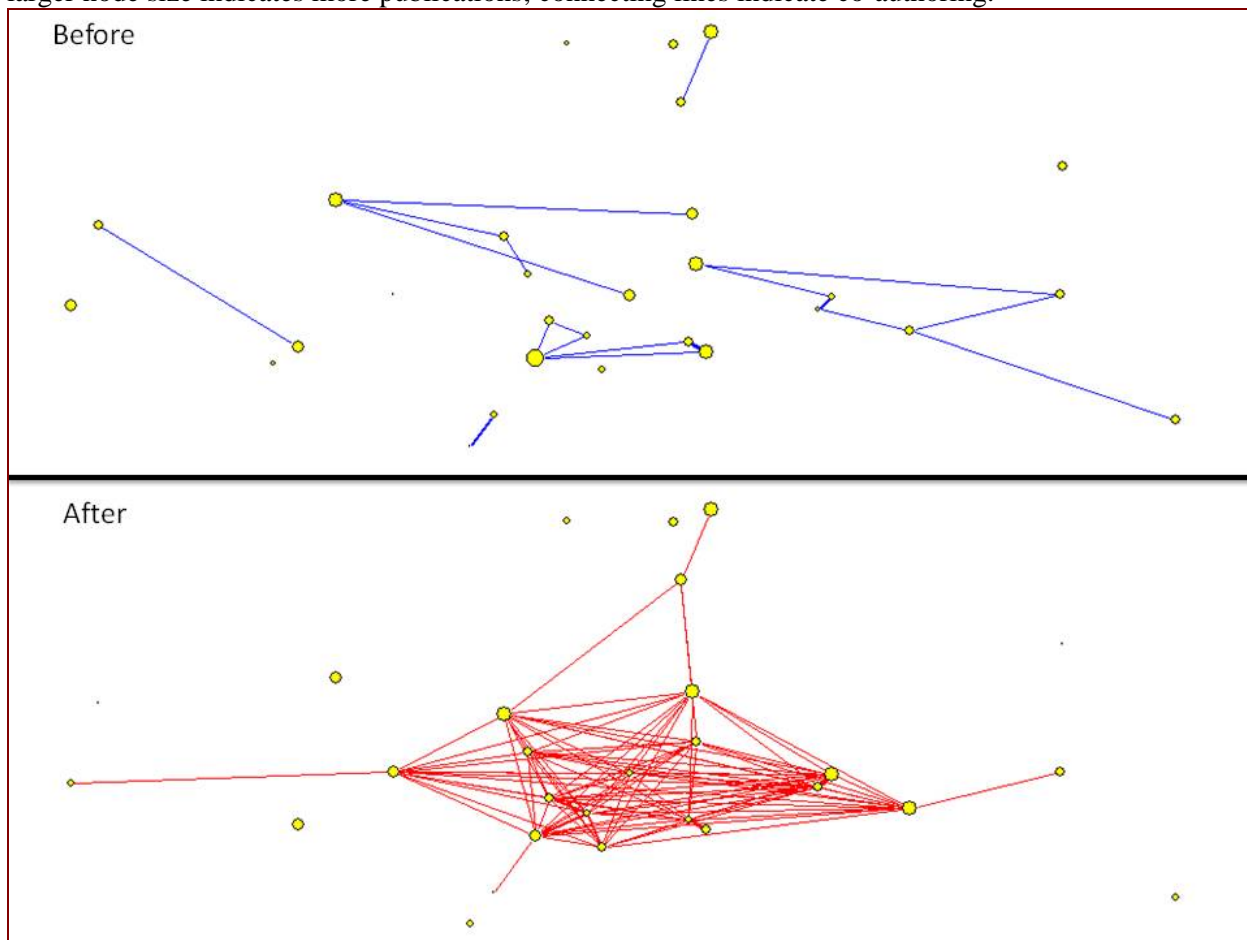


Figure 4. Change in co-authoring among researchers on one RCN project

For most of the projects, the change in linkage density is moderate. Taking the percent of possible links “After,” minus the percent “Before,” only 2 of 13 RCNs show a change greater than 4% -- one increasing

19% and one 39%. Tallying the total observed co-author, within-project two-way links, we see a notable increase from 196 to 314 for the 13 RCNs pooled. However, 96 of those additional links occur in those same 2 projects.

Note that 8 of the 13 RCNs show an increase in co-authoring; while 4 decrease and 1 is unchanged. The ratio of “After” to “Before” links increases by 60%, but without the 2 outliers, the increase is only 7%. “Average degree” for the RCNs Before vs. After shows an increase from 1.55 to 2.31 (one-tail t test, $p = 0.07$). [Comparing with the CG on this measure is not fair in that these groups did not receive funding to pursue network building, but note that only one CG project evidenced co-authoring of any significance.]

Cross-citation gives an alternative indicator of effective networking. It may well be that two people are active members of a research network, but still don’t have occasion to co-author a paper. The fact that one person cites some of the others’ research provides an indication of intellectual interchange -- awareness and contribution. For the RCNs, the extent of citing each other’s work (excluding self-citation), increases for 8 of 13 projects. Overall, the number of within-project citation links increases from 1124 to 1303 (16%) indicating network growth. Unlike co-authoring, cross-citation is not so concentrated in 2 projects – 7 projects show at least a 10-link increase. Figure 5 shows the percentage gain or loss in both types of links for each of the RCN and CG projects. [Wilcoxon tests, Before vs. After, are not significant.]

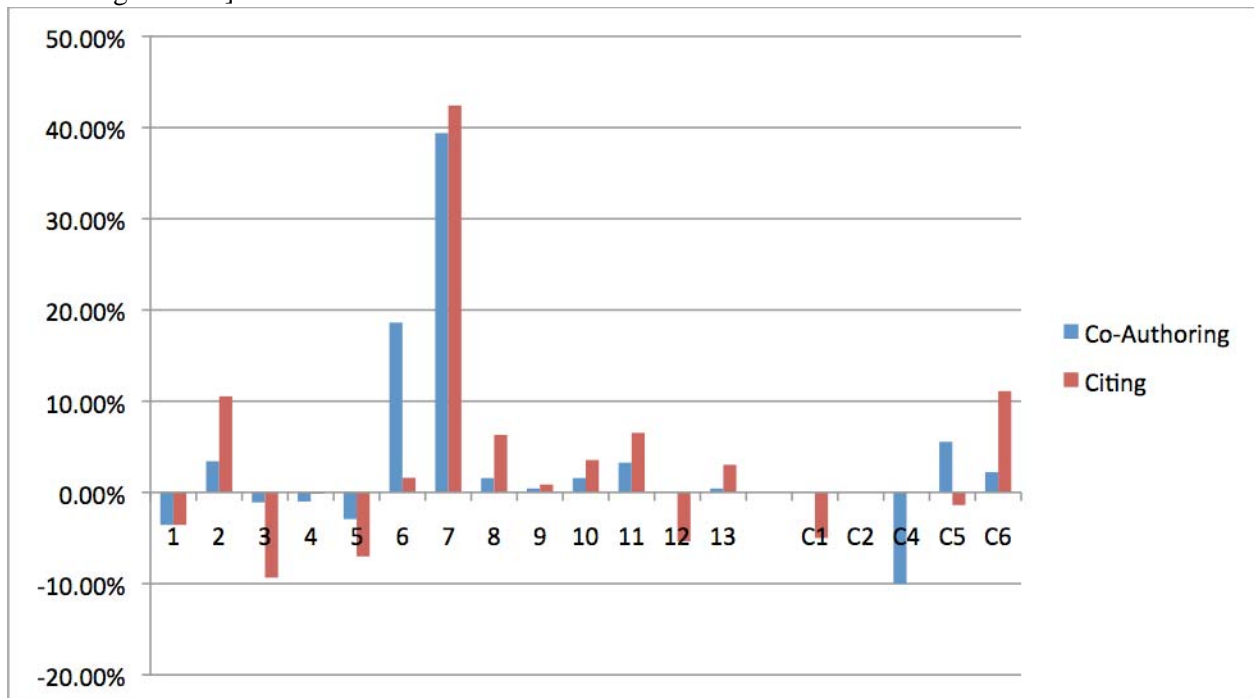


Figure 5. Percent change in co-authoring and in cross-citation among project researchers from the Before to the After period

Might the extremes in Figure 5 reflect different fields? Not particularly – the boldest outlier, RCN#7 publications concentrate most heavily in Ecology (the leading SC for both RCN and CG – see Table 4).

In sum, considerable evidence shows that the RCN awards lead to enhanced networking among the participating researchers. The extent of co-authoring increases from their publications Before to those After (as does cross-institutional and cross-national collaboration) – but unevenly across the 13 RCN projects. The CG also shows increase, Before to After, albeit not statistically significant (but that should be discounted somewhat due to the smaller N of 5 CG groups).⁷ On balance, the RCN gains are notable.

Citation of each other’s research also increases for the RCN group (and more so than for the CG), again unevenly (Figure 5), but reinforcing the evidence of increased intellectual interchange.

Question 3: Does RCN facilitate interdisciplinarity? As discussed, our prime measure of Interdisciplinarity is “Integration score.” This focuses on the diversity of the references which a paper or collection of papers cite. Figure 6 shows the mean Integration score for the indicated sets of articles. For instance, RCN project #1 researchers published 105 “Before” articles, with an integration score averaging 0.396.⁸ The corresponding “After” score (based on 128 articles) is 0.44. [As noted, Integration score scales from 0 to 1.] From experience in other research assessments (c.f., Garner and Porter, 2011), effecting change in this for a given set of researchers continuing to publish in the same field is not easy. So, this approximately 10% boost is quite notable.

Eleven of 13 RCN projects show an increase in Integration score after RCN funding; as do 3 of 5 CGs (Figure 6). Paired comparison t-tests for the 13 RCN projects, Before vs. After, indicate that the increase in Integration per paper is significant ($p = 0.0013$), whereas that for the Comparison group is not ($p = 0.99$).

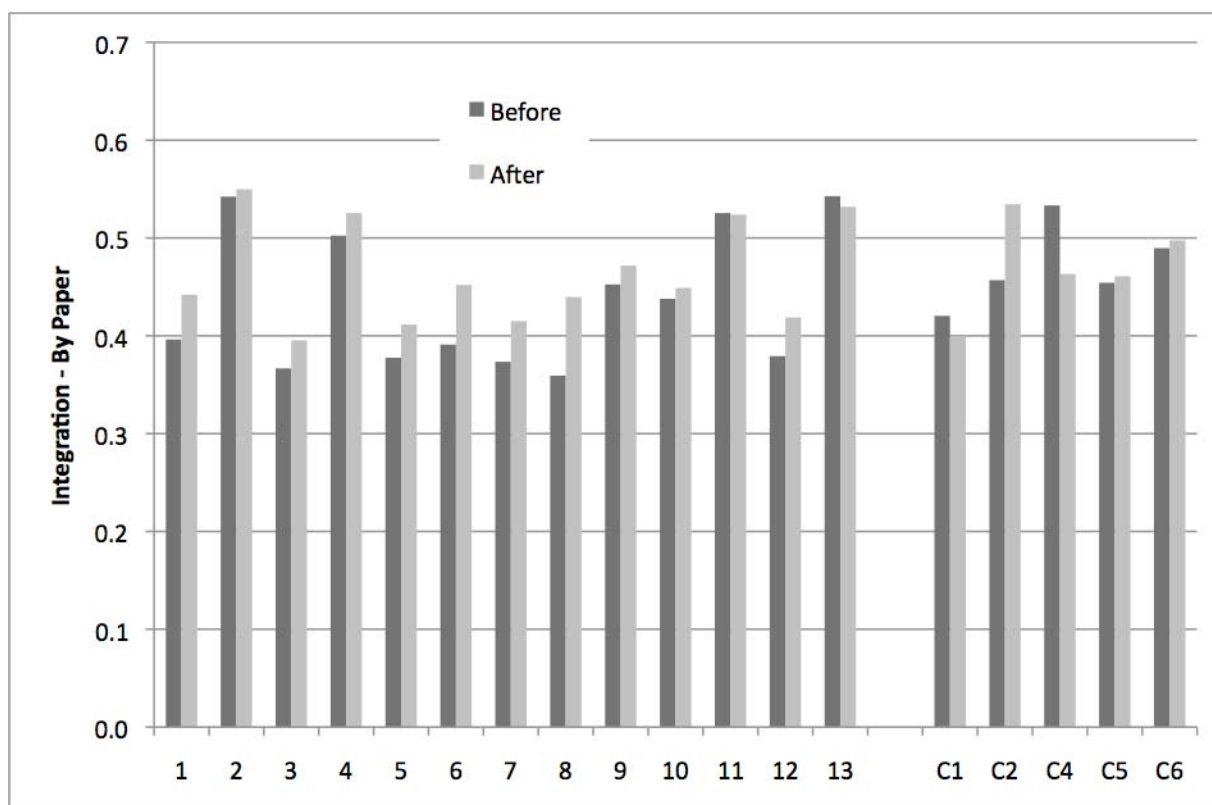


Figure 6. Average Integration Scores (per article) by Project

Aggregating all papers together for a given project (e.g., the 106 for project #1) and calculating a single Integration score based on their collective cited SCs, scores tend to be higher (e.g., for project #1 “Before,” collective $I = 0.477$). The average of the 13 project-collective I scores “Before” is 0.55 vs. 0.56 “After.” This difference is not significant. Neither is the CG decrease in project-collective I scores from 0.58 to 0.55. [While not our point of emphasis, this does suggest that these bioscience research groups are notably multidisciplinary as constituted.] RCN awards are evaluated and awarded based on the breadth of geographic, institutional, and ethnic diversity, and stage of career. Because of this, the RCN groups necessarily represent very broad representation over many axes. It should be noted that because of

the required breadth, these teams are not solely late-stage, big-reputation scientists, though the project leaders often do fit that description.

It appears that at the project-collective level, the research knowledge resources of these researchers (as reflected by their referencing patterns) did not change much. On the other hand, on a per paper basis, there is a modest, but statistically significant, broadening of their referencing. This suggests that their extent of Integration was enhanced by the RCN networking experiences exposing them to their colleague's intellectual resources.

“Specialization score” reflects the diversity of the SCs in which a body of research is published. Since any given paper is published in only one journal (which only sometimes is categorized in more than a single SC by WOS), this is not a meaningful “per paper” metric. On a project-collective basis, Specialization averages 0.49 for the 13 RCNs Before and 0.49 After. So, RCN support does not appear to lead to researchers expanding their publication Subject Categories. Nor would it have been anticipated to do so – the aim of this support is to foster a network addressing a common research theme or focus (e.g., amphibian decline in the neo-tropics). This is consistent with the stability of interests suggested by the Subject Categories tabulation in Table 3.

Returning to Integration scores – NSF also wanted to see how this research compares to other such endeavors. One can look at the CGs. A representative result is the overall per paper Integration score average (i.e., taking all the RCN-After papers' average – 0.48 – vs. all the CG-After papers' average – 0.48 – no difference). [This is slightly higher than the average of the project averages – 0.46 vs. 0.47; see Figure 6).

To address whether the degree of interdisciplinarity varied by the disciplines engaged, consider the RCN-After papers. We chose 6 SCs based on a goodly number of RCN-Derived publications and seeking diversity among the SCs. In that Biochemistry & Molecular Biology data (1485 articles in 2005) were at hand as one of the benchmark SCs (Porter and Rafols, 2009), we included this SC (9 RCN-Derived articles; 116 RCN-After). Additional ~random comparison article sets were drawn from WOS-2009 for the other 5 SCs (Figure 7), ranging in sample size from 126 to 509 articles. The RCN-Derived sets include 13 Meteorology articles (146 RCN-After) and from 29-62 articles in the other four SCs (with RCN-After article sets ranging from 110 to 622). To provide focus, Figure 7 doesn't include CG data (quite similar to RCN), nor RCN-Before (averaging about 0.02 less than RCN-After – consistent with the upward trend in Integration scores witnessed over time (Porter and Rafols, 2009).

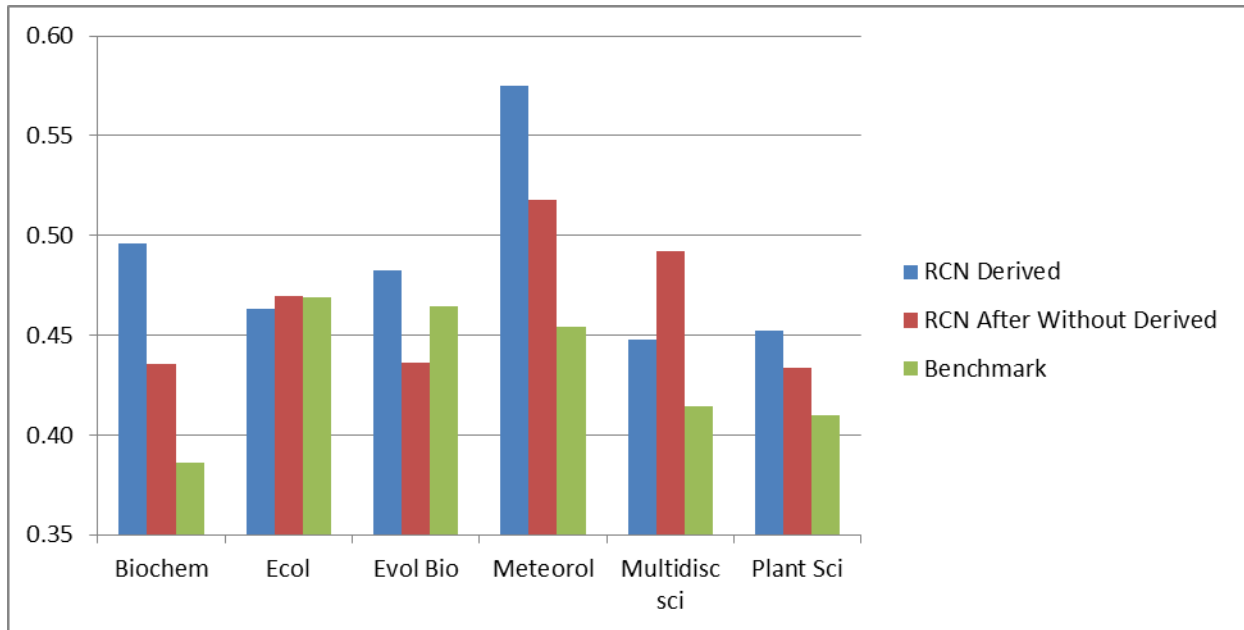


Figure 7. Benchmarking RCN article Integration Scores by Subject Category

Figure 7 suggests moderate variability in the Integration scores by discipline. RCN-Derived Integration scores range from a mean of 0.45 for Plant Sciences to 0.58 for Meteorology. For the larger RCN-After samples, the Integration scores show a similar pattern across these SCs (e.g., papers appearing in Meteorology journals tend to be the most interdisciplinary by this measure).

Recognizing that the RCN-Derived article sets per SC are limited in size (so not highly robust), they do stand out as especially interdisciplinary, based on Integration scores. Note in Figure 7 that they exceed the RCN-After papers in 4 of 6 SCs, trailing in one and essentially tied in the sixth (Ecology). Note there as well that the RCN-derived papers' Integration scores exceed the Benchmark papers published in the same SC in 5 of 6 SCs, essentially tied in Ecology. This finding prompted us to look further at the RCN-Derived papers (see Question 4).

So, does RCN support increase interdisciplinarity? The main evidence (Figure 6) supports this assertion, although increases are modest. They are statistically significant gains from Before to After RCN; also, the CG does not really show notable gains from Before to After. The papers deriving directly from RCN support (Figure 7) stand above the benchmarks.

Question 3: How influential is RCN-Derived research? As per Figure 1, the present experimental design focuses on Before-vs.-After comparisons for the RCN program – seeking to find out how RCN support for networking affects later research behavior. As described, project reports located 235 RCN-derived articles. [Of which 57 appear in the RCN-After search; the rest being published prior to 2006.] This section shares some interesting findings about these RCN-Derived articles.

Table 5 compares the RCN-Derived articles with others in terms of:

- Times Cited/Year – Citations are an indicator of impact (albeit recognizing varied bases for citation); these are given “per year” to normalize for the varying time available for cites to accrue since publication.
- Journal Impact Factor – An alternative measure of article influence that avoids the uncertainties and unevenness of tallying citations to the particular article, but Impact Factor reflects the relative citation intensity to the journal (not just to the given paper).

Results show that the RCN-Derived articles are quite special. One-way ANOVA results confirm that the groups are not equivalent on Times Cited per Year or on Journal Impact Factor; and clearly RCN-Derived is the standout on both. Using normalized (“per year”) data does not unduly favor the RCN-Derived group as the citation rate holds throughout the time period.⁹

Table 5. Article Influence

	# of Articles	Times Cited per Year [average +/- standard deviation]	Journal Impact Factor [average +/- standard deviation]
RCN Before	1657	4.4 +/- 6.4	4.3 +/- 5.9
RCN Derived	235	7.0 +/- 10.2	6.1 +/- 7.7
RCN After without Derived	2039	2.0 +/- 3.6	4.0 +/- 4.7
CG Before	184	3.1 +/- 3.2	3.7 +/- 4.4
CG After	236	2.1 +/- 2.7	3.3 +/- 3.1

Of the total RCN and CG articles analyzed, relatively few have accrued over 100 citations. Dividing the articles into two subsets provides a sharp contrast, showing that RCN-Derived papers are more than twice as likely to have achieved this elite level:

- 18 of 235 RCN-Derived articles have received over 100 cites -- 7.7%
- 125 of the other 4116 articles by these RCN and CG researchers have reached that level of recognition -- 3.0%

Separating out the 18 RCN-Derived articles with over 100 cites each, the remaining 217 RCN-Derived articles average 21.3 citations each. Of those 217, 162 have been cited 5 or more times. Put another way, 77% of the 235 RCN-Derived articles have received 5 or more citations.

In terms of journals, 21 of the 235 RCN-Derived papers were published in *Science* or *Nature* (these are especially high impact factor journals). [Before RCN, the RCN researchers did publish more than did CG researchers in *Science* and *Nature* (4.3% vs. 2.2%); but their RCN-Derived papers show a much higher rate still (8.9%).] Another 58 were published in journals with impact factors above 5 (very high). So, the directly supported research outputs seem to be notably influential.

In sum, publications deriving from RCN support stand out as notably more influential (in terms of citations received) than other papers by the RCN or CG researchers. Several related measures reinforce this interpretation – propensity to be published in high impact factor journals, percent with over 100 cites, and percent in premier journals.

Question 4: How broadly does this research knowledge diffuse? As introduced earlier, Diffusion score measures the extent to which a paper is cited across diverse Subject Categories (Carley and Porter, to appear). Table 6 shows the Diffusion scores. RCN vs. CG differences are not significant. The only pattern of notable difference is between the Before and After RCN-award scores. It appears that Diffusion score increases with time since publication (i.e., favoring those earlier publications).¹⁰ The logic for this might be that upon publication, a paper is first recognized by researchers close to the area; over time, recognition and interest may spread to more distant SCs, resulting in a higher Diffusion score.

Table 6. Diffusion Scores

	# of Articles	Diffusion Score [average +/- standard deviation]

RCN Before	1599	0.44 +/- 0.13
RCN Derived	217	0.41+/- 0.12
RCN After without Derived	1765	0.40+/- 0.15
Comparison group Before	173	0.46+/- 0.13
Comparison group After	197	0.40+/- 0.17

To better understand what is happening, we explored a subset of the RCN-Derived papers' influence patterns in more depth. The largest concentration of these publications was in the "Ecology" SC (75 papers). We drew an approximately random sample of WOS Ecology publications in 2009 for comparison. Integration scores for these show very similar patterns (Figure 7). Diffusion score comparisons for these Ecology papers are somewhat confounded by the different times since publication, contributing to different extent of citations received (mean of 32 for RCN-Derived Ecology vs. 11 for Random Ecology papers). The Diffusion scores do not indicate broader outreach for the RCN papers (mean of 0.41 vs. 0.43 for the Random Ecology – non-significant difference). However, other statistics hint otherwise –

- 55% of the RCN-Derived paper citations were received in other than the publication SC(s), compared to 47% for the Random Ecology set
- 54 total (different) SCs citing the RCN-Derived publications vs. 23 different SCs citing the Random Ecology SC publications.

On balance, RCN-Derived and RCN-After papers appear to be cited about as broadly as other papers in their fields.

Conclusions and Discussion

Conclusions

To summarize the key findings:

- The NSF RCN program engages a generous swath of bio- and environmental science disciplines.
- RCN funding has induced increased collaboration, as attested to by multiple measures.
- RCN funding has fostered somewhat more interdisciplinary research publications, especially for the RCN-derived publications, with indications of modest increase when comparing RCN-After vs. RCN-Before.
- While not an express program aim, RCN-derived papers show as especially influential (publication in more prestigious journals and higher citation rates).
- Research knowledge Diffusion associated with RCN-derived or RCN-After publications was not notably different than RCN-Before or the Comparison Group.

Prior sections have noted study design and data limitations in interpreting the results, so will not be repeated. Rather, consider the implications for research evaluators. These bibliometric analyses offer some notable strengths. The empirical results are reproducible and don't depend on direct peer judgment. The indicators and visualizations provide multiple perspectives on what has been achieved by program support. These results also pose some notable weaknesses. The low number of groups means the comparisons are relatively weak in terms of statistical "power" to detect actual differences. The extensive data gathering, especially of citing papers (~80,000 here) requires considerable resources.

Future Research and Research Evaluation Applications

This section explores how the approach and methods described might be applied in other research assessments.

The Non-Equivalent Control Group (NECG) design (Figure 1) is more interpretable than 1) simple case studies, 2) Before vs. After comparisons without a Comparison Group, or 3) comparisons with a non-equivalent control group (without the Before data). Research assessors should consider the potential of NECG designs. However, these pose challenges in identifying one or more suitable comparison groups. In this case, unfunded proposals to the program in question were used, but that required careful protection of researcher identities. Furthermore, one cannot rely on statistical significance per se to conclude there was or was not a program-induced effect – one must consider the potential selection and other biases present in the NECG comparisons afforded.

Methodologically, Integration score for a project was calculated two ways – 1) as the average of the Integration scores of each associated publication, and 2) as a single Integration score calculated on the diversity of the project’s total set of references (for all papers together). The former is the standard mode used, and it shows a moderate increase for RCN-After over RCN-Before, whereas the latter is essentially stable. Our interpretation is that the sum total of research fields engaged by a given set of researchers participating in an RCN does not, and should not be expected to, increase. But, the RCN researchers are “integrating” that knowledge more within their research work, as reflected by the diversity of sources cited within papers.

Social network analysis (SNA) is being widely used to understand research dynamics. Results indicate that within-project connectedness goes up from Before to After for most of the RCN projects. We report on average degree, extent of co-authoring, and extent of cross-citation. This study also presents a novel composite network analysis model that combines co-authoring with cross-citation. The rationale is that cross-disciplinary interchanges among a diverse research network do not necessarily lead to “everyone” co-authoring. Rather, exposure to fresh ideas and research papers could reflect in more diversified referencing. Evidence is presented here for enhanced citation of each other’s work. Another measure (not presented) is bibliographic coupling – to what extent do papers cite the same items (or authors)? Such measures can also enrich assessment of research centers in terms of ways in which researchers “networking” can reflect in changed outputs (Rogers, 2011).

NSF sought this assessment to provide evidence of the effects of a novel program intended to spur development of targeted research networks. The results support the conclusion that the funding did spur network building, albeit to a moderate degree. That finding serves NSF research policy-making in considering extension of the RCN program model to other fields. In such considerations, tabulating and mapping the disciplinary participation in RCN can help consider generalizability beyond this biology-based program.

One could imagine that there might be backgrounds that are “too diverse” -- or -- not sufficiently diverse to attain notably impactful research. The NSF HSD program, noted in the Introduction, sought to bring together quite diverse researchers and to generate papers that drew interest beyond the social, behavioral & economic sciences. By our reckoning, they have succeeded to a considerable degree (Garner and Porter, 2011). An interesting extension would be to further characterize the participating researchers – e.g., their training disciplines, departmental homes, appointments (research positions; teaching faculty), institutional types, seniority – and relate to research outputs as studied here.

Interdisciplinarity has emerged as a popular theme for research program development. The concept is that bridging disciplinary boundaries can afford richer prospects of novel and important scientific advances – i.e., great strides are apt to occur at the “interstices” among the disciplines. There is also a belief that interdisciplinary research is especially conducive to making progress on applied problems of societal importance. The present results do not speak directly to the merits of interdisciplinary research, but they do support the notion that programmatic parameters can alter researcher activities toward interdisciplinarity. This study provides evidence of cross-field collaboration and of publications that draw

upon more diverse research knowledge sources (higher Integration score). And, there is also the tantalizing result that the RCN-derived publications stood forth as more influential, suggesting that this program aiming at stimulating networking had also generated especially high impact (i.e., highly cited) research.

This assessment focuses on research publications and citations. Publications are the most prominent “output” measure available for such scientific endeavors; citations provide an “outcome” measure. Citation analyses offer windows to research knowledge transfer processes. Herein, analyses of the citations *by* the RCN and CG journal articles were used to gauge the intellectual diversity of those articles (Integration scores). Citations *to* those articles were studied to understand the diversity of researchers drawing upon such research knowledge (Diffusion scores) and the degree of influence (Times Cited). Such tools enrich research assessment by helping track research knowledge transfer patterns. One could extend those analyses by further categorizing citing types – e.g., basic vs. applied; biomedical vs. clinical; and so forth.

R&D program managers would surely like to know what factors contribute to the effectiveness of such research. Further research on the research “system” involved would be quite desirable. Funding acknowledgements catalogued by WOS since 2008 offer another means to associate particular programs with research publications. Investigation of infrastructures and *institutional arrangements* associated with special program funding could help understand what combinations are most efficacious. For example, certain institutional types may work together more effectively in RCN research networks, leading to higher productivity or more interdisciplinary outputs. Collaborative mechanisms, team science behaviors, and other *process* attributes certainly affect the generation of research outputs in important ways. Wagner et al. (2011) review methods for measuring interdisciplinary attributes – a field of study with important early work as well (c.f., Mar et al., 1985; Klein, 1996). A number of resources are available to pursue analyses of interdisciplinary research processes and team science.¹¹

References

- Achen, C.H. (1986), *The Statistical Analysis of Quasi-Experiments*, Berkeley, CA: The University of California Press.
- Balakrishnan, A., Pena, V., and Lal, B. (2011), *Measuring Interdisciplinarity: A Unique Comparison Between the Researcher and Research Proposal*, American Evaluation Association, Anaheim.
- Campbell, D. & Stanley, J. (1963), *Experimental and Quasi-experimental Designs for Research*, Chicago: Rand-McNally.
- Cook T., and Campbell, D.A. (1979), *Quasi-Experimentation*, New York: Houghton Mifflin.
- Carley, S., and Porter, A.L. (in press), A Forward Diversity Index, *Scientometrics*.
- De Bellis, N. (2009), *Bibliometrics and Citation Analysis*, Lanham, MD: The Scarecrow Press.
- Falk-Krzesinski, H.J., Hall, K., Stokols, D., and Vogel, A. (2010), Science of Team Science. In *Wikipedia: The Free Encyclopedia* (Wikimedia Foundation, Inc).
- Garner, J., and Porter, A.L. (2011), Assessing the Human and Social Dynamics Program—Exceptional Cross-disciplinarity, *Atlanta Conference on Science and Innovation Policy, 2011*.

- Judd, C.M., and Kenny, D.A. (1981), *Estimating the Effects of Social Interventions*, Cambridge, UK: Cambridge University Press.
- Klavans, R., and Boyack, K. (2009), Towards a consensus map of science, *Journal of the American Society for Information Science and Technology* 60 (3), 455-476.
- Klein, J.T. (1996), *Crossing Boundaries: Knowledge, Disciplinarity, and Interdisciplinarity*, Charlottesville, VA: University Press of Virginia.
- Leydesdorff, L. and Rafols, I. (2009), A Global Map of Science Based on the ISI Subject Categories. *Journal of the American Society for Information Science and Technology*, 60 (2), 348-362.
- Mar, B.W., Newell, W.T., and Saxberg, B.O. (eds.) (1985), *Managing High Technology*, New York: Elsevier Science Publishers.
- Martin, B., and Irvine, J. (1983), Assessing basic research: Some partial indicators of scientific progress in radio astronomy, *Research Policy*, 12 (2), 61-90.
- National Academies (2005), *Facilitating Interdisciplinary Research*, Washington, DC: National Academies Press.
- Porter, A.L., and Rafols, I. (2009), Is Science Becoming more Interdisciplinary? Measuring and Mapping Six Research Fields over Time, *Scientometrics*, 81 (3), 719-745.
- Porter, A.L., Roessner, J.D., and Heberger, A.E. (2008), How Interdisciplinary is a Given Body of Research?, *Research Evaluation*, 17 (4), 273-282.
- Porter, A.L., Cohen, A.S., Roessner, J.D., and Perreault, M. Measuring Researcher Interdisciplinarity, *Scientometrics*, Vol. 72, No. 1, 2007, p. 117-147.
- Porter, A.L., Schoeneck, D.J., Roessner, D., and Garner, J. (2010). Practical research proposal and publication profiling, *Research Evaluation*, 19 (1), 29-44.
- Porter, A.L., Garner, J., and Crowl, T. (to appear). The RCN (Research Coordination Network) experiment: Can we build new research networks? *BioScience*.
- Porter, A.L., Newman, N.C., Myers, W., and Schoeneck, D., Projects and Publications: Interesting Patterns in U.S. Environmental Protection Agency Research, *Research Evaluation*, Vol. 12, No. 3, 171-182, 2003.
- Porter, A.L., Schoeneck, D.J., Roessner, D., and Garner, J. (2010). Practical research proposal and publication profiling, *Research Evaluation*, 19(1), 29-44.
- Rafols, I. and Leydesdorff, L. (2009), Content-based and Algorithmic Classifications of Journals: Perspectives on the Dynamics of Scientific Communication and Indexer Effects. *Journal of the American Society for Information Science and Technology*. 60(2), 348-362.
- Rafols, I., Porter, A.L., and Leydesdorff, L. (2010), Science overlay maps: A new tool for research policy and library management, *Journal of the American Society for Information Science & Technology*, 61 (9), 1871-1887.

Rafols, I. and Meyer, M. (2010), Diversity and Network Coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, 82(2), 263-287.

Rogers, J. (2011), Program Level Assessment of Outcomes and Impacts of Research Performance of Centers, American Evaluation Association, Anaheim.

Stokols, D., Hall, K.L., Taylor, B.K., and Moser, R.P. (2008), The Science of Team Science: Overview of the Field and Introduction to the Supplement. *American Journal of Preventive Medicine* 35, S77-S89.

Stirling, A. (2007), A General Framework for Analysing Diversity in Science, Technology and Society, *Journal of the Royal Society Interface*, 4 (15),707-719.

Tang, L., and Walsh, J.P. (2010). Bibliometric fingerprints: name disambiguation based on approximate structure equivalence of cognitive maps, *Scientometrics* DOI 10.1007/s11192-010-0196-6.

van Raan, A.F.J. (1988), *Handbook of Quantitative Studies of Science and Technology*, (North-Holland).

van Raan, A.F.J. (2000), R&D evaluation at the beginning of the new century, *Research Evaluation*, 9(2), pages 81–86.

Wagner, C.S., Roessner, J.D., Bobb, K., Klein, J.T., Boyack, K.W., Keyton, J., Rafols, I., and Borner, K. (2011), Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature, *Journal of Informetrics*, 5165, 14-26.

Appendices

1. RCN Data Summary Table

	RCN													Mean	Std Dev
	1	2	3	4	5	6	7	8	9	10	11	12	13		
# researchers in proposal	8	27	14	38	19	30	12	20	22	23	18	8	22	20.08	8.58
# authors Before	8	25	14	36	18	29	10	16	18	26	19	4	17	18.46	8.84
# authors After	8	27	15	36	17	28	13	17	21	28	19	6	21	19.69	8.50
# papers Before	106	180	138	359	135	218	37	57	139	176	153	40	88	140.46	85.88
# papers After	128	342	135	440	145	191	78	86	182	248	214	43	101	179.46	111.78
# authors Derived (RCN projects only)	N/A	12	5	13	14	18	12	13	11	7	9	0	9	10.25	4.69
# papers Derived (RCN projects only)	N/A	9	3	37	50	26	38	36	22	7	11	2	4	20.42	16.64
# authors/paper Before	6.07	5.90	5.41	4.91	3.73	3.53	4.97	3.59	3.45	5.36	5.57	2.88	3.64	4.54	1.09
# authors/paper After	5.84	7.01	6.51	5.86	4.77	5.58	7.68	6.07	4.76	5.36	6.81	3.95	4.33	5.73	1.10
# institutions/paper Before	1.86	3.33	2.33	3.25	2.35	2.33	3.43	2.21	1.77	2.94	3.09	1.73	2.34	2.53	0.60
# institutions/paper After	2.71	4.09	3.07	3.93	2.66	3.25	4.91	3.77	2.67	3.17	3.64	2.14	2.93	3.30	0.75
# countries/paper Before	1.43	1.80	1.36	1.50	1.61	1.57	2.05	1.31	1.30	1.78	1.63	1.10	1.50	1.53	0.25
# countries/paper After	1.57	2.14	1.62	1.88	1.94	1.83	2.89	1.98	1.70	1.92	1.90	1.26	1.60	1.86	0.38
# Times Cited/year Before	6.91	4.86	5.77	4.82	3.78	2.72	4.42	1.86	2.40	6.14	5.03	3.73	3.05	4.27	1.52
# Times Cited/year After	2.50	2.24	2.79	1.86	1.57	1.82	2.66	2.30	1.66	2.35	2.43	2.54	1.10	2.14	0.50
# Times Cited/year Derived (RCN projects only)	N/A	9.08	4.33	5.40	7.03	7.22	5.73	8.58	5.70	14.67	2.07	2.20	7.12	6.59	3.36

2. Control Group Data Summary Table

	CG						Mean	Std Dev
	1	2	4	5	6			
# researchers in proposal	5	4	5	9	10	6.60	2.70	
# authors Before	5	3	5	6	9	5.60	2.19	
# authors After	3	4	5	9	10	6.20	3.11	
# papers Before	28	36	20	31	68	36.60	18.49	
# papers After	24	44	40	31	96	47.00	28.48	
# authors/paper Before	3.64	4.06	5.35	3.42	3.85	4.06	0.76	
# authors/paper After	2.92	5.36	5.23	5.10	5.33	4.79	1.05	
# institutions/paper Before	2.43	1.94	1.65	1.45	2.29	1.95	0.41	
# institutions/paper After	1.88	2.91	2.23	2.26	3.03	2.46	0.49	
# countries/paper Before	1.50	1.44	1.05	1.19	1.24	1.28	0.19	
# countries/paper After	1.33	1.80	1.18	1.39	1.68	1.47	0.26	
# Times Cited/year Before	2.52	1.78	3.06	3.37	3.86	2.92	0.80	
# Times Cited/year After	1.78	1.02	2.09	2.63	2.39	1.98	0.63	

¹ Two EPA STAR programs were compared with each other, but there was no separate control group and no statistical testing per se.

² No suitable comparison group was identified and no statistical tests were presented.

³ Comparing our benchmark data using a similarity matrix based on a smaller sample of co-citation data (previously reported, Porter et al., 2007) vs. this full WOS similarity matrix, results correlate strongly ($r = 0.98$), however, the new benchmark values are lower than previously (0.43 vs. 0.62 for five benchmark SC sets).

⁴ We continue to generate variations on the science overlay maps. Leydesdorff has generated a 2009 SC-x-SC similarity matrix, from which Rafols has extracted an 18-factor solution – available at www.leydesdorff.net/overlaytoolkit. Also, there is a new 4-factor (Meta-Discipline) mapping option. Those interested in making their own science overlay map can also do so at www.idr.gatech.edu; see approaches presented at: www.interdisciplinaryscience.net. The 2009 base maps are not used here. Differences between the 2007 and 2009 base maps are small.

⁵ Wilcoxon (non-parametric) tests yield essentially the same results.

⁶ An astute reviewer noted analytical options. One could convert these results to change scores (which show slightly larger gains for RCN than for CG), or to After/Before ratios (in which case the CG equals the RCN). On the other hand, one could argue that the RCN collaboration level being higher to begin, makes it more challenging to increase.

⁷ Because of our highly unbalanced design (13 vs 5), traditional statistics tests are very unreliable and hold little power. In order to determine if there is statistical evidence for changes in co-authoring and cross-citation, we ran two-way ANOVA's on bootstrapped subsets of the RCN projects and the controls. Random sets of five RCN projects were selected and then compared to the five controls. Repeating 100 times, we find that approximately 22% of our comparisons showed significant differences in either co-authoring or cross-citations. No interactions were significant.

⁸ We calculate Integration scores for the 105 of 106 articles that had at least 3 instances of a cited SC. Those SC instances could accrue from fewer than 3 references, if the cited journals happen to be associated with more than a single SC (about 40% of WOS journals are categorized in multiple SCs). On the other hand, a paper with more than 3 references could be excluded if those did not include sufficient WOS journals and corresponding SCs.

⁹ A reviewer wondered if normalization per year might favor the RCN-Derived papers in that citation intensity may peak a few years after publication? This does not seem to be happening in this case. Citing record counts for the RCN-Before and CG-Before papers published 1999-2001 do not taper off through 2009 – e.g., RCN-Before papers receive 5063 cites in 2004, increasing modestly each year through 2008 (5856), with almost as many in 2009 (5673). CG-Before papers receive 515 citations in 2004 and 514 in 2009 (ranging from 475 to 525 in between) – i.e., quite stable.

¹⁰ One-way ANOVA finds the Diffusion score differences among the five groups highly significant. Wilcoxon tests show RCN-Before higher than RCN-After, and CG-Before higher than CG-After. Differences between RCN and CG in the same time periods are not significantly different.

¹¹ Some resources of note – Science of Team Science initiative and conferences [www.scienceofteams.org/]; science overlay mapping and analyses of interdisciplinary research [<http://www.idr.gatech.edu/>]; and <http://www.interdisciplinaryscience.net/>].