Journal of Informetrics 8 (2014) 43–58

Contents lists available at ScienceDirect

Journal of Informetrics

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

Conceptualizing the interdisciplinary diffusion and evolution of emerging fields: The case of systems biology

Alyssa W. Goldman*

Concept Systems Inc., 136 East State Street, Ithaca, NY 14850, USA

ARTICLE INFO

Article history:
Received 25 June 2013
Received in revised form 22 October 2013
Accepted 25 October 2013
Available online 16 November 2013

Keywords:
Diffusion
Interdisciplinarity
Network analysis
Systems biology

ABSTRACT

This paper contributes to the longitudinal study and representation of the diffusion of scholarly knowledge through bibliometrics. The case of systems biology is used to illustrate a means for considering the structure and different roles of journals in the diffusion of a relatively new field to diverse subject areas. Using a bipartite network analysis of journals and subject categories, a core–intermediary–periphery diffusion structure is detected through comparative analysis of betweenness centrality over time. Systems biology diffuses from a core of foundational, theoretical areas to more specific, applied, practical fields, most of which relate to human health. Next, cluster analysis is applied to subject category co-occurrence networks to longitudinally trace the movement of fields within the core–intermediary–periphery structure. The results of these analyses reveal patterns of systems biology's diffusion across both theoretical and applied fields, and are also used to suggest how the dynamics of a field's interdisciplinary evolution can be realized. The author concludes by presenting a typology for considering how journals may function to support attributes of the core–intermediary–periphery structure and diffusion patterns more broadly.

1. Introduction

Diffusion of scholarly knowledge is a complex process to measure, as can be inferred by the various aspects and methods through which researchers have investigated this topic. Inquiries mapping the diffusion of ideas through specific articles (Liu & Rousseau, 2012), research institutes (Boner, Penumarthy, Meiss, & Ke, 2006), medical subject headings (MeSH) (Leydesdorff, Rotolo, & Rafols, 2012), and essential science indicator (ESI) fields (Liu & Rousseau, 2010) are some of the perspectives through which dissemination and diffusion patterns have been studied. Diffusion indicators also vary among studies, and have included citations (Boner et al., 2006), publications (Liu & Rousseau, 2010) and collaborations between entities (Liu, Rousseau, & Guns, 2013). Developing a journal diffusion factor has also been explored (Frandsen, Rousseau, & Rowlands, 2006) as a means of capturing the breadth of influence of a particular journal title across the literature.

Diffusion and movement of ideas across academic or cognitive space (Liu & Rousseau, 2010; Liu, Rafols, & Rousseau, 2012) continues to be a focal area in informetrics. This paper contributes to the development of frameworks for studying the diffusion and evolution of knowledge across disciplines. The approach described in this article contributes to the ability to model and describe diffusion across disciplinary borders, considering the relationships among disciplines publishing on a particular topic or field as representing the field's interdisciplinary landscape. By investigating novel means of modeling the interdisciplinary trajectories of ideas through this analysis, we enhance our understanding and conceptualization of the

---

* Tel.: +1 857 225 2054.
E-mail address: alyssa.w.goldman@gmail.com

1751-1577/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.10.009

processes that facilitate the spread, or movement, of scientific knowledge to diverse areas where further application of this knowledge can potentially enhance growth and discovery across a range of fields.

The purpose of this article is to extend the longitudinal study and representation of diffusion of scientific knowledge through the lens of journals and their Web of Science (WoS) subject categories. Specifically, this paper explores how certain journals' affiliate subject categories, in the context of a field's broader publication database, might be described as mechanisms for facilitating an evolving subject's diffusion into diverse disciplines. The Institute of Scientific Information (ISI) assigns subject categories (SCs) to journals that are included in the *Science Citation Index (SCI)*, accessible through Thomson Reuter's Web of Science (WoS) database. A journal's subject category assignment is based on various qualities, including journal title and citation patterns (Leydesdorff & Rafols, 2009).

As this paper is concerned with how knowledge diffuses into different disciplines via journals, subject categories provide an appropriate lens for parsing journals into disciplinary boundaries that are useful for studying development in the structure of a publication set. Subject categories provide a more detailed representation of the qualitative disciplinary distinctions among journals than the broader 22 *ESI* fields (Liu & Rousseau, 2010). Whereas journals are assigned to only one *ESI* field, journals can be assigned to one or more of the 200+ WoS subject categories depending on its scope, allowing for a more nuanced examination of interdisciplinary diffusion structure and its evolution.

For the present inquiry, systems biology is used as an empirical example to explore changes in disciplinary structure over time. Systems biology was selected based largely on the breadth of its potential interdisciplinary application, as well as its relatively recent post-genomic inception within bioscience, providing a focused time frame for the present inquiry. The sequence of analyses used in this case study reconcile prior efforts to describe the nature of a topic's diffusion with an interest to also explain the movement of ideas or change in an interdisciplinary publication structure over time. These results suggest how a network perspective of subject categories may yield insight into the overall structure of a relatively new field, and how the attributes of certain journals may contribute to structural changes in the network over the course of a field's development. Specifically, this paper expands consideration of journal attributes relevant to diffusion beyond subject categories alone, in order to suggest how such attributes may be used to categorize the functional role of a particular journal at a more macro level, i.e. in relation to all other journals in the network at a given point in time. In this paper, functional role refers to how one might broadly characterize the facilitation pattern of a particular journal in the diffusion of a scientific concept to other disciplines. This approach extends prior work on diffusion structure by suggesting a means for realizing how a journal longitudinally contributes to knowledge diffusion in the context of the overall network using an attribute of interest (in this case, subject categories), and proposing insight that might be gained from this more macro level approach. The output of this work contributes to efforts in information science to comprehensively and longitudinally describe the diffusion of scientific ideas, and suggests novel ways of considering journals' functions in supporting aspects of a field's interdisciplinary structure and evolution.

This paper is structured as follows: First, a brief background of systems biology and its interdisciplinary potential support the use of this field as an empirical example and the methods undertaken. Second, descriptive statistics of system biology's publication base provide an overview of the field's scholarly development to date. Next, a bipartite network analysis explores changes in the centrality of journals' affiliate subject categories over the 12-year time span considered. The results of this analysis suggest a core–intermediary–periphery diffusion pattern that is then more formally studied in the context of the journals facilitating this structure and their changing positions within this structure over time. The paper concludes by offering a typology for considering the role of journals in the structure of a topic's interdisciplinary diffusion that can be tested and expanded with future research.

2. Background

2.1. The case of systems biology

This paper uses systems biology as a case study for exploring the roles of journals in the diffusion of a field or subject area across disciplines over time. Systems biology is a field of biological science that focuses on the study of whole systems of biological components, including the interactions and emergent properties of their constituent parts and the implications of those properties for enhancing our understanding of biology and related fields. Systems biology is grounded in the fundamental concept that biological components exist in a nested hierarchy (cell, tissue, organ, organism, etc.), and that adopting a network perspective in examining the interrelationships within and across hierarchical levels can generate novel insight about biological processes and their structural and functional dynamics (Potters, 2010). Built on the tenets of molecular biology, systems biology aims to use principles of information transfer among molecular components to study the human organism as a complex adaptive system constantly interacting with developmental, ecological and environmental variables.

Traditional reductionist approaches to molecular biology have focused largely on single parts, i.e. specific genes, or collections of parts in isolation from the broader context in which they exist. Although these methods have produced a tremendous amount of data on individual genes, proteins, metabolic factors and short sequences of DNA, the systems approach has introduced a new paradigm emphasizing the holistic interrelationships and collaborative functioning of those elements in the context of the whole system. The inherent complexity of adopting a large, multi-scale approach to biological systems demands methods and technologies capable of analyzing significantly more data than fields like molecular biology have typically considered. Systems biology represents the fusion of biological thinking and mathematical modeling, encompassing the sophisticated, complex measurement of the multiple interacting parameters that govern the homeostasis or balance of biological systems, and hence holds great potential for enhancing fields of medicine and human health, particularly in understanding disease.

As such, the emergence of systems biology has led to the development of new machineries and techniques capable of modeling the vast data sets generated through single molecule sequencing. Technological advancements to expedite, analyze and model extraordinary amounts of data such as high throughput sequencing, coupled with the broad potential application of systems biology theory, have necessitated a certain degree of interdisciplinary partnership. The intricacies of complex data sets, for example, have required collaboration among biologists, mathematicians and physicists in order to create robust equations and appropriate models for summarizing data sets and predicting the activity of cellular networks. Collaboration among computational engineers and chemists has been required to develop the devices and software needed to produce accurate and reliable quantitative measures over time. Much of the foundational, theoretical progress in systems biology can therefore be attributed to interdisciplinary work that has to a certain extent established an initial infrastructure for the field.

More recently, the field has seen calls for ongoing and increased interdisciplinary collaboration as a critical conduit for further advancements, particularly as they relate to the application of theory to practice (Facciotti, 2009; Friboulet & Thomas, 2005). Whereas collaboration across biology, computation, mathematics and other relevant fields must continue in order to sustain this infrastructure, the potentially broad applicability of systems-level findings also require interdisciplinary research, education and translation. In addition, cross-disciplinary collaboration is needed to create a shared language for the field that will allow researchers from various disciplines to effectively communicate in solving systems problems (Carney, 2003).

For the present inquiry, systems biology was selected based on the potential breadth of its disciplinary application. Further, by selecting a relatively new field with interdisciplinary applicability, the movement of ideas through journals and disciplines may be more apparent than in older, more established topics that are less likely to be studied by a still broadening spectrum of subject areas. Systems biology was therefore hypothesized to be an illustrative example for exploring mechanisms for tracing and representing the evolution of a field through journals and their various disciplinary affiliations.

2.2. *Timeframe for analysis*

In the analyses pursued in this paper, concepts from systems biology are assumed to be transmitted by way of carriers (papers) that are linked via journals to various disciplines (Leydesdorff, Rotolo, & Nooy, 2013). Although the theoretical foundations of systems biology developed in part prior to the year 2000 (Mesarovic, 1968), the impetus for significant disciplinary progress followed the near completion of the Human Genome Project (HGP) at the turn of the 21st century (Lander et al., 2001; Venter et al., 2001). HGP is particularly noteworthy in this context, in that it was believed to provide scientists with the parts list of all elements of the human genome that could then be studied and modeled at the systems-level using high-throughput technologies (Ideker, Galitski, & Hood, 2001). In this regard, the data necessary to adopt a systems level perspective was not available for computational infrastructure development and research until 2000. The analyses pursued in this paper therefore include articles on systems biology published between 2000 and 2011, spanning the time when the concept first gained significant traction in the biosciences through the year of Thomson Reuter's recent publication of the *Journal Citation Report*.

In light of the emphasis on journals and subject categories as the elements of focus in these analyses, it is important to note that journals dedicated specifically to systems biology emerged only several years into the time span considered (2000–2011). A preliminary analysis of the journals most frequently represented in the publication data revealed that *Molecular Systems Biology* and *BMC Systems Biology*, established in 2005 and 2007 respectively, collectively comprise ten percent of all publications in the data set and fifteen percent of publications in five most recent years (2007–2011). These statistics provide an insightful backdrop for the subsequent analyses. In addition to underscoring the emergent post-genomic state of the field, the percentage of total publications appearing in these journals suggests that much of the progress in systems biology has been (and to some extent continues to be) published by outlets less specifically dedicated to this area. The role of these two journals in the network evolution overall is revisited later in the results interpretation.

2.3. *Subject categories as unit of analysis*

Whereas diffusion by citation is one common means of measuring the trajectory of knowledge, the act of publication itself and the new subject areas in which publication occurs is also indicative that the knowledge of interest has been obtained by or studied in a formerly unchartered territory of academic space (Liu & Rousseau, 2010). In other words, publication on a topic or idea by a given discipline can be considered indicative that the topic has indeed "spread" into that particular discipline or field. Additionally, when relatively recent timeframes are explored like that in this study, using a publication database can mitigate challenges of a citation database, as a publication can face significant delays between when it is published, cited, and then appears in a citation database (Bollen, Van de Sompel, Hagberg, Bettencourt et al., 2009). Hence, the subject categories of the journals publishing on a relatively new topic (in this case, systems biology) can be appropriate indicators of the breadth of disciplinary diffusion to date.

Given the present interest in describing and representing the evolution of interdisciplinary diffusion over time, a relational perspective among disciplines by way of journals is needed in order to conceptualize how ideas might move or spread to diverse fields. Network analysis therefore provides an appropriate analytic lens for considering relationships among disciplines by virtue of journals. Not only can a network perspective capture the complexity of the diffusion process, but it can also demonstrate the extent to which disciplines are interconnected through publications on the topic or field of interest. The following section describes the sequence of statistical and network analysis techniques used to first assess the overall structure of the diffusion, and to then more carefully consider the evolution of that structure and the various mechanisms by which journals facilitate interdisciplinarity.

3. Data & methods

Publication data was extracted using Thomson Reuter's ISI Web of Science (WoS) citation index. This index was selected based on its notable multidisciplinary coverage, including 12,000 of the highest impact journals worldwide and more than 250 disciplines. WoS is frequently used in bibliometric and scientometric inquiries due to its comprehensive cited reference search that provides inclusive databases for subsequent analyses of published work. For the present inquiry, the term "systems biology" was entered in the topic search field, generating all records containing "systems biology" in the title, abstract or keywords. The search was limited to the years 2000 through 2011, and yielded 4446 articles total. Complete output records of the articles were retrieved, including journal title, year of publication and journal subject categories.

3.1. Descriptive statistics

First, basic trends in the publication set were assessed over the course of the 12-year time span, including overall progress and expansion of disciplinary breadth. A consistent year-over-year increase in the number of publications on systems biology (Table 1) is evident. The increasing number of articles published each year can be described well as a linear function ($y = 94.08x - 241.0$, $R^2 = 0.950$). The cumulative number of publications in each year of the 12-year database can be described well as a power law ($y = 0.926 \times {}^{0.342}$, $R^2 = 0.977$). It should be noted that for these fitting curves, $x = 0$ theoretically refers to the year 2000, as this is the start ("time zero") for this inquiry; however, because the regression line would be negative for years 2000, 2001, and 2002 ($x = 0$, 1, and 2), I present these lines only as general trends to describe the data and not as true fits.

These results indicate very generally that increased progress in knowledge development has been made each year. Further, these results indeed suggest that the field of systems biology acquired presence as a focused topic for scientific research and academic study at the turn of the 21st century, confirming the year 2000 as an appropriate temporal search parameter for this inquiry.

A similar but less consistent trend was found in examining the number of subject categories represented by the journals of the publication set each year (Table 1). These results indicate that the disciplinary breadth of system biology's diffusion has in fact increased over time. This trend can be described well by a third order polynomial ($y = -0.176x^3 + 3.130x - 4.343$, $R^2 = 0.989$), though this trend is purely descriptive and carries no theoretical value. The overall growth in the number of categories by which the journals publishing on systems biology are classified is indicative of increasing disciplinary diversity (Rafols & Meyer, 2010). It is worth noting, however, that despite an increase in the number of categories represented, there is an uneven distribution of publications across these categories. A relatively large proportion of those categories include only 1 or 2 publications, as shown in Table 1. This distribution may be somewhat expected in emerging scientific fields, under the assumption that the new scientific concepts are slowly applied to distinct or more specialized fields within those in which they are originally conceived. Theoretical, foundational categories such as Mathematical Computational Biology occupy a large percentage of the total publications each year (10–18% between 2005 and 2011), whereas more specialized, applied

Table 1
Progress in systems biology from 2000 through 2011. This table includes the number of publications each year, cumulative number of publications each year, subject categories overall, and subject categories including only 1 or 2 publications.

Year	# of publications	Cumulative # of publications	# Subject categories represented by publications	#Subject categories representing 1 or 2 publications
2000	2	2	5	5
2001	4	6	6	5
2002	11	17	21	16
2003	68	85	44	29
2004	161	246	57	25
2005	250	496	81	44
2006	367	863	83	34
2007	444	1307	97	44
2008	564	1871	118	58
2009	729	2600	121	47
2010	906	3506	118	43
2011	940	4446	117	44

or conceptually distinct categories such as Tropical Medicine, Rheumatology, and Engineering Manufacturing include one or two publications and are not consistently represented in each year's publication set. Clinical categories such as Oncology, Immunology and Toxicology are first represented in the dataset a few years after 2000, and then steadily grow in the number of total publications each year, suggesting that the potential application of systems biology's theoretical basis to medicine is increasingly realized.

It is also worth noting the potential discrepancy in how less represented subject categories over time may signify diffusion phenomena. In some instances, these 1 or 2 publications represent the inception of systems biology into fields in which it was then more substantially studied in later years (e.g. Pathology, Statistics Probability). In other cases, these 1 or 2 publications represent less sustained or singular (perhaps even random) instances, such as those publications in journals with subject categories that are less obviously related to systems biology (e.g. Sociology, Anthropolgy). These instances, in which the diffused concept appears in journals with less overtly relevant subject categories, may also speak to a broader issue of JCR subject category accuracy or "correctness" that has been discussed in prior work (Jansenns, Zhang, Moor, & Glanzel, 2009; Rafols & Leydesdorff, 2009). In this sense, while some of these less represented categories may reflect initial diffusion, in other cases these categories may represent instances of weaker concordance between journals' fields and their JCR classifications. (In the bipartite network analysis that follows, these less relevant subject categories emerge as network isolates, which may also suggest the use of the presented methods in detecting those subject categories for which apparent diffusion may in fact be more of a random phenomenon due to JCR classifications).

Nonetheless, the overall increase in the number of categories indicates that systems biology's potentially broad applicability is under study to varying extents by an increasing number of disciplines. The apparent interdisciplinarity of the field to date lays the groundwork for investigating whether a more generalized pattern in this longitudinal, multidisciplinary trend can be realized to describe systems biology's diffusion. Specifically, to what extent are these disciplines interconnected, representing systems biology's diffusion as a coherent whole versus multiple, separate clusters of concentration? How might journals and their subject categories be used to conceptualize trends in disciplinary connectivity, and how might the evolutionary integration of these categories be modeled? I explore these questions in the subsequent analyses.

3.2. Bipartite network analysis

A bipartite network analysis was used to explore connectivity among subject categories by way of journals in order to better understand the year over year integration (or lack thereof) of different disciplines publishing on systems biology, and to examine topical changes in the structure of the publication database over time. A bipartite network (also referred to as a bimodal or two-mode network) is a network in which the nodes can be partitioned into two distinct sets of entities, A and B, such that each edge in the network connects a node from set A with a node from set B (nodes from set A are never directly connected to other nodes in set A, and nodes from set B are never directly connected to other nodes in set B). For example, one classic use of a bipartite network has been in describing individuals and their relationships to organizations, where individuals represent one set of nodes and the organizations represent another set of nodes. An edge exists between an individual and an organization if that individual is a member of that organization. Individuals are only indirectly connected to one another by way of a shared connection to an organization. Bipartite or two-mode networks can be useful as a first step in assessing the strength of relationship between two nodes of the same set by virtue of their shared relationship to a common node of the different set. In other words, bipartite networks can be helpful in understanding the extent to which nodes in set A are connected to one another via their common relationships to nodes in set B (Borgatti & Everett, 1997).

For the present analysis, relationships among subject categories by way of journals provide a means for understanding the extent to which journals are serving as bridges or connectors between disciplines, as well as which disciplines are more strongly connected to one another via the sources publishing on the topic or idea under study, in this case systems biology. Because a primary aim of this analysis is to assess the interdisciplinary evolution of a field over time, it is necessary to longitudinally evaluate the extent to which diverse disciplines publishing on systems biology become increasingly integrated with one another. Ties between subject categories and journals are inherent due to JCR journal classifications. The dynamic movement of systems biology across various fields is driven by new publications on this topic, which add new journals and new subject categories to this network over time. The diffusion patterns explored in these networks therefore enable exploration of the relative movement of systems biology across a continuously changing frame of reference.

Within the systems biology database extracted from WoS, journals' subject category assignments can represent the disciplines into which the field has diffused, such that journals are considered the carriers through which knowledge on the topic (systems biology) moves into new and diverse disciplines. As an initial step in understanding the interdisciplinary structure of systems biology's diffusion, bipartite (two-mode) networks were constructed using journals and their subject categories as the two distinct sets of nodes. Specifically, the networks were constructed as follows:

1. Ten subsets of data were extracted from the larger, 12-year database of 4446 publications for longitudinal comparison. The first subset included data related to all publications from years 2000 to 2002. (Data from the years 2000, 2001 and 2002 are combined for the purposes of this analysis, given the relatively small number of journals – 17 total – represented by publications during those years. Appendix A includes additional information specific to the bipartite networks of 2000–2001 publication data.) The second subset included data related to all publications from years 2000 to 2003; the third subset included data related to all publications from years 2000 to 2004, and so on, with the tenth subset including

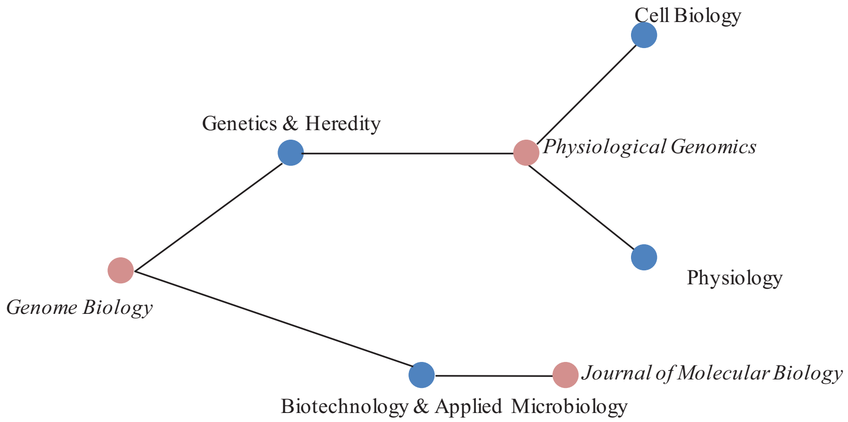

Fig. 1. Bipartite network construction. Schematic representation of bipartite network construction, using journals in a given subset of the extracted database of systems biology publications as one set of nodes, and the subject categories affiliated with those journals as the other set of nodes. In the diagram below, journals are depicted in *italics* and subject categories in non-italics. An edge exists between a subject category and a journal if the journal is affiliated with that subject category. The journal *Physiological Genomics*, for example, is affiliated with subject categories Cell Biology, Physiology, and Genomics & Heredity, and therefore has a distinct tie to each of those subject categories.

data related to all publications from years 2000 to 2011, i.e. all articles in the database. In this regard, the ten subsets of data included the cumulative publications on systems biology from the year 2000 through each year thereafter up to and including 2011.

2. For each of the ten subsets of data, a bipartite network was constructed with one set of nodes consisting of all journals represented by the publications in the given subset, and a second set of nodes consisting of all subject categories represented by the journals in the same subset. An edge exists between a given journal and subject category if that journal is affiliated with that subject category. Many journals are assigned multiple subject categories, denoting that their publication focus spans multiple disciplines. For this analysis, multiple subject categories were separated so that journals sharing one but not all subject categories could be accounted for as related by the shared subject category in the network analysis. For instance, a journal with the subject categories Immunology and Toxicology would have a distinct tie to each of the two categories. Fig. 1 provides a schematic representation of how these bipartite networks among journals and SCs were constructed and can be visually represented.

In order to understand the structure of disciplinary integration in this network, the betweenness centrality of subject categories and journals was compared among each year's bipartite network. Betweenness is a common centrality measure used in social network analysis as a means of determining the most influential people in the network by virtue of the number of shortest paths that pass between a node, also considered indicative of who has the greatest control over the flow of information (Newman, 2001). In theory, the higher the betweenness value of a given actor in the network, the more central they are in the network, the more information passes through them from other actors in the network, and the greater the distance between other actors in the network if that actor were removed. The betweenness index for n_i can be represented as: $C_B(n_i) = \Sigma g_{jk}(n_i)/g_{jk}$ where $g_{jk}(n_i)$ is the number of geodesics linking to actors that contain actor i (Wasserman & Faust, 1994).

Extending the concept of betweenness to the present inquiry, a higher betweenness centrality of a subject category can be considered indicative of its importance or influence in the network by virtue of its ability to integrate or connect other disciplines. A higher betweenness value of a subject category signifies its greater access to pathways among other subject categories, relative to the other subject categories in the network. While information flow in this network is not conceptualized with the same agency as it often is in social networks, a journal's ties to multiple subject categories can be thought of as knowledge pathways through which developing fields like systems biology may spread. The greater the betweenness centrality of a given subject category, the more efficiently systems biology may be able to "reach" other disciplines in the network by virtue of that subject category's connections to other SCs via journals. This theoretical basis has been applied in prior work using betweenness centrality as an indicator of interdisciplinarity (Leydesdorff & Rafols, 2011). In other informetric studies, the higher the betweenness centrality of a given actor in a co-authorship network, the more that individual is considered to be the "middleman" or connector among other individuals in the network (Otte & Rousseau, 2002). Applying this concept to the present inquiry, the higher the betweenness centrality of a subject category, the more that discipline functions as a connector or integrator among other disciplines in the network that are publishing on systems biology.

With regard to diffusion, betweenness centrality of subject categories in this type of bipartite network can suggest the role of that discipline in the diffusion process. If we consider that researchers of a particular discipline tend to refer to and publish in journals of a particular subject category, the extent to which that subject category is linked to other subject categories through sources (journals) publishing on systems biology may have implications for the likelihood that the

Table 2
Subject category-journal bipartite network characteristics by year. Bipartite network size is determined by the total number (sum) of nodes (SCs and journals) in each year's cumulative network (i.e. SCs and journals publishing on systems biology from the year 2000 through each subsequent year). "#SCs" refers to the number of nodes in the bipartite network that represent subject categories. "# Journals" refers to the number of nodes in the bipartite network that represent journals. The percentage of SCs and journals with betweenness centrality of zero is included to demonstrate the proportion of those SCs and journals located on the network periphery.

Database	Bipartite network size	# SCs	# Journals	% SCs with betweenness centrality 0	% Journals betweenness centrality 0	*M* SC betweenness	*M* Journal betweenness
2000–2002	25	18	17	0.412	0.611	0.036	0.039
2003	85	30	55	0.400	0.545	0.042	0.019
2004	176	43	133	0.465	0.579	0.042	0.010
2005	293	64	229	0.469	0.520	0.030	0.006
2006	432	73	359	0.425	0.529	0.027	0.004
2007	560	80	480	0.375	0.535	0.026	0.003
2008	722	95	627	0.389	0.547	0.022	0.002
2009	877	103	774	0.350	0.574	0.021	0.002
2010	1050	112	938	0.345	0.596	0.020	0.002
2011	1174	115	1059	0.322	0.608	0.020	0.001

concept is further diffused into diverse disciplines. Relationship patterns in this network can reveal dynamics of the diffusion process, including scientific progress or discoveries in certain areas that have implications for a range of other fields. Subject categories with relatively high and low betweenness values may be relevant in this analysis, as a high betweenness value can indicate a centrally important area of the field, while low betweenness values can indicate relatively new areas into which systems biology has only recently diffused, or that have not applied systems biology in a way that is particularly pertinent to other disciplines. Over time, an increasing betweenness value might suggest a subject category's increasing integration or importance as a discipline into which the principles of systems biology have been applied.

For this analysis, the normalized betweenness values of journals and SCs in each of the ten networks was calculated, adjusting for the maximum number of ties possible for each of the two modes. The results displayed in Table 2 report the size of the bipartite networks each year, the number of SCs and journals, separately, in the network each year (which sum to the total size of the bipartite network), the percentage of SCs and journals with betweenness centrality of 0, and the mean betweenness of SCs and journals each year. It is interesting to note that by 2011, the cumulative bipartite network forms a single component with the exception of seven journals and their affiliate SCs (nine total) that are not attached to the larger component (i.e. seven journals and their SCs are isolated from the large component). These SCs include disciplines less obviously related to systems biology such as Cultural Studies, Philosophy, Anthropology and Sociology, so it is not surprising that they are not connected to the larger component. The development of a single large component by 2011, however, signifies that a relatively coherent, integrated interdisciplinary structure emerges relatively soon (12 years) into the field's emergence. The data in Table 2 and the analyses that follow include all SCs and journals in the database, regardless of whether they are part of the emerging large component.

The results in Table 2 are further indicative of relationships between journals and SCs in this context. A steadier increase in the number of SCs, accompanied by a more pronounced increase in the number of journals, suggests that as systems biology is under study by new disciplines each year, an even greater number of journals affiliated with these SCs are publishing on this topic.

The betweenness values lend further insight into the evolution of the network structure. Nodes with a betweenness of zero are nodes that lie on the periphery of the network, and therefore have no centrality and do not function as links to more than one SC or journal. As is evident in Table 2, the percentage of SCs with a betweenness value of 0 decreases over time, indicating that SCs overall become increasingly integrated into the network each year; however, the percentage of journals with a betweenness value of 0 is maintained between 52 and 60% over the course of the time span. Additionally, while the mean betweenness for SCs and journals are fairly equivalent in 2002 and then both decrease over the time span, the mean journal betweenness decreases at a significantly greater rate than that of the SCs. As indicated in Fig. 2, the betweenness ranges for journals and SCs also begin as somewhat similar (a difference of 0.039 in 2002); over time, however, the journal betweenness range decreases dramatically while the SC betweenness range increases through 2005 and then begins to decline at a much steadier rate than the journals' betweenness range.

Collectively, the results of the betweenness analyses may reflect several phenomena taking place during system biology's interdisciplinary evolution. First, one might interpret the higher percentage of journals on the periphery as suggesting that multidisciplinary growth in the field is likely taking place at the periphery of the network, since in order for a journal to have a betweenness value of 0, it must be linked only to a single SC and have no other ties in the network. Comparing the higher percentage of journals with the lower percentage of SCs with a betweenness of 0 suggests that over time newer disciplines (SCs) are becoming increasingly central. This increasing centrality may be driven by the increasing number of journals publishing on systems biology that are affiliated with the newer, more peripheral SCs. As new disciplines are introduced to the network by virtue of affiliate journals publishing on systems biology, an increasing number of journals affiliated with that discipline also come to publish on systems biology over time.

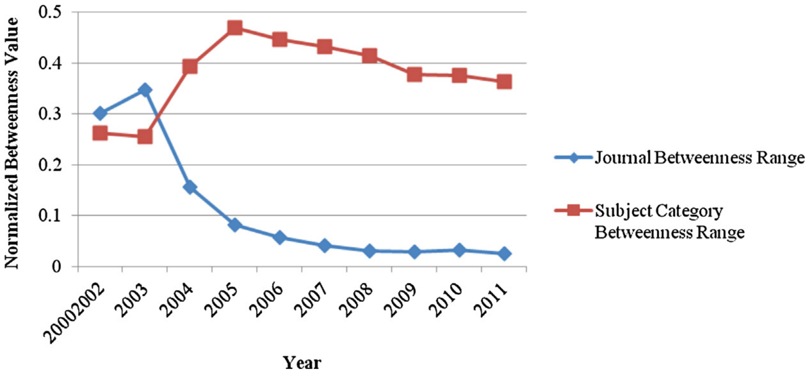

Fig. 2. Betweenness value range of journals and SCs each year.

This evolutionary pattern is further supported in comparing the mean and range of SC and journal betweenness over time. In a two-mode network, a given node's betweenness centrality increases to the extent that nodes of the other set are affiliated only with that node (Faust, 1997), and to the extent that it is serving as a bridge between other nodes of its same set. In other words, an SC's betweenness increases as an increasing number of journals affiliated exclusively with that SC publish on systems biology, and to the extent that it sits on the shortest path between other SCs. Therefore in the present context, the increase in SC betweenness range and decrease in journal betweenness range between 2002 and 2005 indicates that during this time period a greater number of journals affiliated with only 1 SC were publishing on systems biology. This pattern suggests that early in system biology's evolution, advancements in the field were focused within a more concentrated set of disciplines as the theoretical foundation developed. The subsequent steady decrease in SC betweenness range after 2005 suggests that the field indeed becomes more interdisciplinary as a result of fewer journals having exclusive affiliation with a single SC, and therefore linking distinct disciplines more frequently. Put differently, between 2005 and 2011, more journals publishing on systems biology have an interdisciplinary focus (i.e. more journals were affiliated with multiple SCs).

Nonetheless, the decrease in SC betweenness range is fairly steady during this time period, indicating that as systems biology becomes more interdisciplinary, there remains substantial publication activity in journals affiliated with a single SC. One can postulate that these journals are likely those initially responsible for the increase in SC betweenness prior to 2005 and are likely affiliated with some of the core, theoretical foundations of systems biology (e.g. Computational Biology, Mathematics, Molecular Biology, etc.) that newer developments and practical applications are based upon. It is also reasonable to assume that the journals and disciplines in which early publications appear continue to publish on and advance that topic over time.

The results of the bipartite network analysis collectively suggest that the scholarly structure of systems biology may be generally characterized as a core–periphery configuration of journals and subject categories (Borgatti & Everett, 1999). In this context, the disciplinary core can be broadly considered as those subject categories that maintain presence in the network since 2000 and that maintain a relatively high normalized betweenness values each year despite growth in the network size. These include Biotechnology & Applied Microbiology, Biochemistry & Molecular Biology, Computer Science, Mathematical & Computational Biology, Biophysics, and Genetic & Heredity. Progress occurring at the core is evident through an increasing number of journals related to these theoretical, fundamental areas of systems biology, which are also those fields responsible for the initial research and establishment of the field in the late 1990s and early 2000s. Initial growth followed by relatively steady decline in the mean and range of SC betweenness values also indicates that these core SCs maintain somewhat central positions in the diffusion landscape. Growth at the periphery appears to occur by way of interdisciplinary journals (those with multiple affiliate subject categories) that introduce and connect more specialized areas to the network (e.g. Pediatrics) or those less overtly relevant to applied systems biology (e.g. Geography) that have betweenness of or close to zero.

In addition to the core and periphery positions, some subject categories appear to hold more intermediary places in this network, which can be defined in this context as those subject categories that are consistently present in the network since their introduction sometime after 2000, maintain a non-zero betweenness value (i.e. they are more integrated in the network than peripheral SCs), but are not part of the core, foundational areas. These SCs are positioned between peripheral and core nodes. This position can be considered analogous to what has been termed the "semi-periphery" in world systems theory and international relations, whereby the actors (countries) or nodes between the core and peripheral positions play a critical role in mediating relations between actors or nodes in the core and peripheral regions of the network (Wallerstein, 2004). A review of the SCs that fit this intermediary or semi-periphery description based on the bipartite network analysis reveals that many of these disciplines represent frontiers in the practical application of systems biology, particularly in those areas intended to advance human health. Intermediary subject categories such as Pharmacology & Pharmacy, Toxicology, Immunology, Nutrition & Dietetics, and Neurosciences & Neurology signify that principles of systems biology and knowledge derived from its infrastructure has been effectively translated to the study of medical prevention and treatment. Increasing betweenness values of some intermediary categories over time, for example Oncology, further suggests that the realization of systems biology's applicability is effectively diffusing across a range of publication outlets in certain disciplines, and that those disciplines are becoming increasingly embedded into the network.

The core–intermediary–periphery configuration that emerges from the bipartite network analysis lends important insight into the structure of the content areas that represent system biology's diffusion by way of journals that span multiple subject categories. A central group of strongly interacting subject categories seem to constitute a core, while peripheral subject categories appear to be added to the network each year by way of journals that connect to existing categories and also introduce new subject categories to the network. Indeed, core–periphery structures are not new to science, and have been realized in modeling patterns of international relations and scientific cooperation (Hwang, 2008; Wagner & Leydesdorff, 2005). Whereas core–periphery structures commonly define peripheral nodes as interacting mainly with the core and only minimally with one another (Borgatti & Everett, 1999), the systems biology case appears to also include what could be considered intermediary, or semi-periphery nodes – journals that interact somewhat with one another by way of common subject categories, but that also bridge peripheral categories with those at the core. This type of bilateral region of a core–periphery structure has also been thought to exist and play an important role in defining the core–periphery evolution of global scientific collaboration (Zelino, 2012).

As this paper intends to consider the roles of different journals as knowledge carriers across time and academic space, the core–intermediary–periphery structure considered here becomes the focus of the subsequent analysis, in which journals – as the disciplinary integrators – are explored in greater depth. The following analysis adds insight into how one might model what constitutes the core, intermediary and periphery over time, and how one might characterize the types of journals that drive temporal changes in this structure as system biology diffuses into new disciplines.

3.3. Subject category co-occurrence analysis

Trends in betweenness values observed in the bipartite network analysis suggests that systems biology's diffusion permeates from a core of highly connected subject categories into more specialized peripheral fields. Whereas network structure can be generally inferred based on relative betweenness values of SCs, more formally defining the structural boundaries of the core, intermediary and periphery can allow for parameters within which positional movement can be compared and assessed over time. Without defining structural bounds of the landscape, it becomes difficult to assess the meaning of a given SC's centrality in the context of the field's interdisciplinary growth.

Inquiry into how journals facilitate the development of a core–intermediary–periphery pattern from year to year may also shed light on how interdisciplinary diffusion occurs by way of different journals. By exploring how journals may operate as different types of integrators within this structure over time and in the context of all prior publications on systems biology, distinctions among journals as diffusion carriers can be made and the evolution of systems biology's landscape better understood. This type of distinction differs from traditional bibliometric indicators often used to assess or draw conclusions about a journal's influence such as impact factor (Garfield, 1972). Prior work has evaluated and compared other measures of scientific impact, including both traditional measures such as impact factor, as well as network measures such as betweenness and degree centrality of citation- and usage-derived networks (Bollen, Van de Sompel, Hagberg, & Chute, 2009). This work has indicated that scientific impact is indeed a multi-dimensional construct, assessable through various lenses including network analysis centrality measures. The analysis explored here offers another network perspective by which the multi-dimensional construct of scientific impact can be considered.

The present analysis uses a network approach to assess impact from the perspective interdisciplinary diffusion over time, and the extent to which a given journal contributes to the development of the particular diffusion structure. The influence of interest in this case is the integration of diverse disciplines rather than citation efficiency. Whereas prior work has developed diffusion factors to quantify the disciplinary breadth of a given journal (Frandsen et al., 2006), this analysis attempts to add qualitative description to the function of a given journal in the context of the diffusion of a specific idea or topic and within given temporal and structural parameters.

The subject category co-occurrence analysis now described is used to illustrate a means of more formally differentiating subject categories in the core, intermediary and periphery positions of this SC network structure, creating parameters for comparing SC position each year. In this context, co-occurrence refers to the frequency with which two subject categories are paired together in association with one or more journals. For example, if a given journal in the dataset has SCs "X" and "Y" and includes 6 publications in the dataset, and another journal with 4 publications has SCs "X", "Y" and "Z", X and Y would co-occur 10 times in the network.

In theory, subject categories at the core are those with the oldest and strongest relationships. These categories were expected to be those related to the theoretical foundation of systems biology, including those disciplines related to its analytic infrastructure and conceptual roots in genetics and "omics". Newer disciplines at the periphery were expected to be more specialized applications of concepts at the core. Distance from the core was therefore hypothesized to be a function of both age in the network (newer disciplines farther from the core) and relative content specificity (more specific applications of foundational disciplines farther from the core).

Given that a core is typically considered a group of highly interacting nodes of a network (Borgatti & Everett, 1999), agglomerative hierarchical clustering of the SCs based on their co-occurrence in the network provides a means of determining the range of interaction among SCs each year. Johnson's single-link hierarchical clustering method was performed, such that

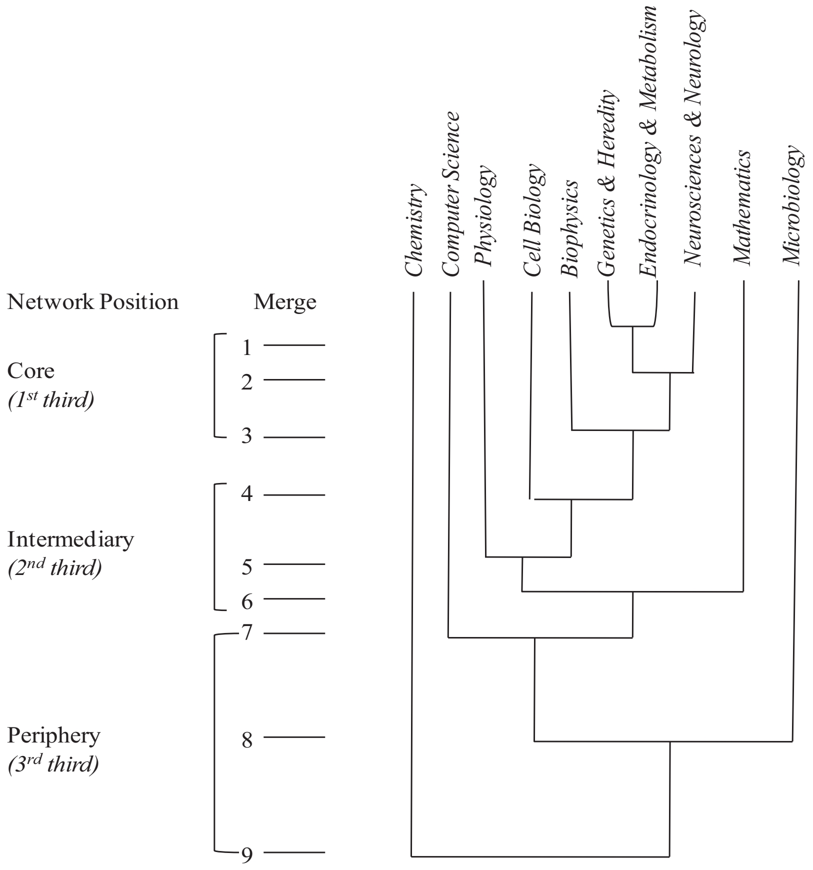

Fig. 3. Schematic representation of the hierarchical cluster analysis of the SC co-occurrence networks. Each line of the diagram represents an SC. SCs in the first third of the merges (partitions) constitute the core (highest frequency of co-occurrence), the second third constitute the intermediary, and the final third constitute the periphery. In this illustration, *Genetics & Heredity*, *Endocrinology & Metabolism*, *Neurosciences & Neurology*, and *Biophysics* would constitute the core, merging in the first third; *Cell Biology*, *Physiology*, and *Mathematics* would constitute the intermediary, merging in the second third; *Computer Science*, *Microbiology*, and *Chemistry* would constitute the periphery, merging in the final third.

those SCs that co-occur most frequently are considered as having the greatest similarity or strongest interaction and therefore merge earliest in the cluster analysis (Johnson, 1967).

As the clustering progresses and other SCs merge with those more frequently co-occurring, those SCs that co-occur least with others in the network will merge last. The partitions or merges of the cluster analysis can therefore provide markers by which the co-occurrence, or strength of interaction between the SCs can be determined. Those SCs that merge first are those with the strongest relationships; strength of relationship is progressively weaker as the clustering progresses. For the purposes of defining structural parameters in the network, those SCs that merge in the first third of the clustering can be considered as constituting the “core”, the next third the “intermediary”, and the final third the “periphery”. The schematic representation shown in Fig. 3 demonstrates how subject category position in the network was delineated based on hierarchical clustering of SC co-occurrence.

In order to track SC movement within these three positions over time, hierarchical cluster analysis was conducted for each of the ten co-occurrence networks. Like the bipartite network analyses, these analyses used cumulative subject category co-occurrence networks for the years 2000 through 2011. (The 2004 network, for example, included all publications from 2004 and the years prior.) The hierarchical clustering of each of the ten co-occurrence networks was then divided into thirds based on merges (partitions) to identify the SCs in each network as positioned in the core, intermediary or periphery. This method of parsing the co-occurrence hierarchy defined structural boundaries against which an SC’s position could be determined and compared across time in the co-occurrence networks.

A database (Appendix B) was created to score each subject category based on its age, original position in the network, and any movement closer to or farther from the core. Movement closer to or father from the core was determined by whether a given SC changed position in the core–intermediary–periphery delineation of the cluster analysis from year to year. For example, an SC positioned in the periphery 1 year and in the intermediary the next year would be considered as having moved closer to the core, whereas an SC positioned in the core or intermediary 1 year and the periphery the next year would be considered as having moved farther from the core. In this context, a category considered “young” or “old” refers to age in

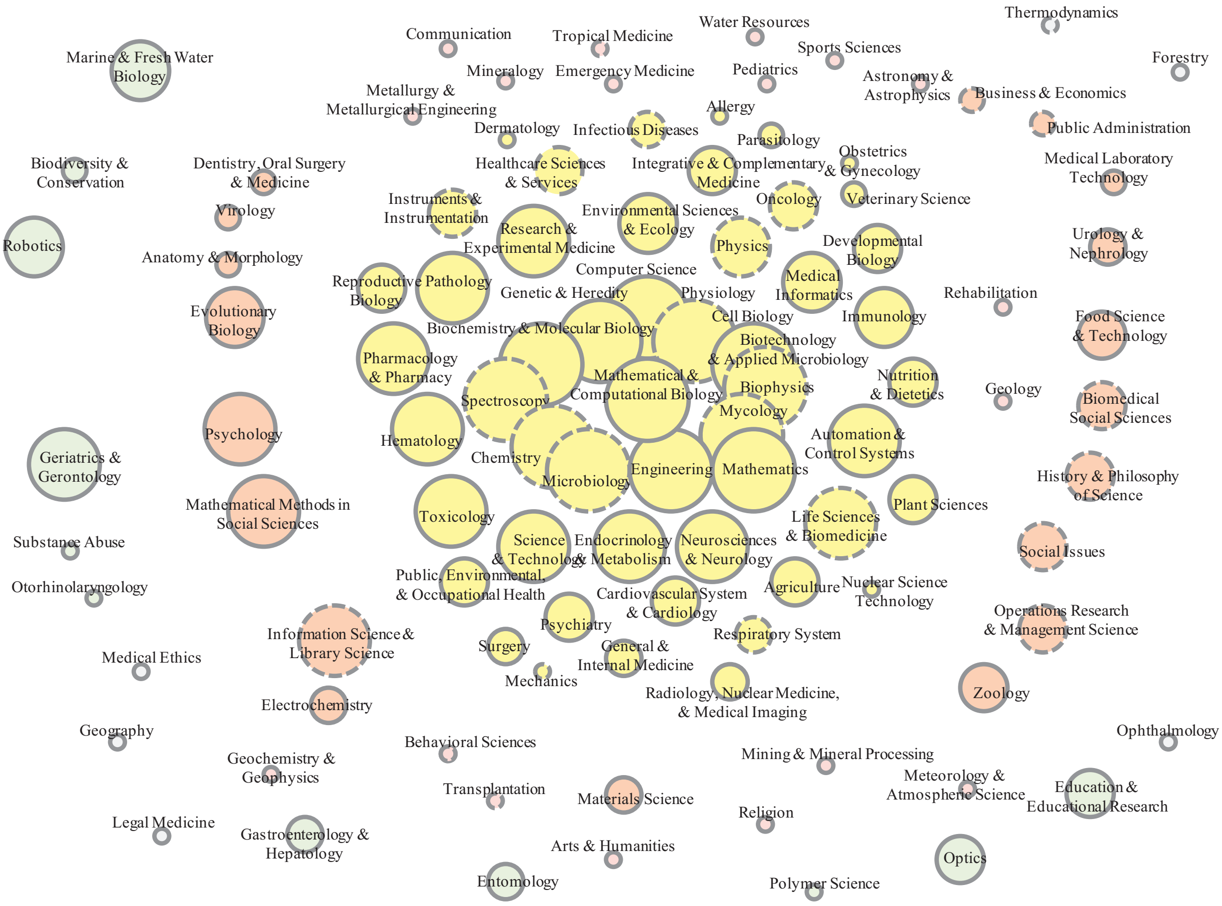

Fig. 4. Subject category positions as of 2011. Network positions are based on the position of each category after the hierarchical cluster analysis of the 2000–2011 co-occurrence network. Yellow categories constitute the core, orange the intermediary and gray the periphery. Circle size indicates the category's score in 2011, which accounts for both age in the network and movement toward and/or away from the core throughout its existence in the network. Larger circles represent higher scores and smaller circles represent lower scores. Circles with dotted borders signify categories that changed network position at least twice since being introduced into the network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the network only, not age as a discipline. Subject categories that first appeared in the network in between 2000 and 2002 began with a score of 10 (data from 2000 to 2002 was again combined for this analysis). Each subject category introduced every year thereafter began with a lower score, i.e. 9 in 2003, 8 in 2004, etc. Original positions for each category were noted as "c" (core), "i" (intermediary) or "p" (periphery). For each year that a given category moved a position closer to the core, a 1 was added to the original score; for each year that a category moved a position farther from the core a 1 was subtracted from the original score. For categories that maintained position in the structure, the score was not changed.

Fig. 4 is a visualization of the results of this analysis, indicating both the final score (size) and position (color) of each subject category as of 2011. This representation shows the disciplinary landscape of system biology's diffusion to date, using the core–intermediary–periphery structure as a lens for conceptualizing the field's interdisciplinary evolution over the 12-year period.

Using the sequence of methods described, subject categories that move from one position to another can be traced to specific journals that may drive this movement to different structural positions in the network. Because this analysis uses co-occurrence to determine ties between SCs, an increasingly strong relationship between two categories can only be attributed to an increasing number of publications on systems biology in journals with those affiliate SCs. In order to understand which journals may be responsible for a given SC's movement closer to the core in a given year for example, one can refer back to the publication database to see which journals introduced to the network that year are affiliated with that SC. By cross-referencing movement of specific SCs from year to year with journals assigned to those SCs and the years in which they publish- information found in the originally extracted WoS database – we can begin to understand which journals facilitate or hinder diffusion to diverse disciplines within this core–intermediary–periphery structure. In the following section (Section 4.2), the results of the systems biology example are used to illustrate the types of inferences that can be made about specific journals' functions in this structure, cross-referencing the results of the analysis with the publication database to describe this field's interdisciplinary diffusion in the context of the landscape mapped in these analyses.

4. Results interpretation

4.1. Bipartite network analysis results interpretation

Interpretation of the bipartite network analysis can be used to confirm or enhance what is already known or predicted about systems biology as a field. For example, increasing betweenness values in the core subject areas, suggestive of ongoing advancements, are not surprising. The models and methods needed to manage systems-level analyses continue to be refined and expanded. The only moderate decrease in mean and range of subject category betweenness over time in light of the increasing number of SCs added to the network each year may also be reflective of the cyclical diffusion pattern advocated by Aderem (2005) and other leaders of the Institute for Systems Biology.[1] Aderem predicted that ongoing, systems-level discoveries in biology will dictate the necessity for new technological and computational tools; the advent of new tools and methods will in turn facilitate biological breakthroughs and application to new areas and disciplines. As the betweenness centrality of core subject categories remains the highest of all other SCs in the network, the number of journals also increases. This pattern suggests that an increasing number of journals in the dataset (i.e. an increasing number of publications) are tied to the core subject categories, either exclusively or through their ties to multiple subject categories, hence continuing to sustain a higher betweenness of those foundational, theoretical disciplines while the network as a whole also continues to expand.

It is quite possible that these distinct patterns of growth at the core and periphery are the result of increasing efforts to develop technological and computational capabilities (core) in response to new questions, concepts and data produced by experimental and applied research. As theory from systems biology is applied to better understanding different diseases, novel computational and graphical tools are needed to integrate relevant information and elucidate emergent system behavior. For example, in applying systems biology concepts to better understanding prostate cancer development, the need to develop a comprehensive mRNA database became evident, as did the ability for tools such as mass spectrometry to be incorporated into biomarker detection for various stages of this disease (Hood, Heath, Phelps, & Lin, 2004). Generally speaking, as systems biology theory is applied in order to better understand certain diseases and conditions, measurement and detection techniques must be developed specifically to model and estimate the pertinent biological parameters that emerge from this research.

It is this type of positive feedback loop between theory and application that will further enhance the ability to realize a predictive and preventative field of medicine. This pattern is also reflected in the diffusion studied in this bibliometric analysis. Ongoing growth at the core strongly suggests that systems biology is still very active in terms of technology and computational development, with much potential and specificity in applied areas yet to be discovered.

Nonetheless, movement of several clinical disciplines toward the core over time, such as Immunology, Healthcare Sciences & Services and Oncology, underscore the progress that has been made in translating systems biology theory to practice. A review of the younger clinical categories represented in Fig. 4 include some more specific clinical areas such as Allergy, Infectious Diseases and Parasitology. It is possible, consistent with the core–intermediary–periphery diffusion pattern evident to date, that intermediary subject categories also represent the broader clinical categories within which more specific clinical application is beginning to take place. Allergy, for example, represents a sub-field of Immunology. One can postulate that the application of systems biology theory to the study of immunology has generated sufficient insight for the appropriate parameters to now allow for a focused study on application to allergy. Increasing betweenness of disciplines related to medicine and human health is reflective of their progressive integration as sub-disciplines within systems biology. Systems pharmacology and systems medicine, for example, are emerging fields that are realizing the advantages of adopting a systems approach in targeted selection for intervention. The ability to model genetic and cellular regulatory networks under different conditions of perturbation has provided the foundation for relating the structure and function of the human genome to health and disease states, illuminating network-level drug targets versus the traditional molecular-level targets (Auffrey, Chen, & Hood, 2009).

Some systems biologists envision a medical paradigm shift, in which medicine and healthcare eventually become increasingly predictive, preventive and personalized as systems-level inquiries provide deeper insight into the complex pathways and dynamics involved in disease processes (Hood et al., 2004). Clinical application of systems medicine has already suggested that such an approach provides far more individualized, synergistic and spatially and temporally sensitive treatment than traditional reductionist approaches, specifically for complex diseases like Diabetes (Ahn, Tewari, Poon, & Phillps, 2006). Over time, it is possible that systems medicine and systems pharmacology gain sufficient prominence as separate fields to become a distinct core with strong ties to this network core.

Perhaps an exemplar of such efforts to bridge theory and practice is the creation of the human diseasome, a network of disorders and disease genes that has engendered a platform for further exploring the complexity, commonalities and interconnectedness among disorders and associated genes (Goh et al., 2007). While the diseasome map represents a significant milestone in the application of foundational systems biology concepts to the study of medicine, it simultaneously marks an expansion to the theory of using networks to explore the link between genotype and phenotype. The human diseasome

---

[1] www.systemsbiology.org (Institute for Systems Biology).

continues to further evolve in its application as a tool for understanding patterns and relationships among diseases and genetics, signifying an intersection of theory and application in system biology's diffusion.

4.2. Subject category co-occurrence analysis results interpretation

At the core of the network structure represented in Fig. 4, the earliest and strongest tie emerges between Biotechnology & Applied Microbiology and Genetics & Heredity. This relationship can be traced to journals including *Genome Biology, Genetic Engineering News*, *BMC Genomics*, *Genome Biology*, and *Omics – A Journal of Integrative Biology*, in which systems biology work was published early on in the time frame examined. More recently, publications in *Disease Markers* has strengthened the tie between Biotechnology & Applied Microbiology and Genetics & Heredity while also relating more applied disciplines including Toxicology, Pathology, and Research & Experimental Medicine. Other consistently present strong ties include those between Biochemistry & Molecular Biology, Mathematical & Computational Biology, Computer Science and Mathematics. Much of the intersection between these disciplines can be attributed to journals dedicated to bioinformatics, specifically *BMC Bioinformatics* and *Journal of Computational Biology*. These strong disciplinary relationships align with the interdisciplinary foundation that has allowed systems biology research to flourish, and its continued growth suggests that this foundation is still in a period of progress and ongoing development.

Whereas the visual representation of this analysis indicates that the infrastructure-related disciplines exist at the center of the core, those disciplines also at the core but slightly younger (smaller circles) suggest promising directions in which systems biology has been successfully applied. Toxicology, Parasitology, Pharmacology & Pharmacy, Neurosciences & Neurology and Immunology, for example, represent disciplines in which the conceptual foundations or infrastructure of systems biology is more purposefully used. In the bipartite network analysis, these categories could be considered intermediary based on their betweenness values. Although the co-occurrence analysis positions these categories as part of the core, the journals with which they are affiliated can have intermediary roles in the network. The roles of the journals associated with these younger, core subject categories and similar others (6- to 8-year old at the core) can be hypothesized as existing within two broad functional categories.

The first category includes those journals that bridge the oldest disciplines of the core with more applied areas. In the case of Toxicology, for example, these journals include *Mutation Research, Biomarkers* and others whose subject categories consist of more foundational disciplines like Biotechnology & Applied Microbiology in addition to Toxicology. A review of these publications' abstracts indicates that these articles introduce the application of systems biology concepts to model phenomena related to toxicology, or to study gene–environment interaction or individual response to certain proteins at the systems level. In another example, Pharmacology & Pharmacy is represented by journals such as *Combinatorial Chemistry and High Throughput Sequencing* and *ChemBioChem.* As with the prior example, these journals' subject categories include older core categories such as Biochemistry & Molecular Biology and Chemistry, as well as Pharmacology & Pharmacy. These publications broadly consider building cellular models of disease and demonstrating the potential of systems views to provide insights for drug discovery.

The second category includes journals that bridge subject categories at the core that are slightly younger than the central, foundational categories with those even younger at the core or in the intermediary. These relationships tend to represent even more specifically applied discoveries in systems biology. For example, the 14 publications from 2007 through 2011 in *Pharmacopsychiatry* consider the insight derived from a systems level understanding of cellular processes in mental disorders, and the implications for diagnosis, treatment and drug development. Here again, the application of systems biology is still explicit, yet occurs in a more focused and specific context. Psychiatry's presence in the core is likely due to the fact that *Pharmacopsychiatry*'s subject categories include Psychiatry and Pharmacology & Pharmacy. Because Pharmacology & Pharmacy is a more centrally connected and older discipline in the network, the tie between Pharmacology & Pharmacy and Psychiatry that this journal creates is hypothesized as facilitating Psychiatry's position in the core. In this case, a less central SC joins the core by virtue of its co-occurrence with a more central and embedded discipline.

Younger disciplines in the intermediary and periphery (smallest circles) tend to include publications in journals that represent multiple, more specific applications of systems biology, and for which there is a small number of articles (3 or fewer) in the dataset. Tropical Medicine, for example, is represented by a small circle in the intermediary with a dotted outline, indicating that it was in the periphery at least one time in its 2-year existence in the network. A review of those journals in the database with subject category Tropical Medicine reveals that its current state in the intermediary of the network was likely driven by its inclusion as a category with Infectious Disease and Parasitology in the journal *PLOS Neglected Tropical Diseases.* Only 2 articles of the 4446 included in the database were published in this journal. This small proportion, combined with the multidisciplinary nature of *PLOS Neglected Tropical Diseases*, suggests that Tropical Medicine's position as an intermediary discipline is by virtue of its ties to two other subject categories that are more prominently integrated into the network. Infectious Disease and Parasitology are not only older disciplines in the network (6 and 5, respectively), but are also associated with a greater number of journals in the network, some of which have affiliations with older subject categories at the core. The intermediary position of some newer subject categories may therefore be the result of stronger second or third degrees ties that to a certain extent are "pulling" a subject category toward the core even though it is relatively weakly represented in the dataset as a whole.

In a similar example, Mechanics is a younger discipline (4 years in the network) that began in the periphery in 2008, advanced to the intermediary in 2009, and to the core in 2011; nonetheless, Mechanics appears as a subject category affiliated with only 7 articles in the database. In considering what might account for this relatively quick and recent transition, those journals affiliated with Mechanics such as *Advances in Applied Mathematics and Mechanics* and *International Journal of Nonlinear Sciences and Numerical Simulation* are also affiliated with well-integrated, older categories such as Mathematics, Engineering and Physics. In this regard, a relatively underrepresented category in the publication dataset as a whole can become part of the core based on its co-affiliation with other categories that comprise the conceptual basis of systems biology.

Finally, addressing the journals that constitute ties to categories in the periphery can provide insight into why certain categories remain farther from the core despite their age in the network. Geriatrics & Gerontology, for example, is one of the older disciplines in the network but remains in the periphery throughout its entire 9 years. A review of those journals affiliated with this discipline reveal that its initial tie to the subject category network in 2003 occurred by way of publication in *Mechanisms of Aging and Development*, with Cell Biology and Geriatrics & Gerontology as subject category affiliations. After 2003, publications applying systems biology to geriatrics appeared in a number of journals with Geriatrics & Gerontology as a subject category (*Age*, *Experimental Gerontology*, *Biogeronotology*, *Journals of Gerontology Series*). These journals, however, are not assigned any additional subject categories other than Geriatrics & Gerontology, suggesting that the absence of ties to other subject categories beyond the single tie to Cell Biology limits Geriatrics & Gerontology from advancing toward the core over time.

5. Discussion & conclusion

This paper suggests that when studying diffusion of emerging scientific fields through bibliometrics, identifying a structural pattern of the field's growth and then examining movement within that structure can enhance our conceptualization of and ability to describe interdisciplinary diffusion. As the example of systems biology demonstrates, journals as publication carriers and their specific SC affiliations provide a useful lens for examining the spread or application of concepts to diverse disciplines, and the extent to which journals publishing on the topic of interest act as integrators (or not) of those disciplines within the context of the diffusion structure more broadly. Whereas the use of betweenness centrality in the bipartite network analysis provided insight into the general diffusion pattern, the co-occurrence networks and subsequent cluster analysis allowed for identification of structural boundaries across which changes in disciplinary position can be tracked over time and interpreted in conjunction with the publication database.

With regard to diffusion of emerging fields, these results suggest that more specialized disciplines may be exposed to new knowledge by way of those journals whose focus spans the seminal foundation of the field as well as the more specific or exploratory applications. As new subject categories are introduced to the field, they may be more likely to become part of the core over time if they have a second or third tie to subject categories closer to the core center. Subject categories that lack at least two first or second degree ties to the foundational core disciplines are more likely to remain in intermediary or peripheral positions over time.

A notable caveat to this approach is the potential for journals with multiple classifications to assume a seemingly influential role in the diffusion network, despite having some categories that are only weakly related to categories in which the topic of focus has been more prominently studied. For instance, Psychiatry is a subject category that appears in the core in Fig. 4, however this field is more weakly related to systems biology. *Pharmacopsychiatry* includes both Psychiatry and Pharmacology & Pharmacy as subject categories, the later of which emerges as a prominent application of systems biology. As a result, journals with only Psychiatry as a subject category such as *Schizophrenia Research* may be thought of as highly influential in the network since its classification appears at the core, even though this journal includes only one publication on systems biology. One should therefore exercise caution when using this approach to draw conclusions about the influence of a particular journal in the diffusion network, and consider which specific categories of a multidisciplinary journal may be most significantly responsible for the journal's role in the broader diffusion landscape. Future studies may also refine the approach presented in this paper in order to account for those journals that may seem to exert more influence on the network due to association with categories that are more prominent core areas.

While journals with multiple subject categories are clearly important for disciplinary integration, many categories in the present example such as Geriatrics & Gerontology are more heavily represented in the database by journals with only one subject category assignment. These categories not only exist at the periphery and intermediary, but also at the core. In fact, *Molecular Systems Biology*, *BMC Systems Biology*, and *PLOS One* – the three journals most frequently represented in the database – are assigned only to Biochemistry & Molecular Biology, Mathematical & Computational Biology, and Science & Technology, respectively. Although these journals certainly contribute to the overall number of publications on systems biology, they are not functioning as categorical integrators through the analysis presented here. It is possible, however, that given the explicit focus of *Molecular Systems Biology* and *BMC Systems Biology* on systems biology, and given the broad disciplinary diffusion of systems biology explored in this paper, these journals may be purposefully interdisciplinary in their aims and scope. It is possible that these journals were established to some extent in response to the disciplinary breadth of system biology's diffusion. In other words, these journals may serve as outlets for a range of disciplines adopting tenets of systems biology, capturing the spectrum of subjects in the network. Future research could test this hypothesis through a content analysis of the publication abstracts in these journals, or through analyzing the diversity of journals cited by these publications. It is also likely that the frequent presence of these journals in the data set and their exclusive affiliation with core subject categories are in large part responsible for the relatively higher betweenness values of those core categories observed in the longitudinal bipartite network analysis.

In sum, these results suggest that subject category position in the core, intermediary or periphery of the network is not only a function of age and content specificity. Journal disciplinary breadth is an important factor in determining whether seemingly specific and/or younger applications of systems biology are incorporated into the core over time. For systems biology, journal disciplinary breadth is an important indicator of the extent to which the theoretical basis is successfully applied in more practical areas. It is possible that the sequence of methods explored in this paper may be applicable to assessing the extent to which theory bridges with practice in other fields for which this type of diffusion pattern is an outcome of interest.

A typology of journal roles in these network dynamics may be useful in applying the hypotheses derived from this analysis to characterize a journal's function at a given point in a diffusion process. Some categorizations and definitions might include:

Core Bridges: Journals with ties connecting subject categories at the center of the core with subject categories within the core but not at the center. Core Bridges tend to connect the older, foundational disciplines to those slightly younger areas that have begun to broadly explore application of the new field. These journals may also function as bridges between theory and the frontier of practical application in a given field. For systems biology, these journals connect SCs such as Mathematics and Computational Biology with more clinically relevant disciplines such as Oncology and Immunology.

Intermediary Bridges: Journals with ties connecting subject categories at the core (not at the center) with subject categories at the edge of the core, in the intermediary or in the periphery. Intermediary Bridges tend to connect relatively integrated disciplines close to the core center with younger and more specialized disciplines. These journals may function as bridges between broader practical application and more specific sub-fields of inquiry. For systems biology, these journals connect more general clinical fields such as Oncology with more focused application to specific populations (e.g. Pediatrics) or sub-fields (e.g. Urology & Nephrology).

Reinforcers: Journals with single subject category affiliations that strengthen that subject category's overall presence in the network but do not serve any integrative purpose. Reinforces may exist at the core, intermediary or periphery, and function largely as augmenting the more exclusive study of a particular discipline and its relevance to the topic or field of interest.

These categorizations are only broad suggestions based on the results of this particular analysis. Refinement and new definitions may be appropriate in the future depending on the nature of the field's evolution and the role of journals in governing the structure and dynamics of interdisciplinarity. Additionally, as more subject categories move into the core using the method presented, the parameters for defining what constitutes the core, intermediary and periphery may need to be adjusted. Whereas in 2011 systems biology was still diffusing into unchartered areas, at a certain point in time the disciplinary breadth of system biology's diffusion will expectedly ebb, since there are a finite number of fields that can potentially incorporate knowledge related to systems biology in a sustained and meaningful way.

Adjustment of core–intermediary–periphery boundary identification may take into account the number and frequency of journals serving as "reinforces" of a particular subject category, in conjunction with the frequency of publications in journals that facilitate a subject category's movement toward the core.

Taken together, these results suggest the value of using a combination of non-traditional indicators in certain bibliometric inquiries, particularly when a functional measure of influence is useful to assess the outcome of interest. The bipartite analysis sheds light on the overall structure of the diffusion, including areas of more established adoption and two growth trajectories (at the core and periphery), while the subject category co-occurrence analysis allows for more specific hypotheses to be made regarding the role of journals in driving integration among disciplines in different positions of the network. The implication for diffusion to more specialized areas is that journals with more disciplinary breadth may be considered mediators in the diffusion process. These results may have implications for researchers seeking to broaden the disciplinary exposure of their work, or to contribute to interdisciplinary diffusion of a topic or field.

Future research may use similar combinations of bibliographic perspectives to examine the structure and dynamics of other emerging scientific fields. For example, it may be interesting to review the emergence of nanoscience as a scientific discipline in the 1980s to see whether similar patterns are evident through network analyses that mirror the findings presented here. Like systems biology, nanoscience introduced a new scientific perspective born out of a need for expanding instrumentation and technological capabilities. These types of inquiries may advance our knowledge of how new fields with broad potential applicability tend to diffuse into other more practical, applied areas, especially in understanding mechanisms that may facilitate or hinder interdisciplinary diffusion and uptake.

Acknowledgements

The author thanks the two anonymous reviewers for their suggestions to improve this article, as well as Matthew E. Brashears and Konstantin Cesnulevicius for their comments and guidance on prior versions of this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.joi.2013.10.009.

References

Aderem, A. (2005). Systems biology: Its practice and challenges. *Cell*, *121*, 511–513.

Ahn, A. C., Tewari, M., Poon, C. S., & Phillips, R. S. (2006). The clinical applications of a systems approach. *PLoS Medicine*, *3*(7)

Auffrey, C., Chen, C., & Hood, L. (2009). Systems medicine: The future of medical genomics and healthcare. *Genome Medicine*, *1*, 2.

Bollen, J., Van de Sompel, H., Hagberg, A., Bettencourt, L., Chute, R., Rodriguez, M. A., et al. (2009). Clickstream data yields high-resolution maps of science. *PLoS ONE*, *4*, e4803.

Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE*, *4*, e6022.

Boner, K., Penumarthy, S., Meiss, M., & Ke, W. (2006). Mapping the diffusion of scholarly knowledge among major U.S. research institutions. *Scientometrics*, *68*, 415–426.

Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks*, *19*, 243–269.

Borgatti, S. P., & Everett, M. G. (1999). Models of core/periphery structures. *Social Networks*, *21*, 375–395.

Carney, S. L. (2003). Leroy Hood expounds the principles, practice and future of systems biology. *Drug Discovery Today*, *8*, 436–438.

Facciotti, M. T. (2009). Training interdisciplinary scientists for systems biology. *Journal of Investigative Medicine*, *57*, 471–473.

Faust, K. (1997). Centrality in affiliation networks. *Social Networks*, *19*, 157–191.

Frandsen, T. V., Rousseau, R., & Rowlands, I. (2006). Diffusion factors. *Journal of Documentation*, *62*, 58–72.

Friboulet, A., & Thomas, D. (2005). Systems biology – An interdisciplinary approach. *Biosensors & Bioelectronics*, *20*, 2404–2407.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, *178*, 471–479.

Goh, K., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabasi, A. (2007). The human disease network. *Proceedings of the National Academy of the Sciences of the United States of America*, *104*, 8685–8690.

Hood, L., Heath, J. R., Phelps, M. E., & Lin, B. (2004). Systems biology and new technologies enable predictive and preventative medicine. *Science*, *306*, 640–643.

Hwang, K. (2008). International collaboration in multilayered center-periphery in the globalization of science and technology. *Science Technology Human Values*, *33*, 101–133.

Ideker, T., Galitski, T., & Hood, L. (2001). A new approach to decoding life: Systems biology. *Annual Review of Genomics and Human Genetics*, *2*, 343–372.

Jansenns, F., Zhang, L., Moor, B. D., & Glanzel, W. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Journal of Information Processing and Management*, *45*, 683–702.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M., & Baldwin, J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*, 860–921.

Leydesdorff, L., & 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*, 348–382.

Leydesdorff, L., & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics*, *5*, 87–100.

Leydesdorff, L., Rotolo, D., & Nooy, W. (2013). Innovation as a nonlinear process, the scientometric perspective, and the specification of an 'innovation opportunities explorer'. *Technology Analysis and Strategic Management*, *25*, 641–653.

Leydesdorff, L., Rotolo, D., & Rafols, I. (2012). Bibliometric perspectives on medical innovation using the medical subject headings (MeSH) of PubMed. *Journal of the American Society for Information Science and Technology*, *63*, 2239–2253.

Liu, Y., Rafols, I., & Rousseau, R. (2012). A framework for knowledge integration and diffusion. *Journal of Documentation*, *68*, 31–44.

Liu, Y., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. *Journal of the American Society for Information Science and Technology*, *61*, 340–351.

Liu, Y., & Rousseau, R. (2012). Towards a representation of diffusion and interaction of scientific ideas: The case of fiber optics communication. *Information Processing and Management*, *48*, 791–801.

Liu, Y., Rousseau, R., & Guns, R. (2013). A layered framework to study collaboration as a form of knowledge sharing and diffusion. *Journal of Informetrics*, *7*, 651–664.

Mesarovic, M. D. (1968). *Systems theory and biology*. Berlin: Springer-Verlag.

Newman, M. E. J. (2001). Scientific collaboration networks. II: Shortest paths, weighted networks, and centrality. *Physical Review E*, *64*, 016132.

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, *28*, 441–452.

Potters, G. (2010). Systems biology of the cell. *Nature Education*, *3*, 33.

Rafols, I., & 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*, 1823–1835.

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, *82*, 263–287.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., & Sutton, G. G. (2001). The sequence of the human genome. *Science*, *291*, 1304-+.

Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, *34*, 1608–1618.

Wallerstein, I. (2004). *World-systems analysis: An introduction* (5 print ed.). Durham: Duke University Press.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York and Cambridge: Cambridge University Press.

Zelino, R. (2012). Identifying the core–periphery structure of science. *Scientometrics*, *91*, 601–615.