How small is the center of science? Short cross-disciplinary cycles in co-authorship graphs

Chris Fields

528 Zinnia Court
Sonoma, CA 95476 USA

fieldsres@gmail.com

October 17, 2014

Abstract

Cycles that cross two or more boundaries between disciplines in the co-authorship graph for all of science are used to set upper limits on the number of co-authored papers required to cross 15 disciplines or subdisciplines ranging from macroeconomics to neurology. The upper limits obtained range from one (discrete mathematics, macroeconomics and nuclear physics) to six (neuroscience). The 15 disciplines or subdisciplines examined form a "small world" with an average separation of only 2.0 co-authorship links. It is conjectured that the high-productivity, high average degree centers of all scientific disciplines form a small world, and therefore that the diameter of the co-authorship graph of all of science is only slightly larger than the average diameter of the co-authorship graphs of its subdisciplines.

Keywords: Cross-displinary brokers; Field mobility; Graph centrality; Graph diameter; Nobel laureates; Preferential attachment; Small-world networks

1 Introduction

Numerous studies have now confirmed the finding of Newman (2001, 2004) that the co-authorship graphs of many scientific disciplines can be characterized as "small worlds" comprising many dense clusters of researchers separated by relatively short (mean length

$l \leq 10$) minimum paths (reviewed by Mali, Kronegger, Doreian and Ferligoj, 2012). Despite over two decades of efforts to make the world of science as a whole more interconnected by encouraging inter-, multi- or trans-disciplinary collaborations, however, the research enterprise remains organized into disciplines (Jacobs and Frickel, 2009). Global maps of science constructed from citation data, for example, consistently show clear disciplinary clusters separated by relative distances consonant with traditional divisions between physical, biological and social sciences, with mathematics occupying an uncertain periphery (Börner, Chen and Boyack, 2003; Moya-Anegón *et al.*, 2007; Klavans and Boyack, 2009; Rafols, Porter and Leydesdorff, 2010). Newly-emergent disciplines such as cognitive science or bioinformatics rapidly establish lineage-based hierarchies, specialist journals, distinctive curricula, and other barriers to entry. One common explanation for the resiliance of disciplinary organization appeals not just to institutional structure but to research efficiency. Lambiotte and Panzarasa (2009), for example, suggest that "successful science production will therefore be organized into a disproportionately large number of cohesive collaborations among scientists with homogeneous knowledge within the same community, and relatively few brokered collaborations among scientists with heterogeneous knowledge across different communities" (p. 188); Jacobs (2013) similarly argues that its organization into disciplines is a critically important enabler of the research enterprise. There being relatively few cross-disciplinary collaborations is consistent with the common view that cross-disciplinary brokers are relatively "weak links" between otherwise separate disciplines, and that the sparsity of such links keeps researchers in different disciplines relatively distant, when compared to researchers in the same discipline, in the co-authorship graph spanning all of science.

How much more distant are researchers in different disciplines - psychology and physics, for example - compared to researchers in a single discipline? The size of the co-authorship graph for all of science renders a direct approach to this question combinatorially explosive and hence computationally infeasible. If one restricts consideration to the elite sample comprising Nobel laureates, however, a partial answer is provided by their Erdős numbers, their distances from the late Paul Erdős in the connected component of the co-authorship graph for all of science. Since 1990, the average Erdős numbers for the incomplete sample of Nobel laureates for whom Erdős numbers are available (De Castro and Grossman, 1999; see http://www.oakland.edu/enp/erdpaths/ for more current data) range from four for economists and biomedical scientists to five for physicists and six for chemists (data from http://www.oakland.edu/enp/erdpaths/, accessed January - June, 2014). One can conclude, therefore, that these Nobel laureates in economics and biomedicine, for example, are separated from each other by no more than eight co-authors on average, less than the mean distances between authors within either computer science (Newman, 2001) or sociology (Moody, 2004) during the relevant period. Erdős and his immediate collaborators are all, moreover, mathematicians specialized in discrete mathematics; hence only two or three co-authorship links, on average, separate one of these Nobel laureates in economics or biomedicine from the boundary between his field and discrete mathematics. Given that Nobel laureates are plausibly regarded as "centers" of scientific productivity and collaboration within their disciplines, their proximity to the boundaries of their respective disciplines suggests that when embedded in the co-authorship graph for all of science, these disciplines have at least one dimension, the one connecting the "center" defined by the examined Nobel laureates to discrete mathematics, that is narrow. If all disciplines have such narrow dimensions, and if these narrow dimensions are proximate to each other, the center of science as a whole might be small.

To provide a quantitative measure, let us restrict attention to paths in the co-authorship graph in which every vertex represents a distinct author and every edge represents a distinct publication, and define the "width" of a discipline along such a path P as the number of co-authorship steps and hence distinct authors and publications required to cross that discipline on P. The "minimum width" of a discipline is then the smallest such number, i.e. the discipline's width along the shortest such path. The minimum width of a discipline clearly depends on the number and distribution of cross-disciplinary brokers within that discipline; if brokers are few and mutually distant, the minimum width may be larger than the average distance between researchers in the discipline, but if brokers are numerous or tightly clustered, it may be smaller. Any discipline can, moreover, be wide when measured along some paths and narrow when measured along others. The longest dimension of any graph G is its diameter $d(G)$, defined as the length of the longest of the minimal-length paths between two vertices of G; intuitively, $d(G)$ is the shortest distance between the two most-separated vertices of G (for this and other graph-theoretic concepts, see Diestel, 2010, or for a briefer, application-specific summary, Börner, Sanyal and Vespignani, 2007). Any path in G with length greater than $d(G)$ is clearly not a minimal path between its endpoints; hence $d(G)$ provides an upper limit for the minimal width of any graph G. The diameter of the connected component of mathematics in the early 2000's was about 20 (Grossman, 2005), while the diameters of physics, biology and computer science were 20, 24 and 31 respectively (Newman, 2001); these numbers provide upper limits for the widths of interest here.

The present paper examines the idea that scientific disciplines might have a narrow dimension - one even smaller than the average separation between researchers - by measuring the widths of various disciplines along paths that cross disciplinary boundaries in the giant connected component of the co-authorship graph of all of science. It shows that subdisciplines as diverse as cognitive science, virology and quantum computing can be crossed in three to five co-authorship links, often into an altogether different discipline. On such paths, up to one third of the authors may be cross-disciplinary brokers who co-author papers, often with considerable regularity, in more than one discipline. These cross-disciplinary brokers are, moreover, often Nobel laureates or other prominent researchers who are naturally regarded as "central" to their disciplines. These results suggest a picture of disciplines as narrow and porous near their highly-productive centers, and broader and more firmly bounded away from their centers. As discussed further below, cross-disciplinary brokers create this narrowness and porosity by clustering near centers of productivity. In this picture, the world of science is richly interconnected and hence "small" for researchers near the highly-productive, porous centers of disciplines, but potentially much larger for researchers located far from disciplinary centers. The porosity of and interconnections between disciplinary centers suggest, moreover, that the world of all of science - at least as represented by its co-authorship graph - may be not much larger than the world of any particular discipline.

2 Data and Methods

The present analysis focuses on *closed* paths in the co-authorship graph of all of science; in graph theory, such paths are termed *cycles*. Cycles are chosen for analysis over open paths for two reasons. First, cycles are indicators of robust connectivity within graphs. Any connected graph exists somewhere on the architectural spectrum between a tree, in which any two vertices are connected by exactly one path and hence there are no cycles, and a complete graph, in which every vertex is directly connected to every other vertex and every path is either a cycle or an embedded component of a cycle (compare Fig. 1A and Fig. 1B). Graphs with large numbers of cycles are more resistant to disconnection by the removal of individual edges, and hence are in a natural sense more robust. Second, any co-authorship cycle must, by definition, enter a discipline through one author and leave through another (Fig. 1C). Hence cycles provide width measurements for every discipline they enter, unlike open paths that provide no width information for the disciplines in which they begin and end. An analysis of discipline-crossing cycles is, therefore, more efficient than an analysis of open paths for the graphic presentation of width information.

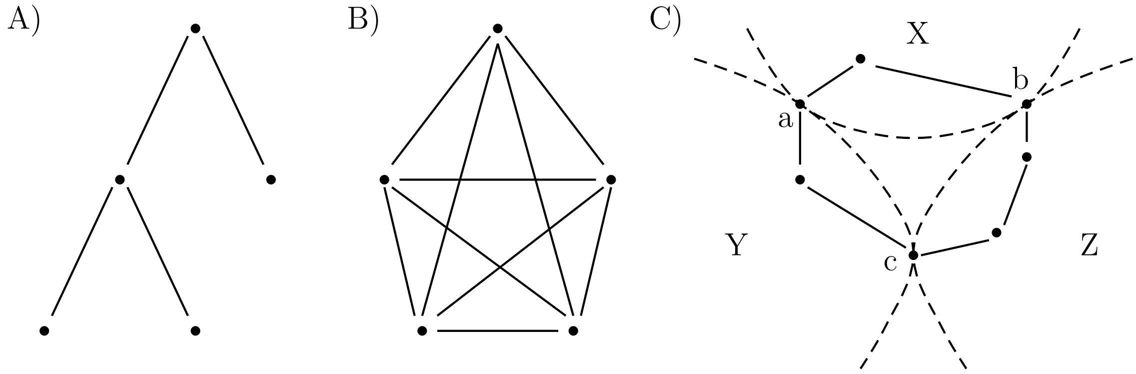

Fig. 1: A) A tree on 5 vertices. B) The complete graph on 5 vertices. C) Schematic co-authorship 7-cycle crossing three disciplines (bounded by dashed lines) X, Y and Z. Vertices labelled a, b and c represent cross-disciplinary brokers. Each broker “straddles the border” between two disciplines.

For the present purposes, a *cross-disciplinary broker* is defined as someone who has published co-authored papers in more than one discipline, and hence has co-authorship connections extending into at least two disciplines. As shown schematically in Fig. 1, moreover, disciplines intersect *only* at such brokers; if there are no cross-disciplinary brokers present, the co-authorship graphs of distinct disciplines do not touch. A thorough analysis of the timing of publications within a given author's career, and hence a rigorous distinction between authors who work in two or more disciplines in sequence and those who pursue work in multiple disciplines in parallel or who collaborate across disciplinary boundaries only occasionally is not attempted; some specific cases are discussed below. Consistent with this focus on cross-disciplinary authors, all papers are assigned to a single discipline in the present analysis, including papers that employ methods or techniques from one discipline to answer a question posed in another. Papers are assigned to disciplines based on their content, not on the disciplinary backgrounds or departmental associations of their authors or the nominal discipline of the journal or proceedings in which they are published. In most cases, for example "High-spin states above 3.5 MeV in ^{91}Nb" (Fields *et al.*, 1978), a paper authored by staff of a nuclear physics laboratory and published in the journal *Nuclear Physics*, these latter indicators suggest a disciplinary assignment consistent with that inferred on the basis of content. In other cases, however, this correspondence between indicators of disciplinarity breaks down. The 1996 paper "Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness" by neurologist Stuart Hameroff and cosmologist Roger Penrose, for example, was published in the journal *Mathematics and Computers in Simulation*, an unusual venue for a paper proposing a structural-biochemical analysis of conscious awareness and relating it to data from anesthesiology. This paper is, however, well-known in its field and has garnered 585 citations. Other cases in which the disciplinary assignment of a paper might be considered unusual or controversial are discussed below, in the contexts of the co-authorship graphs in which they are included. All papers considered are included in the References so that their titles and places of publication may be examined.

The number of cross-disciplinary brokers identified by this method can be expected to increase with the number of distinct disciplines to which papers are assigned; distinguishing disciplines too finely, therefore, carries the risk of artifactually inflating the number of "brokers" identified. To minimize this effect, cycles were sought that cross at least one of the boundaries separating the 16 "consensus" disciplines identified by Klavans and Boyack (2009), for example, from Physics to Biology or Social Sciences to Computer Science. To facilitate the discussion, subdiscipline labels such as "molecular biology" or "nuclear physics" are used; cases in which more than one subdiscipline of a Klavans and Boyack consensus discipline are included in the analysis are discussed explicitly.

Co-authorship data were obtained by manual searches of Google Scholar^{TM}between January and June, 2014. Primary and secondary research papers, review articles, research-based science-policy papers and scholarly books were included in the analysis; otherwise-unpublished technical reports, textbooks, joint editing of collections, and editorial or opinion pieces were not included. As the goal was to efficiently discover short cycles in the co-authorship graph that cross one or more disciplinary boundaries, co-authorship paths leading either from or back to cross-disciplinary brokers known to the present author were followed preferentially. The co-authorship cycles discussed below are not, therefore, claimed to be representative in any way; they are rather illustrative examples of discipline-crossing cycles that allow upper limits to be placed on the minimum widths of the disciplines being crossed.

It should also be noted that any analysis of the co-authorship graph systematically under-reports the extent of inter- or cross-disciplinary research by discounting all single-author papers, and in particular, all single-author papers in a discipline other than the single author's primary discipline. As the proportion of single-author papers in the sciences and mathematics has been decreasing over the past several decades (Porter and Rafols, 2009), the effect of this bias is expected to be small.

The citation counts reported were obtained from Google Scholar^{TM}in early June, 2014.

3 Results

It is widely recognized that all newly-emergent subdisciplinary specializations require cross-disciplinary brokerage as defined here; the scientists who founded the subdiscipline of nuclear physics in the mid-20th century, for example, all began their careers either in some other area of physics, or in a different Klavans and Boyack (2009) discipline altogether, such as Chemistry (e.g. Eugene Wigner) or Mathematics (e.g. John von Neumann). Newly-emerging disciplines are, therefore, good places to look for cross-disciplinary brokers. One of the most prominent subdisciplines to emerge in recent decades is molecular biology, the study of biological outcomes of biochemical manipulations, primarily manipulations of nucleic acids. While molecular biology is generally considered part of the Klavans and Boyack discipline of Biology and its methods are now routinely employed throughout the biological and medical sciences, it contains substantial contributions from Biochemistry and academic departments combining molecular biology and biochemistry are not uncommon. Figure 2 shows overlapping 7- and 8-cycles that cross between molecular biology and nuclear physics three times. Nuclear physics has width four in both the upper and lower cycles and hence a minimal width of at most four. Molecular biology has width four in the upper cycle and three in the lower cycle, showing that distinct paths across a discipline can, as would be expected, confer distinct widths.

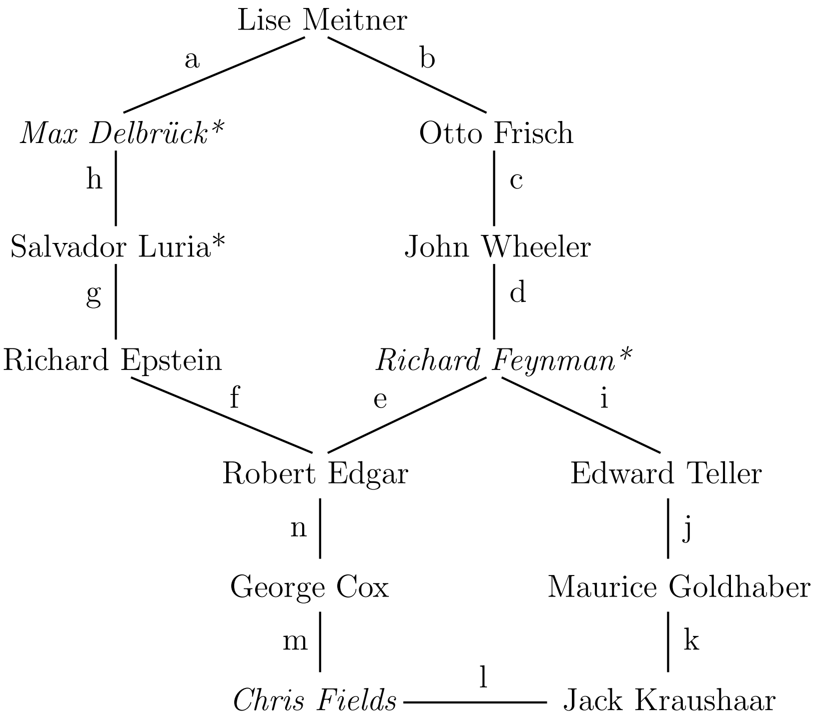

Fig. 2: Overlapping 7- and 8-cycles in which molecular biology has widths of four (upper cycle) and three (lower cycle) and nuclear physics has width four in both upper and lower cycles. Here and below, *italics* indicate a cross-disciplinary broker and a '*' indicates a Nobel laureate. Edges (a), (b), (c), (d), (i), (j), (k) and (l) are nuclear physics papers; edges (e), (f), (g), (h), (m) and (n) are molecular biology papers. Labels are: (a) Meitner and Delbrück, 1935; (b) Meitner and Frisch, 1939; (c) Frisch and Wheeler, 2009; (d) Wheeler and Feynman, 1945; (e) Edgar *et al.*, 1962; (f) Epstein *et al.*, 1963; (g) Wiberg *et al.*, 1962; (h) Luria and Delbrück, 1943; (i) Feynman, Metropolis and Teller, 1949; (j) Goldhaber and Teller, 1948; (k) Kraushaar and Goldhaber, 1953; (l) Fields *et al.*, 1978; (m) Cox *et al.*, 1989; (n) Cox, Kusch and Edgar, 1981.

The cycles shown in Fig. 2 demonstrate that Nobel laureates - here, both Max Delbrück and Richard Feynman - are sometimes also cross-disciplinary brokers. Delbrück is well known in physics as the discoverer of what is now known as Delbrück scattering; however, he published his discovery only as an addendum to Meitner and Köster (1933), a relatively little-known paper with 32 citations. His subsequent work with Meitner is hardly known; Meitner and Delbrück (1935) has received only four citations. Delbrück is better known as one of the founders, together with Luria, of molecular biology; Luria and Delbrück (1943) has been cited 3,024 times. Feynman's early contribution to molecular biology, on the other hand, is considerably less well-known than his work in theoretical physics; Edgar *et al.* (1962) has 48 citations, compared to Wheeler and Feynman (1945) with 1,182 citations or Feynman, Metropolis and Teller (1949) with 618 citations. Nick Metropolis, the second author of the latter paper, was a prominent mathematical physicist who also published papers in discrete mathematics; his Erdős number is two. Hence this paper also links both molecular biology and nuclear physics to discrete mathematics.

The co-authorship graph of all of science is generally considered as a time-independent representation of scientific collaboration; people still talk about Erdős numbers, for example, although Erdős died in 1996. Proximity of authors in the co-authorship graph, therefore, reflects academic lineage as well as shared interests. With its over five-decade timespan, Fig. 2 reminds us both that scientific communities are multi-generational, and that the co-authorship graph represents not just the present, but also the history of scientific, and in particular cross-disciplinary collaboration. This time-independent view does not, however, directly reveal potential changes in interdisciplinarity over time, a point that will be discussed further below.

Like the development of new subdisciplinary specializations, large-scale, goal-oriented projects that require significant development of new methods or technologies for their completion can be expected to involve cross-disciplinary brokerage; indeed such projects often provide the impetus for administrative efforts toward greater interdisciplinarity (Jacobs and Frickel, 2009). The Human Genome Project (HGP), a multinational effort that began in the late 1980s and reached its initial goal of a “draft” sequence of the human genome in 2001, involved collaborations both among molecular biologists and between molecular biologists, mathematicians and computer scientists at a scale unprecedented in biology (for brief histories, see Roberts *et al.*, 2001 or Watson and Cook-Deegan, 1991). Such collaborations created multiple discipline-crossing cycles in the co-authorship graph (Fields, 2014; see http://chrisfieldsresearch.com/erdos.htm for additional data) and led to the development of a new specialization, typically called either bioinformatics or computational biology, within molecular biology. A 5-cycle that crosses between molecular biology and discrete mathematics twice is shown in Fig. 3. This cycle gives molecular biology a minimal width of at most two and Erdős' discipline of discrete mathematics a minimal width of at most three.

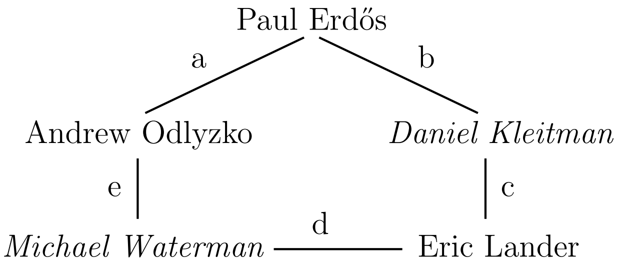

Fig. 3: A 5-cycle in which molecular biology has width two and discrete mathematics has width three. Edges (a), (b) and (e) are mathematics papers; edges (c) and (d) are molecular biology papers. Labels are: (a) Erdős and Odlyzko, 1979; (b) Erdős and Kleitman, 1968; (c) Pachter *et al.*, 1999; (d) Lander and Waterman, 1988; (e) Griggs *et al.*, 1990.

Unlike in Fig. 2, the assignment of papers to disciplines in Fig. 3 requires some close reading. With 604 citations, Lander and Waterman (1988) is by far the best-known. Its authors both began their careers in mathematics and its title, "Genomic mapping by fingerprinting random clones: A mathematical analysis" suggests that it might be a mathematics paper. It was, however, published in *Genomics*, a journal established in 1987 to serve the HGP community, and by 1988 both of its authors had already published multiple papers on biological topics in collaboration with biologists. The paper analyzes physical maps, a kind of data introduced to molecular biology by the HGP, and is clearly written for biologists: it carefully discusses experimental methods, in-progress physical mapping projects and their data, and potential sources of errors and uncertainties. Lander and Waterman (1988) is, therefore, a molecular biology paper, albeit a mathematical one. Pachter *et al.* (1999) describes a practical algorithm for identifying components of genes in DNA sequence data; it is also written for biologists, with minimal formalism and extensive results of analyzing real data sets. Griggs *et al.* (1990), on the other hand, is motivated by a biological problem, but immediately abstracts it and presents its highly-generalized results in the formalism-heavy style of a typical mathematics paper. The two papers co-authored by Erdős are pure mathematics papers motivated by mathematical questions. None of the latter four papers have more than 50 citations. Such order-of-magnitude differences in citation counts are, as also seen in Fig. 2, not unusual for papers appearing in discipline-crossing cycles; papers that report broadly-applicable methods or tools, for example, typically garner more citations than papers reporting "everyday" research results.

While the cycles shown in Figs. 2 and 3 connect two disciplines, cycles crossing multiple disciplines can also be easily found. Figure 4 shows a 10-cycle with linked 4- and 6-cycles that together cross virology, a subdiscipline within Klavans and Boyack's (2009) consensus discipline of Infectious Disease, social-network theory, an emerging blend of Social Sciences and Mathematics, the general theory of computing within Computer Science, and molecular biology. Here virology has a minimal width of at most three, while network theory and the theory of computation have minimal widths of at most two. Like the upper cycle in Fig. 2, the central 10-cycle gives molecular biology (including bioinformatics) a width of four, still smaller than the mean distance of 4.6 ± 0.2 reported by Newman (2001) for authors of biomedical papers listed in MedLine, while the outer loop with edges (a) and (b) gives molecular biology a width of five.

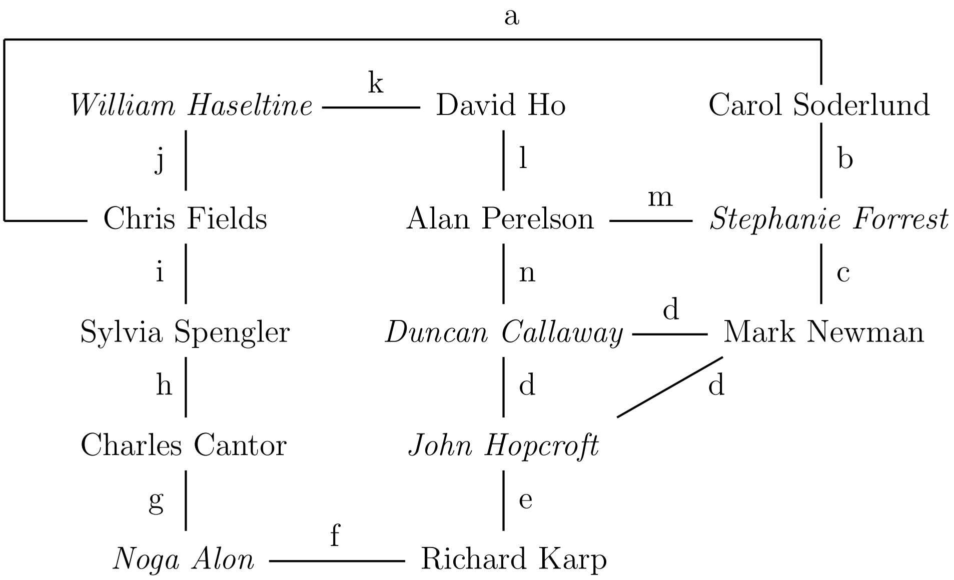

Fig. 4: A 10-cycle with linked 4- and 6-cycles crossing four disciplines: molecular biology (edges (a), (b), (g), (h), (i) and (j)), network theory (edges (c) and (d)), theory of computation (edges (e) and (f)), and virology (edges (k), (l), (m) and (n)). Labels are: (a) Fields and Soderlund, 1990; (b) Burks *et al.*, 1994; (c) Newman, Forrest and Balthrop, 2002 (d) Callaway *et al.*, 2001; (e) Hopcroft and Karp, 1973; (f) Alon *et al.*, 1995; (g) Alon *et al.*, 2006; (h) Murray *et al.*, 1993; (i) Waterman *et al.*, 1994; (j) Adams *et al.*, 1995 (k) Goh *et al.*, 1986; (l) Ho *et al.*, 1995; (m) Forrest *et al.*, 1993; (n) Callaway and Perelson, 2002.

Figure 4 illustrates both the interdisciplinary character of social network theory and the diverse backgrounds of the scientists engaged in it. Burks *et al.* (1994) is a bioinformatics paper describing an algorithm for DNA sequence assembly, a problem that can be formalized in terms of network traversal. Forrest *et al.* (1993) models molecular recognition by the immune system, a critical problem in virology. Stephanie Forrest, a cross-disciplinary co-author of both papers, is a computer scientist, as is John Hopcroft. Duncan Callaway, on the other hand, is a physicist, as is Mark Newman. Callaway *et al.* (2001), of which the latter three are all co-authors, models the outcomes of probabilistic network-assembly processes and was published in *Physical Review E*, an interdisciplinary physics journal. Figure 4 also shows that an individual - here, the present author - can function as a cross-disciplinary broker on some paths (e.g. the lower cycle in Fig. 2) but not on others.

The paper of Ho *et al.* (1995) shown in Fig. 4 reports on the mechanism of HIV infection and has received 4,177 citations, that of Hopcroft and Karp (1973) reports a technical result in graph theory and has been cited 2,004 times, while that of Adams *et al.* (1995) reports an initial analysis of transcription in the human genome and has received 1,015 citations. These can be contrasted with Alon *et al.* (2006), which describes graph-theoretic methods for analyzing multiplexed biological assays and has received only two citations. The other papers represented by edges in these linked cycles fall between these extremes; most have received over 100 citations.

James Watson and Francis Crick shared the 1962 Nobel Prize for Physiology or Medicine, with Maurice Wilkins, for their characterization of the double helix of DNA; Watson and Crick (1953) has been cited 9,703 times. Watson subsequently played a leading role in the HGP, while Crick, another Nobel laureate acting as a cross-disciplinary broker, turned his attention to neuroscience and published several highly-cited papers with Christof Koch, including Crick and Koch (1990) with 1,631 citations. The 13-cycle shown in Fig. 5 is one result; this cycle spans molecular biology, cognitive science, and both experimental and theoretical neuroscience. Neuroscience is a Klavans and Boyack (2009) consensus discipline. Cognitive science, one the other hand, draws on anthropology, artificial intelligence, linguistics, neuroscience and philosophy of mind as well as cognitive and developmental psychology; it thus combines elements of the Klavans and Boyack disciplines of Computer Science, Social Sciences and Humanities with elements of Psychology and Neuroscience. Figure 5 shows that the minimum width of cognitive science is at most four, while that of neuroscience is at most six. Molecular biology here has a width of three.

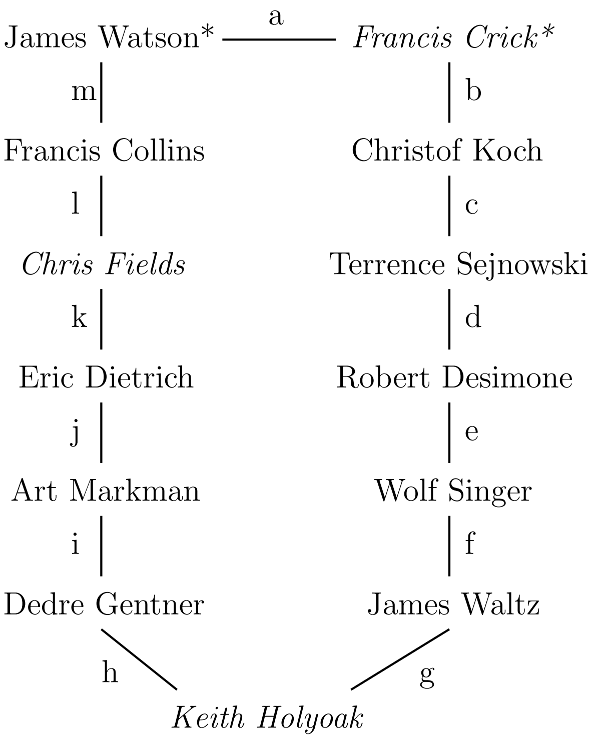

Fig. 5: A 13-cycle spanning molecular biology (edges (a), (l) and (m)), neuroscience (edges (b) - (g)) and cognitive science (edges (h) - (k)). Labels are: (a) Watson and Crick, 1953; (b) Crick and Koch, 1990; (c) Sejnowski, Koch and Churchland, 1988; (d) Lehky, Sejnowski and Desimone, 1992; (e) Womelsdorf *et al.*, 2007; (f) Linden *et al.*, 2003; (g) Waltz *et al.*, 1999; (h) Gentner and Holyoak, 1997; (i) Markman and Gentner, 1993; (j) Markman and Dietrich, 2000; (k) Dietrich and Fields, 1996; (l) McCombie *et al.*, 1992; (m) Collins and Watson, 2003.

The cognitive science papers in this cycle all concern abstract computational modeling of the mind. Both Markman and Gentner (1993) and Gentner and Holyoak (1997) discuss models of analogical reasoning, Markman and Dietrich (2000) defends the use of explicit representations of mental contents in computational models of cognition, while Dietrich and Fields (1996) discusses methodological implications of the frame problem, a classic problem in artificial intelligence. The neuroscience papers, on the other hand, all focus on the implementation of cognition by the brain. Crick and Koch (1990) and Sejnowski, Koch and Churchland (1988) are theoretical papers; the remainder are experimental. Cross-disciplinary broker Keith Holyoak joins the abstract modeling of cognitive science with the implementation details of neuroscience, examining relational reasoning - the core of analogy - in patients with focal lesions in prefrontal cortex in Waltz *et al.* (1999) and with neuroimaging methods in other work (e.g. Cho *et al.*, 2010).

Francis Collins is, at the time of this writing, Director of the U.S. National Institutes of Health and hence a prominent leader of the biomedical research enterprise. His early genome-project work with Charles Cantor, reported in Smith *et al* (1987), together with that of the present author, joins the cycle shown in Fig. 5 to that of Fig. 4.

While the cycle in Fig. 5 depends heavily on experimental papers, heavily theoretical cycles can also be found. The 14-cycle shown in Fig. 6 spans quantum cosmology and quantum computing then crosses via discrete mathematics into molecular biology before returning through neurology. Neurology and quantum cosmology are straightforwardly located within the Klavans and Boyack (2009) disciplines of Medical Specialties and Physics, respectively. Quantum computing, on the other hand, merges the concepts and methods of Physics with those of Computer Science to solve problems motivated mainly by Computer Science, in particular, information security. Quantum cosmology and quantum computing both have widths of three on this cycle, while discrete mathematics achieves a width of one. Molecular biology has a width of four, while neurology has width three.

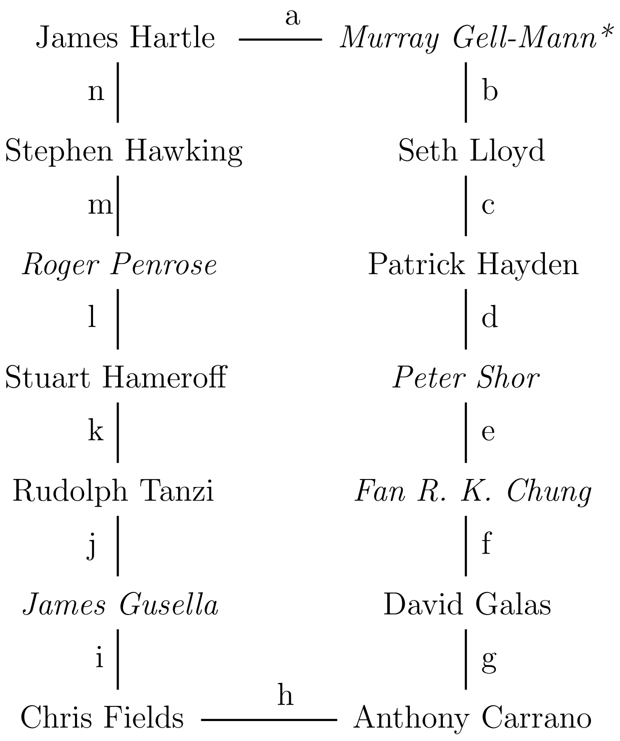

Fig. 6: A 14-cycle spanning quantum cosmology (edges (a), (m) and (n)), quantum computing (edges (b) - (d)), discrete mathematics (edge (e)), molecular biology (edges (g) - (i)) and neurology (edges (j) - (l)). Labels are: (a) Gell-Mann and Hartle, 1993; (b) Gell-Mann and Lloyd, 1996; (c) Guha *et al.*, 2014; (d) Hayden *et al.*, 2004; (e) Bienstock *et al.*, 1991; (f) Chung *et al.*, 2003; (g) Branscom *et al.*, 1990; (h) Martin-Gallardo *et al.*, 1992; (i) McCombie *et al.*, 1992; (j) Rosen *et al.*, 1993; (k) Craddock *et al.*, 2012; (l) Hameroff and Penrose, 1996; (m) Hawking and Penrose, 1970; (n) Hartle and Hawking, 1983.

In this cycle, Nobel laureate Murray Gell-Mann is a cross-disciplinary broker, working both in quantum cosmology and quantum computing. Gell-Mann's Nobel prize, however, honors his work in what would come to be known as high-energy physics. Feynman and Gell-Mann (1958), with 2,409 citations, is an early example of this work and joins, together with the work of the present author, the cycle shown in Fig. 6 with the lower cycle of Fig. 2. Joining these cycles yields paths from quantum computing and quantum cosmology to nuclear physics, as the field represented by Feynman and Gell-Mann (1958) was known at the time, and gives nuclear physics a minimal width of one. The most highly-cited paper in the cycle shown in Fig. 6 is, however, not one of Gell-Mann's but is rather Rosen *et al.* (1993), a report on the genetics of amyotrophic lateral sclerosis ("Lou Gehrig's disease") that has garnered 4,669 citations. The only difficult-to-classify paper in this cycle is Chung *et al.*, 2003, a paper dealing with partial duplication of an existing network as a means of producing daughter networks. It is classified as a bioinformatics and hence a molecular biology paper instead of a network theory or a mathematics paper both for its biological motivation and its extensive discussion of the biological relevance of its results.

As a final example, the 12-cycle shown in Fig. 7 illustrates the close connection between theories of rationality in economics and artificial intelligence as well as confirming the influence of Erdős and his collaborators on molecular biology. Nobel laureate Herbert Simon is a cross-disciplinary broker in this cycle, linking macroeconomics, here represented solely by Hawkins and Simon (1949), with Simon's other major field, artificial intelligence. In this cycle macroeconomics has a width of one, while discrete mathematics expands to width four. As in Fig. 3, molecular biology has a width of two on this cycle, its minimum in these examples, while artificial intelligence has width five.

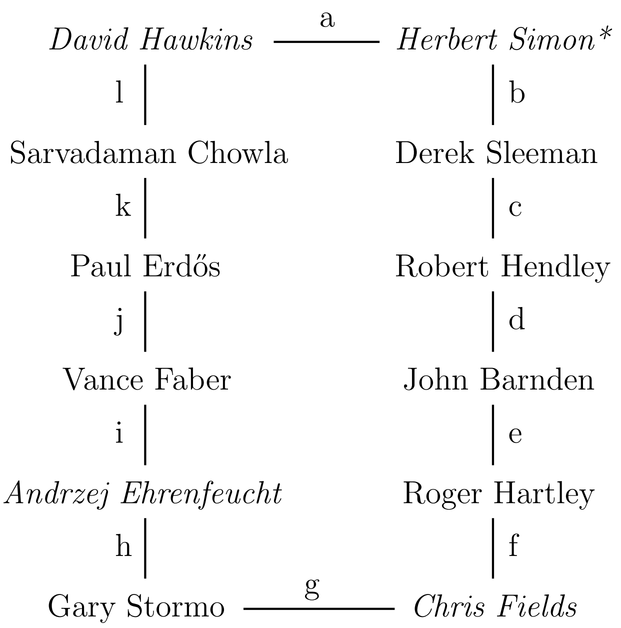

Fig. 7: A 12-cycle spanning macroeconomics (edge (a)), artificial intelligence (edges (b) - (f)), molecular biology (edges (g) and (h)) and discrete mathematics (edges (i) - (l)). Labels are: (a) Hawkins and Simon, 1949; (b) Simon, Valdés-Pérez and Sleeman, 1997; (c) Sleeman and Hendley, 1979; (d) Zhang *et al.*, 2006; (e) Hartley and Barnden, 1997; (f) Fields, Coombs and Hartley, 1988; (g) Mount *et al.*, 1992; (h) Stormo *et al.*, 1982; (i) Ehrenfeucht, Faber and Kierstead, 1984; (j) Erdős, Faber and Larson, 1981; (k) Chowla and Erdős, 1951; (l) Chowla and Hawkins, 1962.

The cycle shown in Fig. 7 illustrates another social network operative in the sciences: David Hawkins and Andrzej Ehrenfeucht were faculty colleagues at the University of Colorado, Boulder, and were professors and mentors of the present author. This network of intellectual lineages is only partially reflected in the co-authorship graph, but plays an important role in forming the "invisible colleges" (Crane, 1972) that hold researchers with similar backgrounds, interests and intellectual orientations together.

The example cross-disciplinary cycles considered here demonstrate three points. First, discipline crossing cycles exist. Such cycles do not just connect pairs of disciplines, but sometimes cross multiple disciplines. The existence of such cycles shows that the co-authorship graph, even when abstracted to a graph of connections between *disciplines*, is not a tree. While the examples considered here do not connect all of Klavans and Boyack's (2009) consensus disciplines, they at least suggest the speculation that no discipline is an island unto itself, populated entirely by researchers who work and publish only in its domain. Second, these examples show that a wide variety of disciplines have small minimal widths; discrete mathematics, macroeconomics and nuclear physics have minimal widths of at most one, molecular biology has a minimal width of at most two, even the clinically-oriented discipline of neurology has a minimal width of at most three. The paths that demonstrate these small widths do not merely cross from one subdiscipline into a different subdiscipline of the same major discipline, but rather into a different discipline even in Klavans and Boyack's very high-level classification, e.g. from Physics to Biology as in Fig. 2. It is important to emphasize, moreover, that these results for minimal widths have no implications for the *diameters* of these disciplines, which as noted above are 20 or more in cases that have been measured. A discipline, even a Klavans and Boyack consensus discipline, can have a large diameter and still have small widths along some paths. Third, and obviously correlated with the second, cross-disciplinary brokers can be found very close to each other. These examples all traverse the present author's primary career discipline of molecular biology, and it is possible that molecular biology, a relatively new discipline that has attracted physicists since its inception, includes more cross-disciplinary brokers than others; indeed the typically central position of the biomedical sciences in maps of science suggests as much. Cognitive science, neuroscience, artificial intelligence and social network theory are also new disciplines populated, at least early on, by researchers whose careers began in a different discipline. This observation, however, has a converse: new disciplines would not form if existing disciplines did not include sufficient individuals willing and able to adopt or at least try out new interests, master new methods and technologies and work with new colleagues. The question of interest is, where are these individuals to be found within a discipline?

4 Discussion

The discipline-crossing cycles shown in Figs. 2 - 7 share a distinctive feature: all include Nobel laureates, mathematicians with Erdős numbers of one, or both. Such researchers are naturally considered "centers" of their respective disciplines; they tend to have many collaborators, to serve as intellectual and often political leaders, and to attract younger researchers by well-studied preferential attachment mechanisms (e.g. Barabási *et al.*, 2002).

"Centrality" in co-authorship graphs, and in social and other networks more generally, has been defined formally in three primary ways. Any vertex separated by at most half of the graph diameter from any other vertex is "distance central" in a graph. The verticies with the largest number of incident edges have "degree centrality," while the vertices through which the maximum number of minimal-length paths between other vertices flow have "betweenness centrality." These three definitions of centrality coincide in a star graph, which consists of a central vertex connected to N other vertices, none of which are connected to each other, but they do not coincide in general (e.g. Fields, 2014, Fig. 2). The extent to which any of these formal definitions captures the intuitive notion that a researcher is "central" to a field is, however, a topic of considerable debate (e.g. Borgatti and Everett, 2006; Newman, 2006; Lambiotte and Panzarasa, 2009; Landherr, Friedl and Heidemann, 2010). Minor contributors to multi-author papers, for example, may have high degree and even distance centrality within a discipline but low betweenness centrality, while "weak links" between disciplines may have high betweenness centrality across disciplines but low degree or distance centrality. The informal notion of "centrality" exemplified by Nobel laureates is, therefore, adopted here. In addition to the Nobel laureates and mathematicians with Erdős numbers of one, many of the other researchers appearing in Figs. 2 - 7 are "central" in this informal sense. As noted earlier, Francis Collins (Fig. 5) is Director of the U.S. National Institutes of Health, a "central" position in biomedicine on any reasonable definition. Edward Teller (Fig. 2) was Director of the Lawrence Livermore National Laboratory, Eric Lander (Fig. 3) is currently Director of the Broad Institute and Christof Koch (Fig. 5) is Chief Scientific Officer of the Allen Institute for Brain Science. As shown by Batagelj and Mrvar (2000) in an analysis of collaboration among mathematicians with Erdős numbers of one or two, such disciplinary leaders tend to work and publish together in cliques that together form a small, within-discipline world; this also holds true for Nobel laureates in the biomedical sciences (Fields, 2014; in prep.) and physics (in prep.). Any metric on the co-authorship graph that did not place such prominent, politically-influential, mutually-collaborating individuals near the "centers" of their disciplines would, indeed, be regarded as of little utility for understanding the social organization of the sciences. With this informal notion of centrality, Figs. 2 - 7 show that scientists who are central to their disciplines are sometimes also cross-disciplinary brokers (e.g. Delbrück, Feynman, Alon, Hopcroft, Crick, Penrose, Gell-Mann, Chung, Simon) or co-authors of cross-disciplinary brokers (e.g. Luria, Teller, Erdős, Lander, Collins, Koch, Watson, Hawking). A researcher can, in other words, be both *central* to a discipline and located on or close to its *boundary* in the co-authorship graph.

The potential coincidence of centrality within a discipline and a location on that discipline's co-authorship border naturally raises the question of *when* a researcher acts as a broker that was posed and deferred earlier. Max Delbrück, for example, was a physicist who became a biologist; such field mobility is not uncommon among physicists (e.g. van Houten *et al.*, 1983), though of course most migrant physicists do not become Nobel laureates in their adopted fields. Francis Crick, similarly, turned to neuroscience well after his 1962 Nobel Prize for co-discovering the structure of DNA. Richard Feynman, however, only briefly explored biology after receiving his Nobel Prize in physics, returning to physics thereafter.

Murray Gell-Mann similarly explored quantum computing after receiving his Nobel Prize, returning to quantum cosmology and foundational work in quantum theory thereafter. Herbert Simon explored economic decision making and artificial intelligence in parallel over the course of his career; his 1978 Nobel Prize in economics only slightly lagged his 1975 Turing Award, the closest analog of the Nobel Prize in computer science. Hence not all cross-disciplinary brokerage among Nobel laureates represents field mobility; some is brief exploration and some is career-spanning parallelism. While a thorough analysis is beyond the present scope, one might expect the careers of cross-disciplinary researchers who are not Nobel laureates to exhibit similar diversity. If this is the case, many researchers may be central to their disciplines and at the same time be cross-disciplinary brokers, i.e. be located on their disciplines' co-authorship borders.

If this idea that centrality and brokerage can co-exist is taken seriously as a hypothesis, the connection data in Figs. 2 - 7 can be summarized by a set of inferred one-step connections between subdisciplines as shown in Fig. 8.

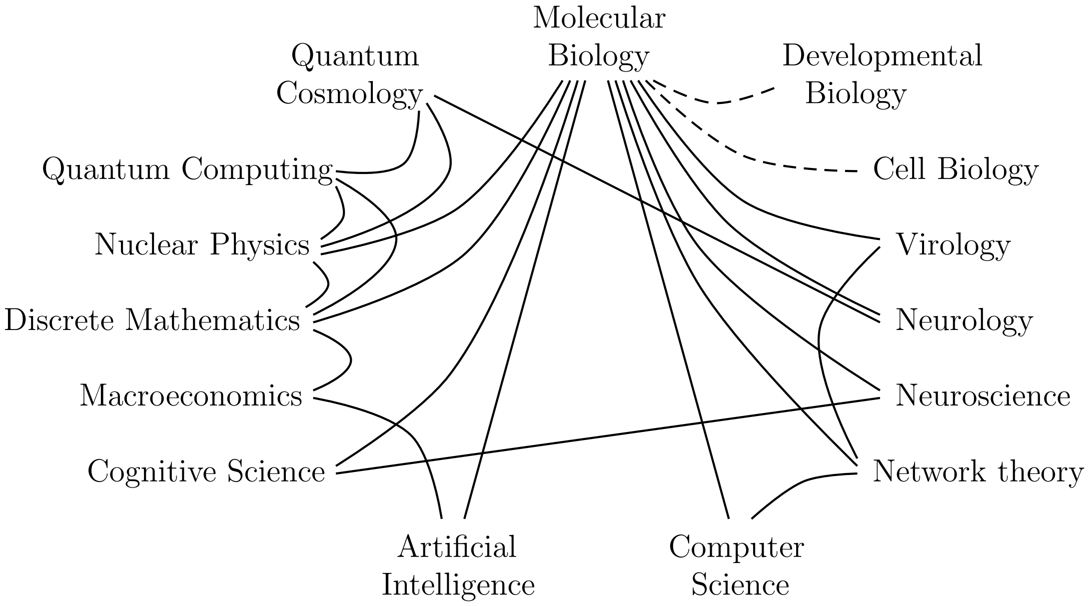

Fig. 8: One-step connections between 15 subdisciplines suggested by the cycles shown in Figs. 2 - 7 (solid lines) or by those in Fields, 2014 (dashed lines). The average degree of the vertices is $z = 2.9$.

The graph shown in Fig. 8 displays significant clustering: 11 of the vertices link to molecular biology and four link to each of discrete mathematics and nuclear physics. The average distance between vertices is 2.0, smaller that the average of $logN/logz = 2.5$ expected for a random graph with $N = 15$ and $z = 2.9$ (Newman, 2002). Assuming the informal notion of a "center" discussed above, the subdisciplinary "centers" in Fig. 8 thus form a "small world" even smaller than expected. Adding obvious links such as those between artificial intelligence and computer science, developmental biology and cell biology, or neurology and neuroscience shrinks this small world even farther.

Universities organize the sciences, including mathematics and engineering, into different numbers of disciplines; MIT lists 20, for example, while the University of California at Berkeley lists 49. On the assumption that many of these could be further divided - both MIT and Berkeley have a single department for all of physics, for example - one might estimate that there are on the order of 100 science and engineering specializations with sufficient cohesion to be considered subdisciplines comparable to those in Fig. 8. Were these 100 subdisciplines to be as well-connected as the 15 in Fig. 8, one would expect an average degree of $z = 19.3$ and an average "small world" distance between subdisciplines of about 1.6. Even with an average degree of only 6 - just twice that of Fig. 8 - one would expect an average distance between subdisciplines of less than three.

The idea that disciplinary centers form a small world makes an immediate, although at this point admittedly somewhat speculative, prediction about science as a whole. The subdisciplines included in Fig. 8 either are or include components of 11 of the 16 Klavans and Boyack (2009) consensus disciplines: Mathematics, Computer Science, Physics, Biology, Biochemistry, Infectious Disease, Medical Specialties, Neuroscience, Psychology, Social Sciences and Humanities. On the assumption that these subdisciplines are at least approximately representative of scientific subdisciplines generally, the diameter of the connected component of the co-authorship graph for all of science can be expected to be the larger of either the diameter of the connected component of the largest-diameter subdiscipline, or the average diameter of the connected components of all subdisciplines plus two or three. Adding disciplines, in other words, is expected not to increase the overall size of science as measured by the diameter of the co-authorship graph. That this might be true is hinted by the data for subdisciplines of physics in Newman (2001): the diameter of the connected component of the arXiv preprint database - 20 - is only one larger than the diameter of its largest-diameter component, hep-th, and only three larger than the average diameter of its components. If the co-authorship diameters of scientific subdisciplines are shrinking with time as the analysis of Barabási *et al.*, 2002 as well as the steady increase in co-authorship within disciplines (Porter and Rafols, 2009; Wallace, Larivière and Gingras, 2012) suggests, one might estimate that the average co-authorship diameter of scientific subdisciplines is currently no more than 25 and the co-authorship diameter of all of science no more than 30.

As a final note, it should be emphasized that while the cycles shown in Figs. 2 - 7 demonstrate the involvement of some "central" researchers in cross-disciplinary brokerage and the proximity of others to cross-disciplinary brokers, the question of the scientific impact of such brokerage must be approached with caution. Some of the papers included in these cycles, including some co-authored by identified brokers, are highly cited and hence *prima facie* high-impact while others are not. Treating such citation numbers as direct measures of the flow of ideas between disciplines, however, may be problematic. Adams *et al.* (1995), for example, reports an initial analysis of transcription in the human genome and has received 1,015 citations as noted above. The work reported in this paper would not, however, have been possible without the previous development of a high-throughput DNA sequencing facility that carefully balanced the biological work of clone preparation and the computational task of data analysis with the capabilities of then-current DNA sequencing machines, all under the management of a process-tracking database system. The paper that described this facility and the engineering considerations underpinning it, Adams *et al.* (1994), was published by the same research group in the same high-visibility journal, *Nature*, but has received only 33 citations despite arguably influencing the designs of all subsequent high-throughput DNA sequencing facilities. Lander and Waterman (1988) provided the first formal criterion for "sufficient" sampling redundancy for biological molecules with a given diversity or sampling error, and hence underlies not only subsequent physical mapping methods but DNA sequence assembly methods as well; its significance for molecular biology is probably far under-represented by its 604 citations. The 351 citations to Hawkins and Simon (1949) may similarly under-represent the significance of "a result that every graduate student in economics must study" (Kuhn, 2004, p. ix).

5 Conclusion

By examining co-authorship cycles that cross two or more disciplines, the present paper demonstrates that cross-disciplinary brokers can sometimes be found in close proximity to each other, and that when they are, they are at least sometimes also close to the Nobel laureates and other highly-productive, well-connected, politically-influential individuals who informally define disciplinary centers. These observations together suggest that disciplinary centers may contain many cross-disciplinary brokers, who render them highly porous and hence tightly clustered. This porosity and clustering of disciplinary centers, in turn, renders the "center" of science a small world, one with an average distance between subdisciplines of distinct major disciplines - and hence between the major disciplines themselves - that is smaller than the average distances between researchers within typical disciplines and much smaller than the diameters of typical disciplines. If this picture is correct, researchers located near the centers of distinct disciplines are in at least some cases closer to each other, as measured by co-authorship distance, than they are to typical colleagues within their own disciplines. Researchers located far from the centers of their disciplines are, conversely, only slightly farther from researchers in other disciplines than they are, on average, from their within-discipline colleagues. Testing this picture clearly requires both systematic examinations of the co-authorship proximity between "central" researchers and brokers across multiple disciplines, and addressing the potentially more-difficult question of how brokers who are located far from disciplinary centers are distributed in the co-authorship graph.

The co-authorship graph is, of course, only one of many ways to examine the structure of science. The robust disciplinary divisions revealed by maps of science constructed from citation data (Moya-Anegón *et al.*, 2007; Rafols, Porter and Leydesdorff, 2010) indicate that citations are predominantly either within-discipline, or between closely-related disciplines such as biology and biochemistry or computer science and mathematics. One might speculate, moreover, that cross-disciplinary brokers would preferentially cite each other, especially during the emergence of new disciplines such as bioinformatics or artificial intelligence, when relatively large numbers of brokers are working at least somewhat cooperatively to build a new area of cross-disciplinary specialization. Such preferential citations would contribute to the appearance of a new citation cluster and hence a new specialization in a citation-based map. This speculation remains to be systematically tested.

The co-authorship graph is, moreover, generally considered time-independent as noted earlier. This naturally raises the question of whether, and if so how, interdisciplinarity changes through time. Nobel-level physicists were, for example, involved in the early development of molecular biology as shown in Fig. 2. Are physicists still "crossing the border" to molecular biology today? A relatively small group of cross-disciplinary brokers founded the new discipline of bioinformatics in the late 1980s. Did most of these scientists stay within their new discipline thereafter, or did they migrate on to even newer cross-disciplinary fields? It is hoped that the present work will spur new interest in such questions.

It is, finally, worth noting that as science evolves over the next several decades, generational forces will inevitably push previous leaders like Erdős, Crick or Feynman out of the centers of their disciplines. They will, however, remain near their respective disciplines' co-authorship borders. How this growing population of well-connected scientists from the past will affect both the local and the global structure of the co-authorship graph of the future remains to be investigated.

Acknowledgments

Many thanks to two anonymous referees for their helpful questions and comments on earlier versions of this paper.

References

Adams, M. D., A. R. Kerlavage, J. Kelley, J. Gocayne, C. Fields, C. Fraser and J. C. Venter (1994). A model for high-throughput automated DNA sequencing and analysis core facilities. *Nature* 368, 474-475.

Adams, M. D., A. R. Kerlavage, R. D. Fleischmann, *et al.* (85 co-authors) (1995). Initial assessent of human gene diversty and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 377, Suppl. 3-174.

Alon, N., R. M. Karp, D. Peleg and D. West (1995). A graph-theoretic game and its application to the k-server problem. *SIAM J. Computation* 24, 78-100.

Alon, N., V. Asodi, C. Cantor, S. Kasif and J. Rachlin (2006). Multi-node graphs: A framework for multiplexed biological assays. *J. Computational Biology* 13, 1659-1672.

Barabási, A. L., H, Jeong, Z. Néda, E. Ravasz, A. Schubert and T. Vicsek (2002). Evolution of the social network of scientific collaborations. *Physica A* 311, 590-614.

Batagelj, V. and A. Mrvar (2000). Some analyses of Erdős collaboration graph. *Social Networks* 22, 173-186.

Bienstock, D., F. R. K. Chung, M. L. Fredman, A. A. Schäffer, P. W. Shor and S. Suri (1991). A note on finding a strict saddlepoint. *American Mathematical Monthly* 98, 418-419.

Borgatti, S. P. and M. G. Everett (2006). A graph-theoretic perspective on centrality. *Social Networks* 28, 466-484.

Börner, K., C. Chen and K. W. Boyack (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology* 37, 179-255.

Börner, K., S. Sanyal and A. Vespignani (2007). Network science. *Annual Review of Information Science and Technology* 41, 537-607.

Branscom, E., T. Slezak, R. Pae, D. Galas, A. V. Carrano and M. Waterman (1990). Optimizing restriction fragment fingerprinting methods for ordering large genomic libraries. *Genomics* 8, 351-366.

Burks, C., M. L. Engle, S. Forrest, R. J. Parsons, C. A. Soderlund and P. E. Stolarz (1994). Stochastic optimization tools for genome sequence assembly. In M. D. Adams, C. Fields and J. C. Venter (Eds) *Automated DNA Sequencing and Analysis.* New York: Academic Press, pp. 249-259.

Callaway, D. S., J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman and S. Strogatz (2001). Are randomly grown graphs really random? *Physical Review E* 64, 041902.

Callaway, D. S. and A. S. Perelson (2002). HIV-1 infection and low steady state viral loads. *Bulletin of Mathematical Biology* 64, 29-64.

Cho S., T. D. Moody, F. Fernandino, J. A. Mumford, R. A. Poldrack, T. D. Cannon, B. J. Knowlton and K. J. Holyoak (2010). Common and dissociable prefrontal loci associated with component mechanisms of analogical reasoning. *Cerebral Cortex* 20, 524-533.

Chowla, S. and P. Erdős (1951). A theorem on the distribution of values of L-functions. *J. Indian Mathematical Society* 15, 11-18.

Chowla, S. and D. Hawkins (1962). Asymptotic expansion of some series involving the Reimann zeta function. *J. Indian Mathematical Society* 26, 115-124.

Chung, F. R. K., L. Liu, T. G. Dewey and D. J. Galas (2003). Duplication models for biological networks. *J. Computational Biology* 10, 677-687.

Collins, F. S. and J. D. Watson (2003). Genetic discrimination: Time to act. *Science* 302, 745.

Cox, G. N., M. Kusch and R. S. Edgar (1981). Cuticle of *Caenorhabditis elegans*: Its isolation and partial characterization. *J. Cell Biology* 90, 7-17.

Craddock, T. J. A., J. A. Tuszynski, D. Chopra, N. Casey, L. E. Goldstein, S. R. Hameroff and R. E. Tanzi (2012). The zinc dyshomeostasis hypothesis of Alzheimer's disease. *PLoS One* 7, e33552.

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

Crick, F. R. C. and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences* 2, 263-275.

De Castro, R. and J. W. Grossman (1999). Famous trails to Paul Erdős. *Mathematical Intelligencer* 21 (3), 51-63.

Diestel, R. (2010). *Graph Theory* 4th Edition. Berlin: Springer.

Dietrich, E. and C. Fields (1996). The role of the frame problem in Fodor's modularity thesis: A case study in rationalist cognitive science. In K. M. Ford and Z. Pylyshyn (Eds) *The Robot's Dilemma Revisited.* Norwood, NJ: Ablex, pp. 9-24.

Edgar, R. S., R. P. Feynman, S. Klein, I. Lielausis and C. M. Steinberg (1962). Mapping experiments with r mutants of bacteriophage T4D. *Genetics* 47 (2), 179-186.

Ehrenfeucht, A., V. Faber and H. A. Kierstead (1984). A new method of proving theorems on chromatic index. *Discrete Mathematics* 52, 159-164.

Epstein, R. H., A. Bolle, C. M. Steinberg, E. Kellenberger, E. Boy de la Tour, R. Chevalley, R. S. Edgar, M. Susman, G. H. Denhardt and A. Lielausis (1963). Physiological studies of conditional lethal mutants of bacteriophage T4D. *Cold Spring Harbor Symposia on Quantitative Biology* 28, 375-394.

Erdős, P., V. Faber and J. Larson (1981). Sets of natural numbers of positive density and cylindric set algebras of dimension 2. *Algebra Universalis* 12, 81-92.

Erdős, P. and D. J. Kleitman (1968). On coloring graphs to maximize the proportion of multicolored k-edges. *J. Combinatorial Theory* 5, 164-169.

Erdős, P. and A. M. Odlyzko (1979). On the density of odd integers of the form $(p-1)2^{-n}$ and related questions. *J. Number Theory* 11, 257-263.

Feynman, R. P. and M. Gell-Mann (1958). Theory of the Fermi interaction. *Physical Review* 109, 193-198.

Feynman, R. P., N. Metropolis and E. Teller (1949). Equations of state of elements based on the generalized Fermi-Thomas theory. *Physical Review* 75, 1561-1573.

Fields, C. A., J. J. Kraushaar, R. A. Ristinen and L. E. Samuelson (1978). High-spin states above 3.5 MeV in ^{91}Nb. *Nuclear Physics* 326, 55-64.

Fields, C., M. Coombs and R. Hartley (1988). MGR: An architecture for problem solving in unstructured task environments. In: Z. Ras and L. Saitta (Eds.) *Methodologies for Intelligent Systems, 3.* Amsterdam: Elsevier. pp. 40-49.

Fields, C. and C. Soderlund (1990). **gm**: A practical tool for automating DNA sequence analysis. *Computer Applications in the Biosciences* 6, 263-270.

Fields, C. (2014). Some effects of the Human Genome Project on the Erdős collaboration graph. *J. Humanistic Mathematics* 4(2), 3-24.

Forrest, S., Javornik, B., Smith, R. E. and Perelson, A. S. (1993). Using genetic algorithms to explore pattern recognition in the immune system. *Evolutionary Computation* 1, 191-211.

Frisch, O. R. and J. A. Wheeler (2009). The discovery of fission. *Physics Today* 20(11), 43-54.

Gell-Mann, M. and J. B. Hartle (1993). Classical equations for quantum systems. *Physical Review D* 47, 3345-3382.

Gell-Mann, M. and S. Lloyd (1996). Information measures, effective complexity, and total information. *Complexity* 2, 44-52.

Gentner, D. and K. J. Holyoak (1997). Reasoning and learning by analogy: Introduction. *American Psychologist* 52, 32-34.

Goh, W. C., C. Rosen, J. Sodroski, D. D. Ho and W. A. Haseltine (1986). Identification of a protein encoded by the trans activator gene *tatIII* of human T-cell lymphotropic retrovirus type III. *J. Virology* 59, 181-184.

Goldhaber, M. and E. Teller (1948). On nuclear dipole vibrations. *Physical Review* 74, 1046-1049.

Griggs, J. R., P. Hanlon, A. M. Odlyzko and M. S. Waterman (1990). On the number of alignments of k sequences. *Graphs and Combinatorics* 6, 133-146.

Grossman, J. W. (2005). Patterns of research in mathematics. *Notices of the AMS* 52 (1), 35-41.

Guha, S., P. Hayden, H. Krovi, S. Lloyd, C. Lupo, J. H. Shapiro, M. Takeoka and M. M. Wilde (2014). Quantum enigma machines and the locking capacity of a quantum channel. *Physical Review X* 4, 011016.

Hameroff, S. and R. Penrose (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and Computers in Simulation* 40, 453-480.

Hartle, J. B. and S. W. Hawking (1983). Wave function of the universe. *Physical Review D* 28, 2960-2975.

Hartley, R. T. and J. A. Barnden (1997). Semantic networks: Visualizations of Knowledge. *Trends in Cognitive Sciences* 1, 169-175.

Hayden, P., D. Leung, P. W. Shor and A. Winter (2004). Randomizing quantum states: Constructions and applications. *Communications in Mathematical Physics* 250, 371-391.

Hawking, S. W. and R. Penrose (1970). The singularities of gravitational collapse and cosmology. *Proceedings of the Royal Society of London A* 314, 529-548.

Hawkins, D. and H. A. Simon (1949). Note: Some conditions of macroeconomic stability. *Econometrica* 17, 245-248.

Ho, D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard and M. Markowitz (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373, 123-126.

Hopcroft, J. E. and R. M. Karp (1973). An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM Journal of Computation* 2, 225-231.

Jacobs, J. A. (2013). *In Defense of Disciplines.* Chicago: University of Illinois Press.

Jacobs, J. A. and S. Frickel (2009). Interdisciplinarity: A critical assessment. *Annual Review of Sociology* 35, 43-65.

Klavans, R. and K. W. Boyack (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology* 60, 455-476.

Kraushaar, J. J. and M. Goldhaber (1953). Direction and polarization correlations of successive gamma-rays. *Physical Review* 89, 1081-1089.

Kuhn, H. W. (2004). Introduction. In J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior, Sixtieth Anniversary Edition.* Princeton: Princeton University Press, pp. vii- xiv.

Lambiotte, R. and P. Panzarasa (2009). Communities, knowledge creation and information diffusion. *Journal of Informetrics* 3 (3), 180-190.

Lander, E. S. and M. S. Waterman (1988) Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics* 2, 231-239.

Landherr, A., B. Friedl and J. Heidemann (2010). A critical review of centrality measures in complex networks. *Business Information Systems Engineering* 6, 371-385.

Lehky, S. R., T. J. Sejnowski and R. Desimone (1992). Predicting responses of nonlinear neurons in monkey striate cortex to complex patterns. *Journal of Neuroscience* 12, 3568-3581.

Linden, D. E. J., R. A. Bittner, L. Muckli, J. A. Waltz, N. Kriegeskorte, R. Goebel, W. Singer and M. H. J. Munk (2003). Cortical capacity constraints for visual working memory: Dissociation of fMRI load effects in a fronto-parietal network. *NeuroImage* 20, 15181530.

Luria, S. and M. Delbrück (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28, 491-511.

Mali, F., L. Kronegger, P. Doreian and A. Ferligoj (2012). Dynamic scientific co-authorship networks. In A. Scharnhorst, K. Börner and P. van den Besselaar (Eds) *Models of Science Dynamics.* Berlin: Springer, pp. 195-232.

Markman, A. B. and D. Gentner (1993). Structural alignment during similarity comparisons. *Cognitive Psychology* 25, 431-467.

Markman, A. B. and E. Dietrich (2000). In defense of representation. *Cognitive Psychology* 40, 138-171.

Martin-Gallardo, A., W. R. McCombie, J. D. Gocayne, *et al.* (17 co-authors) (1992). Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3. *Nature Genetics* 1, 34-39.

McCombie, W. R., A. Martin-Gallardo, J. D. Gocayne, *et al.* (21 co-authors) (1992). Expressed genes, *Alu* repeats and polymorphisms in cosmids sequenced from chromosome 4p16.3. *Nature Genetics* 1, 348-353.

Meitner, L. and M. Delbrück (1935). *Der Aufbau Der Atomkerne: Natürliche und Künstliche Kernumwandlungen. (The Structure of Atomic Nuclei: Natural and Artificial Nuclear Transformations).* Berlin: Springer.

Meitner, L. and O. Frisch (1939). Disintegration of Uranium by neutrons: A new type of nuclear reaction. *Nature* 143, 239-240.

Meitner, L. and H. Kösters (1933). Über die Streuung kurzwelliger γ-Strahlen (On the scattering of short-wave γ-rays). *Zeitschrift für Physik* 84 (3-4), 137-144.

Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review* 69, 213-238.

Moya-Anegón, F., B. Vargas-Quesada, Z. Chinchilla-Rodríguez, E. Correra-Álvarez, F. J. Munoz-Fernández and V. Herrero-Solano (2007). Visualizing the marrow of science. *Journal of the American Society for Information Science and Technology* 58, 2167-2179.

Mount, S. M., C. Burks, G. Herts, G. D. Stormo, O. White and C. Fields (1992). Splicing signals in *Drosophila*: Intron size, information content, and consensus sequences. *Nucleic Acids Research* 20, 4255-4262.

Murray, M. N., H. G. Hansma, M. Bezanilla, T. Sano, D. F. Ogletree, W. Kolbe, C. L. Smith, C. R. Cantor, S. Spengler and P. K. Hansma (1993). Atomic force microscopy of biochemically tagged DNA. *Proceedings of the National Academy of Sciences USA* 90, 3811-3814.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences USA* 98, 404-409.

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences USA* 101, 5200-5205.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences USA* 103, 8577-8582.

Newman, M. E. J., S. Forrest and J. Balthrop (2002). Email networks and the spread of computer viruses. *Physical Review E* 66, 035101.

Pachter, L., S. Batzoglou, V. I. Spitkovsky, E. Banks, E. S. Lander, D. J. Kleitman, and B. Berger (1999). A dictionary-based approach for gene annotation. *Journal of Computational Biology* 6, 419-430.

Porter, A. L. and I. Rafols (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics* 81, 719-745.

Rafols, I., A. L. Porter and L. Leydesdorff (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the Association for Information Science and Technology* 61, 1871-1887.

Roberts, L., R. J. Davenport, E. Pennisi and E. Marshall (2001). A history of the Human Genome Project. *Science* 291, 1195.

Rosen, D. R., T. Siddique, D. Patterson *et al.* (33 co-authors) (1993). Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 362, 59-62.

Sejnowski, T. J., C. Koch and P. S. Churchland (2988). Computational neuroscience. *Science* 241, 1299-1306.

Simon, H. A., R. E. Valdés-Pérez and D. H. Sleeman (1997). Scientific discovery and simplicity of method. *Artificial Intelligence* 91, 177-181.

Sleeman, D. H. and R. J. Hendley (1979). ACE: A system which analyses complex explanations. *International Journal of Man-Machine Studies* 11, 125-144.

Smith, C. L., S. K. Lawrance, G. A. Gillespie, C. R. Cantor, S. M. Weissman and F. S. Collins (1987). Strategies for mapping and cloning macroregions of mammalian genomes. *Methods in Enzymology* 151, 461-489.

Stormo, G. D., T. D. Schneider, L. Gold and A. Ehrenfeucht (1982). Use of the 'Perceptron' algorithm to distinguish translational initiation sites in *E. coli*. *Nucleic Acids Research* 10, 2997-3011.

van Houten, J., H. G. van Vuren, C. Le Pair and G. Dijkhuis (1983). Migration of physicists to other academic disciplines: Situation in The Netherlands. *Scientometrics* 5, 257-267.

Wallace, M. L., V. Larivière and Y. Gingras (2012). A small world of citations? The influence of collaboration networks on citation practices. *PLoS One* 7, e33339.

Waltz, J. A., B. J. Knowlton, K. J. Holyoak, K. B. Boone, F. S. Mishkin, M. de M. Santos, C. R. Thomas and B. L. Miller (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science* 10, 119-125.

Waterman, M., E. Uberbacher, S. Spengler, F. R. Smith, T. Slezak, R. Robbins, T. Marr, D. Kingsbury, P. Gilna, C. Fields, K. Fasman, D. Davison, M. Cinkosky, P. Cartwright, E. Branscom and H. Berman (1994). Genome informatics I: Community databases. *Journal of Computational Biology* 1, 173-190.

Watson, J. D. and R. M. Cook-Deegan (1991). Origins of the Human Genome Project. *FASEB Journal* 5, 8-11.

Watson, J. D. and F. R. C. Crick (1953). Molecular structure of nucleic acids. *Nature* 171, 737-738.

Wheeler, J. A. and R. P. Feynman (1945). Interaction with the absorber as the mechanism of radiation. *Reviews of Modern Physics* 17(2-3), 157-181.

Wiberg, J. S., M. L. Dirksen, R. H. Epstein, S. E. Luria and J. M. Buchanan (1962). Early enzyme synthesis and its control in *E. coli* infected with some amber mutants of bacteriophage T4. *Proceedings of the National Academy of Sciences USA* 48, 293-302.

Womelsdorf, T., J.-M. Schoffelen, R. Oostenveld, W. Singer, R. Desimone, A. K. Engel and P. Fries (2007). Modulation of neuronal interactions through neuronal synchronization. *Science* 316, 1609-1612.

Zhang, L., J. A. Barnden, R. J. Hendley and A. M. Wallington (2006). Exploitation in affect detection in improvisational e-drama. *Lecture Notes in Computer Science* 4133, 68-79.