Journal of Informetrics 3 (2009) 191–209

Contents lists available at ScienceDirect

Journal of Informetrics

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

Towards an explanatory and computational theory of scientific discovery☆

Chaomei Chen[a,b,*], Yue Chen[b], Mark Horowitz[a], Haiyan Hou[b], Zeyuan Liu[b], Donald Pellegrino[a]

[a] College of Information Science and Technology, Drexel University, USA
[b] The WISE Lab, Dalian University of Technology, China

ARTICLE INFO

Article history:
Received 1 September 2008
Received in revised form 12 February 2009
Accepted 17 March 2009

Keywords:
Theory of scientific discovery
Transformative scientific discoveries
Theory of structural holes
Intellectual brokerage
Knowledge diffusion
Information foraging

ABSTRACT

We propose an explanatory and computational theory of transformative discoveries in science. The theory is derived from a recurring theme found in a diverse range of scientific change, scientific discovery, and knowledge diffusion theories in philosophy of science, sociology of science, social network analysis, and information science. The theory extends the concept of structural holes from social networks to a broader range of associative networks found in science studies, especially including networks that reflect underlying intellectual structures such as co-citation networks and collaboration networks. The central premise is that connecting otherwise disparate patches of knowledge is a valuable mechanism of creative thinking in general and transformative scientific discovery in particular. In addition, the premise consistently explains the value of connecting people from different disciplinary specialties. The theory not only explains the nature of transformative discoveries in terms of the brokerage mechanism but also characterizes the subsequent diffusion process as optimal information foraging in a problem space. Complementary to epidemiological models of diffusion, foraging-based conceptualizations offer a unified framework for arriving at insightful discoveries and optimizing subsequent pathways of search in a problem space. Structural and temporal properties of potentially high-impact scientific discoveries are derived from the theory to characterize the emergence and evolution of intellectual networks of a field. Two Nobel Prize winning discoveries, the discovery of *Helicobacter pylori* and gene targeting techniques, and a discovery in string theory demonstrated such properties. Connections to and differences from existing approaches are discussed. The primary value of the theory is that it provides not only a computational model of intellectual growth, but also concrete and constructive explanations of where one may find insightful inspirations for transformative scientific discoveries.

1. Introduction

The intellectual structure of a scientific field is an abstraction of the collective knowledge of scientists in the field, including scholarly publications and other forms of intellectual assets. *Scientific change* refers to profound changes of the intellectual

☆ CiteSpace is freely available at http://cluster.cis.drexel.edu/~cchen/citespace. Color versions of the figures in this article are available at http://cluster.cis.drexel.edu/~cchen/papers/JOI.

* Corresponding author at: College of Information Science and Technology, Drexel University, Philadelphia, PA 19104, USA.
E-mail address: chaomei.chen@cis.drexel.edu (C. Chen).

1751-1577/$ – see front matter
doi:10.1016/j.joi.2009.03.004

structure of a field. In this article, we will focus on the nature and key mechanisms of scientific discoveries that could lead to such fundamental changes—*transformative scientific discoveries*.

The nature of scientific change has been studied from many distinct perspectives, notably including philosophy of science (Collins, 1998; Laudan et al., 1986; Schaffner, 1992), sociology (Fuchs, 1993; Griffith & Mullins, 1977), and history of science (Brannigan & Wanner, 1983). Quantitative studies of the topic can be found in the fields of scientometrics, citation analysis, and information science in general (Chen, 2003; Heinze & Bauer, 2007; Heinze, Shapira, Senker, & Kuhlmann, 2007; Hummon & Doreian, 1989; Small & Crane, 1979; Sullivan, Koester, White, & Kern, 1980; Wagner-Dobler, 1999). Scientific literature has increasingly become one of the most essential sources for these studies. Social network analysis and complex network analysis also provides valuable perspective (Barabási et al., 2002; Newman, 2001; Redner, 2004; Snijders, 2001; Valente, 1996; Wasserman & Faust, 1994).

What do these diverse perspectives have in common and how do they differ in terms of their views of scientific change, scientific discovery, and knowledge diffusion? In this article, we introduce an explanatory and computational theory of scientific discovery as a key step for understanding and explaining the emergence and evolution of the intellectual structure of a field. We are motivated for a number of reasons. First, despite of the perceived role of serendipity and other unpredictable factors, it is evident that an important subset of scientific discoveries shares important and generic properties (Bradshaw, Langley, & Simon, 1983; Simon, Langley, & Bradshaw, 1981). In order to obtain conclusive evidence, one will need a theory of scientific discovery that can provide a unifying conceptual framework so that one can characterize a variety of scientific discoveries from a consistent perspective. Second, given one concrete case of scientific discovery, it may be studied from multiple and often not interconnected perspectives. For example, a philosophical study of a scientific revolution may have little overlap with a sociological study of the same process. Even two philosophical studies of the same scientific revolution could appear to be unrelated in the eyes of laypersons. We need a theory that can not only explain scientific change, but also relate to various existing theories. Third, statistical models of network evolution have been used to identify statistical and topological properties of scientific networks. However, such properties, although generic in nature, do not readily offer further explanations of why scientists in a network behave in a particular way. Motivations, decisions, and interpretations underlying such properties are often detached or left out. Thus, we need a theory that not only identifies statistical and topological properties of scientific networks, but also offers practical insights into the mechanisms that may drive scientists' observed behavioral patterns. The work described in this article is the first step towards this long-term goal.

There are many types of theories, including descriptive, explanatory, generative, predictive, and prescriptive (Bederson & Shneiderman, 2003). Our immediate goal is to develop a simple, descriptive, explanatory, and generative theory of scientific discovery. We are interested in identifying some generic mechanisms of discovery in order to explain transformative scientific discoveries to begin with and other types of discoveries later on. Such generic mechanisms are in fact generative in nature because scientists and computer simulation algorithms would be able to emulate such mechanisms. We have a few expectations of our new theory. First, it should help us to recognize the significance of new discoveries as soon as possible. Second, it should help us to identify as many potential areas of growth as possible. Third, it should help us to explain both the creation of knowledge and its diffusion within a consistent and unified framework.

The rest of the article is organized as follows. We will first review existing conceptualizations of scientific change in the philosophy of science, sociological theories of scientific change, sociological theories of creative ideas, information foraging theory, and a recurring theme among these various views. The recurring theme is, simply speaking, that insights, creative ideas, and transformative scientific discoveries are the work of a broad range of brokerage mechanisms. Next, we will expand the recurring theme and construct a simple theory of scientific discovery to explain the growth of a scientific field. We will then describe conjectures that one can derive from the first principles of the theory, including structural and temporal properties of citation and co-citation networks. We will include a brief analysis of Nobel Prize winning discoveries as illustrative cases. Finally, we will outline ongoing and future work, including large-scale computer simulation and a wider range of high-impact scientific discoveries.

2. Existing conceptualizations of scientific change

2.1. Specialties and scientific change

Specialty is a key concept in the study of scientific change. A specialty is a group of researchers and practitioners who have similar training, attend the same conferences, read and cite the same body of literature (Fuchs, 1993). There are a variety of studies of specialty in the literature (Chubin, 1976; Fuchs, 1993; Morris & Van der Veer Martens, 2008; Mullins, Hargens, Hecht, & Kick, 1977; Small & Crane, 1979). For example, Mullins et al. studied author groups corresponding to co-citation clusters using questionnaires and concluded that co-citation clusters indeed represent the intellectual structure and that co-authors do form social groups (Mullins et al., 1977). Co-author networks have also been studied in complex network analysis of community structures (Girvan & Newman, 2002). These finding provide an empirical basis for the analysis of scientific change based on co-citation networks as we shall introduce later in this article.

The dynamics of the structure of a specialty is a central issue in the context of scientific change. Research has shown that major changes in a variety of disciplines tend to be originated within small, socially coherent groups (Griffith & Mullins, 1977). Kuhn observed that new paradigms are typically initiated by young scientists or newcomers to a crisis-laden field (Kuhn, 1962). In addition, Crane (1972) found that the desire for originality motivates scientists to maintain contacts with