Technological Forecasting & Social Change 81 (2014) 67–81

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Visualizing geo-spatial data in science, technology and innovation

Jan H. Kwakkel [a,*], Stephen Carley [b], John Chase [c], Scott W. Cunningham [a]

[a] Faculty of Technology, Policy, and Management, Delft University of Technology, 2600 GA, Delft, The Netherlands
[b] School of Public Policy, Georgia Institute of Technology, Atlanta, GA 30032, USA
[c] SRI International, Arlington, VA 22209, USA

ARTICLE INFO

Article history:
Received 10 April 2012
Received in revised form 17 August 2012
Accepted 13 September 2012
Available online 24 October 2012

Keywords:
Tech mining
Geo-visualization
Technological forecasting

ABSTRACT

This paper provides a range of alternatives for analysts when dealing with geo-spatial data addressing activities in science, technology and innovation. There are two sets of theory which frame the debate concerning science, technology and innovation, and which drive further methodological advancements. The power of a good visualization in providing insight to decision-makers is well known. Visualizations complete a full cycle of decision-making involving analysis, design, action and then further monitoring. We advance the paper by assessing available geographic information in science and technology databases. The paper then systematically outlines current best practices and alternatives for visualizing geographic data. Different geographic map options provide different possibilities for the display of data. We show some of these options in the paper. Future research is needed into both the available tools and techniques, as well as a more in depth specification of the kinds of decision support needs that exist and have a geo-spatial component.

1. Introduction

Over the last two decades, there has been a renewed and growing interest in the study of innovation at a regional level. This has been named 'the new regional science' [1,2]. Along a different line, there also has been a growing interest in the study of social and economic networks and their influence on innovation [3,4]. Separate from these developments in these respective scientific fields, there is an ever expanding amount of data becoming available to the analyst. In particular with the rapid rise of the internet over the last decade, researchers interested in innovation have gained access to a wide variety of new data sources. Moreover, more traditional data sources like science and patent databases have been improving the amount of data they provide, for example by adding detailed affiliation and funding data to their records.

This paper explores how these new data sources can be of use to researchers studying innovation. The two research fields which frame the debate concerning science, technology and innovation drive further methodological advancements. In particular, the new regional science highlights the importance of the spatial character of innovation with its emphasis on physical distance, localized spillovers, and the local embeddedness of firms and other organizations. This thus necessitates that the researchers explicitly take into account the geospatial aspects of their data, which raises a wide variety of questions related to visualization and analysis. Similarly, the social and economic network literature is moving towards studying the dynamics of network formation over time, more granularity in the character of links between nodes, and accounting for the level of analysis [4]. This also necessitates additional data collection and analysis capabilities.

This paper provides a range of alternatives for analysts when dealing with geo-spatial data addressing activities in science, technology and innovation (STI). We address the theoretical role of regions and districts in innovation policy. We then discuss an

* Corresponding author. Tel.: +31 15 27 88487.
E-mail address: j.h.kwakkel@tudelft.nl (J.H. Kwakkel).

0040-1625/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.techfore.2012.09.007

evolving body of analytical methods for addressing theory and delivering useful policy advice. These analytical methods may be implemented using a range of tools. We discuss open source scripting languages and libraries in this paper.

We begin with a brief survey describing how and why geography shapes and structures innovation systems. From this survey we advance a set of requirements or needs for improved information systems for supporting policy. In Section 3, we confront the policy need with a survey of the available data. Section 4 details the design options available to analysts seeking to introduce visual and spatial quantities into their innovation analyses. Section 5 provides specific examples illustrating the potential range and applicability of visualization and geographical information support for innovation policy. Section 6 contains our concluding remarks, including a discussion on some of the major limitations of geo-visualization.

2. Theories of regional innovation

Before examining in detail these methodological, and data concerns, we first present two major theories of economic and innovative activities, presenting these theories as a framing device for the rest of the paper. There are two sets of theories which frame the debate concerning science, technology and innovation. The first set of theories of note contains theories on national and regional systems of innovation [1]. The second set of theories consists of theories on social and economic networks [3].

2.1. RIS theory

Like any other system, an innovation system consists of discrete units and the structural relationships occurring between these units. Because this is an innovation system under investigation, the system is organized to produce and utilize science and technology for the individual and collective betterment of society. There is a wide variety of concepts used to denote this innovation system, including, national systems of innovations [5,6], regional systems of innovations [1,2], industrial [7] or technological districts [8], learning regions [9,10], industrial atmosphere [11], clusters [12], the French *filière* approach [13], etc. Without intending to suggest that all these concepts are essentially the same, we would like to draw attention to what is shared across this wide and diverse literature. This literature emphasizes the importance of the institutional setting of norms, routines, etc. that provide an organizational-support infrastructure for economic competitiveness, the importance of informal networks grounded in trust as well as more formal organizations and mechanisms for sustaining this trust in reducing transaction costs among organizations, and the importance of institutional and organizational learning [2]. In addition, this literature is interested in the role of geographical proximity as a facilitator for the rapid dissemination of tacit knowledge and other externalities. The literature itself is torn between whether the object of study is the formal relationships between the parts, or the actual operational character of the system. The literature is also torn between whether the system is emergent, or the consciously designed product of policy [1]. This diverse literature draws on various heterodox branches of economics, including evolutionary, institutional, and innovation economics. An important question in this literature is the geospatial boundary that is used, this can be drawn at a national level such as is done in the literature on national innovation systems, but it can also be drawn on a regional level. However, the literature on regional innovation systems remains vague on how a region should be defined. As stated by Cook et al. [2] there are both an administrative and a cultural dimension to a region. They clearly recognize that a region emerges over time. This emergence is structured both by bottom up forces where cultural or other forces in an area result in political demands for regional administrative and/or political organizations (regionalism), and by top down forces where the state or another supraordinate body delimits a supralocal territory (regionalization). They argue that for an area to constitute a region, it should show internal homogeneity, cohesiveness, and identity, and the designation should distinguish an area in a specific way from other areas.

An early and influential study by Storper [14] surveys theories of high technology trade. The author describes the need for unlocking the competitive dynamics of regions and then provides several international case studies to anchor his claims. Keeble and co-authors [15] provide an institutional analysis of the Cambridge region. McEvily and Zaheer [16] argue that regional networks provide firms with economic linkages which sustain industrial competitiveness. A questionnaire, targeted at firms, is developed to better investigate their questions. Both physical distance and regional variables have been used in the investigation of innovation systems and networks. This work lends itself to the use of archival sources of data such as patent databases or firm-level databases. Baptista and Swann [17] investigate the proposition that firms in clusters are more innovative. They use an archival analysis based on a database of British firms, and conclude that there are positive within industry effects. Such benefits do not necessarily extend between industries. Almeida and Kogut [18] use patent data to study the mobility of inventors. They conclude that only certain regions (Silicon Valley) gain localization effects. Very similar in spirit, the study of Rosenkopf and Almeida [19] examines the interaction of firm alliances and inventor mobility in meeting the knowledge requirements of firms. Their hypotheses are anchored by citation analyses of patents. Cunningham and Werker analyze the role of physical and technical proximity in the production of successful collaborations in nanotechnology research.

Acz et al. [20] ask fundamental questions regarding the measurement of innovation at various scales of aggregation, using publication and innovation count data. Carlsson et al. [21] cite three major methodological critiques of the work. The first issue is a matter of scale or level of analysis — is the unit of analysis the technology, the product, the firm, or the geographic area? What are the boundaries of the system in time and space? The second issue is therefore how to determine the population, delineating the actors and the components. The third issue is how to measure the performance of the system, and more particularly, how to evaluate system performance in exclusion of any of the components. How sufficient are these methods and theories for supporting policy-making? Martin and Sunley [22] note the pre-eminence of the "cluster" concept, rooted in the work of Porter [23]. This concept has been a motivating factor driving the formation of local, regional and industrial policies the world over. The authors note the various problems with the theory, and recommend the circumspect application of these theories. Likewise,