ARTICLE IN PRESS

Technovation ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Technovation

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

Using web mining to explore Triple Helix influences on growth in small and mid-size firms

Yin Li [a], Sanjay Arora [a], Jan Youtie [a,b], Philip Shapira [a,c,*]

[a] School of Public Policy, Georgia Institute of Technology, Atlanta, GA, USA

[b] Enterprise Innovation Institute, Georgia Institute of Technology, Atlanta, GA, USA

[c] Manchester Institute of Innovation Research, Alliance Manchester Business School, University of Manchester, Manchester, UK

ARTICLE INFO

Article history:
Received 4 November 2014
Received in revised form
2 December 2015
Accepted 3 January 2016

Keywords:
Triple Helix
Micro-level relationships
Small and medium enterprises
SME
Growth
Website data

ABSTRACT

While broad "Triple Helix" frameworks of industry, government and university collaborations have the potential to enhance innovation and economic development at macro-levels, at the micro-level of the firm it should not be assumed that such relationships are uniform in character or outcomes. Each firm will negotiate and develop its own set of relationships with other innovation system actors based on its capabilities and strategies. To better understand these dynamics, particularly from the perspective of small and medium-sized enterprises, this study probes the micro-level characteristics and impacts of external enterprise relationships. Novel website-based Triple Helix measures are introduced that extend the analytical scope beyond customary indicators (such as patent analysis or entropy measures) to include communication and coordination among all three helices at the micro-level of individual firms. This approach is used to explore the micro-level characteristics and impacts of industry, government and university relations for small and medium-sized enterprises by analyzing a subset of 271 U.S. green goods small and medium-sized manufacturing enterprises. We compare the website-based measures with case study results to authenticate the method. A panel data regression model is then employed to analyze the simultaneous impacts of various combinations of industry, government and university links on firm sales growth (2008–2011), with controls for region, scale, and application domains. The ability of website-based indicators to distinguish the impacts of different mixes of Triple Helix relations is demonstrated. While relationships with all three helices have a positive total marginal effect on firm sales growth, local relationships and relationships that emphasize links with government and industry make particularly notable contributions to growth in the sample green goods enterprises. The implications of these findings are discussed.

1. Introduction

Over the last three decades, there has been much discussion of the role of university, industry, and government collaborations in advancing innovation and economic development. A major stimulator to this discussion has been the concept of the "Triple Helix" of university–industry–government relationships and the importance of such linkages in the development of dynamic, knowledge based innovation systems (Leydesdorff and Etzkowitz, 1996; Etzkowitz and Leydesdorff, 2000; Etzkowitz, 2008). As Etzkowitz and Leydesdorff argue, "university–industry–government network relations are the key to knowledge-based economic development in a broad range of post-laissez-faire capitalist and post-socialist societies" (Etzkowitz and Leydesdorff, 2000).

There is now an array of published work, including in management, economics, regional development, technology transfer, and innovation policy, which references the Triple Helix framework and comparable ideas (see, for example, Carayannis et al., 2012) of relationships among key innovation system actors. This body of literature includes work from both institutional and evolutionary perspectives (Meyer et al., 2014; Ranga and Etzkowitz, 2013). Within each of these strands, there has been a growing interest in quantitative studies of university, industry and government roles and relationships. Efforts have been made to measure specific types of linkages and institutional contributions, such as science–technology linkages (Acosta and Coronado, 2003; Callaert et al., 2006; Meyer et al., 2003) and entrepreneurial universities (Baldini, 2006; Uranga et al., 2007; Acosta et al., 2009; Iversen et al., 2007). Others have developed entropy measures of system level dynamics (Leydesdorff, 2003, 2008; Leydesdorff and Meyer, 2006, 2007). This literature on Triple Helix indicators and empirics, however, has yet to address some of the fundamental

* Corresponding author at: Manchester Institute of Innovation Research, Alliance Manchester Business School, University of Manchester, Manchester, UK.
E-mail address: pshapira@manchester.ac.uk (P. Shapira).

http://dx.doi.org/10.1016/j.technovation.2016.01.002

issues of the Triple Helix framework, including quantitative assessment of the impacts of Triple Helix relationships (Meyer et al., 2014).

In this paper, we conceptualize Triple Helix relationships as sets of communication and coordination networks among business, government, and universities. We investigate research questions concerning how the growth of small firms is influenced by involvement in, and linkages to, Triple Helix actors; and how such influences might differ by sector/helix (e.g., government, university, business) and at different levels of innovation systems (i.e. regional, national and global). We introduce a novel method of exploring indicators of relationships through the analysis of information reported on company and other websites. Our analysis builds from micro-foundations – by investigating the variety and patterns of relationships described by individual firms. The key to our analytical approach is the development of a set of website indicators of the intensity of company signaling relationships with universities, government agencies, and other firms as a measurement of their engagement in Triple Helix communication and coordination relationships. This approach has two advantages over existing Triple Helix indicators based on patent analysis or entropy measures. First, it is able to represent, at the level of micro actors, several otherwise undisclosed information pathways, particularly those important for flows of tacit knowledge and coordination information. This kind of information is not well captured by patents (which offer more formal reports of knowledge relationships but cover only that small portion of knowledge acquisition and collaboration deemed worthy of a patent application). Second, these indicators measure micro-level relations from the perspective of the firm, focusing on the unit of analysis that is central to the process of creating economic benefits such as business growth. As we will see, using web-based indicators of Triple Helix relationships has added power when combined with other available micro-level measures of business performance.

We apply the web mining method to study how Triple Helix links influence the development of small and medium-sized enterprises (SMEs) in green goods manufacturing in the United States. Green goods manufacturers comprise firms who produce and market manufactured items that have environmental or natural resource benefits when used by other businesses, organizations or households (Shapira et al., 2013). The products of green goods manufacturers are varied, ranging from energy and transportation equipment to recycling and environmental control systems. Green goods manufacturers may develop and apply innovations in products, processes, and business models to improve efficiency, quality, performance, and value to users, and to maintain competitiveness. The green goods domain has garnered much recent policy attention not only for its environmental and resource benefits but also for its potential to re-generate regional and national economies, although there are indications that the promise of large numbers of "green jobs" in the near term has not materialized (Muro et al., 2011). Communication and coordination between government, business, and academic actors has been deemed critical for the success of this domain and for the realization of widespread economic benefits (Ogden et al., 2008; Cooke, 2011).

Green goods SMEs present a particularly interesting case for studying the Triple Helix model of innovation. Innovative green goods technologies are often first developed by new start-up companies. Knowledge for innovation may be obtained through linkages with universities, firms, and government agencies, as well as generated internally or acquired from other sources. External linkages are likely to be important since SMEs (by virtue of their smallness) typically lack all the knowledge and resources essential for growth. Yet, while links with entrepreneurial universities may be helpful, it should not be presumed that only these links are sufficient or that university linkages are more dominant. All three helices (industry, university and government) potentially have critical roles. The green goods economy uses a mix of current and new technologies, so that successful innovation requires not only university technology transfer, but also firm linkages with suppliers, customers and other private sector firms in the value chain to integrate technology, access finance, scale up product manufacturing, offer related services, and deal with logistics and marketing. Government and other public-oriented organizations perform roles in the governance and prioritization of green technologies, research sponsorship, business assistance, and demand stimulation (for example through public procurement). To assess the influence of such aspects requires simultaneous micro-level measurements of relations among multiple actors – and we will demonstrate in this paper how such measurement can be aided by website-based indicators.

In the next section of the paper, we explain the use of website-based measures, compare these measures to existing indicators, and validate the measures through comparisons with case study results. We then apply the web-based measures to study our sample of U.S. green goods SMEs, using these measures to explore several of the core questions central to the Triple Helix model, including: how do university, industry, and governmental relations benefit firm growth? How do the roles of the three types of actors differ? And how does geographical proximity mediate the impacts of the relations? A panel data econometric model is used to explore these relationships. The paper closes with a discussion of the advantages of the website-based method, its limitations, and management, policy and research insights.

2. Measuring triple helix relationships: a web-based approach

Recent studies of the Triple Helix concept have pursued two strands: "neo-institutional" and "neo-evolutionary" (Meyer et al., 2014). The neo-institutional approach focuses on networking and exchanges between different institutions and organizations. The neo-evolutionary model is concerned about evolutionary and exchange mechanisms between the three functions of the helices and issues of knowledge production, wealth creation, and normative control.

The quantitative Triple Helix indicator literature with an underlying neo-institutional approach has focused on measuring either the linkages between science and technology or capturing entrepreneurialism in universities. In measuring science–technology linkages, available work uses patent citation analysis as a key analytical tool to explore the correlations between science intensity and technological productivity (Acosta and Coronado, 2003; Callaert et al., 2006; Van Looy et al., 2007). In capturing entrepreneurialism in universities, the literature uses patents, the number of spinoffs, the presence of knowledge transfer practices, and outcomes in regional economic development as measures of the third mission activity of universities (e.g. Baldini, 2006; Uranga et al., 2007; Acosta et al., 2009; Iversen et al., 2007). The neo-evolutionary strand of Triple Helix indicator literature focuses on capturing information, knowledge and meaning flows in the innovation system, primarily using entropy measures (Leydesdorff 2003, 2008; Leydesdorff and Meyer, 2006, 2007). Empirical literature in this strand has used entropy measures in combination with other scientometric and technometric indicators to examine the knowledge base (Leydesdorff and Fritsch, 2006), national systems of innovation (Park et al., 2005) and regional innovation systems (Vilanova and Leydesdorff, 2001).

While these studies offer insights on university, industry, and governmental relationships across a range of topics, the indicators they use have limitations. The two strands in the Triple Helix

Please cite this article as: Li, Y., et al., Using web mining to explore Triple Helix influences on growth in small and mid-size firms. Technovation (2016), http://dx.doi.org/10.1016/j.technovation.2016.01.002

indicator literature under-emphasize questions relating to economic benefits as a result of institutional and organizational interactions over time. Additionally, the neo-institutional approach has a tendency to focus on universities and to use patenting, spinouts, licensing, and other outcome measures. These measures and analyses are not able to fully capture flows of information and knowledge, particularly tacit knowledge and coordination information, among the helices. Bi-directional interactions between universities and firms also tend to be overlooked in studies that rely on these readily-available indicators (Meyer et al., 2014).

In this paper, we apply a novel analytical approach based on the mining of company and other websites to explore industry, university, and governmental relationships. We apply this approach to SMEs in a targeted domain. Our approach investigates the communication, exchange and coordination relations of the individual firm with universities, other firms, and government. The focus on the networking aspects of the Triple Helix model takes us closer to the neo-institutional strand, but our approach is differentiated by the analytical focus on firms as micro-level actors rather centering on the entrepreneurial university or individual patents. For example, to measure the strength of a company's relationship with universities and other higher education institutions, we measure how often the company mentions certain keywords such as *university*, *college*, *institute*, *academy* on its website. Technically, this is implemented through specifying a language model in web mining computer software, inputting the set of keywords to be applied to the downloaded website data, and using the software to compute and generate the indicator *university* (see Section 3 for technical details). To measure firm relationships with other private sector actors, we identify two classes of possible relationships. First, we measure reported inter-firm relationships. The literature on firm networks discusses informal and formal linkages as a way to achieve vertical disintegration and increasing specialization, access complimentary assets, incorporate and build upon knowledge, and experience positive performance outcomes (Ahuja, et al., 2008; Ahuja, 2000; Borgatti and Foster, 2003; Morgan, 2004; Simmie, 2005). This literature suggests a variety of constructs which are operationalized through keywords that convey an industry membership or partnership focus, including *distributor, joint venture, supplier, affiliation, and association*. Second, we recognize the importance of financial relationships in enterprise growth. Entrepreneurship scholars emphasize the role of venture capital and other equity investment as one way for small startups to overcome resource scarcity and the inability to secure collateral-based lending or debt financing (Amatucci and Sohl, 2007; Auersald, 2007). This leads to a range of finance-related keywords including *venture capital, private equity, private placement, and seed funding*.

Our use of website-based indicators adds a range of measures of communication, coordination, and other activities that are reported openly and online by firms. These indicators complement but also go beyond conventional secondary measures. Indeed, one of the motivations to pursue this approach is to overcome the limitations of publications and patenting as proxies for industrial research and innovation processes. Small and medium-sized manufacturing companies in modern technology-oriented sectors often undertake developmental and innovation activities that are not captured by publication or patent measures. For example, in our sample of 300 green goods U.S. companies, only 10% of companies have produced scientific publications, while 19% of companies undertake patenting. In contrast, two-thirds of the companies' websites report text that suggests they are involved in diverse kinds of research and development (R&D) activities, including the use of terms related to science and laboratories, trials, demonstrations and other R&D activities. This difference between formal indicators of publication and patenting and the other R&D activities reported on company websites is consistent with observations that manufacturers in engineering-intensive sectors relying heavily on embedding tacit knowledge in process and products (Wong and Radcliffe, 2000; Asheim and Coenen, 2005). Coupled with the availability of firm-level sales growth data (from business databases), website-based indicators present reports of a series of measures of external relationships, R&D, and other activities that offer the potential to examine the economic benefits of Triple Helix linkages from a micro-level firm perspective.

It is, of course, important to probe the robustness and exactitude of website-based indicators in capturing what firms actually do. As with other methods, there are limitations as well as advantages (see discussion in Gök et al., 2015). Websites are freely accessible, so reports of confidential, proprietary or negative information cannot be expected. At the same time, authentic companies seek to present themselves to potential customers, suppliers, investors, and other partners, and to maintain their reputations, so there are internal drivers and external influences that encourage veracity in web-reported information. There is an emerging area of inquiry that is establishing web mining as a source of information with validity. Experience has been gained in using mining results from current and historical websites for content analysis in various contexts, and it has been found to produce information with content validity (Murphy et al., 2007). Recently, social scientists have explored the methodological aspects of web scraping and its robustness (Youtie et al., 2012, Shapira et al., 2013; Arora et al., 2013, 2015; Gök et al., 2015). Stuart et al. (2007) examined the metrics obtained from hyperlinks on university websites to co-authorship, finding that significant numbers of links did reflect collaborative relationships. Gök et al. (2015) compared web mining with several other conventional ways of obtaining information on business (especially SME) R&D activities. They find that website data offers advantages in coverage, currency, accessibility, quantity and flexibility, although these benefits may be potentially disadvantaged in consistency, interpretability and handling. Overall, this work suggests that website information can yield additional insights not easily obtainable from databases of publications and patents.

We observe that small companies disclose their collaborative relationships with governments, universities and other industrial entities on their websites. This includes information about co-developed projects and products with universities or partnering with other companies. Small companies have sufficient motivations to disclose such relationships because such information represents a positive signal of competitiveness and website disclosure is an available and inexpensive channel of dissemination. Previous studies in the innovation literature have shown that firms are willing to spend corporate resources on similar activities, such publishing R&D results in scientific journals, to signal competences (Allen, 1983; Hicks, 1995; Muller and Penin, 2006). These signaling strategies are not limited to resourceful large firms, but are increasingly adopted by small- and mid-sized firms (Li et al., 2015). A parallel study, using text mining results of company websites over a period of time, demonstrates a relationship between website contents and the strategic behaviors of SMEs (Arora et al., 2015). We acknowledge that there is a stream of marketing and e-commerce literature suggesting companies can manipulate information on the internet through the strategic use of social media and online forums (Dellarocas 2003, 2006; Godes et al., 2005). However, these findings have limited applicability to the present study because of our focus on unidirectional company websites rather than interactive online communications such as social media or online forums. Moreover, many of the actors in these market studies are large corporations while the companies in our sample are SMEs which, by definition, have less power and scale to enable substantial market influence. Large corporations' websites—

Please cite this article as: Li, Y., et al., Using web mining to explore Triple Helix influences on growth in small and mid-size firms. Technovation (2016), http://dx.doi.org/10.1016/j.technovation.2016.01.002

Table 1
Company histories compared with website measurement.
Source: Manual review of company history (including articles in Factiva) compared with website based keyword count measures (see Section 3.2). Match between history and website indicators corroborated (=✓) or not corroborated (=✗).

Relationships	Company A			Company B			Company C			Company D		
	History	Website	Match	History	Website	Match	History	Website	Match	History	Website	Match
University	University Spinoff	High	✓	University Spinoff	High	✓	None	Near Zero	✓	University collaboration	High	✓
Industry	Part of an industry group	High	✓	VC funded; member of a regional association	High	✓	Corporate spinoff, VC funds, IPO, engaged in M&A	High	✓	Corporate spinoff, VC funds, IPO, engaged in M&A	Low	✗
Government	DOE grant	High	✓	DOE grant	High	✓	None	Near Zero	✓	None	None	✓

typically including annual reports, corporate responsibility text, descriptive information, and news articles—are frequently oriented to public audiences, while SME websites often present strategic-oriented text about products, partners, and customers. For example, relationships with a university could result in repeated entries in the company's websites in sections such as “events”, “about the company”, “products” and so on. This suggests that SMEs' websites can be mined to pick up information about their strategies, products, and relationships.

To further confirm the validity of website-based indicators of SME relationships with universities, industry and government, we compared the website-based measurements and results from company case studies in a sample of U.S. green goods manufacturers (see the next section for details about sampling strategy). For four cases, we collected information about company history through reading their corporate websites and searched relevant business news in the Factiva database. In general, the comparison suggests that the website-based measurement and the case studies support one another (Table 1). We illustrate the comparison in the following four representative cases. Company A is a battery manufacturer, a spinoff from Massachusetts Institute of Technology, which became part of an established industry group, and received a Department of Energy (DOE) grant to conduct R&D. The three measures of company A's website variables, “University”, “Industry”, and “Government” are all higher than average. A Georgia-based solar panel manufacturer, Company B, is a university spinoff with funding from venture capitalists. It engages in membership in a regional industry association, and it received a DOE grant. These activities were corroborated by the high scores on Triple Index indicators generated through the website analysis. In contrast, Company C engaged in various activities within the industry sphere, including being a corporate spinoff, funded by venture capital, and engaging in acquisition, while having little connections with the university or government. Not surprisingly, Company C scores high on the Triple Helix “industry” indicator, and nearly zero on the university and government related website measurements. Company D is a particularly interesting case. The company has a history in collaborating with a local university in product development, and it does score high on the “university” Triple Helix indicator. The company does not have any reported government connections. It also scores low on the “industry” indicator. A manual search on the founder's name shows that he is a serial entrepreneur; thus Company D is likely to be connected with various players in the industry through the founder's personal entrepreneurship connections. While website measurement validly captured the firm's position on two other dimensions, industry links were underestimated as the firm did not explicitly signal these relationships of its founder on its corporate pages.

While not without imperfections, these case observations coupled with the evidence from the extant literature indicate that it is feasible and valid to use website-based indicators to identify and measure the intensity of SME's university, industry and governmental linkages. Overall, we anticipate that web-based indicators as signals for communication and coordination activities are likely to be valid in measuring the presence of links depending on a company's willingness to disclose certain relationships on its web pages. The next part of the paper presents our empirical study design, method and results, including how we deploy, measure and interpret these web-based indicators of Triple Helix relationship.

3. Exploring micro-level relations in the growth of U.S. green goods SMEs

The following section considers the background to and context of the green goods manufacturing domain targeted in the study and elaborates our enterprise sampling approach. We then relate further details about how website mining is used to develop micro-level indicators of enterprise Triple Helix relationships and how these measures are coupled with other secondary enterprise data. We then discuss the econometric model used to assess the influence of these relationships on business growth and discussion of the results.

3.1. Background to the U.S. green goods industry

Green growth, according to Organization for Economic Cooperation and Development (OECD, 2011), is “fostering economic growth and development while ensuring that natural assets continue to provide the resources and environmental services on which our well-being resides.” Heightened global concerns about environmental sustainability, energy resources, and climate change have led to increased attention being paid to novel green technologies that can underpin green growth, including manufactured products and processes that use energy more efficiently and help to reduce carbon emissions. Additionally, green technologies have been put forward as generators of new employment and contributors to economic and regional rebalancing. This was particularly evident during the 2007–2008 global financial downturn and subsequent economic recession, when green technologies were highlighted as sources for economic renewal in many developed countries and regions, including the United States. In the U.S., there were calls for improved coordination in the development of innovative green technologies. In 2008, for instance, the former White House Chief of Staff under the Clinton Administration, John Podesta, along with two colleagues, wrote in *Issues in Science and Technology* to warn the White House that while “no one questions the need to develop new energy technologies”, the U.S. government needed “an integrated approach to energy technology innovation” (Ogden et al., 2008). Podesta and colleagues sought a coordinated and integrated progress at multiple stages of research, development and deployment (RD&D) ranging from generic research of energy technologies to specific industrial implementations. They contended that such coordination in U.S. green industry subdomains should include government and industry interactions with universities to reflect trends in research and integrate different fields, joint ventures between public and private sectors in R&D and deployment projects, and indirect support for the private sector to commercialize research in universities and federal laboratories. The new strategy must establish "a role for the federal government, industry, universities and laboratories" (Ogden et al., 2008).

In other words, what Podesta and colleagues proposed was a Triple Helix approach to stimulating U.S. green technologies and supporting the manufacturers who make green goods that deploy these technologies. Effective and enhanced networks of communication and coordination among industry, university and government in green goods domains were seen as necessary, because of the needs of collaboration among upstream developers and downstream implementation, the nature of technology requiring interactions among firms and universities, and the role of the government in supporting demand through public procurement. Combined together, the features of green technology industries suggest that they offer an especially interesting case for the study of Triple Helix relationships and their effects on enterprise growth and innovation, for several reasons. First, green goods manufacturers are diverse and draw upon a series of new and established technologies. The technologies deployed by green goods manufacturers range from cutting-edge solar panel and smart grid technologies to mature water and waste management technologies reoriented to an expanded market (PriceWaterhouseCoopers, 2014). While the role of university technology transfer and spin-off entrepreneurship is important, green goods manufacturing firms may tap into a wider range of innovation sources including government labs and industrial partners as well.

Second, while government has direct and indirect influence on innovation and Triple Helix coordination in all economic sectors (Lazonick, 2008; Mazzucato, 2013), it plays particularly important roles in green technology development and deployment. This includes influence on the demand-side (for example, through energy, environmental, housing, urban planning, and transportation policies) and on the supply-side (through green technology targets prioritized for public R&D spending, financial incentives, and business support). Here, the U.S. case is salient. Following inauguration in January 2009, the Obama administration advanced plans to develop cleaner energy sources, reduce greenhouse gas emissions, and generate U.S. green jobs (Council on Foreign Relations, 2009). Support for green technologies and industries was provided through a range of tools including R&D investments from the American Recovery and Reinvestment Act (ARRA, 2009), of which 12% was allocated to the energy area in Title IV of the legislation to build infrastructure and conduct R&D in energy efficiency and renewables. This included grants and facility funding awards for manufacturers of advanced batteries and related electrical systems and components (ARRA, 2009, Title IV, 123 Stat. 138). High-profile failures of clean energy federal loan recipients (such as solar panel makers Solyndra and Abound and the electric vehicle company Fisker) subsequently raised concerns about the U.S. government's ability to successfully support individual green goods manufacturers, even though overall the federal clean energy loan program has returned a profit (Brady, 2014). While this program has been controversial and visible, the enabling roles of federal as well as state and local governments are played through multiple other mechanisms including federal Small Business Innovation Research (SBIR) awards and various state and local initiatives (see CleanEdge, 2015). These government investments and initiatives enable the work of other Triple Helix actors, such as universities and equity.

While industry, university and government all play indispensable roles in developing green goods technologies and manufacturing, in our micro-level analysis the firm is the key unit. Our web-based indicators are helpful in constructing new measures of relationships at the firm level. We seek to explore three central questions that can benefit from further empirical testing. First, does engaging in Triple Helix relationships benefit SME firm growth? While there are numerous case studies of Triple Helix relationships, the empirical literature discussed in the previous section has thus far not been able to link Triple Helix relations to firm growth in a quantitative manner (Meyer et al., 2014). This leads to a related second question: which Triple Helix relationships or combinations are most beneficial? It is commonly contended that the role of the three helices are complementary, but it is difficult to estimate the impact of each relationship, or estimate an optimal mix of relations, without measuring multiple relations at the same time. A third question is about the appropriate level to situate Triple Helix relationships. Are regional actors more important? Or are national or even global actors more important? For example, a recent analysis of six "eco-innovation" regional systems of innovation in six different countries by Cooke (2011) highlighted the role of actors at different levels, particularly at national and city-regional levels, in developing green technologies. Although the roles differ across different types of market economies and societies (i.e., liberal, coordinated or hybrid market economies), Cooke's case studies suggest that, in general, the role of national governments lies in setting regulatory frameworks, while cities and regions are more direct and proactive in supporting innovations. Although Cooke's study is instructive, it leaves plenty of scope for further work to assess and quantify the relative roles of Triple Helix actors at different spatial levels.

3.2. *Empirical design*

In this section, we develop an empirical design using website-based indicators and secondary data to explore the link between Triple Helix relations and the growth of SME green goods manufacturers in the United States. We focus on the three questions highlighted in the previous section: (1) Do Triple Helix relationships benefit the growth of SMEs? (2) How do Triple Helix relations with actors from different helices and different mixes of Triple Helix relationships influence firm growth? (3) How do relationships with Triple Helix actors from different levels of innovation systems (i.e. regional, national and global) impact firm growth? We target our empirical research to U.S. green goods manufacturing SMEs identified through a sampling strategy discussed below.

3.2.1. *Sampling strategy*

Our analysis focuses on U.S. SMEs that entered green goods manufacturing in the 2003–2007 period. We collect information on these firms in this establishment period, as well as their information in the subsequent period 2008–2011. Rather than using the U.S. Bureau of Labor Statistics definition of green industry or U.S. Patent and Trade Office (USPTO) green patent definitions, which can be too broad, or, if patents are used, too narrow (Shapira et al., 2013), we employ a keyword based approach to search for firms from a major business database, Dun & Bradstreet (D&B) Million Dollar Database. We apply more than 100 keywords to the line of business field in the D&B database to identify relevant companies in three green goods sub-segments: environment (e.g., biological treatment, air pollution, environmental monitoring, land remediation, waste management, water treatment, and recovery and recycling), renewable energy (e.g., wave and tidal, biomass, wind, geothermal, and photovoltaic/solar), and emerging low carbon (e.g., alternative fuel vehicles, alternative fuels, electrochemical processes, batteries, carbon capture and storage, and building technologies) (for details on this method, see Shapira et al., 2013). Additional filters are applied to restrict search results to manufacturers with U.S. headquarters, established in the 2003–2007 timeframe, and starting with 250 or fewer employees. Despite the time-based criterion for establishment, the search results include some older firms which were re-incorporated in the 2003–2007 period as new producers in green goods segments. This search process yielded 2505 potential companies.

To ensure that we could apply the web mining method (described in next section) to companies in our sample, we manually checked each of these companies to determine if they had a current website at the time of our sampling work. The website criterion reduced the number of companies to 700. A four-point coding scheme was then applied to each of these companies to determine the company's relevance to green goods manufacturing. Two coders were involved to ensure inter-coder reliability. Companies that were not involved in manufacturing or lacked a green product offering were removed from the list. In addition, for technical feasibility reasons, company websites primarily using Adobe Flash or other graphical based approaches were not included in the sample. As a result, the search and exclusion process yielded a sample of 300 highly relevant companies. Of these 300 companies, 94 were in the emerging low carbon area, 178 were in the environmental area, with the remaining 28 being classified as renewables.

In contrast to other technology focused industries, the green goods subdomains, as evidenced by our sample, show broad representation across the U.S. (Fig. 1). As expected, there are clusters of firms located in the leading technology-intensive corridors of Northern and Southern California and the Northeast. In addition, we find a number of firms located in the Midwest, Southeast, and Southwest. For example, Atlanta, Georgia, a city better known for services and logistics rather than manufacturing, is home to 17 of the firms in our sample. Florida also stands out as a state that is home to several of the firms. The Midwest has a significant share of the green goods firms, suggesting the domain's potential to reinvigorate industries in this traditional U.S. manufacturing region.

3.2.2. Website mining and indicators of triple helix relationships

We aim to identify the influence of Triple Helix relationships in *earlier* periods on business performance in a *subsequent* period, with appropriate controls. This approach allows a period of elapsed time for potential causes to generate possible effects. We obtain the archived historical websites of the companies from the Wayback Machine. Available via The Internet Archive (http://archive.org), this is a repository of historical web pages saved periodically over time. The Wayback Machine has archived billions of webpages worldwide since 2001. From there, the retrieval of historical company websites provides an unobtrusive representation of prior firm circumstances. The method of extracting website-based indicators from web mining is drawn from Arora et al. (2015), using a six-step process to retrieve business intelligence from the Wayback Machine to ensure social science construct validity. The six steps comprise (1) a sampling process and check for website presence; (2) organizing and defining the boundaries of the web crawl; (3) crawling with computer aided tools; (4) website variable operationalization (to address social science research questions); (5) integration with other data sources; and (6) social science analysis and interpretation.

We extracted historical company websites for the period 2008–2011 from the Wayback Machine for each firm in the sample. To ensure accuracy in capturing Wayback Machine contents, we downloaded and crawled websites multiple times from late 2012 to early 2014. We employed IBM Content Analytics as a software tool for crawling archived websites and extracting variables from website contents. We refer to this tool as ICA (although subsequently IBM has subsequently re-designated the software as Watson Content Analytics). We accessed the program through the

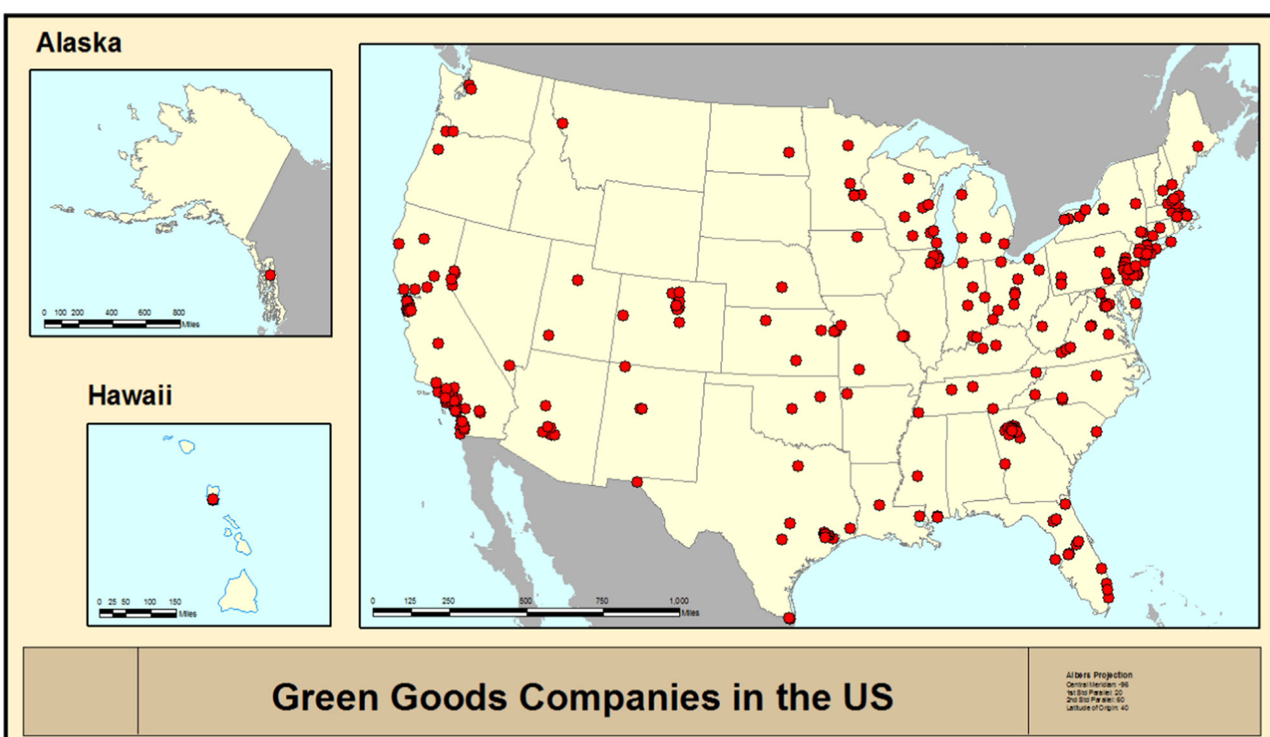

Fig. 1. Geographical distribution of green goods companies in the US
Source: author analysis of U.S. green goods manufacturing SMEs (*N*=300). See text for discussion of method.

Please cite this article as: Li, Y., et al., Using web mining to explore Triple Helix influences on growth in small and mid-size firms. Technovation (2016), http://dx.doi.org/10.1016/j.technovation.2016.01.002

Table 2
Operationalization of Triple Helix relation variables and geographical variables.

Variables	Sources	Operationalization
Relationship with Triple Helix actors		
University	Websites	Counting keywords: university, college, institute, academ*
Industry_financial	Websites	Counting keywords: venture capital, VC, valuation, public offering, IPO, acquisition, merger, liquidity event, seed funding, round of funding, stage of funding, investor, investment firm, private equity, bank, adventure capitalist, bridge loan, buyout, corporate venture capital, corporate venturing, deal flow, debt financing, direct financing, drive-by deal, due diligence, equity financing, financier, full ratchet, fund of funds, institutional investors, lead investor, leveraged buy-out, liquidity event, lock-up period, management buy-in, management buy-out, master limited partnership, mezzanine debt, mezzanine financing, owner-employee, portfolio company, private equity, private placement, raising capital, recapitalization, resyndication, risk capital, investment company, secondary purchase, syndication, term sheet, underwriter, vulture capitalist
Industry_relational	Websites	Counting keywords: partner, stakeholder, distributor, collaborat*, alliance, joint venture, agreement, supplier, parts manufacturer, member, affiliation, association, club, participation, network
Government	Sam.gov	“1” for being registered to qualify for government contracts, “0” otherwise
Triple Helix		Interaction effect of university, industry and government (i.e. University × Industry × Government)
Geographical		
Local	Websites	Counting keywords obtained from named entity recognition and extraction of “places” and aggregation of geographic names to the target SME's metropolitan area
National	Websites	Counting key words obtained from named entity recognition and extraction of “places” and coding of the results as to whether or not the geographical names are not in the target SME's local metropolitan area or outside the U.S.

Note: The selection and meaning of search words can be context-specific. We chose these keywords based on our understanding of the U.S. context. For example, “institute” in the U.S. context is often a variation on “university” (as in California Institute of Technology). See Youtie et al., 2012 for further details.

IBM Academic Initiative. ICA can efficiently scrape websites, allow the user to define keyword definition sets, structure parsing rules with these definitions, deploy parsers to collections of documents (i.e. downloaded websites), and deliver data regarding the amount of times the keywords appeared. XML files generated by ICA were exported to a proprietary Java-based program developed by in-house programmers, where the number of terms was summarized on a per-variable basis. The result of this process is a set of indicators based on website information including *R&D intensity, university, investments, venture capital, membership, partners* and two geographic attributes, *local* and *national*. (Additional information on the operationalization of these terms is available upon request. See also Table 2.) Each indicator was normalized based on the number of pages collected for a given year (with duplicates excluded). Since Wayback coverage of company websites can be sparse from year to year, we construct a panel data set with two aggregate time periods, 2008–09 and 2010–11, by summing each normalized-year variable in the period and dividing by two.

In addition to website-based variables, we also measure the company's contractual relationship with the government using information obtained from the U.S. government's procurement database, the System for Award Management (SAM). The SAM system is accessible online at https://www.sam.gov/. To compete for a federal government contract, the applicant must first register with SAM. We used SAM to create a dummy variable that was set to one if the company appears in the government contracting database as an active registrant in 2008.

We used website-based indicators to construct a set of Triple Helix relationship variables and also geographical variables. The Triple Helix variables measure the intensity of relations between the focal firm and actors in each of the three helices. Similarly, the geographical variables measure the company's focus on local or national links by identifying relevant geographical names on the websites. Table 2 summarizes the method of operationalizing Triple Helix and regional variables.

3.2.3. Econometric model and data

We estimate the impacts of micro-level university, industry and governmental relations on firm growth, measured as sales growth, with a panel data econometric model. The dependent variable is calculated from firm-level gross sales in 2010 and 2012, obtained from the D&B Million Dollar Database. In addition to the website-based Triple Helix relationship variables, we identified four sets of control variables that distinguish firms based on size (*number of employees in 2010 and 2012, logged*), product applications (*renewable energy*, *environmental*, and *low carbon*), website size (*average number of words found on a site, logged*) and business cycle (*industry-level gross output 2009 and 2011, logged*). Data on number of employees and product application were gathered from D&B, whereas the number of words on a site was calculated using a similar process to the one outlined above for the keyword-based explanatory website variables. Industry-level gross output is obtained by mapping a firm's primary North American Industry Classification System (NAICS) code, as identified from D&B, to U.S. Bureau of Economic Analysis (BEA) yearly gross output reports. We aim, with this variable, to control for broader business-cycle effects in these industry classes.

From our initial sample of 300 companies, we obtained a panel balanced with 271 companies and 459 observations after dropping cases with zero sales or employment data points in the D&B database. (These zero values may be due to D&B reporting or measurement issues or because the company had ceased operations subsequent to the web mining period.) Table 3 provides descriptive statistics. Website variables are interval variables starting from 0 with no preset higher caps. “0” means that the company website does not contain any relevant keywords in the time period. Except for the “national” variable, the majority of website variables have means less than one, meaning these keywords on average are mentioned less than once per page.

The Pearson correlation matrix with significance levels is shown in Table 4. As expected, logged sales are positively correlated with *gov*, *univ*, and *inds* variables as well as one geographical variable, *local*.

3.2.4. Model identification

The existence of firm-level heterogeneity leads us to construct a fixed-effect panel model. In the structural model for estimation, as shown below, c_i is the time-invariant part of the unobserved effects (i.e. fixed effect) and u_{it} is the time-variant part. x_{it} is a vector of time-variant explanatory variables and z_{i1} represents time-invariant explanatory variables.

$$\log(\text{sales})_{it} = \theta_1 + \beta_1 x_{it} + z_{i1}\gamma + c_i + u_{it} \quad t = 1, 2$$

A challenge in this panel model is unobserved firm-specific effects, particularly because the financial relationship with industrial actors can be endogenous to sales growth. High growth

Table 3
Descriptive statistics.

	Mean	S.D.	Min	Max	Operationalization
lgsales	14.35	1.76	11.00	21.49	Logged sales 2010, 2012
gov	0.48	0.50	0	1	Whether government contractor
univ	0.15	0.40	0.00	4.63	Website variable "university"[a]
inds	0.42	1.31	0.00	21.48	Website variable "industry"[a]
ptnr	0.55	0.98	0.00	9.66	Website variable "partners"[a]
memb	0.28	0.55	0.00	6.30	Website variable "membership"[a]
tphlex	0.24	2.80	0.00	62.20	Website variable "Triple Helix"[a]
local	0.26	0.90	0.00	13.45	Website variable "local"[a]
national	1.21	1.83	0.00	18.00	Website variable "national"[a]
lgemp	2.85	1.56	0.69	8.09	Logged employment 2010, 2012
lwcarbon	0.32	0.47	0	1	Whether the company is in "low carbon" sector
rnlenrgy	0.09	0.28	0	1	Whether the company is in "renewable energy" sector
lgio	9.80	1.35	7.52	14.12	Logged industry-level gross output 2009, 2011
lgwrds	8.21	2.21	2.89	14.63	Logged average number of words per web page[a]

$N=271$ companies, two time periods.

[a] Web variables.

companies are likely to attract investments and external financial resources that further fuels growth. Geographical variables might have similar endogeneity issues since a successful business strategy for focusing on regional relations would be self-reinforcing with performance. By construction, we avoid such issue in representing relationships with government, since the government variable is a predetermined, time-invariant measurement of the firm's contractual relation with the U.S. government at the beginning of the panel period. We also argue that the university variable is exogenous since relationships with universities are often the result of personal connections or firm R&D activities, not arising from any measure of sales growth.

To proxy the unobserved effects and improve estimation, we introduced two exogenous control variables, number of words and industry-level output. The number of words variable is a measurement of the extensiveness of the company's website, which captures a company's preference (part of c_i) to use websites. The BEA yearly industry-level gross output is a measurement of the business cycle (part of u_{it}), which can impact the company's investment plans and sales performance. Given that our data comprises a mix of time-variant and time-invariant exogenous variables, we use the Hausman–Talyor method to generate consistent estimators (Hausman and Taylor, 1981). A post-estimation test

Table 5
Hausman–Taylor estimation on dependent variable sales (logged).

Independent variables	M1	M2	M3
TV endogenous			
Industry ($inds_{it}$)	0.02	0.036	0.05
	(0.03)	(0.10)	(0.10)
Government × Industry ($gov_i \times inds_{it}$)		−0.06	−0.02
		(0.14)	(0.14)
University × Industry ($univ_{it} \times inds_{it}$)		0.37*	0.35*
		(0.20)	(0.20)
Triple Helix ($gov_i \times univ_{it} \times inds_{it}$)		−0.36*	−0.36*
		(0.21)	(0.20)
Local ($local_{it}$)			0.13***
			(0.03)
National ($national_{it}$)			0.015
			(0.02)
TV exogenous			
University ($univ_{it}$)	−0.060	−0.64**	−0.63***
	(0.073)	(0.29)	(0.28)
Partners ($ptnr_{it}$)	0.037	0.027	0.014
	(0.03)	(0.03)	(0.03)
Membership ($memb_{it}$)	−0.02	−0.02	−0.06
	(0.06)	(0.06)	(0.06)
Government × University ($univ_{it} \times gov_i$)		0.69**	0.65**
		(0.30)	(0.29)
Employment ($lgemp_{it}$)	1.065***	1.061***	1.060***
	(0.026)	(0.027)	(0.027)
Industry-level output ($lgio_{it}$)	0.021	0.021	0.020
	(0.018)	(0.018)	(0.018)
N. of words ($lgwrds_{it}$)	0.035**	0.030*	0.020
	(0.016)	(0.016)	(0.016)
TI exogenous			
Government (gov_i)	0.047	0.03	0.4
	(0.08)	(0.09)	(0.10)
Low carbon ($lwcarbon_i$)	−0.19**	−0.19**	−0.19**
	(0.08)	(0.09)	(0.09)
Renewable energy ($rnlenrgy_i$)	0.13	0.16	0.08
	(0.15)	(0.16)	(0.16)
Constant	10.94***	11.01***	11.07***
	(0.236)	(0.239)	(0.237)
Observations	459	459	459
Number of firms	271	271	271
Test of overidentifying restrictions (p-value)	0.48	0.61	0.53

Note: it in the variable's subscript refers to time varying; *i* refers to time invariant Standard errors in parentheses.

*** $p < 0.01$,
** $p < 0.05$,
* $p < 0.1$

Table 4
Pairwise correlations for all variables.

	lgsales	gov	univ	inds	ptnr	memb	tphlex	local	national	lgemp	lwcarbon	rnlenrgy	lgio	lgwrds
lgsales	1													
gov	0.10*	1												
univ	0.10*	0.12*	1											
inds	0.21*	0.07	0.45*	1										
ptnr	0.30*	0.00	0.17*	0.38*	1									
memb	0.13*	0.06	0.31*	0.43*	0.28*	1								
tphlex	0.05	0.09*	0.38*	0.77*	0.18*	0.27*	1							
local	0.11*	0.00	0.08	0.04	0.05	0.10*	0.02	1						
national	0.07	−0.01	0.18*	0.23*	0.17*	0.35*	0.17*	0.19*	1					
lgemp	0.93*	0.08	0.07	0.21*	0.29*	0.11*	0.04	0.07	0.05	1				
lwcarbon	0.03	0.03	0.01	0.02	0.09*	0.00	−0.03	−0.04	−0.01	0.09*	1			
rnlenrgy	0.17*	0.15**	0.18*	0.22*	0.10*	0.08	0.18*	0.16*	0.09*	0.16*	−0.21*	1		
lgio	−0.03	−0.06	0.02	−0.06	−0.08	−0.01	−0.06	0.04	0.05	−0.04	−0.07	−0.08*	1	
lgwrds	0.27*	0.06	0.23*	0.23*	0.29*	0.31*	0.05	0.14*	0.31*	0.25*	0.05	0.09*	−0.08	1

$N=271$ companies.

* $p < 0.05$.

Please cite this article as: Li, Y., et al., Using web mining to explore Triple Helix influences on growth in small and mid-size firms. Technovation (2016), http://dx.doi.org/10.1016/j.technovation.2016.01.002

based on Sargan–Hansen statistics is used to test for overidentifying restrictions.

3.3. Empirical results

Model results from the Hausman–Taylor estimation are presented in Table 5. For all three models, post-estimation tests cannot reject the null hypothesis that overidentifying restrictions are valid at significant levels ($p=0.48$, 0.61, 0.53, respectively); thus the Hausman–Taylor estimators are consistent.

Model 1 is the base model, before introducing interaction effects and regional variables. Sales growth is positively predicted by firm size and website size. Note that, while the effect of firm size is strong across all three models, the effect of website size becomes less or not significant as we add in more variables in later models. This suggests that the correlation between growth and website size might be spurious: it is the information richness of relations with external actors that predicts both growth and website size. Industry-level output is also positively correlated with individual company growth, yet the correlation is not statistically significant. This finding indicates that the general conditions of the business cycle might influence the green goods companies' investment plans captured by the financial aspects of linkages to external industry actors, but it has little predicting power on growth performance. Companies focusing on low carbon applications have significantly less growth than the base group, environmental applications. None of the individual Triple Helix relation variables (*inds*, *univ*, and *gov*) predict sales growth by themselves.

Model 2 introduces the interaction terms between Triple Helix relationship variables to estimate the simultaneous impacts of multiple relationships. The interaction term is a way to model potential complementary or substitute effects. Government, industry, partners and membership relationship variables still have no significant impact on growth, but stronger university relationship by itself predicts less growth (significant at $p<.05$ for M2 and M3). The three-way interaction term $univ \times gov \times indus$ is statistically significant ($p<.10$), indicating moderating effects among the Triple Helix relationship variables. The presence of significant interaction terms suggest that simultaneous influences of the three types of links are not additive, and the influences of additional relationships intervene within existing mixes of relationships.

Model 3 is the full model with all the Triple Helix relationship variables, their interaction terms and the geographical variables. The geographical variable *local*, a count variable of the times geographical names within the metropolitan area where the company is headquartered, are mentioned on company websites, appears to have a strong positive effect on sales growth (significant at $p<.01$). The other geographical variable *national*, which counts national geographical names, also has positive though statistically insignificant effects. This result suggests that local links are more effective in generating growth. Finally, the Triple Helix relationship variables' estimated coefficients are consistent with Models 1 and 2. The following interpretation of coefficients and marginal effects will focus on Model 3, the full model.

The three-way interaction term appears with a negative coefficient, but this term cannot be straightforwardly interpreted based on the sign of the coefficient. It is appropriate to consider the marginal effects of the three Triple Helix relationship variables. In the upper part of Table 6, we calculate the marginal effects of each of the three Triple Helix relationship variables when the variable increases from 0 to 1 and while holding the conditional variables constant at 1. The conditional variables in this study are the remaining Triple Helix relationship variables. For example, to calculate the marginal effect of university relationships on growth when relationships with government and industry are existent, we hold both the government and industry relation variables at 1, calculate the total effects of university, government, and industry relations at university=0 and university=1, and obtain a net difference between the two total effects. The net difference is the marginal effect of university relationships conditional on given government and industry relations. This marginal effect of university relations is 0.06 in Model 2 and 0.02 in Model 3.

Table 6
Marginal effects of Triple Helix variables on growth.

Variable	Conditional	Marginal effects	
		M2	M3
University	Government	0.05	0.02
	Industry	−0.27	−0.27
	Government×Industry	0.06	0.02
Government	University	0.72	1.05
	Industry	−0.03	0.38
	University×Industry	0.3	0.67
Industry	Government	−0.024	0.03
	University	0.406	0.41
	Government×University	−0.014	0.03
Total marginal effects			
Government×Industry		0.006	0.43
Government×University		0.08	0.42
University×Industry		−0.234	−0.22
Government×University×Industry		0.066	0.45

Note: This table shows the marginal effect of variables on sales growth, when it increases from 0 to 1 and holding the conditional variables at 1.

In a similar manner, we calculate the total marginal effect of multiple relations (Table 6, bottom part). In this calculation, the base group is companies with none of the relationships (university=0, government=0 and industry=0). We calculate the total effect of having both university and government relationships, for example, by summing the coefficients of the two variables and their interaction terms: $univ+gov+univ\times gov$. We also calculate the total marginal effect of having all three types of relations (university=1, government=1 and industry=1) in comparison to having none (university=0, government=0 and industry=0).

The result shows that relationships with all three helices taken together have a positive total marginal effect on firm sales growth, suggesting that Triple Helix relationships are beneficial to firm growth. This effect is relatively large in the full model (M3), where green goods SMEs with all three types of relationships have 45% more growth on average than SMEs without any Triple Helix relationships, after controlling for other characteristics of the firm and industry. More interesting than the aggregate effects of all Triple Helix relationships is the effect of specific mixes of relationships. Across the three models, relationships with government or industry consistently show no significant impact on growth, while relationships with universities alone have negative impacts on growth. Our speculation is that relationships only with universities may signify an SME that is primarily focused on research or early-stage development. The likely longer-term nature and costs of the developmental process may reduce or negatively impact near term returns on sales growth (see also discussion of lower profitability among firms who specialize in technology development without commercialization in Padula et al., 2015). However, adding meaningful government relationships on top of university links is associated with SME sales growth. Thus, while companies with only university relations have 63% *less* growth than the base group (i.e. no relationships), companies with both university and government relationships have 42% *more* growth than the base group. This is likely the case of successful research

Please cite this article as: Li, Y., et al., Using web mining to explore Triple Helix influences on growth in small and mid-size firms. Technovation (2016), http://dx.doi.org/10.1016/j.technovation.2016.01.002

commercialization aided by access to a government procurement-stimulated market. A second scenario of positive impact from having two types of links is concurrent relationships with government and industry, which could be interpreted as a form of public–private joint venture. A third scenario is having relationships with industry and university, which is associated with a negative marginal effect. Although there are a number of possible explanations, we suspect that this negative relationship represents the current status of U.S. green industry where private industry demand for green goods and technologies may be weaker in sectors where government is not an active stimulator. Thus combining university and industry relationship does not yield additional benefits.

While different mixes of relationships result in unequal impacts on growth, different types of relationships also vary in their effects on growth. We can examine the effect of types of relationship by looking at the marginal effect of a single Triple Helix relationship variable when added to the other Triple Helix relationship variables. Relationships with the government appear to be the most valuable since the resulting interaction terms have the largest positive marginal effect (0.67) when added to the existing mix of relationships with universities and industry. In contrast, adding university or industry relations to the existing mix of links has somewhat limited marginal benefits, 0.02 and 0.03, respectively. This result suggests that there are additional benefits from the role of the government, particularly to private sector companies.

4. Discussion

Quantitative studies of the Triple Helix model have been constrained by analytical tools to limited areas of inquiry, such as science–technology linkages, entrepreneurial universities, and system-level entropy measures where patenting and bibliometrics data are available. Drawing on recent developments in using web mining technique in social sciences (Gök et al., 2015; Arora et al., 2015), this paper has advanced a novel approach to analyze micro-level Triple Helix relationships by measuring the intensity of university, industry and governmental relationships as disclosed on the websites of companies and other organizations. These website-based indicators allow us to focus on Triple Helix relationships as communication and coordination networks, and offer the advantage of measuring Triple Helix relationships at the micro-level of a firm as well as measuring the firm's relationships with multiple Triple Helix organizations simultaneously.

We compare the information obtained from the website method with manual reviews of available company information for a small case study set, and find that information from the two sources match reasonably well. We then apply the website method to a data set of U.S. small- and medium-sized green goods companies. The paper draws attention to the difficulties in conducting this type of study using conventional bibliometric and patent measures due to the nature of engineering-intensive technologies in the green goods domain and the complexities in dealings across the three helices. We examine green goods SMEs' external government, academia and industrial organization relationships, and quantitatively assess the impacts of the mixes of Triple Helix relationships on firm growth using a panel data econometric model. Furthermore, we demonstrate that the website-based indicators of Triple Helix relationships combined with conventional firm-level business indicators can contribute to understanding how Triple Helix dynamics enhance economic performance, an important gap in the Triple Helix indicator literature (Meyer et al., 2014).

Revisiting the research questions raised in the beginning of the empirical design (Section 3.2), Triple Helix relationships, measured by website-based indicators, do have positive impacts on the growth of green goods SMEs. However, the positive impacts are joint effects of two or three Triple Helix relationship variables, not individual effects of a single Triple Helix variable. The model can also help us to identify productive mixes of Triple Helix relationships. In the case of US green goods SME growth, two scenarios of relationship mixes (university–government, government–industry) appear to be effective. Additionally, the government relationship variable has the largest positive marginal effect when added to the other two helices. Finally, developing relationships with local or regional Triple Helix actors appears to be beneficial for green goods SMEs.

5. Conclusions

We draw three sets of conclusions from this empirical study of the U.S. green goods SMEs in terms of the varying effects of relationships among different types of Triple Helix actors. First, we find that relationships with organizations from two or more helices increase the rate of sales growth, reflecting the importance of coordination, at least in the green goods sector. Second, as local relationships appear to be more effective in facilitating growth than relationships with national actors, we are able to confirm the importance of the local dimension of Triple Helix interactions often interpreted as tacit knowledge transfer among innovation actors co-located with proximity (Boschma, 2005). These results, while perhaps not unexpected, are not often demonstrated in a quantitative manner. For instance, the pioneering works of the Triple Helix framework (Leydesdorff and Etzkowitz, 1996; Etzkowitz and Leydesdorff, 2000; Etzkowitz, 2008) highlighted the potential of the Triple Helix in enhancing economic growth, but most Triple Helix indicator literature finds it difficult to link Triple Helix relationships with firm-level growth in a quantitative manner. This study thus provides evidence that additional empirical studies can build upon to advance micro-level understandings of the Triple Helix and its effects.

Third, the model identifies various combinations of Triple Helix relationship mixes that might be productive. In the U.S. green goods subdomains, these productive mixes of Triple Helix relations include forms of research commercialization assisted by government programs (co-existing relationships with university and government), public–private joint ventures (co-existing relationships with government and industry), and a deep coordination across all three constituents (co-existing relationships with university, government and industry). Future studies could look further into these mixes of relationships and investigate their mechanisms. There are policy and management implications from these findings, though. On the policy side, for instance, configurations and the mix of Triple Helix relationships should be carefully considered in policy designs, as the mix appears to be as important as the extent of the relationships themselves. Here, our study echoes John Podesta and colleagues who called for "an integrated approach" to the green goods industries (Ogden et al., 2008). Instead of focusing on one particular sector (i.e. university or industry), the effective policy is the one that coordinates and integrates players in the innovation system at various stages of innovation development and implementation. From a management perspective, innovative firms actively manage the mix of relationships by strategically articulating the benefits (and disadvantages) expected from these relationships. In this specific context, it is not hard to see that partnerships with industry and government would substantially reduce innovation risks compared to relying solely on linkages to the university.

We acknowledge limitations. There are caveats associated with the web mining method. Arora et al. (2015) provides a comprehensive review of the potential issues and solutions, such as measurement errors and external validity in using web mining methods and company website data. Particularly relevant to this study is the issue of consistency of website information across firms and across time. We ensure cross-firm consistency through narrowing the sample to firms in the same domains and similar activities (i.e. green goods manufacturing). We maintain cross-time consistency through manual verification of websites and normalization of measures over time. Additionally, website-based indicators are essentially built on public self-reporting by firms, which raises further limitations. In a developed economy such as the United States where internet penetration is high and online communication well established, the use of websites by SMEs to disseminate information is a common practice. This underlying assumption might be more problematic in societies where the internet is less prevalent and where online communication is not as customary. Despite these limitations, this paper has demonstrated the usefulness of website-based indicators to explore Triple Helix dynamics as a complement to traditional methods and data.

The results highlight that Triple Helix relationships with government make a particularly notable contribution to the effectiveness of the overall mix of Triple Helix relations. To some extent, this result underscores the importance of government coordination in the U.S. green goods industry and its impact on firm performance. The result is not surprising since the role of government intervention and support is most valued when a nascent private sector market is still emerging, or where a market exhibits signs of underperformance or market failure. While the role of government is always included in the Triple Helix literature, it is not always measured in Triple Helix indicator studies. This paper highlights that there are now new ways to quantitatively measure the influence of government and other actors in the Triple Helix framework. More generally, further studies can build on, and advance, the methods and findings of this paper to examine micro-level patterns of relationships engaged in by firms and to assess their impacts on performance, innovation, and other dimensions.

Acknowledgment

This work was supported by the Economic and Social Research Council [Grant number ES/J008303/1] through the Project on Sustaining Growth for Innovative New Enterprises. We acknowledge helpful comments from anonymous reviewers. The opinions and conclusions presented in the paper are those of the authors.

References

Acosta, M., Coronado, D., 2003. Science–technology flows in Spanish regions: an analysis of scientific citations in patents. Res. Policy 32 (10), 1783–1803.
Acosta, M., Coronado, D., Fernández, A., 2009. Exploring the quality of environmental technology in Europe: evidence from patent citations. Scientometrics 80 (1), 131–152.
Ahuja, G., 2000. Collaboration networks, structural holes, and innovation: a longitudinal study. Adm. Sci. Q. 45 (3), 425–455.
Ahuja, G., Lampert, C.M., Tandon, V., 2008. Moving beyond schumpeter: management research on the determinants of technological innovation. Acad. Manag. Ann. 2 (1), 1–98. http://dx.doi.org/10.1080/19416520802211446.
Allen, R.C., 1983. Collective invention. J. Econ. Behav. Organ. 4 (1), 1–24.
Amatucci, F.M., Sohl, J.E., 2007. Business angels: investment processes, outcomes, and current trends. In: Zacharakis, A., Spinelli, S. (Eds.), Entrepreneurship: The Engine of Growth 2. Praeger, Westport, CT, pp. 87–108.
Arora, S.K., Li, Y., Youtie, J., Shapira, P., 2015. Using the wayback machine to mine websites in the social sciences: a methodological resource. J. Assoc. Inf. Sci. Technol. . http://dx.doi.org/10.1002/asi.23503
Arora, S.K., Youtie, J., Shapira, P., Gao, L., Ma, T., 2013. Entry strategies in an emerging technology: a pilot web-based study of graphene firms. Scientometrics 95 (3), 1189–1207, http://dx.doi.org/10.1007/s11192-013-0950-7..
ARRA, 2009. American recovery and reinvestment act of 2009. Pub.L. 111–5. ⟨http://www.gpo.gov/fdsys/pkg/PLAW-111publ5/html/PLAW-111publ5.htm⟩ (accessed 30.06.15).
Asheim, B.T., Coenen, L., 2005. Knowledge bases and regional innovation systems: comparing Nordic clusters. Res. Policy 34 (8), 1173–1190.
Auersald, P.E., 2007. The simple economics of technology entrepreneurship: market failure reconsidered. In: Audretsch, D.B., Grilo, I., Thurik, A.R. (Eds.), Handbook of Research on Entrepreneurship Policy. Edward Elgar, Cheltenham, UK, pp. 18–35.
Baldini, N., 2006. University patenting and licensing activity: a review of the literature. Res. Eval. 15 (3), 197–207.
Borgatti, S.P., Foster, P.C., 2003. The network paradigm in organizational research: a review and typology. J. Manag. 29 (6), 991–1013.
Boschma, R., 2005. Proximity and innovation: a critical assessment. Reg. Stud. 39 (1), 61–74.
Brady, J., 2014. After Solyndra loss, U.S. energy loan program turning a profit. NPR 30, November 13. ⟨http://www.npr.org/2014/11/13/363572151/after-solyndra-loss-u-s-energy-loan-program-turning-a-profit⟩ (accessed 30.06.15).
Callaert, J., Van Looy, B., Verbeek, A., Debackere, K., Thijs, B., 2006. Traces of prior art: an analysis of non-patent references found in patent documents. Scientometrics 69 (1), 3–20.
Carayannis, E.G., Barth, T.D., Campbell, D.F.J., 2012. The Quintuple Helix innovation model: global warming as a challenge and driver for innovation. J. Innov. Entrep. 1, 2 ⟨http://www.innovation-entrepreneurship.com/content/1/1/2⟩.
CleanEdge, 2015. U.S. Clean Tech leadership index 2015: State & Metro. June. ⟨http://cleanedge.com/reports/2015-US-Clean-Tech-Leadership-Index⟩ (accessed 07.07.15).
Cooke, P., 2011. Transition regions: regional–national eco-innovation systems and strategies. Prog. Plan. 76, 105–146.
Council on Foreign Relations, 2009. Obama–Biden new energy for America plan, January 2009. ⟨http://www.cfr.org/united-states/obama-biden-new-energy-america-plan-january-2009/p18306⟩ (accessed 30.06.15).
Dellarocas, C., 2003. The digitization of word of mouth: promise and challenges of online feedback mechanisms. Manag. Sci. 49 (10), 1407–1424.
Dellarocas, C., 2006. Strategic manipulation of internet opinion forums: implications for consumers and firms. Manag. Sci. 52 (10), 1577–1593.
Etzkowitz, H., 2008. Triple Helix Innovation: Industry, University, and Government in Action. Routledge, London and New York.
Etzkowitz, H., Leydesdorff, L., 2000. The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. Res. policy 29 (2), 109–123.
Gök, A., Waterworth, A., Shapira, P., 2015. Use of web mining in studying innovation. Scientometrics 1, 653–671, http://dx.doi.org/10.1007/s11192-014-1434-0.
Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., Libai, B., Sen, S., Shi, M., Verlegh, P., 2005. The firm's management of social interactions. Mark. Lett. 16 (3–4), 415–428.
Hausman, J.A., Taylor, W.E., 1981. Panel data and unobservable individual effects. Econom.: J. Econom. Soc., 1377–1398.
Hicks, D., 1995. Published papers, tacit competencies and corporate management of the public/private character of knowledge. Ind. Corp. Chang. 4 (2), 401–424.
Iversen, E.J., Gulbrandsen, M., Klitkou, A., 2007. A baseline for the impact of academic patenting legislation in Norway. Scientometrics 70 (2), 393–414.
Lazonick, W., 2008. Entrepreneurial Ventures and the Developmental State: Lessons from the Advanced Economies (No. 2008/01). World Institute for Development Economics (UNU-WIDER), Helsinki.
Leydesdorff, L., 2003. The mutual information of university–industry–government relations: an indicator of the Triple Helix dynamics. Scientometrics 58, 445–467.
Leydesdorff, L., 2008. On the normalization and visualization of author co-citation data. J. Am. Soc. Inf. Sci. Technol. 59, 77–85.
Leydesdorff, L., Etzkowitz, H., 1996. Emergence of a Triple Helix of university–industry–government relations. Sci. Public Policy 23 (5), 279–286.
Leydesdorff, L., Meyer, M., 2006. The Triple Helix, indicators, and knowledge-based innovation systems. Res. Policy 35 (10), 1441–1449.
Leydesdorff, L., Fritsch, M., 2006. Measuring the knowledge base of regional innovation systems in Germany in terms of a Triple Helix dynamics. Res. Policy 35 (10), 1538–1553.
Leydesdorff, L., Meyer, M., 2007. The scientometrics of a Triple Helix of university–industry–government relations. Scientometrics 70 (2), 207–222.
Li, Y., Youtie, J., Shapira, P., 2015. Why do technology firms publish scientific papers? The strategic use of science by small and midsize enterprises in nanotechnology. J. Technol. Transf. 40 (6), 1016–1033. http://dx.doi.org/10.1007/s10961-014-9391-6.
Mazzucato, M., 2013. The Entrepreneurial State: Debunking Public vs. Private Sector Myths. Anthem Press, London, New York.
Meyer, M., Siniläinen, T., Utecht, J.T., 2003. Towards hybrid Triple Helix indicators: s study of university-related patents and a survey of inventors. Scientometrics 58 (2), 321–350.
Meyer, M., Grant, K., Morlacchi, P., Weckowska, D., 2014. Triple Helix indicators as an emergent area of enquiry: a bibliometric perspective. Scientometrics 99 (1), 151–174.
Morgan, K., 2004. The exaggerated death of geography: learning, proximity and territorial innovation systems. J. Econ. Geogr. 4 (1), 3–21.
Muller, P., Penin, J., 2006. Why do firms disclose knowledge and how does it matter? J. Evol. Econ. 16 (1–2), 85–108.
Muro, M., Rothwell, J., Saha, D., 2011. Sizing the Clean Economy: A National and Regional Green Jobs Assessment. Brookings, Washington DC.
Murphy, J., Hashim, N., O'Connor, P., 2007. Take me back: validating the Wayback Machine. J. Comput.-Mediat. Commun. 13 (1), 60–75.

Please cite this article as: Li, Y., et al., Using web mining to explore Triple Helix influences on growth in small and mid-size firms. Technovation (2016), http://dx.doi.org/10.1016/j.technovation.2016.01.002

OECD, 2011. Towards Green Growth. Organisation for Economic Cooperation and Development, Paris.
Ogden, P., Podesta, J., Deutch, J., 2008. A new strategy to spur energy innovation. Issues Sci. Technol. 24 (2), 35–44.
Padula, G., Novelli., E., Conti, R., 2015. SMEs inventive performance and profitability in the markets for technology. Technovation 41–42, 38–50.
Park, H.W., Hong, H.D., Leydesdorff, L., 2005. A comparison of the knowledge-based innovation systems in the economies of South Korea and the Netherlands using Triple Helix indicators. Scientometrics 65 (1), 3–27.
PriceWaterhouseCoopers, 2014. Cleantech MoneyTree™ Report: Q4 2013. ⟨http://www.pwc.com/us/en/technology/publications/moneytree-cleantech-venture-funding-report.jhtml⟩ (accessed 06.06.15).
Ranga, M., Etzkowitz, H., 2013. Triple Helix systems: an analytical framework for innovation policy and practice in the knowledge society. Ind. High. Educ. 27 (4), 237–262.
Shapira, P., Gök, A., Klochikhin, E., Sensier, M., 2013. Probing "green" industry enterprises in the UK: a new identification approach. Technol. Forecast. Soc. Chang. 85, 93–104, http://dx.doi.org/10.1016/j.techfore.2013.10.023.
Simmie, J., 2005. Innovation and space: a critical review of the literature. Reg. Stud. 39 (6), 789–804. http://dx.doi.org/10.1080/0034340050021367.
Stuart, D., Thelwall, M., Harries, G., 2007. UK academic web links and collaboration-an exploratory study. J. Inf. Sci. 33 (2), 231–246.
Uranga, M.G., Kerexeta, G.E., Campas-Velasco, J., 2007. The dynamics of commercialization of scientific knowledge in biotechnology and nanotechnology. Eur. Plan. Stud. 15 (9), 1199–1214.
Van Looy, B., Magerman, T., Debackere, K., 2007. Developing technology in the vicinity of science: an examination of the relationship between science intensity (of patents) and technological productivity within the field of biotechnology. Scientometrics 70 (2), 441–458.
Vilanova, M.R., Leydesdorff, L., 2001. Why Catalonia cannot be considered as a regional innovation system. Scientometrics 50 (2), 215–240.
Wong, W.L.P., Radcliffe, D.F., 2000. The tacit nature of design knowledge. Technol. Anal. Strateg. Manag. 12 (4), 493–512.
Youtie, J., Hicks, D., Shapira, P., Horsley, T., 2012. Pathways from discovery to commercialisation: using web sources to track small and medium-sized enterprise strategies in emerging nanotechnologies. Technol. Anal. Strateg. Manag. 24 (10), 981–995, http://dx.doi.org/10.1080/09537325.2012.724163.