Entry Strategies in an Emerging Technology:

A Pilot Web-Based Study of Graphene Firms

Sanjay K. Arora[a], Jan Youtie[b], Philip Shapira[a,c,*], Lidan Gao[d], TingTing Ma[e]

[a] School of Public Policy, Georgia Institute of Technology, Alanta, GA 30332-0345, USA

[b] Enterprise Innovation Institute, Georgia Institute of Technology, Atlanta, GA 30332-0345, USA

[c] Manchester Institute of Innovation Research, Manchester Business School, University of Manchester, M13 9PL, UK

[d] Chengdu Library of the Chinese Academy of Sciences, No.16, Nan'erduan, Yihuan Road, Chengdu 610041, P. R. China

[e] School of Economics and Management, Beijing Institute of Technology, No.5, Zhongguancun Nandajie Haidian District, Beijing 100081, P. R. China

[*] Corresponding author:

Philip Shapira
Manchester Institute of Innovation Research
Manchester Business School
Oxford Road, Manchester
M13 9PL, UK
Tel: +44 (0) 161 275 7376
Fax: +44 (0) 161 275 0923
pshapira@mbs.ac.uk

Abstract

We explore pilot web-based methods to probe the strategies followed by new small and medium-sized technology-based firms as they seek to commercialize newly emerging technologies. Tracking and understanding the behavior of such early commercial entrants is not straightforward because smaller firms with limited resources do not always widely engage in readily visible and accessible activities such as publishing and patenting. However, many new firms, even if small, present information about themselves that is available online. Focusing on the early commercialization of novel graphene technologies, we introduce a "web-scraping" approach to systematically capture information contained in the online web pages of a sample of small and medium-sized high technology graphene firms. We analyze this information and use it to devise measures that gauge how firm specialization in the target technology impacts overall market orientation. The results from this pilot study identify three groups of grapheme enterprises - specialized product development, specialized materials development, and integration into existing product portfolios. Country-level factors are important in understanding these early diverging commercial approaches in the nascent graphene market. We consider management and policy implications and discuss the value, including strengths and weaknesses, of web scraping as an additional information source on enterprise strategies in emerging technologies.

Keywords: Emerging technology; graphene; small and medium enterprise; commercialization; market entry; web scraping

MSC Classification: 91

JEL Classification Codes: C81; D22; M13; O32; O32; O57; Z18

1. Introduction

How do novel discoveries evolve into commercial activity? While technological discoveries present new opportunities for business growth and development, the pathways for commercialization of new technologies are typically complex and fraught with challenges. High –technology small-to-medium sized enterprises (SMEs), in particular, are often thought of as primary conveyors of technological disruption as they attempt to capitalize on scientific and market niches (Arora and Gambardella, 1994). From a policy standpoint, analysts and policymakers have positioned these SMEs as key drivers of economic development and job growth (Wessner, 2007). However, not all high-technology SMEs are alike. The growing literature on SMEs and innovation identifies several differentiating characteristics, including export intensity and internationalization, propensity to engage in partnerships and networks, and engineering vs. science-based orientations, to discriminate between different firm strategies. These studies suggest that different types of firms will respond idiosyncratically to commercial opportunities from novel discovery.

A primary goal of this research is to investigate, from a competitive intelligence perspective, what can be learned about enterprise innovation behavior from website data, and how website data compares with more conventional methodological approaches such as bibliometrics or patent analysis. Scholars in management and communication acknowledge the importance of websites as a vehicle for information, promotion, and various business-to-business and business-to-consumer sales activities (Gabrielsson and Gabrielsson, 2011). Here, we engage in a pilot and exploratory study of website text mining to probe the strategies revealed by twenty science-based SMEs producing a novel type of nanotechnology material, graphene. This analysis is conducted within a comparative framework of small startups in three countries: the

US, UK, and China, while examining salient factors that give rise to heterogeneity amongst high-technology SMEs. The results show that instead of traditional distinctions between science versus engineering approaches, three groups of graphene companies are identified - specialized product development, specialized materials development, and integration into existing product portfolios. Country variations across these groups are noted, in that US and Chinese graphene firms tend to be more oriented than UK companies to product strategies.

This paper begins with the motivation for investigating graphene as an appropriate and compelling area of high-technology SME activity. Then, we develop several propositions related to an expected typology of graphene firms based on a new synthesis of streams of scholarly work on new-technology based firms, born global enterprises, and national systems of innovation. Third, we explain our methodology of web scraping and operationalize five descriptive variables based on keyword analysis. Fourth, we present and discuss our results based on our pilot study. We conclude with methodological observations related to the value and limitations of using web-mined data as a complementary source of information on enterprise strategies, consider management and policy implications, and suggest some opportunities for future work.

2. Graphene

Graphene is a new two-dimensional type of material consisting of a single layer of carbon atoms. It is incredibly strong and light, and exhibits high thermal and electrical conductivity. These unique characteristics have attracted significant scientific, business and policy attention to graphene in the past several years. The ebullience surrounding the material hinges on the possibility that it may revolutionize a number of established industries, including the electronics, energy and bio-medical sectors. Graphene is beginning to be deployed in prototype plastic, steel, and ink composites and may soon find its way into higher level technologies such as batteries, displays, transistors and hydrogen storage (Geim and Novoselov, 2007). Another widely anticipated application area is in electronics; graphene could replace silicon in computer microchips and bring considerable performance improvements to processor performance. In 2010, Andre Geim and Konstantin Novoselov won the Nobel Prize in Physics for their pioneering 2004 work on graphene discovery. Acknowledgement of the significance of their initial work and subsequent breakthroughs grew rapidly, with a worldwide takeoff in graphene research publications around 2007 (Shapira et al., 2012).

Although 3,000 related research papers and over 400 patent applications related to the technology were filed in 2010, mass commercialization of graphene may still be years away due to a number of product and process obstacles (Segal, 2009; Van Noorden, 2011). There are significant technical and cost issues related to scaling up graphene manufacturing, although these will likely decrease as process innovations reduce variability in production and as throughput rises. Additionally, the high electrical conductivity of the material means that scientists must identify a way to contain the charge in graphene sheets so that digital signals can be processed properly. It also remains to be seen whether there are significant environmental, health or safety implications related to graphene. Thus, the development trajectory of this technology holds substantial promise, but risks of commercialization are high.

A number of multinationals are active in graphene research and development (e.g. Intel and IBM in computing, Dow Chemicals and BASF as suppliers of basic graphene material, and Samsung in consumer electronics). In addition, several entrepreneurial firms are in the process of translating graphene basic research to applied research and commercialization. It has been argued that young, innovative small to medium sized enterprises (SMEs) may innovate more than larger firms (Michael and Pearce, 2009), but such a generalization may not be true within the context of nanotechnology, where high capital investments, streamlined R&D processes, and well-defined knowledge-acquisition channels may favor larger corporate entities (OECD, 2010). In sum, firms operating in a product market which has not standardized around a single design (e.g. graphene) experience flexible but inefficient production processes, as well as significant information-seeking activities due to frequent product changes (Abernathy and Utterback, 1978). Accordingly, we expect to find heterogeneity in the technological approaches of graphene SMEs. The following section presents our theoretical framework and develops a number of propositions related to these expected differences.

3. Literature Review

3.1 New Technology Based Firms and the Role of Partnerships

Capabilities and resources are distributed heterogeneously across firms (Barney, 1991). Capabilities can become a source of competitive advantage if they enable the firm to more effectively adapt to changes in the external environment (Eisenhardt and Martin, 2000). In nanotechnology, where basic research is a significant input, peripheral forces such as availability of funding, human capital constraints, and environmental and health safety concerns may influence how firms shape their R&D investment portfolios. In other words, the internal wherewithal of a firm to produce superior technologies may be tempered by its inability to effectively cope with business-related factors unrelated to the technology itself (Auerswald, 2007).

This literature on firm capabilities traditionally assumes homogeneous commercial approaches, but homogeneity may be less prevalent in science-intensive businesses. Autio

(1997a, 1997b) deviates from this approach in his framework of new technology-based firms (NTBFs), which articulates two ways of explaining and classifying SME behavior. He classifies NTBFs into two categories, science-based and engineering-based. Science-based firms oversee the advancement of two transformation processes, producing basic technologies and application-specific outputs. Conversely, engineering-based firms focus mainly on product outputs. Whereas products are characterized by features and functions, basic technologies are identified by their treatment of natural phenomena. While engineering firms are driven by market needs, science-based firms are compelled more by core technologies (Autio, 1997a). Ever greater specialization enables each NTBF to find a niche in the "innovation network" (Arora and Gambardella, 1994). Scientific inquiry generates knowledge for the innovation process and helps firms build their own R&D capabilities (Ahuja et. al, 2008), thereby increasing the "absorptive capacity" of the firm, where absorptive capacity refers to a firm's ability to internalize new, externally-derived discoveries through its extant knowledge of science, technology, and R&D processes (Cohen and Levinthal, 1990).

We posit that both science and engineering based NTBFs rely on (academic and research institution) partnerships to search for and synthesize knowledge. However, given engineering-based NTBFs' downstream foci, we hypothesize that partnerships play a more important role in the business strategies of engineering firms than with science based firms. Science firms, for example, may be led by an academic and thus maintain close ties with a local university. Engineering firms, on the other hand, may license technology from a university *and* actively develop close relationships with customers and/or distributors. We formally state our expectation as Proposition 1:

Proposition 1: *Engineering-based NTBFs rely more heavily on partnerships than do science-based NTBFs.*

3.2 The International Born Global and the Role of Uniqueness

While Autio (1997a, 1997b) acknowledges that internationalization can play an important role in the sources and users of a technology, the NTBF framework offers limited explanatory power with respect to how and why some young high-technology SMEs pursue internationalization as a central component of their business strategies. The born global theory posits that small firms operating in niche markets will look to international channels for market demand (Rennie, 1993). The primary driver of an international orientation is the uniqueness of a product (portfolio) with global appeal. Born global firms may experience high profit margins because of the novelty of their products (Gabrielsson et al., 2008).

Given graphene's recent discovery and associated high entry barriers (in terms of capital, knowledge, etc.), we expect many new SME university spin-offs to focus primarily on graphene in their R&D and product portfolios; that is, we would expect university spin-offs to be science-based NTBFs. Furthermore, it is worthwhile to differentiate between *dedicated* graphene firms, which focus exclusively on graphene production or graphene enabled applications, and *other* firms that produce not only graphene but also other nanomaterials such as carbon nanotubes. We posit that NTBFs offering more diverse product portfolios will operate in a broader domestic market than *dedicated* graphene firms. Therefore, such companies may not be inclined to pursue international opportunities with the same level of resolve as *dedicated* graphene firms operating in more niche markets (Proposition 2). In addition, because engineering-based firms are driven more by market opportunities, we expect that such firms will pursue international channels to a greater extent than their science-based counterparts (Proposition 3).

Proposition 2: *Companies specializing in graphene will focus more intensely on internationalization than diversified nanotechnology companies.*

Proposition 3: *Engineering-based NTBFs will pursue internationalization channels more intensely than science-based NTBFs.*

Although born global firms are likely to participate in and coordinate tightly coupled streams of geographically dispersed, inter-organizational activities (e.g. to liaison with distributors), international new ventures need not own significant foreign investments. They instead may rely on strategic partnerships to develop and secure complementary assets (Oviatt and McDougall, 2005). Indeed, in their case study research of 35 Finnish born global SMEs, Gabrielsson and Gabrielsson (2011) find that most firms do not rely on direct Internet-based sales channels, thus indicating that partnerships are important to effectively exploiting international markets. We state this expectation as Proposition 4.

Proposition 4: *High levels of internationalization are positively correlated with high levels of partnership activity.*

3.3. National Systems of Innovation and Country-Specific Path Dependencies

Commercialization strategies differ not only at the firm level but also at agglomerated units of analysis. The National System of Innovation (NSI) offers another theoretical framework from which to view high-technology firm growth and behavior. The NSI assumes that national differences in technology policies and the underlying support infrastructure set the context for technological change (Malecki, 1997). Therefore, country-level differences in science and technology policy can be considered as one source of difference among firms in a particular industry. Even in the case of born global firms, actual engagement with foreign customers may in part be influenced by the home country's export policy (Rasmussen and Madsen, 2002).

Within the domain of nanotechnology, several comparative studies have contrasted the differences between country NSIs, with respect to public versus private funding, publication and patenting output, and type or level of commercialization by firms and/or universities (e.g. see Miyazaki and Islam, 2007; Youtie et al., 2008; Shapira and Wang, 2010; Shapira et al., 2011). By most if not all measures, the US is considered an established leader in nanotechnology R&D and commercialization. Broadly speaking, the US innovation system is recognized for its close collaborative relationships between universities and firms (Mowery and Sampat, 2006). SMEs in particular are an important element in the US, which often relies on start-up firms to transfer discoveries from the university environment to product markets. Conversely, innovation scholars view the UK innovation system as being productive from a research perspective but challenged to effectively commercialize domestically-produced scientific discoveries; historically, firms in other countries such as Germany and the US have been the primary beneficiaries of Britain's most important scientific breakthroughs (Walker, 1993). In their recent work, Cosh and Hughes (2010) find that while UK firms are more likely to maintain university relationships, US firms are more likely to invest in deeper ties that may result in more fruitful partnerships and attendant by-products.

China's rise as a leader in nanotechnology, particularly in terms of overall funding and publication citations, has been well-documented (see Zhou and Leydesdorff, 2006; Lux, 2007; Shapira and Wang, 2009). Nonetheless, China still faces considerable challenges with regard to the commercialization of home-grown, applied research (Sandhu, 2007). Chinese high-technology firms are often funded (in-part) by universities and research institutions, which encourage commercialization of innovations in order to acquire additional revenue streams and promote career advancement (Shapira and Wang, 2009). Chinese patenting activity in nanotechnology is heavily dominated by the academy, and within the realm of graphene, we would expect a similar trend. Furthermore, even though China has invested significantly in science parks and innovation centers, the country's "top-down" approach and historical emphasis on a robust export agenda has resulted in a "policy duopoly" of misaligned incentives (Yu, Stough, and Nijkamp, 2009). In other words, while Chinese innovation centers promote the development and commercialization of indigenous applied research, Chinese science and industrial parks often support the activities and export agenda of multinational firms. Proposition 5 synthesizes expected differences in in each respective country.

Proposition 5: *UK graphene firms are likely to be science-based, Chinese firms are more likely to be internationally focused than domestic, and US firms are likely to be diversified across all relevant attributes (internationalization, partnerships, specialization, and application focus).*

4. Methodology

Several methods are commonly used for understanding the potential commercial trajectory of novel discoveries. These include analysis of research publications (especially those with corporate authors), patents, and surveys of and interviews with firms and/or scientific experts. (Porter et al, 2010). However, each of these methods faces challenges, particularly with respect to the commercial potential of novel discoveries. For example, survey instrumentation suffers from well-known pitfalls such as low response rates and response bias. Publication and patent analysis may not accurately reveal intent to commercialize an innovation; companies commonly engage in strategic patenting behavior to blanket or fence-off a particular domain as one way to deter competitors (Gilbert and Newbery 1982). In addition, publication and patent data sources suffer from time lags and narrow focus; e.g. not all small and medium-sized innovative enterprises publish or patent to a significant extent due to insufficient resources and/or high overhead requirements (Brouwer and Kleinknecht, 1999). For large firms with some minimum threshold of absorptive capacity, bibliometrics is a viable, primary methodological tool to gauge R&D collaborations with academic researchers. However, since smaller, high-technology firms are often start-ups with university principal investigators acting as scientific and business leads, publication activity may be subsumed under the university rather than the individual researcher's company. In this case, bibliometrics may conflate the lead researcher's university contribution with his or her spin-off work.

From the perspective of innovation networks, while commercial business databases may provide an adequate source for identifying large firm relationships, smaller companies may be underrepresented in such data sources. In fact, many SMEs, including the ones in this study, reveal their linkages to other actors on their websites, not through other sources which are then collated by large business database providers.

This paper applies another methodological approach to characterize SME market orientations – web scraping. The idea behind web scraping is to capture information on a company's home page and other pages and convert this information into variables for analysis. Websites are a primary communication platform for SMEs across all three innovation indicators. Of course, companies do not reveal everything about their market orientation on their web site, but neither do they do this in surveys, publications, or patents.

Analysis of website data is an emerging area of study for social scientists. Katz (2006) identifies two ways of classifying web indicators. The first approach explains how innovation systems affect the web via "empirical investigations"; the second concerns how the web affects innovation systems. A web crawler produces the data in this study, so the methodology employed here is clearly situated in the first class of empirical work. (The latter category is explored through qualitative techniques such as interviews.) However, our approach for obtaining and storing web indicator data extends much further than Katz's notion of page counts and linking. Keyword searching is one salient avenue explored in this paper. In his semantic network analysis of industry, government, and university websites, Kim (2012) exploits both relational (hyperlink) and semantic (lexical) aspects of the web to assess the usage of nano-related keywords in these different sectors. With keyword searching, we emphasize the semantic aspect of revealed, online business activity.

To examine early graphene commercialization orientations in SMEs, we profile twelve US firms, five Chinese companies, and three UK SMEs, for a sample database consisting of 790 pages across 20 companies (see Table 1). The companies were identified during the period

October-December 2010, and web-scraping of the sample company sites occurred through to January 2011. Our process for identifying candidate graphene firms included online searching, reviews of publication, patent, and business databases (e.g. FAME for the UK companies), and verification telephone calls to some of the companies. The primary criterion for eligibility required an SME's website to explicitly mention graphene in a meaningful way. Mostly the companies selected develop and produce graphene, but a few also serve as wholesalers. Most of the firms were founded in the 2000s. Some candidate SMEs had a publication or patent with the term graphene appearing somewhere in the publication or patent. However, when we reviewed their web sites, we saw no mention of graphene, so these firms were omitted from the study. Only three of the twenty firms (Vorbeck, Nanotek Instruments, and Graphene Industries) in Table 1 have graphene patents assigned to the company and mention graphene in a significant way online, although some of the other companies license university-assigned patents. In this way, we focus on NTBFs that would have otherwise been overlooked through more traditional identification processes.

[Insert Table 1 Here]

The web-scraping method used in this paper is based on code from the Texas A&M HHAT project, which was originally developed to understand how hypermedia collections change in content over time (Davis, Maslov and Phillips, 2005). HHAT is written in Perl and has been modified by the authors to include several reporting features. Data is stored in a MySQL database, and SQL is the primary querying mechanism for search terms. For three of the five Chinese firm websites, Google Translate was used in order to translate text from Chinese to English. The authors ensured that keyword searching would remain consistent across English language and native Chinese language sites by forward and backward translating keywords from

Chinese to English. Any differences were accounted for in the search strategy. In summary, web scraping consists of two sequential steps: data collection and preliminary processing followed by querying and reporting (see Figure 1).

[Insert Figure 1 Here]

5. Variables

The variables derived from the web data are used to explore the market orientations of the firms in the sample dataset and are largely based on aggregate frequency of occurrence for particular keywords. We compute a variable for *graphene-ness* (Γ) to denote the degree of firm specialization in graphene. This is calculated by the following expression:

$$\Gamma = \gamma / (\gamma + \nu)$$

where: γ = number of mentions of graphene on a firm's website

ν = number of mentions of all other nanomaterials (e.g. carbon nanotubes, nanoparticles, etc.)

The other web indicator variables, *applications, sectors, globalization, and partnerships*, are standardized by the number of web pages in each respective firm's website; that is, these variables reflect the average number of relevant keyword occurrences per webpage by firm. For *applications*, we look for keywords such as transistor, composite, screen, or ink; for *sectors*, keyword examples include electronics, solar, construction, and health; for *globalization*, we search for keywords such as world, international, or Asia; and for *partnerships*, we seek out keywords such as partner, university and distributor. *Applications* and *sectors* capture the extent to which a firm might be engineering-based vs. science-based; i.e. high scores of *applications* and *sectors* indicate that firm is an engineering-based NTBF with a downstream product focus. *Globalization* and *partnerships* capture the extent to which a firm pursues international opportunities and interactions in the innovation network, respectively. We acknowledge that individual keyword occurrences may not convey a particular concept (e.g. internationalization) with high accuracy, but as shown above, our approach combines several different keywords to approximate a general concept.

Following other studies generating typologies of SMEs (e.g. Hagen et al., 2012), we employ hierarchical clustering using Ward's method to identify and visually inspect groups of firms as a whole and by country. Two tests were used in order to validate that a multivariate difference of means test could be applied on the different groups. Firstly, we employed Box's Test for the equality of covariance matrices; the null hypothesis could not be rejected at the .01 significance level ($21.67 < \chi^2 = 23.21$). Secondly, we examined multivariate normality assumptions for the sample dataset as a whole, plotting ordered distance by chi-square percentile. Although the data from the twenty firms did not appear as perfectly multivariate normal, the small sample size was expected to engender some amount of variability which would not produce a straight diagonal line under most circumstances.

6. Results

From a descriptive standpoint, the 2011 website data provides a number of interesting findings. US and Chinese graphene firms are much more likely to mention specific applications and end-use sectors than are UK SMEs (see Tables A1 and A2). Whereas US firm websites contain about 76 application-related keywords on average – and Chinese SME websites maintain an average of 109 mentions – UK sites reflect a far lower emphasis at an average of only eight references per website. For end-use sectors, a similar but more pronounced trend is apparent. US firms refer to industry markets an average of 45 times per website, Chinese SMEs at an average of 87.8 times, and UK companies at an average of only one time per site.

Within the context of US and Chinese firms, top application keywords include battery, composite, and electrode. Chinese firms in particular also mention semiconductors and displays / screen to a large extent. Within the realm of end-use sectors, electronics and energy are the most commonly cited keywords. While Chinese firms mention electronics much more than energy, US SMEs mention energy more than electronics. For the US websites, these results are interesting because of earlier expectations by some US-based researchers about graphene's promise in electronics, with the potential to replace silicon in certain high-end applications (Bullis, 2008).

The hierarchical clustering analysis (Figure 2) attempts to identify differences in the 2011 sample dataset as a whole, in order to gauge whether *dedicated* graphene firms exhibit distinct market orientations. Based on standardized variables with zero means, firms in Groups 1 and 2 exhibit positive *graphene-ness* (Γ scores between .33 and .83) while companies in Group 3 maintain negative *graphene-ness* (Γ scores between -1.91 and -.49). In fact, the differences in *graphene-ness* between Groups 1 and 2 and Group 3 are statistically significant at the .001 level. However, none of the other variables (e.g. globalization and partnerships) show any signs of meaningful divergence. Hence, based on the sample data, *dedicated* firms that appear to focus more exclusively on graphene are *not* more likely to engage in globalization or partnership efforts, nor are they more likely to focus on end-use sectors or applications, than graphene firms with more diverse product / research portfolios. Therefore, we do not find support for

Proposition 2, which states that firms specializing exclusively in graphene operate in exceptionally narrow niche environments and exhibit born global traits.

[Insert Figure 2 here]

With this result in mind, we turn to the differences between Groups 1 and 2. All of these firms appear to be highly involved in graphene, some more obviously so than others considering company name alone. (Company names with graphene in the title were not used to calculate the level of *graphene-ness* for each firm; e.g. for grapheneindustries.com, we subtracted the number of keyword matches for ‘graphene industries’ from the number of matches for ‘graphene’ alone). Although the sample size within each group is small, we estimate a difference of means comparison. Analysis by each group shows that the differences of means for the four variables, *globalization*, *applications*, *sectors*, and *publications*, between Groups 1 and 2 are significant ($T^2 = 45.11 > F = 19.38$) (see Table 2.) Differences for *globalization* and *sectors* between Groups 1 and 2 are significant at $\alpha = .01$, and differences for *applications* and *partnerships* are significant at $\alpha = .05$.

Are the firms in Group 1 engineering-based ventures? The high sector and application scores imply this is the case. Group 2 SMEs, by contrast, appear to more aptly fit the description of science-based firms that invest a greater proportion of their resources in the development of basic graphene technology. Moreover, because Group 1 maintains distinctively higher internationalization and partnership mean values than Group 2, we find support for Propositions 1, 3 and 4, albeit in a surprising way. Although these results must be interpreted with caution, we see that the strategies of graphene SMEs can be categorized according to three continuums, science vs. engineering-based, abundant vs. limited partnerships, and international-seeking vs. domestic-driven, *only when degree of specialization is accounted for*.

We find mixed support for Proposition 5. From an applications and end-use sector standpoint, the three UK firms clearly lag behind their US and Chinese counterparts. A cursory interpretation of the data suggests that UK graphene SMEs, as a whole, may produce graphene not for end-product purposes but rather as an input for other basic and applied research. Such a supposition is somewhat intuitive given that the technology's development breakthroughs have originated in large part from The University of Manchester, UK. Indeed, from the clustering diagram (Figure 2), we can see that UK graphene firms coalesce in the high-graphene, science-based group.

The cluster analysis and difference of means tests do not provide evidence that Chinese firms are (collectively) internationally focused. However, as Proposition 5 predicts, we find that US SMEs differ in the breadth of their product emphasis and extent of specialization in graphene. In fact, Group 3 consists of graphene SMEs with diversified product portfolios, and most of these firms are US-based. A manual review of the US SME websites indicates that some US firms operate in nanotechnology at large, producing graphene, carbon nanotubes, and other nanomaterials. Other US firms, conversely, focus exclusively on graphene as a basic material – or on products that offer distinctive features and characteristics enabled by the use of graphene. Here, we find evidence that US NTBFs appear to be both science and engineering-based, both internationally and domestically focused, and more or less engaged in partnerships.

[Insert Table 2 here]

7. Discussion

The use and analysis of web-reported data for our sample of graphene firms highlights significant country-level variations in strategy between firms in the US, UK, and China. The UK has a strong scientific leadership position (with two Nobel Prize winners associated with graphene); yet, despite scientific leadership (or ironically perhaps because of this), the UK companies in our dataset were less connected with applications and more science-driven than those in the US and China. It suggests that greater attention in UK policy be paid to commercialization of leading-edge research findings. Indeed, in October 2011, the UK government announced that it would invest £50 million (or about US$79 million) into a Graphene Global Research and Technology Hub to focus private industry and venture investment (Department for Business, Innovation and Skills, 2011).

In contrast, graphene firms in the US and China appeared to be more applications oriented, as indicated by application terms and pages on their websites. This is particularly interesting with respect to China. To date, China has few corporate patents in graphene, with none reported for our sample of Chinese SMEs. Shapira et al. (2012) report that only six of 144 Chinese graphene patents are owned by companies; the rest belong to Chinese universities and research institutions. (The situation in the US is much different, where the corporate-to-university graphene patent ratio is closer to 1:2.) As observed on several firm websites, Chinese graphene SMEs supply graphene to and are in turn supported by local universities and institutions. So, although industrialization and commercialization of these Chinese firms *appears* low given patenting output, the findings in this paper tell a different story. Based on the website data, Chinese graphene SMEs are quite engaged in the commercialization process, even if the appropriability of related intellectual property is claimed by other institutional entities (i.e. universities) in the innovation network. Hence, an analysis which depended exclusively on patent filings would likely underestimate the efforts of Chinese SMEs in emerging technologies such as graphene.

In sum, our findings lend support to Shapira, Youtie, and Kay's (2011) conclusion that the characteristics of a country's NSI influences when and under what circumstances corporate actors enter into nanotechnology markets. For instance, a country's research agenda and approach to technology transfer may influence how firms target specific opportunities in a value chain and with which technologies those opportunities are best exploited.

Table 3 summarizes the different strategic orientations of the three clustered groups. Group 1 contains high-graphene, product-focused engineering NTBFs; Group 2 includes high-graphene, materials-focused science NTBFs; and Group 3 is comprised of low-graphene firms. In principle, it is plausible that the differences between Groups 1 and 2 could in part be explained by three factors of interest, as outlined by Autio (2004). Young organizations may intentionally turn to internationalization to attain competitive advantage. The international orientation of a start-up firm amplifies its needs to engage in a network of beneficial partnerships, both for new sources of knowledge and for downstream distribution and sales channels that can result in performance enhancing innovations. We see intriguing research possibilities in examining the type, geographic makeup, and intensity of various partnership linkages to better understand how new knowledge flows into and out from nanotechnology start-up firms.

[Insert Table 3 here]

Four of the twenty firms in the study *not* associated with universities are clustered in Group 3. This implies that *dedicated* graphene firms (Groups 1 and 2) are more likely to leverage university linkages, which may provide access to knowledge, process equipment, and/or other resources. Conversely, Group 3 SMEs may view graphene development as an incremental innovation in their already diversified R&D activities and product portfolios. From a broader perspective, if extant complementary assets are important and new technological breakthroughs do not devalue those assets significantly, then incumbent firms may be more likely to succeed in the long-term (Tripsas, 1997). Thus, this may indicate that Group 3 firms (as well as larger multi-national firms) may have an advantage over *dedicated* graphene firms with high commercialization interests (i.e. those firms in Group 1). Moderating factors on this incumbent advantage include new-entrant technological superiority and extent of embeddedness in the innovation network. Here, we see the need for additional empirical work guided by interviews, surveys, and commercial data sources.

8. Conclusions

This paper has sought to examine how new enterprises form in highly novel science-based areas. Graphene has seen rapid growth over a short period of time, from benchmark discovery in 2004, accelerated research activity from 2007, and Nobel recognition in 2010. This short scientific history has been accompanied by the relatively fast entrance of small companies into the graphene domain. Although the field is in the early stages of commercialization, the visible and early presence of companies suggests that the discovery-to-application path is likely to be accelerated for graphene. This paper employed a novel method – web-scraping – to understand the paths that small companies engaged in the graphene domain are taking. Thus, this paper has conclusions both for the use of web-scraping as an additional method to develop intelligence about enterprise strategies and for the understanding of entry strategies in emerging technologies.

Our experimentation with web-scraping methods to develop new sources of information on firm strategies raises a series of issues and implications. We acknowledge that web sites are "self-reports" and recognize the importance of understanding what can – and cannot – be captured by this method. While processing data from websites offers a new source of information to complement other sources, attention needs to be given to the ways that different firms, sub-sectors and countries use web-sites. Firms vary in what they report, and there may be strategic moments when firms seek to post more or less information about their technology development and business strategies. If firms seek visibility, notice from potential customers or funders, attention to their technology or product catalogue, the chance to demonstrate an edge over potential competitors, or other market opportunities, they are likely to use their web site as one of their communication tools. Conversely, if a firm desires to be more surreptitious about its technological capabilities, research lines, or partnerships, its web site may go dark for a period with limited information reported.

Our study draws data from Chinese as well as UK and US firms. A potential concern is whether Chinese companies systematically differ from or are limited in what they report compared with their Anglo-American counterparts. This paper finds that one can indeed acquire revealing information from Chinese SME web sites. We found that Chinese companies have about twice as many current pages per web site (63) in 2011 as US companies (33) and UK companies (22). In addition, accounting for language barriers, this research used Google Translate to interpret Chinese language websites. While there are limitations with the use of this tool, our experience with translation technologies coupled with manual verification by native Chinese speakers is that of high accuracy in correctly translating the English-equivalent to our search terms.

Our web-scraping method is tuned to capture and analyze raw text on web sites. While we are confident in the method's precision on full text data, we acknowledge that other forms of media, such as images, PDFs, or Flash displays, are not within the scope of analysis. It is possible that this limitation may under-capture the information available on web sites from firms placing greater weight on non-textual methods of communication. Consequently, a future line of work is to develop techniques for capturing and analyzing a wider variety of web-based content. Additionally, we note that web pages change on a frequent and possibly ad hoc basis. In this respect, website data sources are more capricious than patent and journal article data, which rarely change after initial publication. Although our study is based on a cross-sectional analysis for one time period, we see opportunities for further studies to capture and amalgamate firm-level web site data, both retrospectively (see also Youtie et al., 2012) and going forward in real-time, to determine longitudinal shifts in enterprise strategies and technology pathways. We also look to other web sites, including those of media, business support organizations, and universities, which are searchable, that could be "scraped" to gather additional current and historical information about the company. In conclusion, we do not suggest sole reliance on web sources; rather, we position web-scraping as a fresh addition to available bibliometric, patent and business databases information sources.

The literature on NTBFs emphasizes the difference between firms that are oriented toward applications and firms that focus on the scientific aspects of an emerging field. This research confirms the usefulness of the science-versus-application bi-modal continuum: We distinguish between firms oriented toward scientific pursuits and those that pursue applications. At the same time, our findings suggest that a second dimension is important for capturing new enterprise activity in science-intensive domains, that of specialization-integration. This work found that some NTBFs specialized in the production of graphene material, entering at the frontier of research with dedicated strategies ranging from developing new product applications to advancing material applications with a focus on consistency and scale-up of production. Other NTBFs sought to integrate graphene into their existing applications or incorporate them into other nano-material offerings. In addition, we contribute to the literature on born global enterprises by underlining the importance of the uniqueness of novel offerings prior to internationalization.

As a result of our work, we found a set of three pathways for small firm entry into the commercial opportunities offered by graphene. These pathways comprised specialized product development, specialized materials development, and integration into existing product portfolios. While additional modes of commercialization may emerge as graphene development continues, even at this stage it appears that any particular bi-modal typology is insufficient for capturing early small firm strategies. It is likely that further work (including with larger samples of firms) on the specialization-integration orientation of international and domestic NTBFs could be productive in refining perspectives on SME entry strategies in emerging technologies.

Acknowledgements

Support for this research was provided through the Center for Nanotechnology in Society at Arizona State University (National Science Foundation Award 0531194). Additional support was provided by the Manchester Institute of Innovation Research, the Georgia Tech Program in Science, Technology and Innovation Policy, and the Manchester-Atlanta-Beijing Innovation Co-Lab (with support from the British Council UK-US New Partnership Fund). The findings in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation, the authors' institutions, or the other sponsors.

References

Arora, A. & Gambardella, A. (1994). The changing technology of technological change: general and abstract knowledge and the division of innovative labour. *Research Policy*, 23, 523-532.

Abernathy, W. & Utterback, J. (1978). Patterns of industrial innovation. *Technology Review*, 80(7), 40-47.

Ahuja, G., Lampert, C. M., & Tandon, V. (2008). Moving beyond Schumpeter: Management research on the determinants of technological innovation. *Academy of Management Annals,* 2, 1-98.

Auerswald, 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* (pp. 18-35). Cheltenham, UK: Edward Elgar.

Autio, E. (1997a). "Atomistic" and "Systemic" Approaches to Research on New, Technology-based Firms: A Literature Study. *Small Business Economics*, 9(3), 195-209.

Autio, E. (1997b). New, technology-based firms in innovation networks symplectic and generative impacts. *Research Policy*, 26(3), 263-281.

Autio, E. (2004). Creative tension: the significance of Ben Oviatt's and Patricia McDougall's article "toward a theory of international new ventures." *Journal of International Business Studies*, 36(1), 9-19.

Barney, J.B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17, 99 – 120.

Brouwer, E. (1999). Innovative output, and a firm's propensity to patent: An exploration of CIS micro data. *Research Policy*, 28(6), 615-624.

Bullis, K. (2008). TR10: Graphene Transistors. *MIT Technology Review*. http://www2.technologyreview.com/article/409593/tr10-graphene-transistors/. Accessed December 11, 2012.

Cohen, W.M. & Levinthal, D.A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Sciences Quarterly*, 35(1), 569-596

Cosh, A. & Hughes, A. (2009). Never mind the quality feel the width: University–industry links and government financial support for innovation in small high-technology businesses in the UK and the USA. *The Journal of Technology Transfer*, 35(1), 66-91

Davis P., Maslov, A., & Phillips, S. (2005). Analyzing history in hypermedia collections. *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia*. 171-173.

Department for Business, Innovation and Skills. (2011, Oct. 3). £50 million hub to commercialise Nobel Prize winning material. Central Office of Information, News Distribution Service. Retrieved from http://nds.coi.gov.uk/content/Detail.aspx?ReleaseID=421451&NewsAreaID=2. Accessed January 20, 2012.

Eisenhardt, K. M. & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal*, 21(10-11), 1105-1121.

Gabrielsson, M., & Gabrielsson, P. (2011). Internet-based sales channel strategies of born global firms. *International Business Review*, 20(1), 88-99. doi:10.1016/j.ibusrev.2010.05.001

Gabrielsson, M., Kirpalani, V. H. M., Dimitratos, P., Solberg, C. A., & Zucchella, A. (2008). Born globals: Propositions to help advance the theory. *International Business Review*, *17*(4), 385-401. doi:10.1016/j.ibusrev.2008.02.015

Geim, A. K. & Novoselov, K. S. (2007). The rise of graphene. *Nature Materials*, 6(3), 183-91.

Hagen, B., Zucchella, A., Cerchiello, P., & De Giovanni, N. (2011). International strategy and performance—Clustering strategic types of SMEs. *International Business Review*. doi:10.1016/j.ibusrev.2011.04.002.

Katz, J. S. (2006). Web indicators for complex innovation systems. *Research Evaluation*, 15 (2), 85-95.

Kim, J. H. (2012). A Hyperlink and Semantic Network Analysis of the Triple Helix (University-Government-Industry): The Interorganizational Communication Structure of Nanotechnology. *Journal of Computer-Mediated Communication,* 17(2), 152-170. doi:10.1111/j.1083-6101.2011.01564.x

Lux Research. (2007). *The Nanotech Report: Investment Overview and Market Research for Nanotechnology*, 5th ed. Lux Research, New York.

Malecki, E.J. (1997). *Technology & Economic Development*, 2nd ed. Essex, UK: Addison Wesley Longman.

Michael, S. C. & Pearce, J. A. (2009). The need for innovation as a rationale for government involvement in entrepreneurship. *Entrepreneurship & Regional Development*, 21(3), 285-302.

Miyazaki, K. & Islam, N. (2007). Nanotechnology systems of innovation - An analysis of industry and academia research activities. *Technovation*, 27(11), 661-675.

Mowery, D. & Sampat, B. (2006). Universities in National Innovation Systems. In Fagerberg, J., Mowery, D., Nelson, R. (Eds.), *The Oxford Handbook of Innovation* (pp. 209-239). Oxford, UK: Oxford University Press.

OECD. (2010). *The Impacts of Nanotechnology on Companies: Policy Insights from Case Studies*. Paris: OECD Publishing.

Oviatt, B. M. & McDougall, P. P. (2004). Toward a theory of international new ventures. *Journal of International Business Studies*, 36(1), 29-41.

Porter, A.L., Guo, Y., Huang, L., & Robinson, D.K.R. (2010). Forecasting Innovation Pathways: The Case of Nano-enhanced Solar Cells. *International Conference on Technological Innovation and Competitive Technical Intelligence* (ITICTI), Beijing.

Rasmussen E. & Madsen, T. (2002). The Born Global Concept. Presented at the European International Business Academy, 2002. Available at http://www.sam.sdu.dk/~era/EIBA%20Rasmussen%202002%20.pdf. Accessed 10 February, 2011.

Rennie, M. W. (1993). Born Global. *McKinsey Quarterly*, 1993(4), 45-52.

Sandhu, A. (2007). The birth of nanoChina. *Nature Nanotechnology*, 2(1), 11-12.

Segal, M. (2009). Selling graphene by the ton. *Nature Nanotechnology*, 4(10), 612-4.

Shapira, P. & Wang, J. (2009). From lab to market: Strategies and issues in the commercialization of nanotechnology in China. *Asian Business & Management*, 8, 461-489.

Shapira, P. & Wang, J. (2010). Follow the money. *Nature*, 468, 627-8.

Shapira, P., Youtie, J., & Kay, L. (2011). National innovation systems and the globalization of nanotechnology innovation. *The Journal of Technology Transfer*, 36(6), 587-604.

Shapira, P., Youtie, J., & Arora, S. (2012). Early patterns of commercialization of graphene, *Journal of Nanoparticle Research*. Online First. DOI 10.1007/s11051-012-0811-y.

Tripsas, M. (1997). Unraveling the process of creative destruction: Complementary assets and incumbent survival in the typesetter industry. *Strategic Management Journal*, 18(S1), 119-142.

Van Noorden, R. (2011). Chemistry: The trials of new carbon. *Nature*, 469, 14-6.

Walker, W. (1993). National Innovation Systems: Britain. In Nelson, R. (Ed), *National Innovation Systems: A Comparative Analysis* (pp. 158-191). New York: Oxford University Press.

Wessner, C. (2007). Government programs to encourage innovation by start-ups and SMEs: the role of US innovation awards. In D.B. Audretsch, I. Grilo, & A.R. Thurik (Eds.), *Handbook of Research on Entrepreneurship Policy* (pp. 172 – 185). Cheltenham, UK: Edward Elgar.

Youtie, J., Hicks, D., Shapira, P., Horsley, T. (2012). Pathways from discovery to commercialization: Using web sources to track small and medium-sized enterprises strategies in emerging nanotechnologies. *Technology Analysis and Strategic Management*, in press.

Youtie, J., Shapira, P., & Porter, A. L., (2008). Nanotechnology publications and citations by leading countries and blocs. *Journal of Nanoparticle Research*, 10(6), 981-986.

Yu, J., Stough, R. R., & Nijkamp, P. (2009). Governing Technological Entrepreneurship in China and the West. *Public Administration Review*, 69(S1), S95–S100.

Zhou, P. & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, 35(1), 83-104.

Figure 1: The web scraping architecture

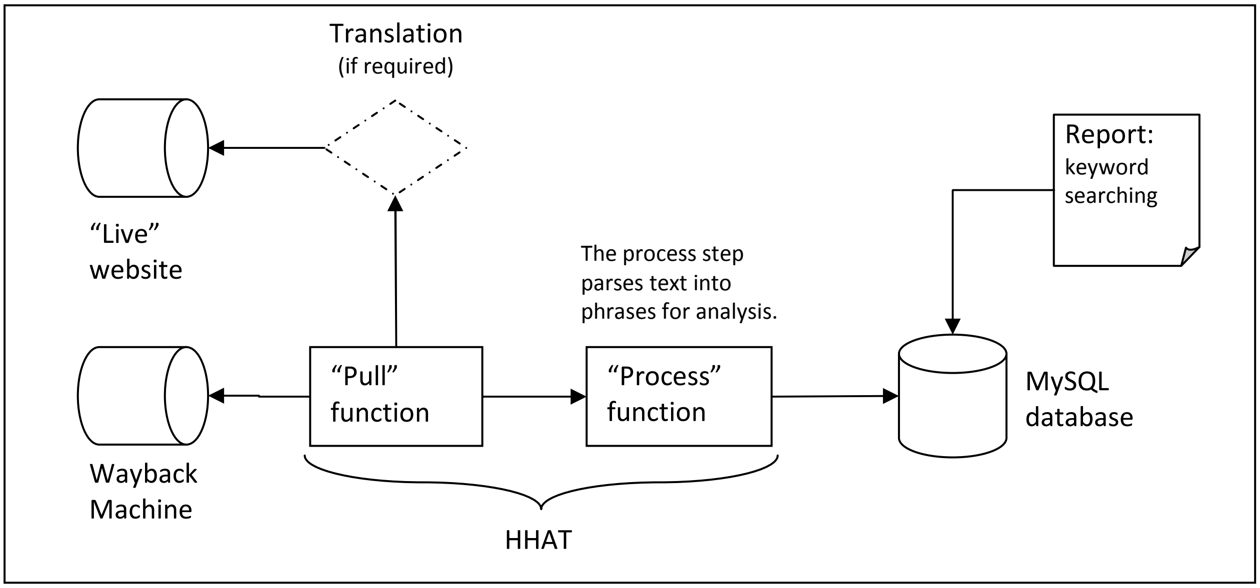
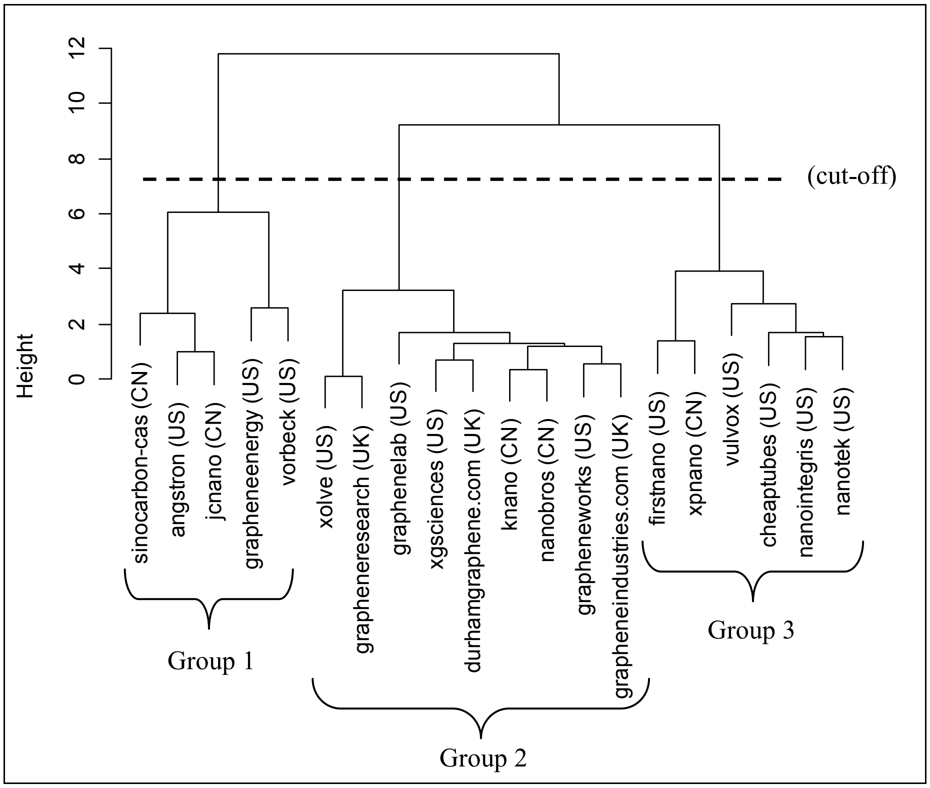


Figure 2. Ward Cluster Dendgrogram



Note: Distance measures are based on the standardized variables, graphene-ness, applications, sectors, globalization, and partnerships.

Table 1: Sample of Graphene SMEs in the US, UK, and China.

Country	Company	Location	University Affiliation	** Patents	† Website Pages
US	* First Nano	Long Island, NY	-	-	139
	Vorbeck	Jessup, MD	Princeton University	6	48
	Cheap Tubes	Vermont	-	-	39
	Nanotek Instruments	Dayton, OH	Wright State University	8	32
	NanoIntegris	Skokie, IL	Northwestern University	-	31
	Angstron Materials (a Nanotek spin-off)	Dayton, OH	Wright State University	-	30
	XG Sciences	Lansing, MI	Michigan State University	-	15
	Graphene Laboratories	Reading, MA	Columbia University	-	14
	Vulvox	Long Island, NY	-	-	13
	Graphene Works	Atlanta, GA	Georgia Institute of Technology	-	10
	Xolve (formerly Graphene Solutions)	Platteville, WI	University of Wisconsin	-	7
	Graphene Energy	Austin, TX	University of Texas	-	5
UK	Graphene Research	Manchester, UK	University of Manchester	-	69
	Graphene Industries	Manchester, UK	University of Manchester	1	7
	Durham Graphene Science	Durham, UK	Durham University	-	8
China	Nano Technology Company	Nanjing, Jiangsu	Nanjing University	-	104
	XP Nano Material Company	Xiamen, Fujian	-	-	71
	Sinocabon Technology Materials ‡	Taiyuan, Shanxi	Chinese Academy of Sciences	-	59
	Xiamen Knano Graphene Technology Company	Xiamen, Fujian	Huaqiao University	-	53
	Nano-Brother Laboratory	Haerbin, Heilongjiang	Harbin Institute of Technology	-	29

Notes: Most university relationships were observed explicitly on company websites, but some relationships were deduced via other channels, including online search and local knowledge networks. * Indicates a subsidiary of a publicly traded company, CVD Equipment Corporation. ** Patent count from Derwent Innovation Index reflects number of patents where firm is the priority assignee (as of 2011). † Websites captured between November 2010 – January 2011. ‡ Follow-up review (early 2012) finds company website no longer online - suggests firm has exited or restructured.

Table 2. Difference of means comparison

Group	Designation	Variable	Mean	s.d.	Min	Max
1	High graphene, product focused	graphene-ness	.49	.21	.33	.83
		globalization	1.36	1.05	.11	2.96
		applications	.97	.44	-.20	2.64
		sectors	1.11	.70	.26	2.65
		partnerships	1.08	.39	-.31	3.01
2	High graphene, materials focused	graphene-ness	.67	.14	.39	.83
		globalization	-.51	.33	-.90	-.02
		applications	-.59	.39	-1.16	-.11
		sectors	-.60	.46	-1.17	0
		partnerships	-.39	.51	-.99	.68
3	Low graphene	graphene-ness	-1.4	.54	-1.91	-.49
		globalization	-.36	.51	-.90	.19
		applications	.08	.96	-.93	1.56
		sectors	-.03	.95	-.92	1.48
		partnerships	-.32	.59	-.97	.54

Variables presented in this table are centered across the 20 firms.

Table 3: A comparison between the three groups of graphene firms

	Group 1	Group 2	Group 3
Description	High graphene, product focused	High graphene, materials focused	Low graphene, diversified nanomaterials
Characteristics	More reliance on university links High application and end-user sector scores convey these NTBFs are engineering-based High globalization and partnership scores convey born global characteristics	More reliance on university links Low application and end-user sector scores convey these NTBFs are science-based Low globalization and partnership scores convey local focus	Less reliance on university links Mixed product and end-use sector scores No evidence to support a domestic or international market focus when compared to Groups 1 & 2
Strategy Implications	Growth-oriented but focused product portfolio may indicate fewer complementary assets	Growth may not be a priority	*If* growth oriented, diverse product portfolio may indicate access to in-house complementary assets

Appendix Table A1: Descriptive Statistics for Application Keywords

Keyword(s)	Firm country	Mean	s.d.	Min	Max
battery	China *	10.8	7.9	1	23
	UK	0.3	0.6	-	1
	US *	17.0	24.0	-	67
composite	China *	32.0	20.4	1	52
	UK	4.7	8.1	-	14
	US *	16.0	22.8	-	76
electrode	China *	10.6	8.2	2	22
	UK	-	-	-	-
	US *	7.6	10.4	-	29
semiconductor	China *	11.0	13.8	1	34
	UK	-	-	-	-
	US	4.7	7.1	-	26
display / screen	China *	16.2	14.9	2	39
	UK	0.3	0.6	-	1
	US	4.7	6.4	-	18
transistor	China	8.4	8.5	-	18
	UK	1.0	1.7	-	3
	US	3.3	7.7	-	27
diode	China	1.0	1.0	-	2
	UK	-	-	-	-
	US	0.7	1.1	-	3
sensor	China	4.8	3.1	2	9
	UK	1.3	2.3	-	4
	US	2.3	3.2	-	9
detector / detection	China	4.2	7.3	-	17
	UK	-	-	-	-
	US	4.7	11.3	-	39
radio / RFID	China	1.2	1.1	-	2
	UK	-	-	-	-
	US	1.6	3.0	-	8
capacitor	China	4.4	3.4	1	9
	UK	0.3	0.6	-	1
	US	4.2	6.4	-	19
circuit	China	3.6	4.3	-	10
	UK	0.3	0.6	-	1
	US	3.2	4.3	-	15
ink	China	0.6	0.9	-	2
	UK	-	-	-	-
	US	5.8	17.4	-	61

Note: * denotes top terms

Appendix Table A2: Descriptive Statistics for Sector Keywords

Keyword(s)	Firm country	Mean	s.d.	Min	Max
electronic	China *	51.8	63.0	5	153
	UK	1.0	-	1	1
	US *	16.2	23.3	-	73
energy	China *	25.6	27.6	5	64
	UK	-	-	-	-
	US *	20.1	16.0	-	44
medicine	China	0.6	0.9	-	2
	UK	-	-	-	-
	US	1.0	1.5	-	5
aerospace	China	5.0	4.8	-	12
	UK	-	-	-	-
	US	2.2	4.3	-	15
automotive	China	2.6	2.1	-	5
	UK	-	-	-	-
	US	4.3	6.1	-	20
construction	China	1.0	1.4	-	3
	UK	-	-	-	-
	US	1.1	2.6	-	9
environment	China	1.2	1.8	-	4
	UK	-	-	-	-
	US	0.6	0.5	-	1

Notes: * denotes top terms. For reporting purposes, many of these terms are comprised of multiple keywords; e.g. "automotive" includes automotive, cars, and vehicles.