Technological Forecasting & Social Change 76 (2009) 1138–1149

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Analysis for radical design

Scott W. Cunningham

Policy Analysis Section, Faculty of Technology, Policy and Management, Delft University of Technology, Postbus 5015, 2600 GA Delft, The Netherlands

A R T I C L E I N F O

Article history:
Received 17 November 2008
Accepted 15 July 2009

Keywords:
Hierarchical random graphs
Architectural innovation
Technology forecasting
Design

A B S T R A C T

This paper examines a technique suitable for monitoring and analyzing systemic change in technology. Technological changes increasingly stem from the novel recombination of existing technologies. Changes are multitudinous. Therefore, new techniques are needed for analyzing technology architecture. A literature review of related work in the field of technology opportunities analysis is presented. We consider a possible, radically decentralized context for the conduct of future design. A case study of new technology architecture in the information technology domain is presented. An analytical method involving mining weighted graphs from technology archives is presented. The role of this new method in a context of distributed decision-making and design is presented.

1. Introduction

This paper examines a technique suitable for monitoring and analyzing systemic change in technology. Technological changes increasingly stem from the novel recombination of existing technologies. Therefore, new techniques are needed for analyzing technology architecture. The work is significant because the analysis of a very significant feature of technological change–the recombination of existing components–is not being supported by most technology opportunity analysis techniques.

A literature review of related work in the field of technology opportunities analysis is presented. A case study of new technology architecture in the information technology domain is presented. An analytical method involving mining weighted graphs from technology archives is presented. The role of this new method in a context of distributed decision-making and design is presented.

A prospective analysis of new technology fundamentally hinges on the concept of novelty. It is the newest and most novel of technologies which presents the greatest challenges for technological forecasting. Fundamental uncertainty surrounds the exploitation and development of new technologies. Much has been made about the convergence of new technologies, particularly in the information, biotechnology and material sectors [1]. The forces impelling convergence at the time are seen as radical, revolutionary, and deeply uncertain. One recent study, for instance, investigated uncertainty and the emergence of dominant designs in aircraft [2]. While in retrospect, the design seemed assured, the actual choices at the time appeared divergent and highly contested.

One approach to the management of technological uncertainty has been to initiate the technological forecasting process only once a dominant design has emerged [3]. Once a dominant design has been selected, uncertainty is fundamentally reduced; processes of organizational and sectoral learning then assist in securing a niche for the new technology. Trend extrapolation approaches, for instance, are based on tracking the emergence of new technologies only once a dominant design is secured [4].

This solution of tracking dominant designs neglects some of the fundamental uncertainty associated with technological evolution. Deep uncertainty characterizes many domains of decision-making in science and technology. In particular, under deep uncertainty, there is little agreement or consensus about system structure. Thus, exploratory modeling is used to explore

E-mail address: S.Cunningham@tudelft.nl.

0040-1625/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2009.07.014

alternative views of the future, seeking decisions robust under a variety conflicting forces [5]. Uncertainty in new design arises in at least two areas [6]. Technological design is an inherently uncertain process which is therefore subject to epistemic uncertainty [7]. Technological design entails the recombination of components in new and often unexpected fashion. New techniques for managing the fundamental uncertainty in technological design and evolution are therefore needed. Previous work has provided technology analysts with a set of techniques for both integrating and decomposing new technologies.

Relevant research has approached the problem of forecasting radical technological change with methods for supporting analysis for both decomposition and integration of new technologies. For the decomposition of technologies, morphological analysis has long been practiced as a technique for recognizing component technologies. Patent studies have used TRIZ to investigate the character of innovative activity [8,9]. Integrative methods also allow for the anticipation of converging technology. Swanson explored knowledge discovery by exploring database links [10]. Swanson demonstrates integrative capability by demonstrating new links between technologies, inherent in the data, which were not readily apparent to the respective scientific communities.

In this paper we examine techniques for exploring emergent structures or architectures of technology. Consider a knowledge base of technology where components of technologies are described, and linkages between the technologies are identified. This description of a knowledge base describes many repositories of scientific and technological information, including: the Internet, science and technology databases, patent databases, newswires, and potentially also newsgroups or other online collaborative environments.

This work is similar in spirit to that of Swanson and Smalheiser reviewed earlier [10]. It extends and elaborates upon the procedures described by these authors for discovering new linkages of knowledge through use of a structured representation of science and technology, which facilitates finding the integrative terms responsible for novel recombination of component technologies. In addition, the paper outlines several important caveat about the use of these models in forecasting new technology: the model lacks a model of the actor; full validation of the model requires a longitudinal analysis; missing links may signal poor quality source materials; and content scoring remains a subjective process.

2. Application to distributed design environments

Our purpose in exploring this topic is to better consider the information needs of designers. Designers may soon be positioned in a new and radically decentralized environment. In this section, the paper explicates the social and technological organizational structures which may permit a new era of open innovation. Chesborough [11] describes a new paradigm of “open innovation” involving the design of technological systems which, in technological requirements, transcends the boundaries of a single firm. Likewise, in terms of knowledge production, researchers form multi-disciplinary teams devoted to specific problems and specific contexts [12].

These authors go further: rather than simply describing a new and distributed environment, they prescribe the manner in which innovative organizations can create an open and porous environment by which to participate in this anticipated new mode of innovation. Given the imperative of new conduct in innovation, so the authors argue, it is necessary for responsive organizations to restructure themselves to exploit this knowledge environment. The actual economic and institutional arrangements necessary to create flexible and distributed networks may have been captured in the regional development literature [13]. Such networks require special technological and infrastructural capabilities to succeed in this emerging environment. In the following paragraphs some ideas about the organization of technological knowledge is described; this knowledge is coupled with the institutional environment of distributed knowledge production. This review suggests an important avenue for research in this article, and in future research: creating software solutions to help innovative organizations develop new technologies within an open innovation environment.

Knowledge is structured hierarchically. A shared knowledge base, developed by multiple, decentralized actors may be stored in a network form. This network, although highly diffuse, may be accessed by all players for personal as well as community betterment. Hierarchies are one form of technological structure confirmed by theories and practice. A hierarchy is one structure of many that have been used for technological integration [14]. The International Institute for Applied Systems Analysis (IIASA) and others have examined the hierarchical embedding of infrastructure systems, and used such hierarchies to examine and anticipate structural change [15]. Clark [16] considers how market forces shape technological hierarchies, and how such hierarchies in turn shape the market Thus, there is a rich basis of theoretical support for structuring technological component data in a hierarchical format.

The knowledge base may be equated with a network structure: the nodes are technologies, and the edges are the component relationships that are present between the respective technologies. The challenge of the technology analyst is to usefully structure this information to anticipate change. The technology designer has a similar challenge in exploring new, heretofore unforeseen, combinations of new technologies. The network data in the raw is not useful for this purpose.

A structured representation of technology is needed for multiple reasons: interpretation, theory, robustness and also the production of actionable results. An unstructured network contains many parameters, which are hard to visualize and interpret. The technology analyst requires structured information, conforming to theories about the organization of science and technology. Without a theory of the data the technology analyst cannot distinguish between meaningful structure and possibly accidental corruption of the knowledge base. Therefore, without a generative model of the data, the interpretation of the data may not be robust. The technology analyst needs to anticipate change. A structured representation of the data provides a principled account of where technological change is most likely to occur.

The article which follows argues that there is a sizeable amount of open source information which is shared between distributed communities of designers and researchers. This knowledge is stored in databases of science and technology. New future-oriented technology analysis techniques, such as the approach suggested here, may contribute to the process and management of radical innovation [17,18]. Radical innovation establishes a new dominant design, eliminating older ideas, and thereby creating a new set of design concepts and a new configuration of components technologies.

The express purpose of these science and technology databases is to research specific existing technologies, and yet the biggest promise of these sources of information may be the diffuse and distributed information they contain about the current state of the knowledge of the community. This diffuse knowledge is networked, and relatively unstructured. What is needed therefore, are techniques for extracting these networks, and accurately structuring the knowledge so that it can be used for analysis, design and forecasting. If this hypothesis of distributed knowledge bases is correct, then it implies certain things about the information and infrastructural needs of organizations engaged in open innovation. Such organizations, not surprisingly, need access to these distributed databases of knowledge. Unlike conventional, disciplinary researchers, these organizations do not necessarily need the database to gain access to individual pieces of information (whether it this be press releases, patents, or research articles). Rather, the participating organizations use the database as a coordination mechanism, enabling them to rapidly respond to the research and development efforts of their peers. Thus, they also require analytic support to gather, structure, and magnify the resident knowledge of distributed communities of peers.

The premise of available, but decentralized, knowledge of science and technology is something which can be tested through the use of machine learning techniques. These concepts are explored further in this paper with a case exploring a software technology known as AJAX. Should the hypothesis of distributed knowledge be demonstrably true, and therefore the value proposition of software for design support be made clear, it is still important to make clear the role of the designer in this process. The primary role of the designer should involve creative recombination of new ideas. If the routine process of exploring for new technological components can be automated, then the designer will be free to spend more time at value-enhancing activities.

The roadmapping activity achieves value by providing a single locus for coordinated research and development activity (see for instance [19]). It is perhaps not surprising then, that many users of technology roadmapping exist in a vertically integrated environment where a few big players have the interests and capabilities to assist in technological coordination. The description of distributed design in this article is, perhaps, somewhat at odds with the stated premise of technology roadmapping. Nonetheless, the purpose of this article is not to advocate flexible networks of innovating firms as a preferred form of innovative activity. Rather, the paper argues that regardless of the particular institutional organizations of innovation which emerge, analysts should avail themselves of a wide variety of techniques appropriate for the task at hand.

3. Methodology

In the following section we develop an analytical method for the representation of emerging technologies in the form of a hierarchical graph. The analytical methods for this approach have emerged from scientometrics, machine learning, graph theory and complexity studies. Systems ecology for instance provides a formal theory of morphological change [20]. A hierarchical random graph is a succinct recipe for generating and encoding a range of related network structures. Using the model is a two-part process of simulating a range of possible networks specified by the model, and comparing and analyzing these results given observed data. Real world evidence is used to tune and parameterize the specific model representations used. In the following section we provide a brief and qualitative account of this model.

Extensive technical details of this data structure are available in the literature on complex networks. A literature on hierarchical random graphs, and their use in managing information about complex networks is emerging [21–24]. Clauset [21] for instance, provides a useful survey article on the random hierarchical graph. We see to complement the technical literature with the following interpretive and example-oriented discussion of the methodology.

The hierarchical random graph consists of a tree-like series of nodes. Tree-like indicates that the structure is fully connected, and that there are no cycles between the nodes. Parent nodes are represented with rectangles and are labeled with probabilities. These probabilities which specify the probability of observed linkages between the right and left side of the tree. Parent nodes are not directly observable in the data. Children nodes can be directly observed in the data, and are given the corresponding labels.

3.1. Example of hierarchical random graph

An example of a hierarchical random graph is presented below. There are four children nodes (A, B, C, D). In this technology analysis application these nodes represent four component technologies of a system. These four nodes may be connected in sixty-four possible networks, representing various combinations of the component technologies. Rather than a full enumeration of links, any observed network of these four component technologies can be compactly represented by introducing three parent nodes, each with their own associated probabilities of linkage. These parent nodes, which cannot be directly observed in the data, represent morphological principles actively at work in structuring the data.

In the example given below there is a 70% chance that nodes C and D are linked with nodes A and B; there are four potential connections here. Contrariwise there is only a 35% chance that node A is linked to B, or that node C is connected to D.

The resultant hierarchical random graph has replaced six bits of information about network connectivity with three probabilities: this is a savings in representation. Whether this is an appropriate exchange depends in part on how effectively we can encode a given network structure, and how concerned we are with a robust representation of the data in the presence of noise.

The hierarchical representation of the data grows more attractive as the network grows larger, and as more noise is introduced into the system. However, we no longer have a single unique specification of the network — the hierarchical random graph describes an ensemble of the sixty four possible graphs. Some of these sixty-four graphs are very likely to be generated by the example hierarchical random graph, while others remain unlikely.

3.2. Simulation of new networks using the graph

The figure above presents two possible realizations of the hierarchical random graph shown in Fig. 1. These realizations are made by randomly establishing the presence or absence of a given linkage given the probabilities specified from Fig. 1. Note that while the generative model is a connected graph, specific realizations of the model may be disconnected.

Hierarchical random graphs such as the one presented in Fig. 1 are generative models, meaning we can use the model to infer the likelihood that any given realization will be generated by the model. Realization 1 is a more likely representation since the most of the high probability links are observed, and few of the low probability links are observed. This network configuration should be expected to be observed 4.3% of the time, or roughly one in 24 times. In contrast, realization 2 is a possible but unlikely realization of the network since most of the high probability connections did not actually occur (Fig. 2). This network occurs 0.8% of the time, or roughly one in 125 realizations.

These likelihoods may be calculated by an equation closely related to the binomial distribution. The binomial distribution is used to calculate the likelihood of a series of results in a set of yes/no experiments given fixed probabilities. The full calculation of the likelihood of the hierarchical random graph depends on a set of probabilities and nodes determined by each parent node (r) in the graph.

The hierarchical random graph is capable of generating a range of results, ranging from perfectly deterministic networks, to highly random networks. According to the specification of the model there may be little or no hierarchical structure, or a network which is richly structured across multiple layers. Fig. 3, below (adapted from [21]), demonstrates how very different networks may be encoded using the hierarchical random graph format.

The left-most graph is a flat network, representing a random network where every node is likely to connect to every other network. The middle network shows clear clustering, and yet is still a flat network. Members of the three groups in this network are likely to connect to each other, but are unlikely to connect to members of other groups. The right most network shows organization at multiple scales: there are clusters, within clusters, within clusters. All three of these networks are assortative: the closer the position on the tree, the more likely two nodes are to connect. Disassortative networks are also possible: progressing down the tree represents a progressive differentiation of members. Thus, the hierarchical random graph is a very expressive formalism capable of capturing many possible network relationships.

3.3. Fitting graphs to data

For each of these structures we must also estimate the associated probabilities of network linkages at each of the parent nodes. The best fit is achieved by fitting probabilities according to the actual proportion of linkages observed in the data. This is the maximum likelihood estimate of the model parameters given the data [21]. With only fifteen possibilities, we can exhaustively search the space of possible network structures. There are only forty-five probabilities which must be estimated; three for each of the model structures. However the combinatorics will explode as we expand our examples to more realistic networks with multiple child nodes. An exhaustive search is no longer possible. We therefore need a way to structure the search to spend most of our time on the most likely network structures.

A derivation of the likelihood [21] is given below in Eq. (1). In this equation the likelihood of a particular observation is dependent on the observed network (D), the specified series of probabilities in the hierarchical random graph (p), the total

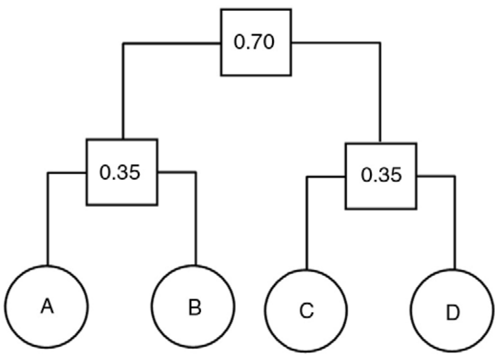

Fig. 1. Example hierarchical random graph.

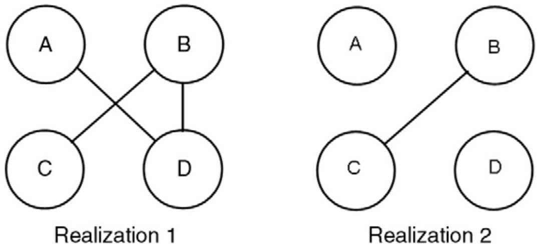

Fig. 2. Two realizations of the example hierarchical random graph.

number of links or edges observed (E) and the number of possible number of interconnections related to the nodes on the left of the tree (L) and the right of the tree (R). The calculations are performed across all the parent nodes (r) in the hierarchical random graph. As noted in the equation likelihoods are calculated using a multiplicative function.

$$\mathcal{L}(D,\{p_r\}) = \prod_{r \in D} p_r^{E_r}(1-p_r)^{L_r R_r - E_r}. \quad (1)$$

3.4. Model search process

The model simulation and fitting process allows a comprehensive search process for small models. The sufficient statistics for the observed network can be calculated. Every possible network consistent with the data can be enumerated, and the likelihood of each network model given the data can be calculated. The analyst can then choose the network or networks which provide the best fit to the data. Larger networks prevent this exhaustive search process. Nonetheless, a systematic technique for searching through the space of models is still necessary.

A Monte Carlo simulation provides a systematic search process which guarantees several desirable properties. The procedure can start with any proposed network, and can be restarted anew at any stage in the process. The search process requires only a limited look ahead. This is accomplished by successive stages of simulation (to evaluate potential alternatives), and then model fitting (to determine the most desirable alternatives). The search procedure includes a measure of network relatedness. A series of step-wise operations can transform one hierarchical graph into any other hierarchical random graph. This guarantees that (subject to randomness) all possibilities are eventually explored. The search procedure guarantees no degradation in model fit as the search progresses. The Monte Carlo simulation invests more computational results in the best available models, thereby driving the search to the most promising possibilities. Gill [25] provides a relevant and comprehensive account of Monte Carlo simulation for data analysis.

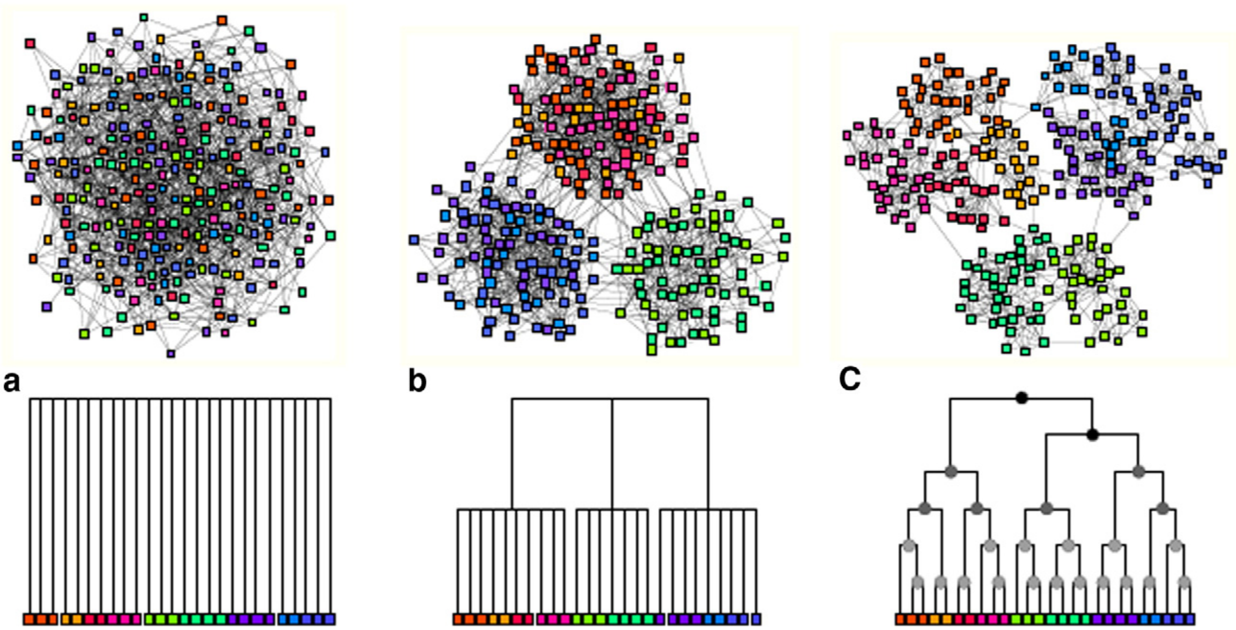

Fig. 3. Different network structures encoded with a hierarchical random graph.

4. Results

In this results section we apply the methodology described in the previous section to a specific system of new technologies. Our previous example of four technologies will be considerably expanded to the analysis of forty-one technologies within an information technology design context. The case study involves Ajax (asynchronous JavaScript and XML).

Ajax is responsible for the smoothly scrolling maps seen in Google Earth and other internet applications (Fig. 5, below). The functionality of such web interfaces, when they first featured, was revolutionary. Ajax is not so much a single technology; rather it is an architecture of technological components. Ajax is therefore a particularly good test bed to test new techniques for anticipating architectural innovation. Recognition of this architecture grew only once the component technologies were given the Ajax name [26]. In this regard of being a novel recombination of source technologies Ajax is not unlike many other modern technologies. Other examples, drawn just from the information technology sector, include grid computing, the iPod and iPhone, virtualization and LAMPP.

4.1. Data collection and comparative analysis

For the case study we collect data about Ajax and component technologies from the Internet. We use Wikipedia as a test-bed, by mining a series of pages and hyperlinks starting from a "seed" page. The seed, in this case, is the computing page on Ajax. At this disambiguation page AJAX the computing technology has been distinguished from other meanings of the word "Ajax." Then, using custom web crawling software written using the scripting language Python, all pages and hyperlinks within two hops from Ajax are collected and analyzed. We eliminate all "meta" pages concerning categorization, images, or talk. We eliminate all external links, and all links to foreign editions of Wikipedia. Finally, we eliminate all links to pages associated with dates and years, as these pages are rarely directly relevant to the topic under consideration.

Many social network analysts are structuralists. Thus, they argue that the behavior of the system is exclusively determined by the structure of the network. There is therefore a corresponding interest in the morphology of the network and in particular network properties. A few network properties have received considerable attention in the literature. First is the average degree of nodes, which is a measure of network connectivity. The second is the clustering coefficient, which is a measure of excess links between closely related nodes. The final measure is average network diameter, which is the minimum number of hops it takes to get between any two nodes in the network. These comparative network measures are used to compare and contrast networks originating from very different social, physical, or natural origins.

In the following few paragraphs a comparative analysis of the Wikipedia morphology is provided. This is useful as a point of comparison between this study and others, even if a strict structuralist account of the data is not adopted. A graphical presentation of a subset of the Wikipedia network near AJAX is given. Network characteristics of this subset of the Wikipedia are provided. Such results are intended as descriptive statistics which can be interpreted only in light of a more elaborate model of the data. A concluding section of the paper reflects upon the sociology of science, in an effort to confront the observed data with sociological theory. In short, some descriptive statistics of the network are provided despite the fact that the author does not endorse a structuralist account of the data.

The network grows rapidly in size. The seed is but a single page; this page contains 41 hyperlinks. These 41 pages connect to another 2932 pages. There are in total over 9600 hyperlinks connecting these pages. This network may be visualized using the Pajek software, with each node representing a single Wikipedia page, and each edge representing a hyperlink. The visualized graph of this network is shown below (Fig. 4).

On the left is the network with the 41 pages one hop away from the Wikipedia "Ajax (Programming)" age. On the right is the network with the 2973 pages two hops or fewer away from the "Ajax (Programming)" page. Many of these nodes in this expanded network are now very remote in content from Internet technologies. Synonymy drives much of this rapid diversification in content. Apparent even in this graph representation of the Wikipedia web is the obvious structuring of the web pages through clustering and local hierarchies.

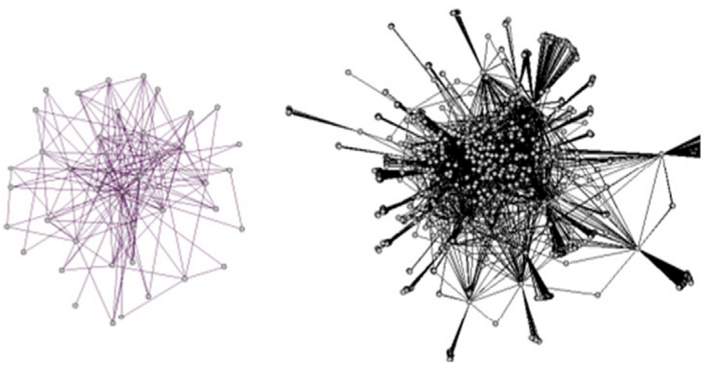

Fig. 4. Expanding network of hyperlinks in Wikipedia.

Table 1
Comparative analysis of network characteristics.

Network	‹k›	C	D
T. pallidum	4.8	0.063	3.69
Terrorists	4.9	0.361	2.58
Grassland	3.0	0.174	3.29
Wikipedia	3.3	0.195	3.45

At least three other networks have been studied in the context of hierarchical analysis of data [21]. These include a metabolic network, a grassland ecology, and a social network of terrorists. Key attributes of the network include average degree ‹k›, average clustering coefficient (C), and average network diameter (d). It is interesting to compare this Wikipedia sample with these other known networks. See Table 1.

The network shares key characteristics with the three previous networks examined. Like the network ecology, this sampled network has a relatively low degree. However, the complete Wikipedia content network probably has a much higher degree — perhaps a degree as high as 10. Like the terrorist network, the nodes are tightly coupled. And like the metabolic network, there is a relatively high average distance between all the nodes in the network. Perhaps unlike these other networks, the Wikipedia network is clearly truncated from a much larger network. The Wikipedia network is mixed in character — although more disassortative than assortative. The network is therefore similar to the grasslands ecology network which has previously been shown to be disassortative [21].

4.2. Fitting the data

The component technologies of Ajax may be represented in hierarchical random graph form. We apply the Monte Carlo simulation procedure of Clauset [21] to fit the 41 pages within one hop of "Ajax (Programming)" into a hierarchical random graph. Our goal in this analysis is to use the hierarchical structure to anticipate new changes in this field of information technology.

The Monte Carlo simulation converges to a final likelihood of −167.938, after a million runs of the simulation. The algorithm runs rapidly on a Pentium 1.73 GHz processor, with a clock speed of 795 MHz and 512 MB of RAM: a million runs are completed in 54 s of processing time. In practice, the algorithm is let run as long as there are consistent improvements in model fitting and likelihood. A visualization of the resultant hierarchical random graph is shown below in Fig. 5.

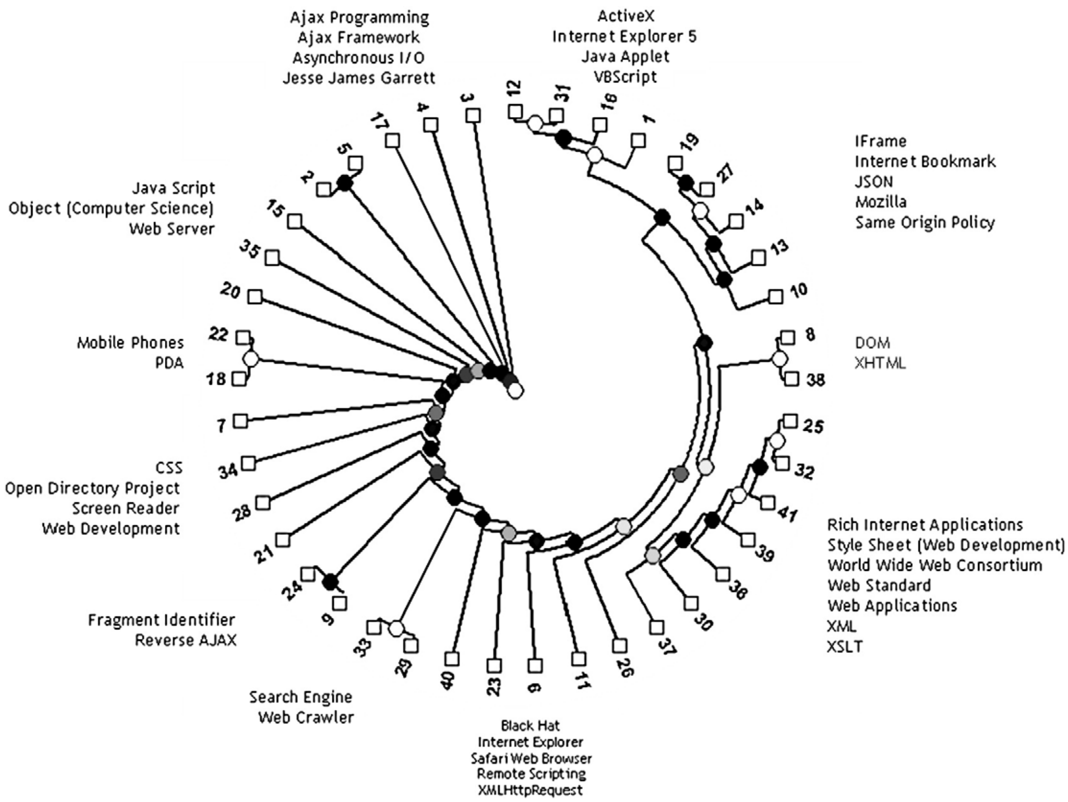

Fig. 5. Maximum likelihood hierarchical random graph for the Ajax data.

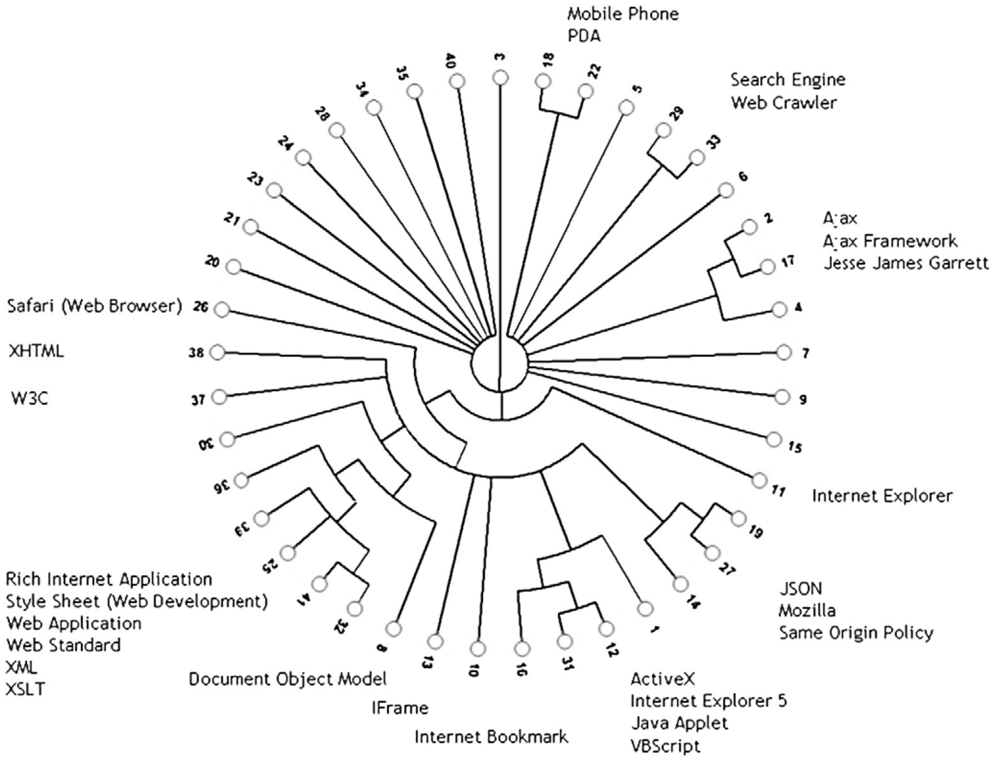

Fig. 6. Consensus diagram for Ajax Technologies.

The resultant hierarchical random graph usefully distinguishes between high-level concepts and more narrowly defined technical concepts. For instance Ajax is placed at the highest level, along with general purpose technologies such as web servers, JavaScript and object-oriented computer science. The hierarchical random graph also productively groups related technologies such as mobile phones and personal digital assistants. Another example of effective groupings of technology induced by the network structure is the combination of search engines and web crawler. At the lowest leaves of the tree are functional groupings of technologies including: scripting languages (ActiveX, Java Applets and Visual Basic Script), document models (DOM and XHTML), alternative implementations of Ajax (using JSON and IFrame), and web application technologies (including key phrases of rich Internet applications, style sheets, web applications).

The Monte Carlo simulation procedure evaluated a range of competing solutions to the model. There are many common features shared among the set of best solutions of the algorithm. It is useful to examine a consensus diagram showing the major shared features shown across multiple solutions. Such a consensus diagram helps identify robust features of the hierarchical structure, structures which may remain stable even as the network changes or expands to include new nodes. This is shown in Fig. 6.

Major features of the maximum likelihood solution are preserved in the consensus diagram. The consensus diagram makes it clear that there are two sets of technologies — technologies which are external and overarching to the Ajax programming paradigm, and technologies which may be considered internal to the Ajax paradigm. The internal technologies constitute a hierarchy in Fig. 6, while external technologies do not reveal much hierarchical structure, at least in this sample of the data.

One challenge to classification revealed by Figs. 5 and 6 is the placement of the various web browsers. Consider that Safari, Internet Explorer, Internet Explorer 5, and Mozilla–all various kinds of Internet browsers–are placed in different locations in the hierarchical random graph. On the face of it, these related technologies should probably be grouped together. However, this is not an artifact of the algorithm, but a real reflection of the Wikipedia coverage of web browsers. It may be that the various browsers become closely associated with specific innovations in media technology.

4.3. Interpreting the results

Fig. 6 does not label the parent nodes with probabilities because of graphic visibility concerns. Nonetheless, the principal virtue of this hierarchical graph approach is the ability to use this probability model to anticipate novel combinations of technologies. We are looking for new combinations which seem logical given the induced structure of the technological system but which have not yet been realized. As noted earlier the parent probabilities provide an explicit hypothesis about the nature of technological linkages. The top ten linkages which are predicted, but not yet observed in the Wikipedia knowledge base are shown below in Table 2.

Table 2
The most likely new combinations of technology predicted by the graph.

No	Topic	Node	Topic	Percent likelihood
11	Internet Explorer	25	Rich Internet Applications	80.7%
11	Internet Explorer	37	W3C	80.7%
25	Rich Internet Applications	37	W3C	75.3%
15	JavaScript	41	XSLT	70.0%
15	JavaScript	38	XHTML	70.0%
15	JavaScript	39	XML	70.0%
15	JavaScript	30	Style Sheet (Web Development)	70.0%
15	JavaScript	13	Internet Bookmark	69.0%
15	JavaScript	29	Search Engines	64.7%
40	XMLHttpRequest	38	XHTML	57.8%

Table 3
Epistemological claims predicated upon the use of a hierarchical random graph.

Claim	Claimant	Data
Scientific and technical knowledge consists of a set of interdependent claims	Popper [31]	Networks of knowledge can be readily structured from science and technology databases using techniques such as hierarchical random graphs
Knowledge claims are heterogenous in character	Derrida [32]	Networks built upon science and technology databases are very heterogeneous in character
Technologists have a wealth of tacit knowledge, built upon practice, which they struggle to encode within the network of scientific progress	Polanyi [35]	Changes in technology in this case are manifested in changes in network structure
Knowledge is built upon the configuration of knowledge claims, and is therefore a distributed characteristic of science	Lakatos [34]	Changes in network structure in this case are diffuse, and not associated with single individuals

As seen in Table 3 the model anticipates a three-fold combination of Internet Explorer, rich Internet applications and the World Wide Web consortium (W3C), with a likelihood of 81%. There are also a host of anticipated JavaScript developments involving style sheets, and the use and conversion of XML documents with standard web pages (HTML). These technology linkages are associated with a 70% likelihood.

Some additional research into the first item reveals that there are indeed developments here: the World Wide Web consortium has recently developed standards for accessible rich Internet applications, and these standards are being incorporated into new capabilities for the Internet Explorer browser. Accessible rich Internet applications (or ARIA) are web pages whose contents are much more accessible to people with disabilities. The standard graphical user interface offered by web pages is a challenge to navigate for people who are not sighted, or cannot use a mouse. The proposed W3C provides a means for including additional information on a web page. The metadata will enable disabled users to identify and manipulate these web components through a range of alternative assistive technologies. An example of an assistive technology is screen readers; screen readers convert plain text HTML messages to audible speech or Braille output. This new standard for rich Internet applications was incorporated in a recent beta version of Internet Explorer 8.

5. Policy impacts

These developments in ARIA are less than a year old at the time this paper was written — the W3C posted a working draft dated 4 February 2008 [27]. Wikipedia documentation of these developments is under 3 months old [28]. The current Wikipedia "Accessible Rich Internet Applications" page is only a stub without the detailed hyperlinks typical of a Wikipedia page. The history demonstrates that the first introduction of the page was 23 May 2008. The first release by Microsoft of a beta version of the browser incorporating the new standards was 5 March 2008.

In summary, the hierarchical random graph did seem to anticipate new technological changes in the area of new standards for accessible rich Internet applications. The graph enabled recognition of changes roughly 200 days after W3C posted a new standard, 200 days after Microsoft released a new web browser beta, and 100 days after Wikipedia editors initiated new content. It recognized these new changes without explicit linkages in the knowledge base of technologies. Thus, the hierarchical random graph approach may provide a new forecasting, analysis and design technique for architectural innovation.

As Henderson and Clark suggest [29] there are really two dimensions of radical change here. There is change in core concepts (or as we identify them in this article the "nodes" of the network), and change in the linkages between the concepts. Thus, as these authors argue, there are two different flavors of radical innovation — architectural innovation (focusing on links), and radical innovation (focusing on nodes). Radical innovation combines both aspects of change.

We have argued in this paper that many previous technology forecasting techniques have focused only on incremental and dominant designs. Particularly needed are new techniques for anticipating architectural innovation: an important component of radical innovation. The consequences ignoring of radical innovation are high. Previous researchers have identified a number of consequences of radical innovation for the poorly prepared firm or country: high costs, high uncertainty, technological inexperience, business inexperience, lengthy time to market, and the general destruction of firm competence [17,29,30].

Techniques such as the link prediction algorithm described here might assist in radical innovation processes by providing rapid anticipation of change, through use of a model which anticipates architectural evolution. Furthermore, the structured representation of the data may help identify areas where competences may need to be further strengthened or even completely restored. Common understanding of technological architecture, as provided by machine learning models and delivered by decision support systems, may contribute to an open innovation paradigm where firms work together as part of an extended technological network [11].

A final note from these results might be directed to assisting innovation theorists. This technological network clearly demonstrated technologies “internal” and “external” to the core technology network. The disassortative character of this network means that architectural innovation is much likely to occur from external technologies. Other technological architectures may be very different: these might be assortative networks which favor the use of technologies which are internal, and therefore already present within the system. The author suggests that the original conception of architectural change, as specified by Henderson and Clark [29], is predicated on external sources of innovation. Certainly, this is the class of technological innovations which are most problematic for forecasting, analysis and design. Nonetheless this leaves the form of architectural innovation whereby innovators explore interactions between routine, highly available components, relatively unexplored in theory and practice. The consequences of assortative and disassortative architectural networks may be very different across firms and industries. We suggest that innovation researchers incorporate this new concept into their theories and case studies.

6. Interpretations from the philosophy and sociology of science

The hierarchical random graph is one possible model of science, technology and innovation data. A more fundamental question is whether such a model is consistent with what is postulated about the sociology and epistemology of science. Towards this end, this section examines features of the hierarchical random graph and relates these features to relevant work in the philosophy of science.

Seven features of the random graph model are problematic and therefore worthy of additional explanation. The section to follow examines the claim that knowledge resides on networks, as a series of claims or propositions. The elements of these claims contain many, diverse elements, only some of which relate directly to technical objects. Some knowledge is directly accessible, while other knowledge is tacit: either unexpressed, or resident in a diffuse way across a network of scientific claims. The role of scientists, engineers, and innovators is to enhance the coherence of this network. The progress of science is such that claims which bridge knowledge and increase coherence between related fields may be increasingly more difficult to formulate. Changes to the network can stimulate revolutionary progress.

These claims are given in Table 3. Corresponding evidence from this case are displayed in this table, and further discussed below. The role of the following paragraphs is to unpack the ideas inherent in the use of a random hierarchical graph, and to compare these ideas with prevalent theories in the philosophy and sociology of science. The goal in doing this survey is neither to validate the use of the model, nor to confirm or deny selected ideas in the philosophy of science. Rather, by confronting observations with the context of discussion provided by philosophy, a stronger basis for interpreting the results of the hierarchical random graph may be formulated.

The status of knowledge is a matter of prolonged and fundamental discussion in philosophy of science. The objectivist would argue that knowledge resides in the world at large, and it is the role of the scientist to absorb this knowledge according to his or her capabilities. This position is increasingly seen as untenable. The primary challenge to the position questions the status of empiricism as a certain route to knowledge. Other philosophers, such as Karl Popper, argue that the practice of naive empiricism is impossible [31]. All observations are conditioned on previous expectations and the formulation of prior theories. This perspective then, suggests that knowledge is a network of interlocked claims. Only some of these claims may be anchored in observation, and even these claims are conditional. These claims, and their interconnected character, are modeled through a hierarchical random graph model.

Such claims are likely to be heterogeneous in character. Semioticians such as Derrida [32] argue that a full accounting of science requires a registration of claims about both about the physical as well as social worlds. A subjective account of knowledge suggests two things. First, many claims cannot be directly anchored in empiricism. By definition then, many claims will have a mixed character, relating to previous theories as well as to physical objects. Second, this implies that it is the claim (and not a specific observation) which is subject to discussion or refutation. Derrida's ideas informed Callon and others who developed actor network theory as a vehicle for research in science and technology [33]. Actor network theory is a material, semiotic method, expressly interested in both concepts as well as objects.

Knowledge about science and technology may come in two forms. Explicit knowledge entails knowledge about specific claims. Because explicit knowledge is encoded in claims, the existence of the claim can be verified through recourse to knowledge bases of science and technology. Tacit knowledge entails knowledge about the configuration of claims [34]. Since the subjectivist perspective on knowledge is conditional on entirety networks of knowledge claims, tacit knowledge is required if the truth or falsity of a specific claim is to be evaluated. Further, since tacit knowledge is based on configuration, it cannot be expressly encoded in a knowledge base. A related account of tacit knowledge has tacit knowledge as the knowledge resident in individuals, and not shared across a scientific community. Polanyi's account of science and technology has technologists laboring at the interface of claims, and actual physical artifacts [35]. The tacit knowledge of actual practice is only partially encoded in the network of competing claims. The raw data of a hierarchical random graph, when applied to a network of science and technology information, is likely to be material as well as semiotic in character.

If tacit knowledge has a character based upon the configuration of knowledge claims, then methods (such as the hierarchical graph) which enable exploration of these networks are needed as support tools. The alternative approach would be to expressly encode the configuration within the database of science and technology. This approach may demand an unrealistic level of unanimity about the status and relatedness of specific knowledge claims.

7. Conclusions

The paper concludes that the proposed method, a hierarchical random graph, is a useful way for structuring diffuse knowledge bases of science and technology. In the case discussed, the technique did appear to anticipate significant standards setting activity, as well as presaging a significant reorganization of the science and technology database to better match technological progress. The success of the method is itself dependent upon the collected, distributed activities of innovators, the activities of which may be monitored by means of science and technology databases. This technique (and other techniques like it) may see application in open innovation, or the "mode 2" of future knowledge production.

The hierarchical random graph model is missing a model of the actor. In other words, it attempts no explanation of the capabilities or interests of the innovator, nor does it attempt to explain how these innovators are institutionally organized. The sociology of science literature, and the innovation policy literature, do offer such discussions, even going so far as to prescribe effective means of social organization. This paper does not attempt such a prescription. Armed solely with the evidence of the graph, it is not possible to recommend the appropriate form of social organization to enhance innovation.

Despite the apparent promise of this approach, the method has only been tried on a single case with limited validation. We use these conclusions therefore to reflect on the shortcomings, and possible future research which might be performed to strengthen the method for technology analysis. The hierarchical random graph approach structured evidence of an existing technological network, recognizing the development of new technological linkages shortly after they actually occurred in the market. Even better would have been to anticipate change before it occurs, rather than recognize change shortly after. A full analysis of this kind may require a longitudinal analysis of network development. Hierarchical models might be built before and after critical time periods, and the resultant predictions compared. Alternatively, a dynamic extension to the hierarchical random graph might be envisaged. In this case it might be possible that the structure of the Ajax (Programming) hyperlink network is relatively stable, and therefore the link prediction may have actually presaged technical developments.

Earlier in the paper there was an important conditional: the hierarchical random graph predicted links which were either undiscovered or undocumented. This is an important distinction: undiscovered linkages may be valuable new innovations worthy of research and development. Undocumented linkages may simply reflect out of date source materials. Nonetheless revealing undocumented links still provides a useful stimulus for technology monitoring efforts.

The procedure proposed in this paper provides an objective method of predicting new technological linkages. It remains subjective in two regards: the character of nodes, and the interpretation of new links. For instance, the nodes used in this study ranged from specific technologies, to people (Jesse James Garrett), to institutions (W3C). While a mix of node types might be desirable (for instance technology as well as process), it may be difficult to establish a uniform definition of the technological components of the network. The case study recognized impending change in nodes related to W3C standard setting, rich Internet technologies, and Internet Explorer. Finding supporting evidence for change was ease. Nonetheless, interpreting the meaning of these changes introduced a component of subjectivity. Such subjectivity may be hard to remove given the epistemic character of uncertainty in new technology.

It is important also to acknowledge that this is only a first demonstration of concept on a relatively limited sample. Repeated trials and a more complete analysis based on an evolving network are desirable items for future research. Use of alternative databases, such as patenting data, would be an interesting item for extending the method.

Acknowledgements

The author gratefully acknowledges the use of C++ code and Matlab scripts as provided by Aaron Clauset on his webpage (Clauset 2008). The author appreciates helpful discussion from Jan Kwakkel on the epistemology of knowledge networks.

References

[1] M.C. Roco, Key note address: governance of converging new technologies integrated from the nanoscale, Paper Presented at the Portland International Conference on Management of Engineering and Technology, Portland, 2007.
[2] A. De Haan, K. Mulder, Sustainable air transport: identifying possibilities for technological regime shifts in aircraft construction, Int. J. Innov. Manag. 6 (3) (2002) 301–318.
[3] P. Anderson, M. Tushman, Technological discontinuities and dominant designs: a cyclical model of technological change, Adm. Sci. Q. 35 (1990) 6045–6633.
[4] A.L. Porter, A.T. Roper, T.W. Mason, F.A. Rossini, J. Banks, Forecasting and Management of Technology, John Wiley & Sons, New York, 1991.
[5] R.J. Lempert, S. Popper, S.W. Bankes, Shaping the next one hundred years: new methods for quantitative long-term policy analysis, Santa Monica: The RAND Pardee Center, 2003.
[6] L. Fleming, Recombinant uncertainty in technological search, Manage. Sci. 47 (1) (2001) 117–132.
[7] W.E. Walker, P. Harremoes, J. Rotmans, J.P. Van Der Sluis, M.B.A. Van Asselt, P. Janssen, M.P. Krayer Von Kraus, Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support, Integrated Assessment 4 (1) (2003) 5–17.
[8] G.S. Altshuller, Creativity as an Exact Science, Gordon & Breach, New York, 1988.
[9] J. Terninko, A. Zusman, B. Zlotin, Systematic innovation: an introduction to TRIZ, Florida: CRC Press, Boca Raton, 1998.

[10] D.R. Swanson, N.R. Smalheiser, An interactive system for finding complementary literatures: a stimulus to scientific discovery, Artif. Intell. 91 (2) (1997) 183–203.
[11] H. Chesbrough, W. Vanhaverbeke, J. West, Open Innovation: Researching a New Paradigm Oxford, Oxford University Press, 2006.
[12] M. Gibbons, C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies, SAGE, London, 1994.
[13] B. Harrison, Industrial districts: old wine in new bottles, Reg. Stud. 26 (5) (1992) 469–483.
[14] N.P. Suh, Axiomatic design of mechanical systems, J. Mech. Des. 117 (Special B) (1995) 2–10.
[15] C. Marchetti, N. Nakicenovic, The Dynamics of Energy Systems and the Logistic Substitution Model, International Institute for Applied Systems Analysis, 1979.
[16] K.M. Clark, The interaction of design hierarchies and market concepts in technological evolution, Res. Policy 14 (1985) 235–251.
[17] J.P. Dismukes, Technologies of thinking' seen key to accelerated radical innovation, Res. Technol. Manag. 48 (4) (2005) 2–8.
[18] A.L. Porter, Tech mining to accelerate radical innovation, PICMET 2007 Proceedings, 2007.
[19] Y. Yasunaga, M. Watanabe, M. Korenaga, Outline of the strategic technology roadmap of METI (Ministry of Trade and Industry of Japan) and trial approaches for technology convergence with the methodology of technology roadmapping, PICMET 2007 Proceedings, Portland, Oregon: PICMET, 2007.
[20] S.P. Hubbell, The Unified Natural Theory of Biodiversity and Biogeography, Princeton University Press, Princeton, 2001.
[21] A. Clauset, Hierarchical Random Graph, 2008, Accessed 25 August 2009. http://www.santafe.edu/~aaronc/randomgraphs/.
[22] A. Clauset, C. Moore, M.E.J. Newman, Hierarchical structure and the prediction of missing links in networks, Nature 453 (2008) 98–101.
[23] M. Sales-Pardo, R. Guimera, A.A. Moreira, L.A.N. Amaral, Extracting the hierarchical organization of complex systems, Proc. Nat. Acad.f Sci. U. S. A. 104 (39) (2007) 15224–15229.
[24] A. Vazquez, R. Pastor-Satorras, A. Vespignani, Large-scale topological and dynamical properties of the Internet, Phys. Rev. E 65 (6) (2002) Article Number 066130.
[25] J. Gill, Bayesian Methods: A Social and Behavioral Sciences Approach, Chapman & Hall, 2002.
[26] J.J. Garrett, AJAX: A New Approach to Web Applications, 2005, Retrieved 15 May 2007, from http://www.adaptivepath.com/ideas/essays/archives/000385.php.
[27] W3C (World Wide Web Consortium), Roadmap for accessible rich internet applications, W3C Working Draft, 4 February 2008, 2008, Accessed 28 August 2008. http://www.w3.org/TR/wai-aria-roadmap/.
[28] Wikipedia, WAI-ARIA, 2008, Accessed 28 August 2008. http://en.wikipedia.org/wiki/Accessible_Rich_Internet_Applications.
[29] R. Henderson, K.M. Clark, Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms, Adm. Sci. Q. 35 (1990) 9–30.
[30] S.G. Green, M.B. Gavin, L. Aiman-Smith, Assessing a multidimensional measure of radical technological innovation, IEEE Trans. Eng. Manage. 42 (3) (1995) 203–214.
[31] K. Popper, Objective Knowledge: An Evolutionary Approach, Oxford University Press, Oxford, 1972.
[32] J. Derrida, Of Grammatology, John Hopkins University Press, Balitmore, MA, 1974.
[33] M. Callon, J.P. Courtial, W.A. Turner, S. Bauin, From translations to problematic networks: an introduction to co-word analysis, Soc. Sci. Inf. 22 (2) (1983) 191–235.
[34] Lakatos, The Methodology of Scientific Research Programmes: Philosophical Papers, Two Volumes, Cambridge University Press, Cambridge, 1978.
[35] M. Polanyi, Personal Knowledge: Towards a Post-Critical Philosophy, University of Chicago Press, Chicago, 1958.

Scott Cunningham received a Ph.D. in Science, Technology and Innovation Policy from the Science Policy Research Unit. He has worked as a data miner for large database companies, developed patents in the fields of pricing and promotion algorithms, been a research fellow at the Technology Policy Assessment Center of Georgia Tech, and helped advise the British and Malaysian governments on their national profiles of scientific activity. He currently works for the Faculty of Technology, Policy and Management, of the Delft University of Technology, as an assistant professor. His book on Tech Mining, co-authored with Alan Porter, was published in 2001 with Wiley.