AAAI Technical Report FS-12-08
Social Networks and Social Contagion

Applied Actant-Network Theory: Toward the Automated Detection of Technoscientific Emergence from Full-Text Publications and Patents

David C. Brock, Olga Babko-Malaya*, James Pustejovsky***, Patrick Thomas****, Sean Stromsten*, and Fotis Barlos*.**

*BAE Systems Advanced Information Technologies, **David C. Brock Consulting, ***Brandeis University, ****1790 Analytics LLC

Abstract

There is growing interest in automating the detection of interesting new developments in science and technology. BAE Systems is pursuing ARBITER (Abductive Reasoning Based on Indicators and Topics of EmeRgence), a multidisciplinary study and development effort to analyze full-text and metadata for indicators of emergent technologies and scientific fields. To define these indicators, our team has applied the primary insights of actant network theory developed within the disciplines of Science and Technology Studies and the history of technology and science to create a pragmatic theory of technoscientific emergence. Specifically, this practical theory articulates emergence in terms of the robustness of actant networks. This applied actant-network theory currently guides our definition of indicators and indicator patterns for the ARBITER system, and represents a novel contribution to the discussion of emergent technologies and fields. Several elements of our theory were validated with 15 case studies and 25 example technologies.

Introduction

Actant network theory provides a vision of science and technology as constituted by networks of heterogeneous elements, interconnected by disparate relationships. Often, this theory is known as "actor network theory," but the alternate label "actant network theory" stresses the heterogeneity of the networks elements, both human and nonhuman. The networks that comprise domains of science and technology contain individuals, institutions, instruments, practices, terminology, materials, funders, meetings, government organizations, laws, journals, patents, publications, and so on. The membership of elements within such a network, and the nature and extent of the relationships between these elements, is dynamic and constantly changing. (Latour, 2005; Law, 1987; Contractor et al., 2011)

New domains in science and technology are formed, and changes in existing domains are produced, by human-directed, open-ended extensions and transformations of these networks. Individuals and groups draw on a variety of conceptual, material, and sociocultural resources in efforts to transform existing domains within science and/or technology against a backdrop of a structured field of intentions. This intention-driven human activity causes networks to form or change. (Pickering, 1993)

Every extension or change of an actant network is a path-dependent novelty, that is, an open-ended evolution of an existing network or networks. Hence, every novelty in the development or creation of the actant networks of science and technology has antecedents within some network or networks. Many contemporary history scholars, Science and Technology Studies practitioners, technology forecasters, machine intelligence researchers, and market-research firms can be seen, in this perspective, as attempting to identify and characterize particular novelties that are relevant for his or her specific interest.

The BAE Systems team is developing ARBITER, an automated system whose purpose is to identify and characterize emerging technologies and emerging fields in science. ARBITER will provide its users with candidate emerging fields and technologies, and characterizations thereof. It will do so by processing very large collections of technoscientific publications and patents more rapidly than with current methods. The collections contain tens of millions of documents from many sources in multiple languages, licensed and sourced from the major commercial and public providers of full-text publications and patents. In particular, we are processing data from Elsevier, Thomson Reuters' Web-Of-Science®, Pub-Med Central as well as both granted and filed patents from multiple national patent offices.

In the idiom of actant network theory, ARBITER's task is to identify, characterize, and evaluate over time certain actant networks that comprise emerging technologies and emerging fields. More specifically, this task is to use indicators from the metadata and full-text of publications and patents in multiple languages in order to identify, characterize, and evaluate over time the actant networks of science and technology.

"Emerging technologies" and "emerging fields" are not precise concepts. The phrases are commonly used to denote new domains within technology and science that are attracting increased interest -- gaining momentum -- and which may eventually have significant economic and sociocultural impact (either through their direct application or via their impact on other domains in science and technology). Often, emerging technologies and emerging fields are taken to have the potential to open up large new areas of technology and science: biotechnology and nanotechnology are paradigmatic of this stance. That is, these emerging technologies and emerging fields are held to be highly generative. (Cozzens et al., 2010)

A major challenge in developing an automated identification system has been this lack of precision in the meaning of emergence, and even greater ambiguity about how it occurs (Goldstein, 1999). The situation is further complicated by the possibility that real-world users might change their understanding of or interests in emergence over time in connection with use of the system.

To address these definitional challenges, the ARBITER team is developing tools that will not only identify potential emerging fields and technologies, but will also allow users to identify and examine developments using characteristics of actant networks that they can specify and customize. More specifically, the ARBITER system will provide users with candidate emerging fields and technologies by assessing the related document groups (RDGs) associated with actant networks within a very large corpus of publications and patents, in multiple languages, from a range of years. The customizable RDG nomination by ARBITER is carried out at the following four levels:

1. General emergence level: overall depiction of emergence (or non-emergence) associated with a given technology or scientific field;
2. Pattern level: general characteristics and dynamics of technologies and fields; for example, robust, mature, growing, attracting attention, scientific vs. technological, and so forth;
3. Indicator level: characteristics of networks, such as growth in publications and patents, consolidation of organizations, and number of authors;
4. Feature level: allows users to generate their own indicators, based on a list of available features and properties that the system is able to extract from the full text and metadata.

ARBITER is designed for flexible use. It will allow users to present the system with an exemplary emerging field or technology of interest. The system will identify characteristics of the exemplar and then nominate a set of RDGs sharing these characteristics. Alternatively, a user could begin with a particular characteristic of interest, and the system will then nominate RDGs that exhibit this specific characteristic. For example, at the indicator-level, a user could find networks that show diversification in geographic location, or that attract many new researchers and organizations. A user may also wish to experiment with a new indicator that can be derived by selecting a feature (an element of an actant network) and some general characteristic, such as growth, persistence, or consolidation.

ARBITER Definition of Emergence

Guided by actant network theory, we have created a pragmatic definition of emergence: *growth in the robustness of an actant network*. The robustness of an actant network involves its resiliency -- the ability of the network to maintain its connectivity and continue to function, despite the failure or removal of crucial actants (e.g. star researchers or major funders). More robust actant networks are more likely to persist and grow in the future and, thus, to have significant effects on science, technology, and society. A field or technology that is emerging is one for which the constitutive actant network is increasing in robustness.

To measure robustness, we define a function of the number and diversity of actants and the relationships between these actants -- the traffic between the network elements. The diversity of actants is their heterogeneity and can be measured directly -- i.e. the number of different types of actants in the relevant network. For example, a network with several different types of funders – civilian government agencies, military agencies, large corporations, venture capitalists, and private foundations – is regarded as more robust than a network with a single funder type. Carbon nanotubes is a good example of a technology with such a robust funding network, with resources from many different sources being committed to this high profile, emerging technology. (Schultz and Joutz, 2010)

Our definition of network traffic integrates the rate of information flow (e.g. the spread of terminology) with other factors that characterize the dynamic behavior of actants in the networks. These include star scientists moving from one organization to another, researchers meeting at newly organized workshops or participating in new conferences, and young researchers entering the field.

It is important to note that robust actant networks may exist not only in emerging domains, but also in mature technologies and scientific fields. What differentiates emerging from mature domains is that the robustness of the network associated with an emerging development will be increasing, while the network associated with a mature domain will be relatively stable. Hence, it is not just robustness, but rather *increasing* robustness in the actant network that is associated with emergence. This reflects the difference between the dynamic nature of emerging developments and the relative stability of mature fields. For example, there are robust networks in both lead-acid and lithium-ion batteries, but only the latter network has increased significantly in robustness over the past two decades, as discussed below.

Our measure of robustness is therefore constructed so that it largely plateaus for mature fields. Increasing robustness identifies scientific and technological domains during a particular phase of their development – their initial emergent phase. This phase is differentiated from other phases such as initial discovery (something which would be very interesting to detect, but is notoriously difficult to identify), and maturity (which is relatively easy to identify, but is generally of less interest).

Our robustness measure is also constructed so that it is relatively unresponsive to changes in small actant networks. For example, if just one paper is published in a field at time t, followed by two papers at time $t+1$, this is 100% growth, but still not evidence of a burgeoning field. Given the paucity of the network, this growth should not be taken as a doubling in robustness.

We define robustness as a non-decreasing, bounded, non-linear function on measures of extent, diversity, and traffic, subject to two constraints: (a) is insensitive to changes near zero, which might be random, and are certainly not strong indications of robustness; and (b) *saturates* at high values of these inputs, so that (for instance) an increase in the number of researchers pursuing work in an already established field will *not* yield a large change in robustness. The first constraint captures the fact that extremely small fields are intrinsically non-robust, because they rely crucially on very few actants, while the second constraint allows us to focus on fields that have not already matured.

Emergence Definition and Speciation

Our articulation of technological and scientific emergence therefore becomes:

Is the actant network of <concept> increasing in robustness during <time period>?

Finding that a field or technology is emerging at time t is, in part, a judgment about the resilience of the network *after* t. That is, a determination that a field is emerging at t relies on observing the changing states of robustness of the actant network *before* t, leading to a certain level of confidence in the resilience of the network *after* t.

Changes in the robustness of a network do not necessarily follow a set progression from emergence to maturity, and then decline. Some networks emerge and then mature, and do not decline. Some emerge, decline, reemerge, and disappear. Others emerge, decline, reemerge and then mature.

Given that the actant networks that comprise emerging technologies and emerging fields always have antecedents in existing actant networks and given that the development of the actant network is always the human-driven, open-ended extension or transformation of the network, we expect the appearance and development of emerging technologies and fields will closely resemble the biological and evolutionary process of *speciation*. In its earliest stages, an emerging technology or field may appear very much like a subspecies or variation within some existing technological or scientific species. Later as the emerging technology or field increases in robustness, it becomes identifiable as a new species in its own right. If the emerging technology or field continues to increase in robustness, it will likely develop subspecies, adapted to particular local aims and conditions. Emerging technologies and fields that achieve a considerable robustness may even have one or several of its subspecies transform into distinct species – new emerging technologies or fields – in their own right.

Centrally important to actant networks and our definition of emergence is the development and use of field-specific terminologies. Major outputs of science and technology are terminological: publications, patents, and, more generally, the generation and communication of information. Within actant networks perhaps the primary type of traffic - the most common form of relationship - is the production and exchange of terminological information. Indeed, for human actors within such a network, their understanding of the technology or field is generated by their consumption of this linguistic traffic. (Latour, 2005)

A primary axiom of our development of the ARBITER system is that for any particular technology or scientific field (including sub-varieties thereof) there exists a *characteristic terminology*. There also exist discoverable patterns of terms that specifically characterize – that is, identify and bound – the field, technology, or sub-variant thereof. Following discovery of a characteristic terminology, additional measures and characterization of it are possible: novelty, growth, persistence, diversification, consolidation, and connectivity. These features of the characteristic terminology constitute measures of the existence and robustness vector of the actant network constitutive of the particular technology or field.

Consider, for example, DNA microarray technology. In the early stages of this technology, there did not exist a common terminology, and "DNA microarray" was not an accepted technological term. Early patents and papers instead referred to the constituent elements of microarray technology: peptide libraries, substrate attachment, and light-directed synthesis. Later, as the actant network surrounding the technology became more robust, and the elements of a DNA microarray became established and standardized, the term "DNA microarray" gained widespread acceptance, and could be used as shorthand for this particular piece of diagnostic equipment.

Litspace Indicators of Emergence

In order to identify, characterize, and evaluate over time the actant networks that comprise emerging technologies and fields, the ARBITER system employs a number of indicators derived from the metadata and full text of publications and patents: what we and other researchers in this area have termed *litspace*. Litspace is partially constitutive of the actant networks that comprise technologies and fields; i.e. litspace is a portion of the actant network. Litspace indicators are thus direct indicators of the actant networks, and will provide analysts with information about network elements, relationships between them, and the temporal dynamics and trends of the networks. In particular, several litspace indicators and their combinations characterize the robustness of actant networks.

To discover indicators, we define a set of entities (litspace actant-network elements), entity types, and relations between entities, which can be extracted from full text and metadata features:

Entities: ORG, PERSON, PUBLICATION, PATENT, PATENT_APPLICATION, FUNDER, PUBLICATION, WORKSHOP, COUNTRY, SENTIMENT, DOCUMENT SECTION, TERM, TOPIC, EVENT…

Entity Types: ORG_TYPE, PERSON_TYPE, PUBLICATION_GENRE, …

Relations: CONTRAST(PUBLICATION, PUBLICATION), CONTRIBUTOR(PERSON, TECHNOLOGY), RELATED_WORK(PUBLICATION, PUBLICATION), FUNDING(FUNDER, ORGANIZATION), …

Unlike previous approaches to detecting emergence, which are based on the citation analysis of papers and patents (e.g. Bettencourt et al 2009, Thomas and Breitzman 2006), we are extracting information from the text of publications and patents, identifying authors, their affiliations, addresses, as well as classifying types of organizations and publications. Moreover, we apply natural language processing technologies to extract scientific terminology from the full text of the documents, to identify different types of relationships between citations, authors, terms, and organizations, including sentiment, contrast, and related work, and to classify documents into genres.

Our indicators measure different properties of these extracted entities, including relationships between them and their dynamics. These properties include:

EXTENT: Count of an entity or a relation between entities.
GROWTH: Change in the EXTENT of an entity, positive or negative.
DIVERSITY: Entropy of the distribution of types of an entity. Count of the types of an entity.
DIVERSIFICATION: Change in the DIVERSITY of an entity, positive or negative.
CONSOLIDATION: Negative diversification.
NOVELTY: The first occurrence of an entity, or combinations of entities.
PERSISTENCE: The continued occurrence of a unique entity over time, with or without gaps.
CONNECTIVITY: Change in the association of particular entities with one another.

We distinguish two types of indicators: *basic* and *complex*. Basic indicators are created by measuring the properties of a specific entity or entity type, such as extent of researchers, extent of patents, diversity of funder types, growth of organizations, connectivity of researchers, consolidation of terminology, etc. Complex indicators are created by defining constraints on combinations of basic indicators. For example, impact indicators identify crucial or important members of network, such as high profile researchers, high impact patents, and highly cited publications.

Characteristics of Actant Networks and Indicator Patterns

Having defined basic and complex indicators of actant networks, we identify particular indicator patterns. These patterns are related to network properties such as resilience, maturity, growth, scientific vs. technological, etc. Examples of these patterns, and the indicators used in these patterns, are outlined below.

Pattern 1.1. Extent of an Actant Network

This pattern measures the size of the actant network. The indicators assembled in this pattern are all counts of actant network elements including researchers, organizations, patents, and characteristic terminology.

Basic Indicators: Extent of Actants

1.1.1. Extent: unique researchers
1.1.2. Extent: unique organizations
1.1.3. Extent: unique conferences
1.1.4. Extent: unique workshops
1.1.5. Extent: unique patents

1.1.6. Extent: funders
1.1.7. Extent: researchers who publish on the same fine-grained topic
1.1.8. Extent of characteristic terminology
1.1.9. Extent: unique countries

Complex Indicators: Extent of High-Impact Actants

1.1.2.1 Extent: star researchers (i.e. researchers with high impact)
1.1.2.2 Extent: high impact patents
1.1.2.3 Extent of a set of publications that all cite one document in common (measures impact of a seminal paper)
1.1.2.4 Extent: high impact organizations
1.1.2.5 Extent: significant/established publications

Pattern 1.2. Resilience of Actant Network

The robustness of an actant network lies in its resilience – its ability to maintain its connectivity and continue to function despite the failure or removal of crucial actants. To measure resilience, we evaluate the number and diversity of central actants and also the number and diversity of the relationships between these actants. The diversity of actants can be measured by heterogeneity - i.e. the number of different types of actants in the relevant network. For example, as noted earlier, a network with several different types of funders – civilian government agencies, military agencies, large corporations, venture capitalists, and private foundations – is regarded as more robust than a network with a single funder type.

Basic Indicators:

Diversity (Extent of Entity Types)

1.2.1 Extent: funder types
1.2.2 Extent: technologies referenced in patents
1.2.3 Extent: types of patent assignees
1.2.4 Extent: scientific document genres
1.2.5 Extent: organization types
1.2.6 Extent: patent genres (stages of development)
1.2.7 Extent: event types (workshops, conferences, etc.)
1.2.8 Extent: patent disciplines
1.2.9 Extent: review-type articles

Pattern 1.3. Connectivity of Actant Networks

This pattern measures the number of relationships and the density of interconnections in a network. To analyze connectivity, we evaluate traffic between actants within the network. Our definition of traffic integrates the rate of information flow (e.g., the spread of terminology) with other factors that characterize the dynamic behavior of actants in the networks.

Basic Indicators:

Connectivity

1.3.1 Extent: publications with authors from different organizations (measures collaboration)
1.3.2 Extent: RELATED WORK relations
1.3.3 Extent: collaborator relations
1.3.4 Extent: contrast relations (such as sentiment)
1.3.5 Extent: author affiliation changes
1.3.6 Extent: author topic changes
1.3.7 Extent: authors entering domain from another
1.3.8 Spread of characteristic terminology

Diversification/Consolidation

1.3.9 Characteristic terminology spreads into different domains
1.3.10 Existence of a set of terms used by everyone in the network

Novelty

1.3.12 Extent: brand new researchers

Pattern 1.4. Novelty of Actant Networks

This pattern is indicative of new domains within technology and science that are attracting increased interest and activity, including newly forming and changing technologies. It evaluates the extent to which people and resources are rushing into a technology or field, in order to distinguish emerging technologies/fields from others. This pattern also evaluates domains where many people are importing terminology and resources from the emerging area into existing areas.

Basic Indicators:

Novelty

1.4.1 Extent: brand new researchers
1.4.2 New journal
1.4.3 New organization working on the topic
1.4.4 New conference
1.4.5 New workshop
1.4.6 New terminology

Growth

1.4.7 Rapid growth of researchers
1.4.8 Rapid growth of inventors
1.4.9 Rapid growth of funding
1.4.10 Rapid growth of funder types
1.4.11 Rapid growth of terminology use
1.4.12 Rapid growth of patents

Pattern 1.5. Growth of Actant Networks

In order to identify growing technologies and fields, we evaluate growth of the number and diversity of actants, as well as growth of traffic between these actants.

Basic Indicators:

Growth

1.5.1 Growth of researchers
1.5.2 Growth of inventors
1.5.3 Growth of researchers from industry
1.5.4 Growth of funding
1.5.5 Growth of funder types
1.5.6 Growth of genres
1.5.7 Growth of publications per genre

1.5.8 Opposite trends in number of patents (and citations to patents) in established and emerging technologies
1.5.9 Growth of terminology use
1.5.10 Growth of patents
1.5.11 Growth of star researchers
1.5.12 Growth of high impact patents
1.5.13 Growth of funder types
1.5.14 Growth of types of patent assignees
1.5.15 Growth of genres
1.5.16 Growth of organization types
1.5.17 Growth of patent genres
1.5.18 Growth of event types
1.5.19 Growth of patent disciplines
1.5.20 Growth of people changing affiliations
1.5.21 Growth of researchers changing topic of interest
1.5.22 Growth of characteristic terminology

Beyond patterns directed to different characteristics of an actant network in terms of its size, growth and connectivity, there are also patterns directed to specific aspects of actant networks. Examples are provided below.

Pattern 2.1. Practical Application of Concept

This pattern is indicative of a technological field, as opposed to scientific fields. The differences between science and technology are reflected in differences between the actant networks that comprise them. Even though the fundamental process of emergence is the same in science and technology, there are differences in the character and evolution of the fields' actant networks. There are also differences in the prevalence of various kinds of actants in the technological network and the scientific network. For instance, the funders will tend to be different; commercial enterprises (particularly small companies without access to extensive funding) are typically more interested in technology rather than science. At the extreme, investors (e.g., venture capital, private equity) do not typically fund science, but rather fund technology.

The consumers of technology are also typically different from those of science. Technologies developed by commercial enterprises are more likely to face an additional actant - i.e., the commercial marketplace. The adoption of a technology may be subject to many external features beyond the quality of the technology itself - cost, packaging, marketing, etc. From the point of view of the investors, the market is often the ultimate arbiter of the success of the technology.

The direct role of the market may be less prevalent in scientific networks. One effect of this is that debates may be more short-lived (and possibly less important) in technology, compared to science. For example, a scientific debate may last decades with no resolution, whereas once the market chooses a "winning" technology, the debate typically ceases. Crudely, one does not have to win a technological debate; one simply has to sell more products (cf. VHS vs. Betamax).

The presence of the marketplace in the technological network affects the types of outputs required of the human actors - i.e., there is a greater interest in patents (which offer the prospect of a monopoly over a given technology) rather than papers (which offer no such monopoly). Again, at the extreme, venture capital and private equity companies are often very interested in companies that have protected their technology via patents, or have concrete plans to do so.

Basic Indicators:
Extent
2.1.1 Extent: patents
2.1.2 Extent: funders
2.1.3 Extent: industrial researchers
2.1.4 Extent: funding from non-government sources
2.1.5 Extent: funding from agencies that fund applied research
2.1.6 Extent: researchers from small companies
2.1.7 Extent: linguistic cues for practical application
2.1.8 Extent: company names in the text of RDGs
2.1.9 Extent: sections or document that discuss practical applications
2.1.10 Extent: trademarks
2.1.11 Extent: patent assignees that are not universities (and non-government)
Novelty
2.1.13 Extent: high impact patents
Connectivity
2.1.14 Extent: patents changing ownership
2.1.15 Extent: small companies bought by large companies
2.1.16 Extent: researchers moving from universities to industry
2.1.17 topics/terminology change from describing experiments to more applied
2.1.18 Terminology appears in claims section of patents (a sign that the terminology is legally recognized)
2.1.19 Terminology spread into different domains
Growth
2.1.20 growth of patents
2.1.21 growing number of researchers from industry
2.1.22 trends in growth of patent assignees
Complex Indicators:
2.1.1.1 Patents that changed ownership include high impact patents
2.1.1.2 Funders fund industrial researchers
2.1.1.3 Companies have high-impact patents
2.1.1.4 Industrial researchers have high-impact patents
2.1.1.5 Researchers moving from universities have patents

Pattern 2.3. Characteristic Terminology

As noted earlier, the production, stabilization, and circulation of *common terminologies* are distinctive activities for an emerging field, giving cohesion within and boundaries about the actant network. As this network increases in robustness, so the usage of field-specific, shared terminologies becomes more robust, i.e. resilient to the failure of crucial actants. What this means is that these terminologies are used more frequently by multiple human and organizational actants, in more forums (e.g. in different document types), and in more contexts.

In our definition of emergence, we referred to an actant network that is growing in robustness. The same requirement applies to our analysis of terminologies – i.e. the number and diversity of connections of emerging terminology to the other actants in the network must be increasing. This includes creation of new terminology as well as the degree of adoption of a terminology (sharing common language), captured as a distributed property of the network.

Basic Indicators:

Novelty

2.3.1 New terminology

2.3.2 Novel patterns of use for existing terms

Growth

2.3.3 Rapid growth of terminology use

Persistence

2.3.4 Terminology use does not change over time

2.3.5 No clustering of term use in the community

Diversification/Consolidation

2.3.6 Terminology spread into different domains

2.3.7 Existence of a set of terms used by everyone in the community

Mobility

2.3.8 Terminology appears in claims section of patents (a sign that the terminology is legally recognized)

Preliminary Findings

The ARBITER system is in an early development stage, and is evolving. Nevertheless, we have observed some interesting preliminary results, particularly related to our definition of emergence in terms of the growing robustness of an actant network. These preliminary results are based on an initial set of 25 technologies, 15 of which we take to be examples of emerging technologies, and 10 of which we consider as non-emerging examples. Examples of the former include carbon nanotubes, ultracapacitors, speech recognition, and photovoltaics; examples of the latter include beds and cribs, asphalt, papermaking, and toothpaste. Our selection of these 25 example technologies is, in the end, subjective. They are not, however, arbitrary and do represent the most typical uses of the phrase "emerging technology."

We paired several of our example technologies, one emerging and one non-emerging, as in the representative cases below. The first case relates to diagnostic testing (where DNA microarrays are regarded as emerging, and sphygmomanometers are regarded as mature and non-emerging), while the second case concerns electrochemical battery technology (with Lithium-ion batteries as emerging, and lead-acid batteries as non-emerging).

Figure 1a shows the number of U.S. patents describing DNA microarray and precursor technologies issued in each five-year period between 1981 and 2010. These patents are broken down according to type of assignee (i.e. owner), between companies, academic/government/non-profit, and individuals. There are a number of interesting features of Figure 1a, for our theory that increasing robustness of actant networks is associated with emerging technologies. First, the actant network associated with DNA Microarrays appears to be growing in size, as reflected in the increasing number of patents issued in each five-year period. Also, the network appears to be relatively robust, with patents filed by multiple types of organizations.

Figure 1a – Number of DNA Microarray US Patents by Assignee Type and Time Period

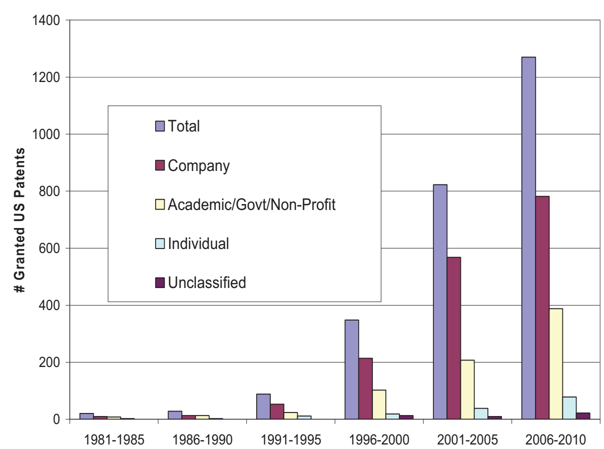
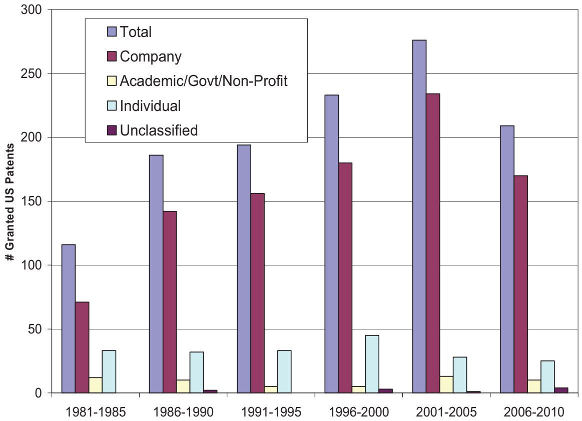


Figure 1b – Number of Sphygmomanometer US Patents by Assignee Type and Time Period



In contrast, Figure 1b shows a much steadier, less rapid, increase in growth in the actant network for Sphygmomanometers. In addition, the network appears to be much more homogenous, with a predominance of commercial -- and a relative dearth of academic, government and non-profit -- patent assignees. Both growth and diversity in this network are low, which leads to low robustness. This is suggestive of a relatively mature technology, in which companies are making incremental improvements on existing technologies, but there is little research interest from either government agencies or academic institutions.

A similar pattern can be identified in the case of battery technology, comparing lithium-ion and lead-acid batteries. While the actant network in lithium-ion batteries is growing rapidly (see Figure 2a), there is no such growth in the lead-acid network (see Figure 2b).

Figure 2a - Number of Lithium-Ion Battery US Patents by Assignee Type and Time Period

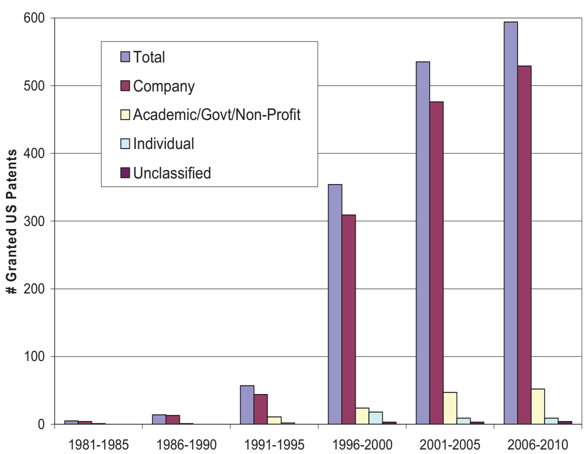
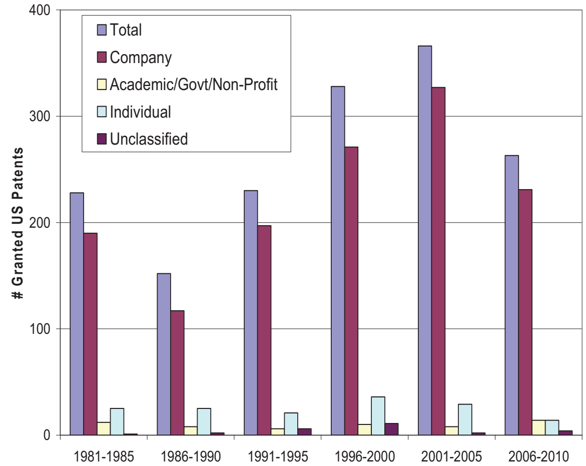


Figure 2b - Number of Lead-Acid Battery US Patents by Assignee Type and Time Period



Comparing Figures 2a and 2b against the earlier Figures 1a and 1b is also instructive, in that it highlights the importance of allowing for differences in networks across different types of technology. Specifically, academic institutions may form a much greater part of actant networks in life sciences than in mechanical technologies. Hence, it may appear at first glance that the actant network in lithium-ion batteries is not particularly robust, since it is dominated by companies, with academic/government/non-profit patents being only 9% of the total. However, this is over twice as high as the percentage in lead-acid batteries (4%), suggesting that, for a battery technology, lithium-ion has developed a robust and growing actant network over the past two decades.

Similar results were also found for the other emerging and non-emerging technologies in our initial sample. The actant networks associated with technologies in the emerging list consistently exhibited growing robustness and diversity, to a greater degree than the technologies in the non-emerging list.

These preliminary results, while promising, are based only on a small sample of 25 technologies, selected manually and categorized subjectively as emerging and non-emerging. This presents a significant risk of sampling bias in terms of the selection of the technologies. Sampling is less of an issue once the technologies are selected, since the RDGs are designed to contain the population of all patents and papers relevant to a given technology, rather than a sample. To address this potential source of bias, we plan to analyze large numbers of technologies (well beyond the 25 analyzed thus far) with input from subject matter experts in the technology selection and categorization process.

Another issue with the validation of ARBITER and our theory is the general ambiguity about the meaning of emergence. To address this issue, we also validate our indicators against more specific test questions that (1) are relevant to emergence as defined by Goldstein 1999 and

(2) that can be answered consistently by subject matter experts. Our test question validation data currently includes 15 case study technologies that are categorized by whether or not there exists a practical application of the technology and whether or not there exists a community of practice for the technology within a certain time period. All of these case studies were judged and categorized by subject matter experts. For each of the technologies and a certain time period, one or more human experts provided a positive or a negative answer to the practical application and community of practice questions. We currently have validated 29 indicators against these two test questions. In each case, we found at least half a dozen indicators that appear to be to be highly informative about the answers to the test questions. These promising indicators include: the number, rate of growth, and diversity of the organizations with which authors are affiliated; the number of countries represented; rate of growth in the number of industrial researchers; and the number and rate of growth of researchers and inventors.

These preliminary validation efforts follow standard quantitative metrics developed for the social sciences (Cook and Campbell, 1979, Trochim and Donnelly, 2007) in that we apply several modes of evaluation to our system. Particularly, the preliminary results speak to both content and to face validity. The former concept measures how closely the indicators derived from our pragmatic theory of emergence fit unseen data presented to our system. Concerning face validity, our system attempts to answer test questions designed by human subject matter experts, and judged thereby, related to whether or not a field or technology is emergent in a given period. The system addresses these test questions by the generation and analysis of these same indicators.

Concluding Remarks

We have developed a practical theory of scientific and technology emergence based on actant network theory from Science and Technology Studies and the history of science and technology. Individuals and groups draw on a variety of conceptual, material, and sociocultural resources in efforts to transform existing domains within science and/or technology against a backdrop of a structured field of intentions. In such a dynamic network, we define emergence as the increase of robustness of a network, which can be measured by extent, diversity, and traffic in the network. We validated several elements of our theory with 15 case studies and 25 example technologies.

Acknowledgement

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20154. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

References

Bettencourt, L. , Kaiser, D., Kaur, J., Castillo-Chávez, C., and Wojick, D. 2008. Population modeling of the emergence and development of scientific fields. *Scientometrics*, 75(3):495–518.

Cook, T. D. and Campbell, D. T. 1979 *Quasi-experimentation: design and analysis issues for field settings*. New York: Rand Mc Nally.

Cozzens, S. et al. 2010. Emerging Technologies: Quantitative Identification and Measurement. *Technology Analysis and Strategic Management* 22: 361-376.

Goldstein, J. 1999. Emergence as a Construct: History and Issues. *Emergence: Complexity and Organization*. 1: 49-72.

Latour, B. 2005. *Reassembling the social: An introduction to actor-network theory*. Oxford, UK: Oxford University Press.

Law, J. 1987. Technology and heterogeneous engineering: The case of the Portuguese expansion. In W.

E. Bijker, T. P. Hughes, & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology.* Cambridge, MA: MIT Press. 111–134.

Contractor, N. 2011. Multidimensional Networks and the Dynamics of Sociomateriality: Bringing Technology Inside the Network. *International Journal of Communication.* 5: 682-720.

Pickering, A. 1993. The Mangle of Practice: Agency and Emergence in the Sociology of Science. *American Journal of Sociology.* 99: 559–89.

Schultz, L. and Joutz, F. 2010. Methods for identifying emerging General Purpose Technologies: a case study of nanotechnology. *Scientometrics*. 85: 155-170.

Thomas P. and Breitzman A., 2006 "A Method for Identifying Hot Patents and Linking them to Government-funded Scientific Research", *Research Evaluation*, 15(2), 145-152.

Trochim, W. and Donnelly, J., 2007. *The Research Methods Knowledge Base, 3rd Edition*. Cincinnati, OH: Atomic Dog Publishing.