

High-Content Words in Patent Records Reflect Key Innovations in the Evolution of Technology

Devin Chalmers¹, C. Cooper Francis¹, Noah Pepper¹, and Mark A. Bedau^{1,2,3,*}

¹Center for Advanced Computation, Reed College, Portland, Oregon, USA

²FOLSATEC, European School of Molecular Medicine, Milan, Italy

³Initiative for Science, Society, and Policy, University of Southern Denmark, Denmark

*Contact author: mark.bedau@reed.edu

Abstract

We study the evolution of technology as reflected in the US utility patents granted in the period 1976-2009. Previous work by Skusa and Bedau (2002) and Buchanan et al. (2010) used cumulative citation statistics to identify the inventions that most affect the course of evolution (those with the highest innovative impact). Here we examine the text of patent records (specifically, titles and abstracts) to identify which features are responsible for the high impact on later innovations. We use the TFIDF metric (term frequency times inverse document frequency) to identify which words best convey a patent's explicit content. Because a new patent is required to cite all important earlier patents ("prior art") that introduced innovations on which the new patent depends, we use the TFIDF scores of words in citing patents to identify a patent's *emergent content*. A patent's emergent content explains its impact on subsequent inventions; it reflects what traits in an invention actually led to a significant number of subsequent innovations. We illustrate two ways to visualize the explicit and emergent content of patents: word arrays and clouds. Examining the emergent content of populations of patents issued during different epochs reveals when important new ideas appear in the evolution of technology and how they affect its subsequent evolution.

Introduction

This paper presents a method to quantify and visualize certain aspects of the evolution of technology as reflected in patent records. Previous work by Skusa and Bedau (2002) (summarized by Bedau (2003)) used citation statistics to visualize and quantify one specific subset of cultural evolution: the evolution of technology as reflected in patent records. Buchanan et al. (2010) developed and extended this use of patent citations to identify which new inventions over the past three decades have seeded the greatest number of further innovations, termed patent "superstars." They concluded that three of the most important inventions in the past three decades were ink-jet printing, PCR, and stents, and they further showed that many superstar patents are "door-opening" inventions that spawn an especially wide range of further types of innovations.

This previous work highlights the importance of answering the following questions:

1. How can we identify which features characterize the core content of an invention?
2. In particular, which features make superstar patents so successful at spawning future inventions?
3. How have the key features driving technological innovation changed over the past few decades?

This paper aims to answer these three questions.

First, following the approach of Skusa and Bedau (2002) and Buchanan et al. (2010), we use citation statistics to identify how the key inventions driving technological evolution (patent superstars) have changed over the past few decades. To determine the content of these patents, a human can simply examine and interpret its title and abstract, but this process is labor intensive and introduces an element of subjectivity. We want to automate the process and make it objective, but this requires a method for identifying which terms in a document from a corpus especially indicate the distinctive content of that document. The TFIDF metric (term frequency times inverse document frequency, described below) is commonly used for precisely this purpose, so we identify the high-content terms in a patent record as those terms with high TFIDF scores. This method can naturally be generalized to identify high-content sequences of terms, or *n*-grams.

There is a complication that must be discussed. The high-content terms in a patent tend to reflect what the *inventor* believes are the important features of the invention; below, we term this the invention's *explicit* content. However, the features of an invention that actually play the biggest role in spawning further innovation might not be anticipated by the inventor, so they might not be well reflected in the patent's explicit content. Instead, they might be only *implicitly* reflected in the terms in the patent's title and abstract. Accordingly, to determine what features *actually* are important for an invention's fecundity, we look to the high-content words in the patents that cite the invention; below, we term this the invention's *emergent* content.

The explicit and implicit content of sets of patents can be visualized by two complementary methods: word arrays

and word clouds. By applying these methods to patents from successive epochs, we visualize how the explicit and emergent content of key inventions have changed over time. Our results described below indicate that innovation in the later half of the 1970s was especially active concerning automobile emissions and personal electronics. In the 1980s, the dominant technology drivers shifted to zeolites and semiconductors. The 1990s and 2000s were both dominated by a range of further technologies, especially inkjet printing, PCR, stents, e-commerce, wireless communication, and solid-state storage.

Our work here illustrates how citations and key terms in patent records provide a rich empirical foundation for the study of the evolution of technology. Since technology is one aspect of culture, this work helps illuminate the similarities and differences between cultural and biological evolution. As the papers in Wheeler et al. (2002) indicate, a variety of approaches are being applied to the study of the evolution of culture. The application of the concept of *memes* from Dawkins (1989) is especially hotly disputed, as illustrated by comparison of Sperber (1996), Fracchia and Lewontin (1999), Dennett (2006), and the papers in Aunger (2000). Rather than adding to these polemics, we provide an empirically grounded account of the actual evolution of one important aspect of culture—patented technological innovations—and we develop a method for identifying the key features in inventions that make their impact on new innovations especially big. This line of research might eventually help resolve some of the controversies about cultural evolution, including those about memetics.

The patent record

Patents are granted to inventions only if the patent’s examiners are satisfied that the invention is novel, non-obvious, and useful. A patent’s novelty is documented by citing the previous patents (and sometimes published papers) on which it depends and builds; these are known as the patent’s “prior art.” Perko and Narin (1997) and Hall et al. (2005) explain that the patent examiner is the ultimate referee of what patents must be cited, and can add citations that were neglected or omitted on the application.

Our data set consists of records of all the utility patents granted between 1976 and 2009 in the US. (That time window was chosen because of the ready availability of patent data for that period.) In this study, a patent’s title and abstract are concatenated to constitute its “record.” (A natural generalization of our methods would add further text to a patent’s record, such as its claims. Our analysis also uses certain other information about a patent, such as its unique identifying number and, most importantly, the previous patents which it cites—its “prior art.”)

Our corpus of 3,630,466 patent records contains 459,232,327 individual word tokens, employing a dictionary of 993,544 word types. Our analysis relies crucially on ci-

tations among patents. The patents in our data set bestowed a total of 38,893,014 citations, of which 30,198,227 (about 80%) hit patents in our dataset. Our patents on average cite 10.97 earlier patents and are cited 8.25 times, but 87,695 (2.4%) cite no previous patents.

Our investigation of the evolution of technology is motivated by an analogy with biological evolution. A patented invention is viewed as an organism, and different inventions compete for adoption by users in various niches. The spread of inventions in niches is analogous to the Darwinian process of natural selection (we make no assumptions here about how close that analogy is). When a new patent cites prior art (i.e., earlier patented inventions on which it depends and builds), we consider the earlier patent to have spawned an incipient daughter species.¹ Those inventions that spawn especially many incipient daughter species and so are most heavily cited, are the inventions that drive the course of the evolution of technology.

From patent citations, it is possible to reconstruct the entire phylogeny of the evolving network of patented inventions. The entire set of patent records is analogous to the entire fossil record, except that the patent record is virtually complete and mostly accurate and unambiguous.² Accordingly the phylogenies that can be reconstructed are stunningly complete, covering every patent (organism in the population). It would be a biologist’s dream to work with empirical phylogenies that are this dense and accurate.

Shadow patents

In order to test whether the citation patterns that we observe in the patent data could have been created by a random process that ignores the content of the patents involved, we construct a system of “shadow” patents. By construction, shadow patents mirror (or “shadow”) many aspects of real patents.

The precise mechanism for generating shadow patents is as follows: If p real patents were granted in year y , then p shadow patents are also granted that year. If a particular patent, i , is granted in year y and cites c earlier patents, then the shadow patent, i_s , is also granted in year y and cites c earlier shadow patents. However, whereas a real patent cites its prior art, a shadow patent cites earlier patent chosen at random (with replacement) from the patents cited by real

¹For simplicity of exposition and when no confusion should result, we will sometimes speak of a patent when we mean to refer to the invention that is patented.

²It is worth noting that the patent record is somewhat “dirty.” Cleaning the data involves various ad hoc and approximate procedures, and raw data is sometimes corrupted or lost. It should be noted in addition that simple citation metrics can draw an incomplete picture of what is happening in the patent data. We know from Cohen et al. (2000) that patent value, citation rate, patent frequency and citation methodology vary greatly in different industries. This should prompt a salutary dose of skepticism about simplistic sweeping interpretations of citation patterns.

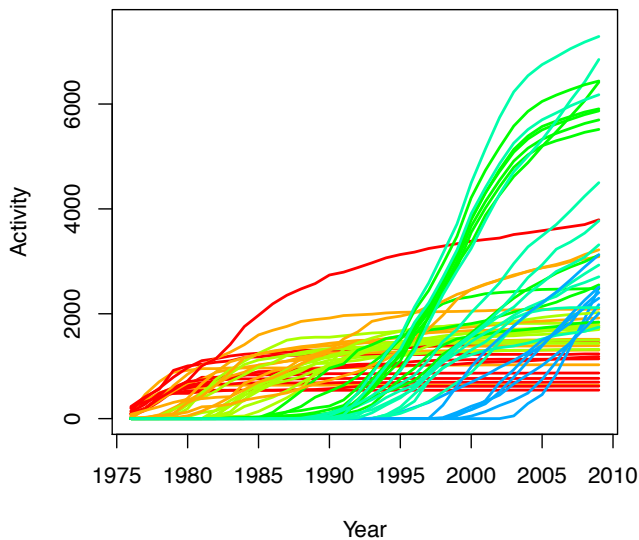


Figure 1: Cumulative citations (or “activity”) of the twenty most heavily cited patents from each decade (see Table 2), divided by the prior expected probability of being cited.

patents granted in year y .

The system of shadow patents is a null hypothesis against which we measure whether the citation patterns we observe in real patents could have been created by a random process that ignores the content of the patents.

Highly cited inventions

Following Skusa and Bedau (2002) and Buchanan et al. (2010), we begin by examining the most highly-cited patents, for their high citation counts show that they have an especially great influence on the subsequent evolution of technology. Because of variation in the citation rate and size of the patent corpus each year, we normalize citation counts to make them comparable across epochs, as follows: In a given year, each incoming citation count is divided by the a priori expected probability of a patent being cited at a given time. Assuming that all patents have an equal probability of being cited, this prior probability of being cited at t is calculated as the number of citations given by all the patents issued at t (the number of citations given out) divided by the number of patents issued up to t (the number of patents that could be cited). Exploration of different normalizations is available in Buchanan et al. (2010).

First we examine the twenty patents that received the most citations from all of the patents issued in each of the last few decades. Table 2 describes most of the main innovations covered by those patents. While some heavily cited patents fall outside of the kinds of innovations we list, most do fit in our list. Since our data starts in 1976, relatively few of the

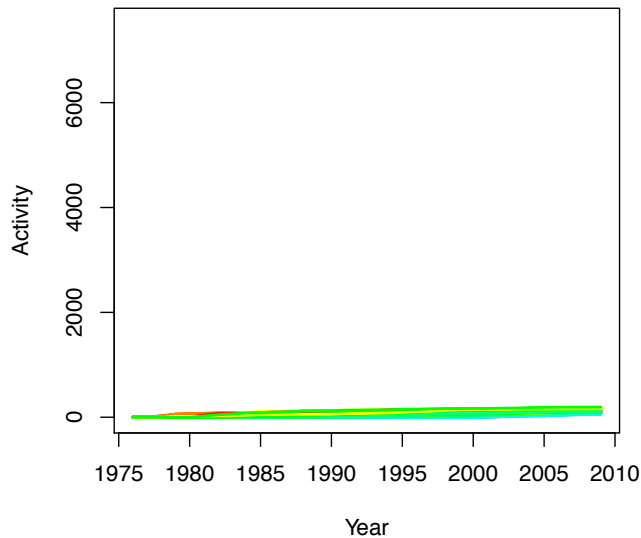


Figure 2: Cumulative citations (or “activity”) of the twenty most heavily cited “shadow” patents from each decade. Compare with Figure 1.

citations from the first decade contribute to our analysis.

Figure 1 shows the cumulative citations received by the twenty most heavily cited patents in each decade, colored by the year in which the patent was granted. These cumulative citation counts dramatically illustrate which patents are most influencing the evolution of technology at any given time. Analysis of the patent titles and abstracts reveals that the most “fecund” innovations of the past three decades fall into the following technology sectors: automobile emissions, personal electronics, zeolites, semiconductors, inkjet printing, PCR and stents. This decade-by-decade analysis corroborates and extends the results reported by Buchanan et al. (2010).

Figure 1 can be directly compared with Figure 2, which shows the cumulative citation counts of the most heavily cited *shadow* patents. (Real and shadow patents are normalized identically.) Note that the most heavily cited real patents receive two orders of magnitude more citations than their shadow counterparts. This indicates that heavy citation counts observed in the real patents are not merely an artifact of the numbers of patents giving and receiving citations. Randomly distributed citations would never produce the high citation counts observed for the most fecund inventions.

Many details about the evolution of technology can be read off from Figure 1. For example, the most highly-cited patents in the 1970s (concerning automobile emission and personal electronics) are never cited after the 70s and become dormant (indicated by flat lines). In addition,

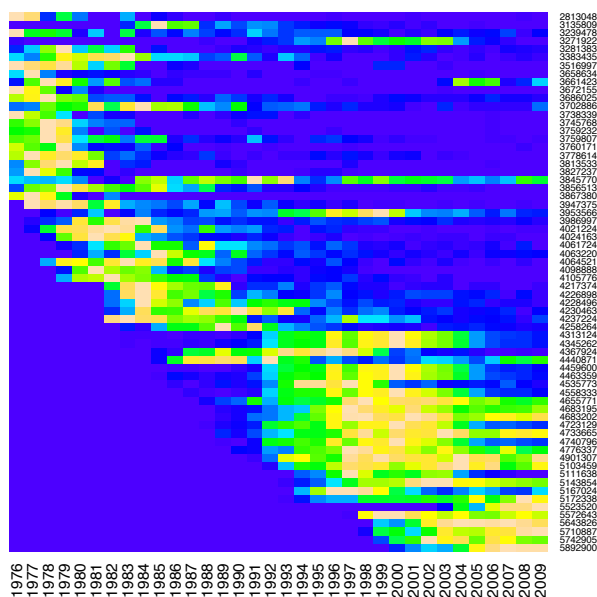


Figure 3: The citation rate for the twenty most heavily cited patents from each decade (see Table 2). Citations are normalized as in Figure 1, and scaled to the interval $[0, 1]$.

one patent (concerning zeolites) is especially heavily cited through most of the 1980s, but its influence subsequently is dominated by a new group of patents (about inkjet printing, PCR, and stents) from the late 1980s, which eventually achieve the highest citation counts overall.

Figure 3 plots the citation rate time series for each of the patents depicted in Figure 1, scaled to the range $[0, 1]$. (Mathematically, this corresponds to the *slope* of the patents shown in Figure 1.) This heatmap shows each patent at each moment, with hotter colors indicating patents that are spawning more new inventions. The heatmap shows that citation rates for most of the most heavily cited patents have cooled off by 2005, and a new crop of patents (about, e.g., genetically modified organisms, e-commerce, and solid-state storage) are heating up today.

The TFIDF measure of high-content words

In this paper, we identify the words that best capture the content of an invention by applying the TFIDF metric to the words in the invention’s patent record. TFIDF scores are a standard way to measure the significance of a word in a given document within a corpus, as Spärch Jones (1972) and Salton and McGill (1983) explain. The intuitive idea behind the TFIDF metric is that the most significant words in a document are used frequently within that document, but are not widely used in other documents from the corpus. Accordingly, the measure has two components: term fre-

quency (TF), and inverse document frequency (IDF). Term frequency is just the frequency of a word w in a document d :

$$\text{TF}(w, d) = \frac{|\{w' \in d : w' = w\}|}{|\{w \in d\}|}.$$

The inverse document frequency of a word w in a corpus D is simply the logarithm of the inverse of the fraction of documents in D which contain w :

$$\text{IDF}_D(w) = \log \frac{|D|}{|\{d \in D : w \in d\}|}.$$

Then the TFIDF score for a word w in a document d in a corpus D is just the product of these two measures:

$$\text{TFIDF}_D(w, d) = \text{TF}(w, d) \times \text{IDF}_D(w).$$

To illustrate the TFIDF metric in the patent record, consider the title and abstract of US patent number 4683202 (granted 28 July 1987), which happens to be the most cited patent in the last decade:

Process for amplifying nucleic acid sequences

The present invention is directed to a process for amplifying any desired specific nucleic acid sequence contained in a nucleic acid or mixture thereof. The process comprises treating separate complementary strands of the nucleic acid with a molar excess of two oligonucleotide primers, and extending the primers to form complementary primer extension products which act as templates for synthesizing the desired nucleic acid sequence. The steps of the reaction may be carried out stepwise or simultaneously and can be repeated as often as desired.

The title and abstract contain 90 word tokens and 56 word types. The most frequent word is ‘the’, appearing seven times, for a term frequency of $\text{TF} = 0.0778$. However, the ubiquitousness of ‘the’ gives it a very high document frequency within the patent corpus, and so a low inverse document frequency, $\text{IDF} = 0.009$, which shrinks its resulting TFIDF score.

The words in the title and abstract of Patent 4683202 with the highest and lowest TFIDF scores appear in Table 1. Note that words with the highest TFIDF scores convey a lot of information about the topic of this patent; for example, ‘nucleic’, ‘acid’, ‘primers’, and ‘amplifying’ all have high TFIDF scores. By contrast, words with the lowest TFIDF scores (‘the’, ‘and’, ‘a’, ...) convey virtually no information about the patent. Instead, they are so-called “stop words” that reflect grammar and logic rather than content.

The emergent content of patents

The evolution of technology that we study consists of the rise and fall of superstar patents that dominate different epochs. This raises a question: What is the *content* of the

Rank	Term	Count	TF	IDF	TFIDF
1	nucleic	5	0.0556	2.3167	0.1287
2	acid	5	0.0556	1.4203	0.0789
3	primers	2	0.0222	3.2907	0.0731
4	amplifying	2	0.0222	2.6341	0.0585
5	complementary	2	0.0222	2.2645	0.0503
⋮	⋮	⋮	⋮	⋮	⋮
51	in	1	0.0111	0.1187	0.0013
52	is	1	0.0111	0.1151	0.0013
53	of	3	0.0333	0.023	0.0008
54	the	7	0.0778	0.009	0.0007
55	and	2	0.0222	0.0217	0.0005
56	a	3	0.0333	0.0135	0.0004

Table 1: TFIDF values for words in the title or abstract of patent no. 4683202, *Process for amplifying nucleic acid sequences*.

innovations in the superstar patents? Which of their features make them superstars? People can often glean such information by reading superstar patents’ titles and abstracts. For example, personal inspection of Table 2 reveals a lot about the content of the most highly cited patents during recent decades. Here we develop methods for determining a patent’s content without human intervention. Specifically, we use TFIDF profiles of the words in a patent to measure the patent’s content.

We start with some definitions. We write $C(p_1, p_2)$ if patent p_1 cites patent p_2 , and we let $\overleftarrow{C}(p)$ be the set of patents that cite p , i.e., p ’s “incoming” citations:

$$\overleftarrow{C}(p) = \{p' : C(p', p)\}.$$

Then, the number of patents that cite p , or $|\overleftarrow{C}(p)|$, can be used to identify the superstars of a set of patents, or superstars $_N(P)$, as the N most heavily cited patents in P , ranked by $|\overleftarrow{C}(p)|$.

Let the representative (or high-content) words of a patent p in the patent record \mathbb{P} be the set of words w in the patent with TFIDF above a given threshold, θ :

$$\text{TFIDF}_\theta(p) = \{w \in p : \text{TFIDF}_{\mathbb{P}}(p, w) \geq \theta\}$$

(For this paper, we typically use a threshold of $\theta = 0.05$, which eliminates most stop words and typically picks out just a few words from each patent.)

These concepts easily extend to a set of patents, P . We can identify their citers,

$$\overleftarrow{C}(P) = \bigcup_{p \in P} \overleftarrow{C}(p).$$

and their high-content words,

$$\text{TFIDF}_\theta(P) = \bigcup_{p \in P} \text{TFIDF}_\theta(p).$$

A central hypothesis in our paper is that the high-TFIDF words in a patent, or set of patents, are key to revealing their content. We consider $\text{TFIDF}_\theta(P)$ to be the explicit content of a set of patents, and we consider the *emergent content* of a set of patents, P , to be the high-content words in the set of patents that *cite* patents in P , or

$$\text{TFIDF}_\theta(\overleftarrow{C}(P)).$$

This content is “emergent” because it is implicit; it depends on what subsequent inventions “see” in the inventions in P , and how the inventions function as prior art. Analogously, $\text{TFIDF}_\theta(\overleftarrow{C}(\text{superstars}(P)))$ is the emergent content of the superstars of a set of patents, P . We give examples of both kinds of emergent content below.

Visualizing emergent content with word arrays

The evolution of the emergent content of the patents consists of a list of words with various associated numerical values. A word’s value can include such things as the word’s TFIDF score, its frequency in the corpus, or the number of patents that contain the word. The evolution of the emergent content in a set of patents can be visualized in various ways, once two things have been determined: (1) Which words contribute to the content? (2) How is the word’s numerical value calculated? The visualization methods described here work for any evolving list of words with associated numerical values.

Word arrays are simply lists of words in some fixed, meaningful order, each associated with its numerical value in a given time period. Word arrays are analogous to gene chips, which visualize the expression profile of protein-producing genes. Since word arrays can be represented in one dimension, aligning word arrays from successive snapshots of a population of patent records yields a two-dimensional “movie” of the evolving meaning of a given period of the evolution of technology.

Figure 4 shows the raw time behavior of the emergent content of the superstar patents in Table 2. The words were selected from the citers of the patents in the table. Word frequencies were computed over all of the abstracts of patents issued in each year. For each word, a time vector of values is computed, with each entry $c_{w,t}$ the word w ’s raw frequency in year t :

$$c_{w,t} = \frac{\sum_{p \in P_t} |\{w' \in p : w' = w\}|}{\sum_{p \in P_t} |\{w \in p\}|}$$

In Figure 4, each word’s vector has been scaled to fit the range $[0, 1]$, in order to show each word’s rise and fall relative to itself. Figure 4 provides one perspective on the evolving content that is driving innovation in the evolution of technology.

Successive columns in a word array indicate successive moments of time. Figure 4 is like a “film strip” of the evolu-

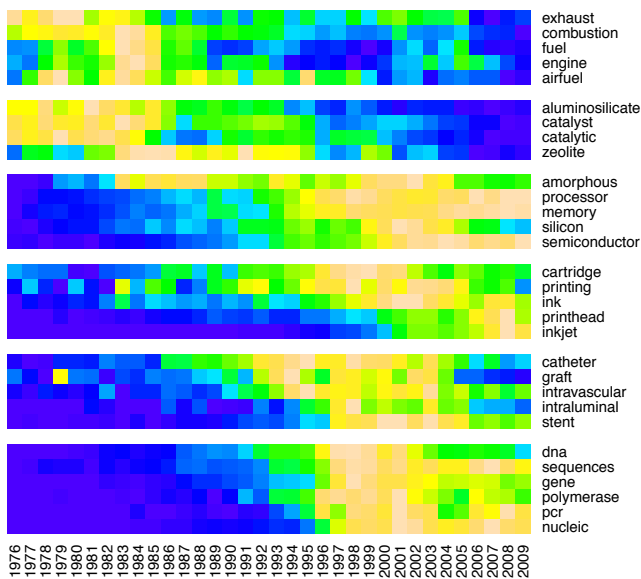


Figure 4: The relative frequency over time of a subset of the emergent content of the top technology patents identified in Table 2. The x -axis is years, and the y -axis is individual high-content words.

tion of certain high-impact players in the evolution of technology; each single column is a single frame in the film. It is evident that the main innovation drivers of the 1970s (automobile exhaust and personal computing) are almost completely dormant today. Similarly, the main technology drivers of the 1990s and 2000s (inkjet printing, PCR, stents, and semiconductors) were almost completely dormant for all of the 1970s and 1980s. Furthermore, inspection shows that stents have been cooling off recently, while key components of the PCR and semiconductor genealogies remain very hot.

Visualizing emergent content with word clouds

The word clouds described in this section are another way to visualize how the content of inventions changes over the decades. A word cloud is a two-dimensional agglomeration of the high-content words in some patents, with the words sized according to their numerical value. Since the most important words are the largest, people can easily read the key content in word clouds.

The algorithm for calculating word clouds from a set, P , of patents in a decade has three steps, illustrated in Figure 5:

1. Determine the decade's superstar patents (colored blue in the diagram), $\text{superstars}(P)$; these are the patents most heavily cited by the patents issued in the decade.
2. Determine all the patents (green stars) that cite any of the decade's superstars, including patents granted after the decade in question: $\overleftarrow{C}(\text{superstars}(P))$.

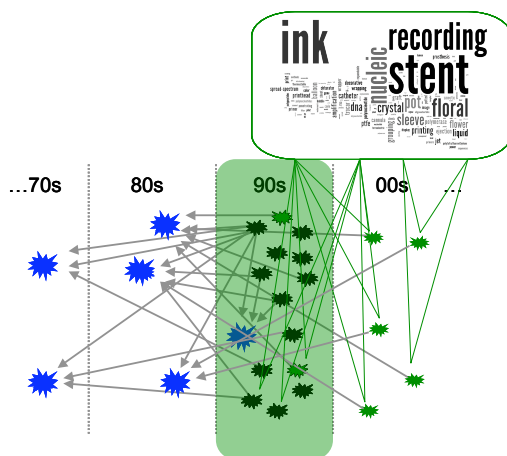


Figure 5: Cartoon sketching the three stages by which word clouds emerge out of a set of patents (e.g., those issued in the 1990s). First, the superstars (blue stars) of the patents issued in the 1990s are identified, then their citers (green stars) are identified, and finally the emergent content of the superstars is identified: $\text{TFIDF}_\theta(\overleftarrow{C}(\text{superstars}(\text{patents}_{1990s})))$. Gray lines are citations between patents.

3. Identify the emergent content of the superstar patents, $\text{TFIDF}_\theta(\overleftarrow{C}(\text{superstars}(P)))$, arrange the words in a cloud,³ and size each word w by the number of patents in the decade that contain the word: $|\{p \in P : \text{TFIDF}_P(p, w) \geq \theta\}|$.

We illustrate word clouds by focusing on the superstar patents in each decade, and extracting the emergent content of superstars in the familiar way. In this case, we choose to size the words in a word cloud by the number of patents in the corpus that contain the word.

Figure 6 shows the word clouds that emerge from the patents in each decade in our data set: the 1970s (starting with 1976), 1980s, 1990s, and 2000s. Collecting and smoothly connecting these snapshots yields a movie of how the key innovations in patented technology evolve over time.

Conclusions

There are many differences between biological evolution and the evolution of technology, but there are also important similarities. The most important similarity here is the non-randomness or adaptive quality of the key features of the entities that have the greatest impact on new innovations. Comparison with shadow patents confirms that citation rates of the most heavily cited patents would virtually never occur if patents were cited at random and irrespective of their

³Word cloud layout algorithm by Jonathan Feinberg, Wordle.net and IBM Research, <http://www.wordle.net/credits>.

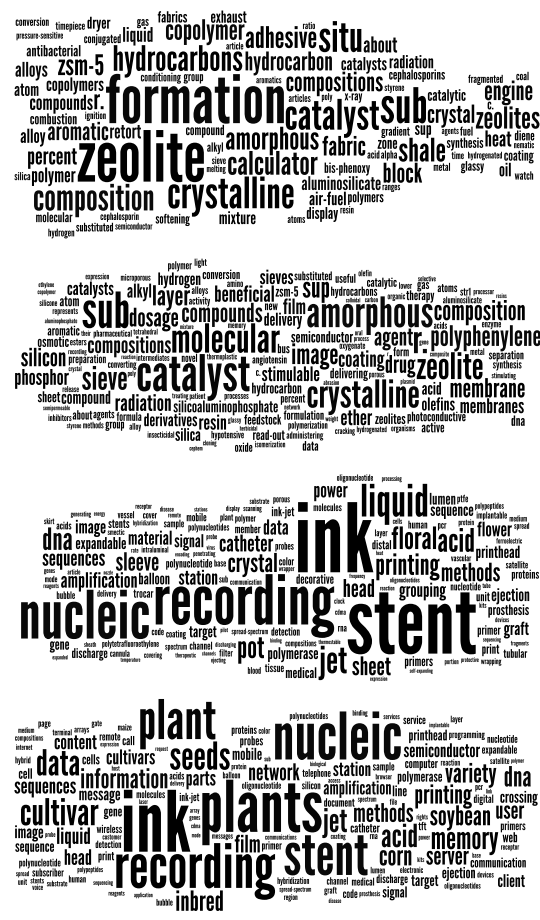


Figure 6: The emergent word clouds for top cited patents in the 1970s, 1980s, 1990s, and 2000s (from top to bottom). The word clouds are still shots from a movie of the evolving meaning of the main technologies driving the evolution of technology.

specific features.

We identify the “emergent” content of sets of patents as the “explicit” content of the patents that cite the patents in the set, measured by high TFIDF scores. We use word arrays and word clouds to visualize the evolution of the key features of patents that have an especially high impact on new innovations. This brings us closer to understanding what makes superstar patents so heavily cited.

Here, the environment that drives adaptation is the technological and economic context of an epoch. If patents and inventions are significantly analogous to biological organisms, then we have created a new way to identify and visualize the emergent semantics of technological evolution through time. Whereas the citation record of patents provides a phylogeny of patented inventions, word arrays and clouds represent the changing emergent content of the drivers of technological innovation through time.

Acknowledgments.

For helpful comments on the topics discussed herein, thanks to Andrew Buchanan, Bobby Gadda, Norman Packard, Emily Parke, and Federico Vaggi, and to the anonymous referees for Artificial Life XII.

References

Aunger, R. (2000). Oxford University Press, Oxford; New York.

Bedau, M. A. (2003). Objectifying values in science: A case study. In P. Machamer, G. W., editor, *Science, Values, and Objectivity*, pages 190–219. University of Pittsburgh Press, Pittsburgh, PA.

Buchanan, A., Packard, N., and Bedau, M. (2010). Adaptive innovative impact on the evolution of technology in the patent record. In *Artificial Life XII (this volume)*. MIT Press, Cambridge, MA.

Cohen, W. M., Nelson, R. R., and Walsh, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why u.s. manufacturing firms patent (or not). *NBER Working Paper Series*, w7552.

Dawkins, R. (1989). *The selfish gene*. Oxford University Press, Oxford; New York, 2nd edition.

Dennett, D. C. (2006). *Breaking the spell: Religion as a natural phenomenon*. Viking.

Fracchia, J. and Lewontin, R. C. (1999). Does culture evolve? *History and Theory*, 38(4):52–78.

Hall, B. H., Jaffe, A., and Trajtenberg, M. (2005). Market value and patent citations: A first look. *Rand Journal of Economics*.

Perko, J. S. and Narin, F. (1997). The transfer of public science to patented technology: A case study in agricultural science. *Journal of Technology Transfer*, 22: 65–72.

Salton, G. and McGill, M. J. (1983). *Introduction to modern information retrieval*. McGraw-Hill.

Skusa, A. and Bedau, M. A. (2002). Towards a comparison of evolutionary creativity in biological and cultural evolution. In Standish, R., Bedau, M. A., and Abbass, H. A., editors, *Artificial Life VIII*, pages 233–242. MIT Press, Cambridge, MA.

Spärch Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28:11–21.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Blackwell, Cambridge, Mass.

Wheeler, M., Ziman, J., and Boden, M. A., editors (2002). *The evolution of cultural entities*. Oxford University Press.

Table 2: Major innovations (or technology “superstars”) as reflected in citation patterns from each decade.

Selections from the twenty patents that received the most citations from patents issued in 1976-1979	Citations
Automobile emissions	
3827237: Method and apparatus for removal of noxious components from the exhaust of internal combustion engines	69
3759232: Method and apparatus to remove polluting components from the exhaust gases of internal combustion engines	44
3745768: Apparatus to control the proportion of air and fuel in the air-fuel mixture of internal combustion engines	44
Personal electronics	
3760171: Programmable calculators having display means and multiple memories	69
3672155: Solid state watch	40
3947375: Liquid crystal materials and devices	39
3813533: Clock calculator	39
Selections from the twenty patents that received the most citations from patents issued in 1980-1989	Citations
Zeolites	
3702886: Crystalline zeolite ZSM-5 and method of preparing the same	196
4061724: Crystalline silica	120
4440871: Crystalline silicoaluminophosphates	93
Semiconductors	
3856513: Novel amorphous metals and amorphous metal articles	119
4226898: Amorphous semiconductors equivalent to crystalline semiconductors produced by a glow discharge process	115
4217374: Amorphous semiconductors equivalent to crystalline semiconductors	109
4064521: Semiconductor device having a body of amorphous silicon	108
Selections from the twenty patents that received the most citations from patents issued in 1990-1999	Citations
Ink-jet printing	
4723129: Bubble jet recording method and apparatus in which a heating element generates bubbles in a liquid flow path to project droplets	753
4463359: Droplet generating method and apparatus	677
4740796: Bubble jet recording method and apparatus in which a heating element generates bubbles in multiple liquid flow paths to project droplets	663
4558333: Liquid jet recording head	637
4345262: Ink jet recording method	630
4313124: Liquid jet recording process and liquid jet recording head	612
4459600: Liquid jet recording device	599
PCR	
4683195: Process for amplifying, detecting, and/or-cloning nucleic acid sequences	620
4683202: Process for amplifying nucleic acid sequences	597
Stents	
4733665: Expandable intraluminal graft, and method and apparatus for	349
4655771: Prosthesis comprising an expansible or contractile tubular body	277
4776337: Expandable intraluminal graft, and method and apparatus for	268
Selections from the twenty patents that received the most citations from patents issued in 2000-2009	Citations
Ink-jet printing	
4723129, 4740796, 4463359, 4558333, 4345262, 4313124, 4459600 (see above)	6518
PCR	
4683202, 4683195 (see above)	2526
E-commerce	
5572643: Web browser with dynamic display of information objects during linking	839
5892900: Systems and methods for secure transaction management and electronic rights protection	770
5710887: Computer system and method for electronic commerce	655
Wireless communication	
5103459: System and method for generating signal waveforms in a CDMA cellular telephone system	802
5742905: Personal communications internetworking	762
4901307: Spread spectrum multiple access communication system using satellite or terrestrial repeaters	665
Solid-state storage	
5643826: Method for manufacturing a semiconductor device	831
5172338: Multi-state EEPROM read and write circuits and techniques	629
Stents	
4733665: (see above)	940