Fitting in or starting new? An analysis of invention, constraint, and the emergence of new categories in the software industry

Elizabeth G. Pontikes

Stanford Graduate School of Business

Acknowledgements: I would like to thank Bill Barnett, Mike Hannan, Hayagreeva Rao, Jesper Sørensen, Jerker Denrell, Martin Ruef, Woody Powell, Greta Hsu, and Glenn Carroll for thoughtful comments and suggestions, and Stanford University for support.

ABSTRACT

New inventions present something that is different from what already exists, and so they often do not fit within our classifications systems. The emergence of novel technologies can disrupt structures and lead to the creation of new market categories, but sometimes technical novelty does not change the existing system. I argue that this depends on whether the existing structure is more or less constraining to organizations. To the extent that categories constrain members, they also provide contrast within the environment. The combination of constraint and contrast makes existing categories less accommodating to novel invention. I propose that organizations that are both different in their knowledge creation and in constraining categories are likely to create a new category label. However, for those that are inventing differently, being in a less constraining category will decrease the likelihood that a firm creates a new label. I study these ideas in the context of the software industry, for the years 1990-2002. I use patent data to identify novel innovation, and created a data set using software press releases to identify market categorization. Results support hypotheses. Firms that are unique in their knowledge development are more likely to create or join new market categories only when the existing category structure is constraining. When existing categories lack constraint, technological difference does not make firms more likely to join new categories.

When does technological invention lead to new organizational categories? Scholars have historically pointed to technical development as the seed for new markets (e.g. (Schumpeter 1934:2006), but we know that many novel inventions do not generate new organizational categories. Even the microprocessor – which eventually led to one of the most important product markets of modern times – remained a mere calculator component for a year before Intel began to promote it as a "computer on a chip" (Aspray 1997). Other important inventions never lead to new categories. One literature employs a technology perspective to explain this. In these studies, researchers typically distinguish among different types of innovation, such as competence-destroying compared to competence-enhancing innovations (Tushman and Anderson 1986), process versus product inventions (Levin and Reiss 1988; Gilbert 2006), or exploration as opposed to exploitation (March 1991). All of these use technological differences to explain why some new knowledge has more impact than other development in disrupting the existing market structure with new organizational categories. At the same time, another literature has developed from the categorization perspective, where researchers explain new category emergence using characteristics of the existing structure (Ruef 2000; Rao, Monin and Durand 2003). Left out of the technology perspective is that new categories are constructed by people based on their perceptions of whether organizations are similar or different, and left out of the categorization perspective is how external forces – such as new technologies – disrupt the system. In this paper, I argue that whether a technological invention becomes the basis for a new category depends both on how novel the development is, and also on the existing category structure.

Inventions are introduced into a social space and business marketplace, where people agree, to some extent, that certain features of organizations are relevant for categorization.

When an organization introduces a new technology, potential customers will evaluate the claim to novelty by comparing it to familiar items that are "in" existing categories. New categories can emerge when the innovation is perceived to be too different to fit into existing categories. For example, Sonic Solutions, one of the earliest companies to provide digital music editing, was founded in 1986 by former employees of the Lucasfilm computer division. Their software used digital technology to edit noise out of old tracks, which previously was done in an analog method of tape splicing. In the analog method, a sound technician would "shuttle" the tape back and forth to find the editing point, and then mark the exact point where the tape would literally be cut and then taped back together. The digital method, which is easier to use and more accurate, displays the signal visually, and the tape is digitally spliced.[1]

Digital audio used a radically new technology to achieve similar objectives more efficiently than the analog technology. The differences between the digital and analog methods of tape splicing ended up being relevant to consumers, and a new organizational category of digital audio editing evolved. It is important to note that the new technology had originally been pioneered while the founders were at Lucasfilm, an established company in the entertainment business, so it is not that a new category emerged because the new technologies were created by a start-up firm. Rather, the features of the new digital technology were perceived as inconsistent with the expectations of what analog editing should be. People expected traditional tape splicing to use sound to detect the splicing points, and for the splicing to be manual. This created a situation where observers differentiated the new technology, and a new category formed.

New inventions do not always result in the emergence of a new organizational category. Similar to music editing's digital shift, automobile manufacturers migrated to electronically controlled engines starting in the 1970's. This was a major technological innovation targeted at

[1] From an interview with Mary Saur, co-founder of Sonic Solutions. January 19, 2006.

the core component of the automobile (Robinson 1976). The innovation caught on; today most automobiles have computerized engines. However, as it diffused it received little attention from the standard customer, and no new category for the automobile emerged. What is the difference between digital music and computerized engines? It is difficult to derive a purely technical explanation for why one spurred the creation of a new organizational category, whereas the other was quietly adopted within an existing category. Although technological innovations are often the sources of new product and organizational categories, the technical novelty of an innovation is not adequate to determine whether a new organizational category will emerge. To understand how new organizational categories come about we must also take into account the existing environment. In my dissertation, I propose that whether the existing category structure imposes strong or weak constraints on its members influences the likelihood that an invention will become the basis for a new organizational category.

Knowledge space and market space

To study whether new inventions create new categories on the market, I distinguish two planes in which an organization acts: knowledge space and market space (figure1). Knowledge space is the arena where an organization invents and develops new knowledge. Market space is where an organization creates and markets products. Knowledge space is populated by many different actors: universities, not for profits, individual inventors, and organizations from many different industries. Market space is populated by all organizations selling products relevant to the industry; some of these organizations develop knowledge, and some do not. Organizations link knowledge space and market space when researching organizations incorporate their developments into products. By defining knowledge space distinctly from market space, I can determine whether an invention is very different or very similar to others' knowledge

developments, based only on knowledge space difference. At the same time, I can measure characteristics of the market structure in market space. This allows me to investigate the extent to which knowledge space difference alone, or knowledge space difference under specific market space conditions, leads organizations to try and form new market categories.

Technical novelty and new category formation

Previous research in the technology perspective shows that the more different a new invention is, the more likely it will be labeled as something new in the marketplace. This perspective defines novelty within knowledge space, as a property of the technology, and investigates what types of novel technologies will change structure of the current environment. Studies show that that different environmental conditions such as firm concentration (Levin, Cohen and Mowery 1985; Gilbert 2006) or technological opportunities across industries (Scherer 1967; Levin, Cohen and Mowery 1985; Jaffe 1986) affect whether organizations develop new technologies. Other research demonstrates that new (or young) firms tend to create more novel innovations, whereas existing (or old) companies innovate in ways that build on their competencies (Tushman and Anderson 1986; Sorensen and Stuart 2000). When a disruptive innovation does emerge, existing firms that have developed features that are complementary to this technology are more likely to adapt to the new environment (Teece 1986). These studies ask what types of environments encourage novel innovation, what types of companies will create novel inventions, and which organizations are most likely to adapt to changes caused by innovation. However, they tend to assume that the same type of technical difference will have the same types of effects on the market, regardless of the current market structure.

Yet we know that the effects of technological change depend on the nature of the current environment. In early telephony, technological change favored either advanced or primitive

firms, depending on technological standards and organizational differentiation (Barnett 1990). In biotechnology, an organization's position in a research network affected the organization's subsequent research, which in turn affected its position in the network (Powell, Koput and Smith-Doerr 1996). In the computer industry, organizations that researched in areas similar to technological communities that use the same CPU were more likely to join that community and adopt the respective CPU (Pontikes and Barnett 2007). These studies indicate that the effects of innovative activity arise not from properties of the innovation itself, but from an interaction between the innovative activity and the environmental structure.

I extend this notion and argue that the creation of a new organizational category, which is often interpreted as an indicator of technical novelty, actually depends on both knowledge difference and the existing category structure. What we generally consider to be a novel technology cannot be measured independently of an environment's category structure. I propose that novelty is partially determined by perceptions of people who are steeped in an existing organizational context, and suggest that the lack of consistent findings about when inventions create new markets stems from this assumption that technological newness is an adequate measure of novelty. Rather than attempt to refine the definition of technical novelty by creating more fine-grained technical distinctions, I suggest that we consider both the technical newness of an innovation as well as the existing category structure.

Categories and varying constraint

Previous research from the category perspective investigates how the existing classification structure gives rise to new organizational categories. Categorization has significant implications for organizations (White 1981; Zuckerman 1999; Zuckerman 2000; Hannan, Pólos

and Carroll 2007). When organizations are classified in meaningful ways, potential customers can more easily find and purchase the products or services offered by the organization, regulatory bodies can determine which rules to enforce, and potential financiers can decide where to invest. Organizations that are not easily categorized may even suffer economically (Zuckerman 1999). But any number of categories might evolve given differences among organizations – so how do existing categories form?

Existing institutions, forms, and categories are the platform on which new meanings are constructed. Dimensions that are relevant to existing categories often form the basis for new categories. Sometimes this happens through the combination of elements from existing categories, such as "biomedical engineering," which applies engineering to the needs of health care, or the recently emerged label "cosmeceuticals," which combines cosmetics with pharmaceuticals. Because new categories often draw on definitions of existing categories, they emerge more frequently in areas where there are many existing categories with similar identities (Ruef 2000). Existing categories can also provide the basis for new categories by providing a point of contrast, when a new category arises in opposition to existing structures (Carroll and Swaminathan 2000; Barnett 2004). For example, the new category of Nouvelle French cuisine came about in this way, rebelling against the long tradition of Classical French cuisine. The well recognized schema for Classical French cuisine specified in detail the appropriate culinary rhetoric, rules of cooking, ingredients, the role of the chef, and even the organization of the menu. In this situation, a new category of cuisine, Nouvelle, emerged using the same dimensions in its schema, but requiring opposite values along those dimensions. Where Classical cuisine called for high game and shellfish as ingredients, Nouvelle cuisine used fruits, vegetables, and sea fish (Rao, Monin and Durand 2003).

In these examples, the meaning for a new category is based on the schemata of existing categories, either by combining relevant dimensions of two different categories, or by establishing a category that is the "opposite" of an existing category, on the same relevant dimensions. This process can only occur because the agreed upon codes and boundaries for existing categories provide contrast for those that are emerging. If there were no agreement that the science of biology was separate from engineering, then biomedical engineering would likely be part of biology. Or, if the schema for Classical cuisine allowed the use of all types of ingredients, then connoisseurs would not recognize dishes using fruits and vegetables as anything new. Indeed, in the disk array market, although there were activists setting standards and building associations around the disk array product, because many organizations from other categories were also selling disk arrays, there was no standard against which "disk array" could be differentiated, and the set of disk array producers did not evolve into a distinct organizational category (McKendrick and Carroll 2001). The fact that existing categories have defined schemata provides negative space into which new categories can emerge.

Hannan, Pólos and Carroll (2007) formalize the emergence of organizational categories using interactions of audiences, or groups that take an interest in the domain. In this conception, consistent with empirical research on categorization, individual audience members first try to cluster organizations by similarity and label these clusters. In the next stage an extensional consensus might develop, where audience members agree on which organizations belong to a label, but do not agree about what the label means, in terms of the specific features that organizational members must have to be included. Audiences agree on a schema for the label in the next stage, when they come to a consensus about the specifics of the meaning of the label. In this stage people agree that organizations identified with that label should possess a specific set

of features. At this point an organizational category develops, and if conformity to a category schema becomes taken-for-granted throughout the audience, the category becomes an organizational form. This conception distinguishes between an unschematized "label" for a cluster of similar organizations, and a schematized "category." Here, labeled clusters are the seeds for new categories.

Previous work from the categorization perspective tends to presume that only categories and forms have a large impact on organizations, and when a cluster of organizations are labeled but not categorized, that labels do not have widespread recognition and do not impact the general environment. But an interesting case comes about when a label for a class of organizations gains widespread acceptance, but where audiences do not develop a schema for the label. In this case, the label becomes well known, is adopted by other organizations, discussed by analysts and the press, and often it holds promise to revolutionize the domain in some way. The fact that the label is catching on may influence new organizations to affiliate with it, as in the diffusion of a fad (Strang and Soule 1998), and there may be a few prominent businesses that are associated with the label that have had great success. An example of this is the label for e-business in the late 1990's, which was widely promoted in the press and adopted by other organizations, but never developed a meaning outside of its actual definition, which was doing business electronically. As the label became part of the public vernacular, businesses increasingly began to claim to be "e-businesses," and that affiliation provided legitimacy from customers and financiers. The widespread use of a label also has second order effects: a person might not have a schema for a particular label, but if he has read articles about it and heard of other organizations claiming the affiliation, he might assume that a schema for the label exists, even if he is not aware of it. If he is a manager at a business, he might be more likely to listen to a sales

pitch or provide funding for a company because of the claimed label affiliation. These labels are not especially constraining, but do confer legitimacy on affiliated members, and are important parts of the classification structure.

In this study, I characterize categories and labels by the extent to which they constrain their members, and to simplify nomenclature I refer to these as categories that have or lack constraint. I then study new category creation by investigating when organizations attempt to use a new label for a category. Rarely are categories schematized in their initial inception, so these new "categories" may be better termed labels that are seeds for new categories. Again, to simplify nomenclature, I refer to these labels as new categories.

Constraint and Knowledge difference: Influence on New Category Emergence

The extent to which existing categories provide or lack constraint will influence how a novel invention will be received. Classification systems make it easier for people to navigate the organizational world, but on the other hand they create expectations about what types of activities are appropriate within the category boundaries. In order to derive benefits associated with membership in a category, organizations must comply with these expectations. When a category is clearly defined and schematized, there is a widespread consensus of what a member should or should not do. This type of category presents constraint; if an organization is to continue to be associated with such a category, its activities will be restricted. For an organization in a constraining category, the creation of very different types of knowledge could violate expectations and push an organization out of a category. Categories which are constraining also create contrast within the environment. By clearly defining what a member ought to do, relevant observers can infer that other activities are distinct from existing structures. For an organization that is different in knowledge space, if it attempts to create a new category,

environmental contrast will help audiences recognize this difference in market space. In this way, for organizations creating new types of invention, categories that are well defined exert constraint that *pushes* the organization into a new category, and they create contrast which can *pull* the organization into a new category. Therefore I expect that when an organization has both knowledge difference, and is in a constraining category, that it will be more likely to create a new category.

On the flip side, when there is not a widespread consensus about what a member of a category should or should not do, members are less constrained. They will be able to engage in a wide range of activities, including knowledge development, and will still be accepted as a member of the category. In addition, when categories have a low degree of constraint, because there is not a clearly accepted notion of what a member should do, there is also less contrast within the environment, which makes it less likely that a new label will be distinguishable from existing structures. Therefore I do not expect that organizations that are very different in knowledge space will be more likely to create a new market label if they are affiliated with categories that lack constraint.

> Hypothesis 1: When organizations are both different from others in knowledge space, and are in constraining categories or labels, they are more likely to create a new market category or label.

Software Industry

I investigate this hypothesis within the empirical context of the software industry. The software industry has generally been elusive to researchers; it developed under the public radar, was shaped by many independent vendors, and its products are intangible. Software is difficult

to classify, and software companies have traditionally taken the lead in creating and validating labels and categories. Industry analysts take on the nomenclature a little bit later, and financial markets have yet to catch up.[2] This industry provides a lot of variation in the constraint exhibited by organizational categories. In addition, software companies were extraordinarily innovative, with vendors across the industry creating new and original products.

The software industry was not referred to as such until the late 1960's but software has been around since just after computers were commercialized, in the 1950's. The first software programs were custom coded in machine-based languages for general-purpose computers. In 1957 FORTRAN, the first higher-level programming language, was created, allowing programmers to code software to run on many different machines. The stock boom of the late 1960's, or the "go-go years," gave rise to thousands of entrants into the software industry, but most of these have not been documented. In 1968, IBM announced that it was unbundling its hardware and software, providing an opportunity for independent vendors in the software industry (Steinmueller 1995; Campbell-Kelly 2003).

In the 1970's the software industry had slow growth and few new entrants, and it was in this decade that classification system for software companies began to emerge. Although the general business press did not pay much attention to software in the 1970's, some providers published industry reports that were available (but expensive). These were INPUT, Aurbach, IDC, Frost & Sullivan, International Resource Development, Business Communications, Data Pro, and Data Decisions, and these analysts helped create initial industry classifications. The main division was between system software and application software, and within applications, software was defined either by industry, or as a cross-industry application. But beneath this

[2] Software stocks are currently divided into only five sectors: application software, business software & services, internet software & services, security software and services, and technical and system software.

classification system was a fragmented industry. Although there were a few dominant software companies, they left room for hundreds of specialized firms who created new niche markets (Campbell-Kelly 2003).

With the rise of the personal computer in the 1980's, the software industry again began to grow, but remained fragmented. According to the US Department of Commerce's landmark 1984 study, *A Competitive Assessment of the United States' Software Industry*, the top 3% of firms accounted for 59% of revenues, the next 23% for 30% of revenues, and the bottom 75% accounted for 21% of all revenues. The popular press and most software histories tend to focus on one organization – Microsoft – but the software industry has not been dominated by one or even a few firms. Rather, it has been shaped by many independent, small-scale vendors. Paradoxically, the fact that there are so many independent software vendors that are hard to track, may have contributed to the historical focus on a few dominant firms, because data on these were accessible, and documenting the entire industry was difficult (Campbell-Kelly 2003).

Despite the growth of the software industry throughout the 1980's, it was not discussed in the business press. From 1966 until 1980, *Businessweek* did not publish one article about the software industry, and thereafter the next article on the industry was published in 1984. Meanwhile, the software industry continued to develop. By the end of 1983, there were about 35,000 PC software products offered by about 3,000 vendors. VisiCalc, the "killer app" that is often attributed with unleashing the PC revolution, became available in 1979, and diffused quickly in the early 1980's. In 1982 the term "productivity application" originated, which referred to the most commonly used personal computer applications, such as the spreadsheet, word processor, and personal database. Later other applications were added, such as drawing packages, scheduling programs, and communications software (Campbell-Kelly 2003).

The software industry progressed through innovation, but this also developed under the official radar. Perhaps because software was so innovative, outsiders had a difficult time recognizing these innovations for a number of years. In *Gordon v. Benson* in 1972, the courts upheld that software programs were merely algorithms that could not be patented. This decision was mostly overturned in *Diamond v. Diehr* in 1981, which decided that a software program could be patented if it was embedded within an apparatus. The debate about whether software should be patentable raged among officials and researchers, but in the mean time software companies continued to patent in large numbers, receiving patent approval for inventions that were questionable at the time, such as pure data structures, methods for performing calculations in a data processor, data compression algorithms, and the like. After a series of cases that increasingly supported the patentability of software, in 1998 the last barrier to patenting pure software was overturned (Cohen and Lemley 2001). Despite the official discussion however, as Cohen and Lemley (2001) note in the *California Law Review*, the approval of software patents was a routine practice long before the courts recognized it. Ironically, software patents may even have been granted too broadly, because the PTO did not hire software experts who could adequately evaluate the patents. After the 1981 case, software was cleverly patented as part of various apparatuses, and so the PTO did not see the expected surge of software patenting after the official ruling in 1998 (Cohen and Lemley 2001).

Software development in the early to mid 1980's created the fundamentals for software as we know it. Around 1982, a consensus emerged that the best way for a computer to multitask was to develop a windowing system, and soon after several firms developed windowing systems for the IBM-compatible PC released from 1984-1985, including VisiCorp, Digital Research, Microsoft, IBM, and Quarterdeck (Campbell-Kelly 2003). Important productivity applications

in the 1980's included spreadsheets and word processing. Growing hardware markets for hard disks, display monitors, modems, and printers gave rise to software markets for utility software, such as products that managed the growing number of files, created more sophisticated graphical displays, and provided communication programs. Improvements in printing led to desktop publishing. By the end of the 1980's, software vendors offered thousands of programs for specialized applications, and dozens more for general purpose applications (Steinmueller 1995).

In the early to mid-1980's, with the initial acceptance of software patents, discussion of software in the popular business press, and the rise of the personal computer, the software industry was emerging on the public radar, and by 1990, the software industry had matured. Electronic documentation of press releases – one of the main forums where software companies described themselves and shaped the industry – are available after 1985, and press releases began to be used across the industry around 1990. Press releases allow me to track the activities of the many small and elusive vendors that have historically made the industry difficult to trace. Therefore I focus this research on the modern software industry, from 1990 – 2002.

An important advance in information technology, the linking of computers in networks, emerged in the late 1980's and spurred software applications for electronic mail, file transfer, and workgroup applications in the early 1990's (Steinmueller 1995). The mid-1990's saw a consolidation at the top of the software industry. According to *Software Magazine*, in 1994 the top 10 of their top 100 software companies accounted for 63% of the revenues of the top 100 (not of the entire industry). Still, this did not slow the rest of the industry; in the same review 23 companies achieve more than 50% growth over the previous year (Bucken 1995). In the mid 1990's companies operated in a number of software categories, including relational databases, Network management tools, ERP, security software, object management software, networking

applications, middleware, financial applications, human resource management, CAD, and integrated voice response systems (Frye and Melewski 1995).

In the mid- to late- 1990's the World Wide Web brought another opportunity for software vendors. Existing companies shifted their focus to creating client/server products that could be used over the World Wide Web (Geppert 1998). The World Wide Web seemed to emerge suddenly, but allowed for an easier and more scalable answer to client/server computing that software companies scrambled to provide. At MicroStrategy, Inc., a business intelligence software company, it was a recent hire out of college who suggested creating a web version of their client/server product; this was the first the executives had even heard of the World Wide Web. They implemented his suggestion, and in 1996 released "DSS Web," which allowed their clients to access their software over the web.[3] The boom of the late 1990's fueled the growth of software companies. They began to focus on data mining, OLAP (On-Line Analytical Processing), and object-oriented programming (Comerford 1998). Less constraining categories emerged including Customer Relationship Management (CRM), e-commerce software, and e-business software; companies affiliated with these created a wide range of different types of software (Hayes 2000).

The history of the software industry is unique and complex. Its technologies did not fit into standard ways of measuring innovation, and the importance of thousands of small independent vendors did not conform to standard ways of measuring industries. As a result, it was overlooked by mainstream business for many years, and much research on the industry has not been comprehensive, opting to focus on a few prominent firms rather than try and understand the industry in its entirety. Nevertheless, independent software firms continued to innovate and

[3] From personal communications during my employment with MicroStrategy. Executives used to recount the company's history.

create their own organizational and product distinctions, supporting a technical infrastructure that became so important to society that it could no longer be ignored. Innovation in this industry was important and categorization was organic, providing a good context to study the relationship between innovation and category creation.

Data and Measures

To test the above hypotheses, I use data on software innovation and the software industry's categories and labels, from 1990 – 2002. I create knowledge space using patent data from the U.S. Patent Office. To identify software organizations, I scan through all software press releases issued from 1990 – 2002. As discussed above, the software industry contained many small, independent vendors, who may not have been documented in official data but who contributed to shaping the industry. In the modern software era, software companies used press releases to announce new products, customers, partners, patents issued, and financing. Press releases were an important public face for software companies, and they are not especially difficult or costly to produce, so this data source captures a wide range of small and young organizations that are otherwise difficult to track. Further, within the press releases, companies classify themselves into categories or labels. Below, I describe in detail the data I use to study innovation and classification in the software industry.

Innovation: Knowledge Space

Building on previous studies (Podolny, Stuart and Hannan 1996; Pontikes and Barnett 2007; Pontikes 2007), I create knowledge space for software innovation, using patent and patent citation data from the United States Patent Office, which is associated with the National Bureau of Economic Research patent data project (Hall, Jaffe and Trajtenberg 2001). The U.S. patent

and trademark office issues patents for inventions that are new, unique and non-obvious. All patents cite relevant 'prior art' on which the new invention is based, and this indicates knowledge foundation for the patent at hand. Two patents that cite the same patent as prior art are more similar in knowledge space than two patents that have no common citations.

For patents to be defensible, they must be specific and accurate, and the patent office is active in requiring that inventors' claims be focused and narrow. Inventors include patents they are aware of in a patent's citations, and the patent examiner will also add citations to ensure that prior art is comprehensively cited. In some studies, it is important that the prior art accurately reflect the knowledge that the inventor is actively aware of, and so citations added by the examiner are problematic. Here, I am interested in locating software organizations in knowledge space in order to determine if their innovations are especially different, or if they are similar to other inventions created within the industry. Patent citations locate an innovation within knowledge space, and so citations added by an examiner are not only unproblematic, but they actually help refine the patents' positions.

Patents are assigned to a class and a subclass when they are issued. There are about 400 classes, so to make the database more tractable, the National Bureau of Economic Research has created a higher level classification system of 36 subcategories and six categories: Chemical, Computers & Communications, Drugs and Medical, Electrical and Electronics, and Other (Hall, Jaffe and Trajtenberg 2001). I use all patents granted in the Computers & Communications category to construct knowledge space for the software industry. This is a broad classification that contains knowledge relevant to software. Note that this includes patents of software organizations, as well as patents issued to individual inventors, universities, and non-software

organizations. Therefore this creates a knowledge space that is distinct from the software business market.

I create knowledge space for each year from 1990 – 2002 using a five-year window of all patents applied for in current year and four years prior. Using citation overlap, I measure the similarities between all patents applied for in the five-year window, and link them together, as illustrated in figure 2, which depicts a simplified knowledge space comprised of ten patents. The dots represent patents in the relevant categories, and the lines indicate that two patents are similar, marked by their similarity coefficients. In this example, patent 1 has a first order similarity to patents 2, 3, and 4, and second order similarities to patents 5, 6, 7, 9, and 10. Once knowledge space is constructed, I look for patents that were created by software organizations as identified from the press releases, and can place these organizations in knowledge space.

To measure similarity between patents, I take every patent in the five-year window, and compare it to every other patent in these data. Following Podolny et al (1996), I measure the number of overlapping citations between these patents, and normalize by the number of citations made by the focal patent:

$$\alpha_{ij} = \frac{s_{ij}}{s_i} \qquad (1)$$

Where s_{ij} is the number of shared citations between patent i and j, and s_i is the total number of citations by patent i. $\alpha_{ij} \in [0,1]$, and is the first-order similarity between patent i and all j's with which patent i shares at least one citation. Using this measure, I construct knowledge space at the patent level for each year from 1990 – 2002. Figure 3 shows knowledge space for selected

years, using a contour graph to show depth in areas where there is heavy patenting.[4] In 1992 there are four clusters of active patenting, which are relatively spread out. In 1995, these clusters have come together toward the center, and there is increased research activity on the periphery. By 1998, two of the existing clusters are starting to move closer together, and new areas of research are emerging. These figures represent the landscape of knowledge that software organizations research within.

With knowledge space constructed, I compute second order similarities by looking at all patent k's that each patent j is similar to (for each j such that $\alpha_{ij} > 0$). The similarity measurement between any patent i and patent k is constructed by:

$$\sigma_{il} = \max_{i,j,k} \{\alpha_{ij} \cdot \alpha_{jk}\} \qquad (2)$$

This measure multiplies the weighted citation overlaps for first and second order links. For example in figure 2, patent 1 has a similarity of 0.5 to patent 2, and has a second order similarity of 0.9*0.6=0.54 to patent 7.

Next, I use first and second order similarities to measure how different an organization's patenting is, as compared to all other patents, all other software companies, and all software companies outside its category. To do this, I take every patent applied for by the focal organization in the current year, and compute its similarity to every other patent in knowledge space within the five-year window. This creates a set of all patents to which the focal patent has non-zero similarity. These similar patents may have been issued by other software companies that are also in the same category as the focal company, but other software companies that are in different categories, or by non-software organizations. To measure how different an organization is in knowledge space with respect to these all other patents, I compute a knowledge

[4] I would like to thank James Moody for the program to create contour graphs from network data.

difference metric. To do this, I first compute how *similar* each patent i belonging to organization A is to patents belonging to all other patents, C:

$$sim_{i,C} = \underset{i \in A, l \in C}{sum}(\sigma_{il}), \qquad sim \in [0, \infty) \qquad (3)$$

I then compute how *different* organization A is from C by summing the inverse of $sim_{i,C}$, plus 1, to bound the measure, and normalizing by the number of A's patents:

$$diff_{A,C} = \sum_{i \in A} \frac{1}{1 + sim_{i,C}} \Big/ npat_A, \qquad diff \in [0,1] \qquad (4)$$

By constructing this measure in this way, I bound the difference measure between 0 and 1. If none of A's patents have non-zero similarity to group C, then $diff = 1$. As A's patents are increasingly similar to group C, $diff \to 0$.

Categorization

I create a data set of software companies, labels, and categories using press releases. Software companies actively issue press releases to distribute news, and they feature these releases on their Web sites. Within each press release, a company will claim an affiliation with a category or label. Figure 4 lists some example classifications. I use press releases issued from 1990 through 2002 to first identify software companies, and then to identify the categories and diffuse labels with which these organizations identify.

-- Insert Figure 5 here --

Businesswire, *PR Newswire*, and *Computerwire* are the three publications that software companies use to release news, and they are available in electronic format for the time period of this study. Organizations in other industries besides software also use these publications, and so

I filter on press releases that mention "software" at least three times. For my raw data, I collected every press release in these three publications with at lease three mentions of "software" released from 1990 – 2002. There are 268,963 of these. At least once in a press release, companies will refer to themselves by their full name, such as Oracle, Corp. To scrape the names of software organizations from the press releases, I wrote a program to automatically pull out words before Inc, Corp, Co, LLP, or capitalized Software. These rules cast a very broad net and returned both the names of software companies and extra "junk," like sentences or phrases. The initial output contained over 300,000 rows of potential Software firms. To filter out the "junk," I ran a series of cleaning steps, resulting in a list of 11,390 phrases that were potential software company names. I then ran manual searches through the raw press release data for each of these to determine if the name represented a software company. This resulted in 5063 potential software firms.

Next, I coded a program to automatically search through the press releases for the sentences where organizations make claims to be affiliated with specific categories and labels. Software companies use fairly standardized language when they describe themselves, stating that the company "is a leader in" a category or label, "is a provider of" the category or label, "develops" the category or label, etc. I created a program to search through the press release data for the each software company followed by these keywords, and to extract the press release date and the descriptive sentence for each company.

In the next step, I created a set of terms that identify categories and diffuse labels used to classify organizations during this time period, using both external sources and the descriptive sentences extracted from the press releases. First, I compiled a list of categories and labels that are used to classify companies in industry publications *Software Magazine* and *Computerworld*.

Then, I manually examined the press releases and compiled a list of popular labels. In total, I created a list of 479 terms identifying categories or diffuse labels. I then matched each firm with a category or label over time. Due to the nature of these data, there is a possibility that an organization might mention a category or diffuse label in a different context, and so it is possible that this first match would cast too wide a net in terms of including organizations in categories. To minimize this, I only include organization-category entries that are mentioned in multiple years. This means that organizations which exist for less than one year are not included in these data. Nevertheless, this data set contains a sample of organizations and categories within the software industry that is much more inclusive than any alternative data set of which I am aware. The final data include 3,381 distinct software organizations and 393 distinct categories and labels, over 14,380 organization-years, and 3115 category-years. Figure 6 shows the number of software firms over time, and figure 7 shows the number of categories and labels over time.

--- Insert Figures 6 and 7 here ---

Next, I measure how constraining the classification is, to determine the extent to which it is a category or a label. Recall from the discussion above that organizational categories emerge when audiences develop a common schema for what type of organization belongs in a category. To measure the extent to which a category lacks constraint, I measure whether organizations are affiliated with multiple categories or labels. The extent to which members identify with emerging labels is critical to whether a category develops and is recognized by outsiders (DiMaggio 1987; Rao, Monin and Durand 2003; Hannan, Pólos and Carroll 2007). Previous research shows that when organizations are members of multiple different categories, they suffer from lower performance (Hsu, Hannan and Kocak 2007), and when members of an emerging

categories also identify with other groups, the category is less likely to become taken for granted (McKendrick and Carroll 2001; Bogaert, Boone and Carroll 2006).

In this analysis I take into account whether organizations are dedicated members of a particular category, or whether they are partial members, in order to measure the extent to which a category or label constrains its members. I build on Hannan, Pólos and Carroll (2007)'s formalization of category emergence, where organizations vary in the degree to which they identify as members of a category or label. I assign organizations a full or partial grade of membership depending on the number of times they self-identify with the respective category in their press releases, divided by the number of times they identify with any category, and I weight counts of the number of members of a given category by these grades of membership. A category's weighted membership is:

$$wmemb_C = \sum_{i \in C} w_i \qquad (5)$$

where w_i is the degree to which an organization is a member of a category or label, $0 < w_i \leq 1$, summed over all organizations with a non-zero membership. I then can differentiate between categories where all members are dedicated – in that they belong primarily to that category or label – and those that only have partial members. Categories that have all dedicated members (where $w_i = 1$) are more sharply defined than those with all partial members. This can be captured by the contrast of the category, which divides the weighted membership by the potential membership, if all organizations were dedicated members:

$$contrast_C = \frac{wmemb_C}{\sum_{i \in C} 1} \qquad (6)$$

A high level of contrast indicates that most members of the category or label are dedicated members, which signifies that the label or category is more likely to have an agreed

upon schema that legitimates the category (Hannan, Pólos and Carroll 2007). Categories with high contrast are more constraining, which restricts members from branching out and identifying with other groups. In an environment like the software industry, there are many categories and diffuse labels that substantially overlap, but that also have a large proportion of dedicated members. The distribution of contrast in the software industry is illustrated in figure 8. The contrast of labels and categories in these data resembles a normal curve, where the majority have contrasts around 0.5, which indicates that there are both dedicated and partial members that claim to be affiliated with the category or label.

When there is a moderate or low level of contrast, this may indicate that it is not clear what it means to be a classified as that category. However, it also may mean that there are overlapping labels or categories. If most of the members of one category also identify with one other group, the category will have a medium level of contrast, but it might be highly restrictive. For instance, the labels "mobile" and "wireless" show moderate levels of contrast (between 0.3 and 0.6 over the date range of this study), but this is mainly due to the overlap between these categories. On the other hand, the label "enterprise" has a contrast of around 0.5, but this is due to overlap with 284 other categories. The "enterprise" classification is an example of a diffuse label; it was used from 1991 through 2002 and has many organizational members, with a good portion of dedicated members, and is widely recognized by customers and analysts as a type of software. However, claiming to be affiliated with the label does not strongly constrain an organization, and as a result software companies claim to be members of this and almost every other label or category. Because diffuse labels present fewer restrictions than do categories, it is more likely that many organizations that are affiliated with a diffuse label will also be affiliated with other labels, or members of other categories.

When organizations that are affiliated with a label also affiliate with a variety of other labels, this indicates that the focal label does not strongly constrain member organizations. Conversely, if all organizations that identify with a term only identify with the focal term, it is a more constraining classification. I argue that taking into account how many other categories member organizations identify with measures the extent to which the term constrains its members. Therefore, I create a measure of "permissiveness" that builds on a category's contrast (defined in (6)) and also takes into account whether members of a category or label are affiliated with many other groups.

A category's contrast is how well-defined it is with respect to its environment, so the extent to which it is not well-defined, or its "fuzziness," is one minus its contrast: $fuzz_C = 1 - contrast_C$. Fuzziness represents the extent to which members identify with at least one other group. But since high fuzziness might simply indicate that a meaningful category overlaps with another meaningful category, I multiply this by the number of *distinct* other labels or categories to which members of the focal category or label also belong, to create a "lack of constraint" metric. This measure represents the extent to which a classification term represents a category or a diffuse label.

$$lack\_const_C = fuzz_C \cdot N_{ocat} \tag{7}$$

Here, N_{ocat} is the number of distinct other labels or categories that members of label or category C also belong to. Figure 8 shows the distribution of the lack of constraint metric. It is a skewed distribution with most categories and labels at relatively low levels of permissiveness, and a long tail to high levels of permissiveness.

Figure 9 illustrates the relationship between the fuzziness of a label or category and its lack of constraint. At low levels of fuzziness lack of constraint is also low. As fuzziness

increases, the range of lack of constraint also increases, and the highest levels of lack of constraint are found for categories or labels with medium levels of fuzziness (or contrast). In the software industry, there are many labels with a relatively high proportion of "dedicated" organizational members, and also with a relatively high proportion of partial members. Of these, some have partial members that are scattered in terms of their alternate identifications, and are also identified with a large number of other categories and labels. Others are more restrictive, where partial members are only identified with a handful of other groups.

A mapping of the categories and labels of the software industry for selected years is provided in figure 10. These figures show the extensive classification system of the software industry over time. The dimensions of these figures are not meaningful, but the distance between the categories/labels indicates how strongly they overlap. Each category or label is represented by a circle, the size of which is based on its lack of constraint. There are a number of diffuse labels, with high permissiveness that overlap with many other labels and categories, clustered in the center. More constraining categories and labels are both in the center and toward the edges. In many cases, two, three or four labels or categories form an isolated cluster, indicating high overlap, but high constraint. These figures illustrate the complexity of classification in the software industry over time. There are few categories where boundaries are absolute; most that have a substantial number of members also have a non-trivial amount of overlap.

Model

To test my hypothesis, I model how a firm's knowledge difference, and knowledge difference interacted with lack of category constraint, affects its likelihood to start using a new

category label. I model this at the firm level, with entry into a new category (defined here as entry into a category in its first or second year of existence) as the event. This rate can be operationalized as:

$$r_d(t) = \lim_{t \to t'} \frac{\Pr(t \le T < t' | T \ge t)}{(t' - t)} \qquad (10)$$

I model this rate as a function of characteristics of the firm, the categories of which the firm is a member, and environmental variables:

$$r_d(t) = r_d(t) \cdot \exp(\alpha_f \cdot \mathbf{x}_f + \alpha_c \cdot \mathbf{x}_c + \alpha_e \cdot \mathbf{x}_e) \qquad (11)$$

I use piecewise continuous hazard rate models to estimate this model.

Results

Descriptives of the data are contained in Table 1. Results of the model of new category creation are listed in Table 2.

--- Insert Tables 1 and 2 here---

Models 1-4 are piecewise continuous hazard rate models on firm entry into a new category, in the first or second year of the category's existence. In these data, organizations that "enter" categories upon their inception are the pioneers who are attempting to use a new label for market classification, or who are creating a new category. Models 2 and 3 test hypothesis 1. Model 2 shows that when the knowledge difference of a firm is added alone, it does not have a significant effect on the likelihood of a firm to start a new category. However, when we include the interaction of the firm's (knowledge difference) x (lack of category constraint), Model 3 shows that when this interaction is included, the knowledge difference of the firm is positive and significant at the $p<0.05$ level, and the interaction is negative and significant at the $p<0.05$ level. Model 3 also is an improvement in fit over model 1, for two degrees of freedom, at the $p<0.05$

level. This provides support for hypothesis 1: when firms are in constraining categories, knowledge difference increases the likelihood that a firm will start a new category, but when firms are in categories that lack constraint, this effect is diminished.

Figure 11 plots the effects of both knowledge difference and lack of category constraint on the multiplier of the entry rate into new categories. When firms are in categories where *lack* of constraint is low (with high levels of constraint), increasing levels of knowledge difference increase the multiplier of the rate of a firm's likelihood to create a new category. However, when firms are in categories where *lack* of constraint is high (constraint is low), knowledge difference does not increase the rate at which firm's create new categories. Model 3 also shows that there is a main effect of lack of constraint, where the less constraining the category in general, the more likely any organization is to move into a new category. This is also the case for a firm's propensity to move into any category (discussed in detail below), and indicates that lack of category constraint gives organizations more room to explore market space, without incurring penalties. A lincom test indicates that the combined effect of lack of constraint, and the interaction (knowledge difference) x (lack of constraint) is positive, and the combined effect of knowledge difference, and the interaction is positive. The negative effect of the interaction does not reduce the rate of entry into new categories, but it tempers the positive effect on rate of entry resulting from knowledge space difference, as illustrated in figure 11.

Table 3 shows results of models on the likelihood that a firm enters an existing category (a category that has existed for more than two years).

-- Insert table 3 here --

These models provide a base level comparison, to investigate whether the effects explored above have to do with the creation of new categories and a potential market space disruption, or if they

merely predict organizational movement. Models 5-6 test whether the hypothesized effects explain new category creation. Model 5 shows that knowledge difference has a negative and significant effect on the likelihood that an organization will move into an existing category, significant at the $p<0.05$ level. Model 6 includes the interaction, which does not have an effect significant at the $p<0.05$ level and which adds noise to the model. A statistical test comparing the coefficients in the two models indicates that we can reject the hypothesis that either the main effect of knowledge difference, or the effect of the interaction, are equal across the two models, significant at the $p<0.05$ level. Taken together, models 2-3 and 5-6 provide support for hypothesis 1. Firms that are different in knowledge space are more likely to create new categories when the categories they are already in are constraining. The more existing categories lack constraint, the less a firm's knowledge difference increases its likelihood to enter new categories. In addition, difference in knowledge space increases the likelihood of starting a new category, not of moving into an existing category.

Discussion

This paper to two literatures that have developed independently, each describing how new organizational categories emerge. One examines innovation as the source of new category formation, and another looks at how existing structures encourage new category formation. Both literatures provide important insights into how new categories are formed, but when innovation is separated from the categorical context, its effect on new category creation cannot be comprehensively understood. Results not only support the notion that both knowledge space differences and market structure affect new market category creation, but show that knowledge space difference interacts with market space structure to affect new category creation. When

organizations are different in knowledge space by the same amount, as measured by technical distinctions, they will be more or less likely to create a new market category depending on characteristics of existing categories. When existing categories lack constraint, the propensity for different types of knowledge to translate into different types of market structures is diminished.

I argue that this effect arises because categorical distinctions are comparative. Evaluating that something is different, or novel, depends on the context. Once a new category has emerged, looking back, the differences between that new category and existing categories are salient and seem obvious. But this is because – by definition – categorization highlights differences. It is not as easy to predict the formation of a new category when looking forward; innovations that are technologically very novel do not always inspire the formation of a new category. I propose that the existing category structure provides varying degrees of constraint and contrast, which affect our perceptions of difference. This study indicates that existing categories structure and technological newness are fundamentally linked in how we recognize novelty.

References

Aspray, W. (1997). "The Intel 4004 Microprocessor: What Constituted Invention?" IEEE Annals of the History of Computing **19**: 4-15.

Barnett, W. (2004). "From Red Vienna to the Anchluss: Ideological competition among Viennese newspapers during the rise of national socialism." American Journal of Sociology **109**: 1452-99.

Barnett, W. P. (1990). "The organizational ecology of a technological system." Administrative Science Quarterly **35**: 31-60.

Bogaert, S., C. Boone and G. Carroll (2006). "Contentious Legitimacy: Professional Association and Density Dependence in the Dutch Audit Industry 1884-1939." Stanford University Working paper.

Bucken, M. (1995). State of the Industry (the 13th Annual Software Magazine top 100). Software Magazine. **15:** 4.

Campbell-Kelly, M. (2003). From Airline Reservations to Sonic the Hedgehog: A History of the Software Industry. Cambridge, Massachusetts, The MIT Press.

Carroll, G. R. and A. Swaminathan (2000). "Why the microbrewery movement? Organizational dynamics of resource partitioning in the U.S. brewing industry." American Journal of Sociology **106**: 715-762.

Cohen, J. and M. Lemley (2001). "Patent Scope and Innovation in the Software Industry." California Law Review **89**: 1-57.

Comerford, R. (1998). "Software Engineering." IEEE Spectrum.

DiMaggio, P. J. (1987). "Classification in art." Annual Review of Sociology **52**: 440-55.

Frye, C. and D. Melewski (1995). Wide Window Beckons to Suppliers: Just Do It. Software Magazine. **15**.

Geppert, L. (1998). "Technology 1998: Analysis & Forecast." IEEE Spectrum.

Gilbert, R. (2006). Looking for Mr. Schumpeter: Where are we in the Competition-Innovation Debate? Innovation Policy and the Economy. A. Jaffe, J. Lerner and S. Stern. Cambridge, The MIT Press.

Hall, B., A. Jaffe and M. Trajtenberg (2001). The NBER patent citations data file: Lessons, insights, and methodological tools. NBER Working Paper Series. Cambridge, MA, National Bureau of Economic Research.

Hannan, M., L. Pólos and G. Carroll (2007). Logics of Organization Theory. Princeton, Princeton University Press.

Hannan, M., L. Pólos and G. Carroll (2007). Logics of Organization Theory: Audiences, Codes and Ecologies. Princeton, Princeton University Press.

Hayes, I. (2000). Application First, Delivery Second! Software Magazine. **20:** 63-64.

Hsu, G., M. T. Hannan and O. Kocak (2007). "Multiple Category memberships in Markets: A Formal Theory and Two Empirical Tests." Working paper.

Jaffe, A. (1986). "Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value." The American Economic Review **76**: 984-1001.

Levin, R., W. Cohen and D. Mowery (1985). "R&D Appropriability, Opportunity, and Market Structure: New Evidence on Some Schumpeterian Hypotheses." The American Economic Review **75**: 20-24.

Levin, R. and P. C. Reiss (1988). "Cost-Reducing and Demand-Creating R&D with Spillovers." The RAND Journal of Economics **19**: 538-556.

March, J. (1991). "Exploration and exploitation in organizational learning." Organization Science **2**: 71-87.

McKendrick, D. and G. Carroll (2001). "On the genesis of organizational forms: Evidence from the market for disk drive arrays." Organization Science **12**: 661-682.

Podolny, J., T. Stuart and M. Hannan (1996). "Networks, knowledge, and niches: Competition in the worldwide semiconductor industry, 1984-1991." American Journal of Sociology **102**: 659-689.

Pontikes, E. and W. Barnett (2007). "Technical Change Among Organizational Communities." Working Paper.

Pontikes, E. G. (2007). "Knowledge sharing or competition? How knowledge space crowding affects knowledge progress and market survival." Working Paper.

Powell, W., K. Koput and L. Smith-Doerr (1996). "Interorganizational collaboration and the locus of innovation." Administrative Science Quarterly **41**: 116-45.

Rao, H., P. Monin and R. Durand (2003). "Institutional Change in Toque Ville: Nouvelle Cuisine as an Identity Movement in French Gastronomy." American Journal of Sociology **108**: 795-843.

Robinson, A. (1976). "Automotive Electronics: Computerized Engine Control." Science **194**: 414-415.

Ruef, M. (2000). "The emergence of organizational forms: A community ecology approach." American Journal of Sociology **106**: 658-714.

Ruef, M. (2000). "The emergence of organizational forms: A community ecology approach." American Journal of Sociology **106**: 658-714.

Scherer, F. M. (1967). "Market Structure and the Employment of Scientists and Engineers." American Economic Review **57**: 524-31.

Schumpeter, J. (1934:2006). The Theory of Economic Development. New Brunswick, Transaction Publishers.

Sorensen, J. and T. Stuart (2000). "Aging, Obsolescence, and Organizational Innovation." Administrative Science Quarterly **45**: 81-112.

Steinmueller, W. (1995). The U.S. Software Industry: An Analysis and Interpretive History. The International Computer Software Industry. D. Mowery, Oxford University Press.

Strang, D. and S. Soule (1998). "Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills." Annual Review of Sociology **24**: 265-90.

Teece, D. (1986). "Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy." Research Policy **15**: 285-305.

Tushman, M. and P. Anderson (1986). "Technological Discontinuities and Organizational Environments." Administrative Science Quarterly **31**: 439-465.

White, H. (1981). "Where do markets come from?" American Journal of Sociology **87**: 517-547.

Zuckerman, E. W. (1999). "The categorical imperative: Securities analysts and the illegitimacy discount." American Journal of Sociology **104**: 1398-1438.

Zuckerman, E. W. (2000). "Focusing the corporate product: Securities analysts and de-diversification." Administrative Science Quarterly **45**: 591-619.

Tables

Table 1. Descriptives.

	N obs	Mean	St Dev	Min	Max
K-diff of firm x mean lack of constraint	14380	0.1867	1.7032	0	101
K-diff of firm x mean fuzziness	14380	0.0033	0.0211	0	0.5933
Knowledge difference of firm	14380	0.0070	0.0431	0	1
No. oth cat x lack of constraint	14380	7816.9780	9588.7150	0	45773
No. oth cat x fuzziness	14380	109.9424	79.6155	0	282
Mean lack of constraint of firm's cat	14380	24.6051	28.0328	0	129
Mean fuzziness of firm's cat	14380	0.3774	0.2407	0	0.8026
No. cat firm entered last year	14380	0.3082	0.7603	0	11
No. cat firm existed last year	14380	0.2162	0.6465	0	10
Firm patented last year dummy	14380	0.1213	0.3265	0	1
Cumulative no. patents over firm's history	14380	22.8459	326.9613	0	14706
No. other categories in software	14380	217.9573	135.9612	0	354
No. categories firm is in	14380	1.7150	1.8836	0	29
Sum density of categories firm is in	14380	64.1338	101.1087	0	1101
Tenure in software*	14380	2.4787	2.5515	0	13

Table 2. Piecewise continuous models on firm's entry into a new category[5]

	Model 1		Model 2		Model 3		Model 4	
Log pseudolikelihood	-1743.58		-1743.58		-1739.56		-1691.57	
Degrees of freedom	18		19		20		23	
Number of observations	14380		14380		14380		14380	
K-diff of firm x mean lack of constraint					-0.1696	**	-0.1488	*
					(0.0792)		(0.0841)	
K-diff of firm x mean fuzzy							1.269	
							(3.345)	
Knowledge difference of firm			0.0699		2.2799	**	1.424	
			(0.8566)		(0.7733)		(1.1327)	
No. oth cat x lack of constraint	-0.0002	**	-0.0002	**	-0.0002	**	0.0000	
	(0.0001)		(0.0001)		(0.0001)		(0.0001)	
No. oth cat x fuzzy							-0.0264	**
							(0.0031)	
Mean lack of constraint of firm's cat	0.0719	**	0.0719	**	0.0735	**	-0.0134	
	(0.0183)		(0.0183)		(0.0182)		(0.0207)	
Mean fuzziness of firm's cat							8.304	**
							(0.7858)	
No. cat firm entered last year	-0.5252	**	-0.5253	**	-0.5284	**	-0.6367	**
	(0.0942)		(0.0940)		(0.0949)		(0.1163)	
No. cat firm existed last year	-0.8278	**	-0.8280	**	-0.8132	**	-0.7620	**
	(0.0713)		(0.0712)		(0.0722)		(0.0718)	
Firm patented last year dummy	0.3243	**	0.3211	**	0.3677	**	0.2870	**
	(0.1256)		(0.1304)		(0.1342)		(0.1285)	
Cumulative no. patents over firm's history	-0.0002	**	-0.0002	**	-0.0002	**	-0.0001	**
	(0.0000)		(0.0000)		(0.0001)		(0.0000)	
No. other categories in software	0.0056	**	0.0056	**	0.0054	**	0.0092	**
	(0.0008)		(0.0008)		(0.0008)		(0.0011)	
No. categories firm is in	0.3476	**	0.3477	**	0.3458	**	0.2519	**
	(0.0375)		(0.0375)		(0.0377)		(0.0445)	
Sum density of categories firm is in	0.0014		0.0014		0.0015	*	0.0035	**
	(0.0009)		(0.0009)		(0.0009)		(0.0009)	
Tenure in software*	-0.0181		-0.0182		-0.0134		-0.0207	
	(0.0242)		(0.0243)		(0.0243)		(0.0254)	
Dummy: years 1994 - 1997	0.2775		0.2780		0.2737		0.4474	**
	(0.2312)		(0.2307)		(0.2279)		(0.1799)	
Dummy: years 1998 - 2000	0.4124	**	0.4128	**	0.4093	**	0.9385	**
	(0.1769)		(0.1766)		(0.1750)		(0.1625)	
Dummy: years 2000 - 2002	-0.6903	**	-0.6906	**	-0.6763	**	-0.4263	**
	(0.1721)		(0.1721)		(0.1702)		(0.1692)	
0-1 year since last cat entry	-4.840	**	-4.840	**	-4.854	**	-6.126	**
	(0.2381)		(0.2377)		(0.2346)		(0.2595)	
1-2 yrs since last cat entry	-5.424	**	-5.425	**	-5.438	**	-7.020	**
	(0.3171)		(0.3163)		(0.3126)		(0.3372)	
2-5 yrs since last cat entry	-5.398	**	-5.398	**	-5.416	**	-7.007	**

[5] Category in its 1st or 2nd year of existence. Firms that enter in these years create the new category on the market

	(0.3136)		(0.3130)		(0.3089)		(0.3353)	
5-10 yrs since last cat entry	-5.844	**	-5.845	**	-5.868	**	-7.379	**
	(0.4465)		(0.4457)		(0.4430)		(0.4663)	
10+ yrs since last cat entry	-4.527	**	-4.530	**	-4.674	**	-6.069	**
	(0.9700)		(0.9699)		(1.0000)		(0.9987)	

** $p<0.05$ * $p<0.10$

Results clustered on firm id

Table 3. Piecewise continuous models on firm's entry into an existing category

	Model 5		Model 6		Model 7	
Log pseudolikelihood	-4926.81		-4924.00		-4924.00	
Degrees of freedom	18		19		20	
Number of observations	14380		14380		14380	
K-diff of firm x mean lack of constraint					-0.0002	
					(0.0186)	
Knowledge difference of firm			-1.171	**	-1.167	
			(0.5242)		(0.6991)	
No. oth cat x lack of constraint	-0.0001	**	-0.0001	**	-0.0001	**
	(0.0000)		(0.0000)		(0.0000)	
Mean lack of constraint of firm's cat	0.0190	**	0.0191	**	0.0191	**
	(0.0085)		(0.0085)		(0.0085)	
No. cat firm entered last year	-0.3463	**	-0.3449	**	-0.3449	**
	(0.0486)		(0.0486)		(0.0486)	
No. cat firm existed last year	-0.6769	**	-0.6739	**	-0.6739	**
	(0.0367)		(0.0366)		(0.0366)	
Firm patented last year dummy	0.3164	**	0.3716	**	0.3716	**
	(0.0551)		(0.0588)		(0.0587)	
Cumulative no. patents over firm's history	-0.0001	**	-0.0001	**	-0.0001	**
	(0.0000)		(0.0000)		(0.0000)	
No. other categories in software	0.0078	**	0.0078	**	0.0078	**
	(0.0004)		(0.0004)		(0.0004)	
No. categories firm is in	0.2621	**	0.2608	**	0.2608	**
	(0.0193)		(0.0194)		(0.0194)	
Sum density of categories firm is in	-0.0004		-0.0004		-0.0004	
	(0.0004)		(0.0004)		(0.0004)	
Tenure in software*	0.0116		0.0123		0.0123	
	(0.0089)		(0.0089)		(0.0089)	
Dummy: years 1994 - 1997	0.6261	**	0.6191	**	0.6191	**
	(0.1108)		(0.1108)		(0.1109)	
Dummy: years 1998 - 2000	0.2694	**	0.2637	**	0.2637	**
	(0.0697)		(0.0697)		(0.0697)	
Dummy: years 2000 - 2002	-0.3910	**	-0.3852	**	-0.3852	**
	(0.0685)		(0.0687)		(0.0687)	
0-1 year since last cat entry	-3.904	**	-3.898	**	-3.898	**
	(0.1336)		(0.1335)		(0.1332)	
1-2 yrs since last cat entry	-4.192	**	-4.182	**	-4.182	**
	(0.1600)		(0.1601)		(0.1596)	
2-5 yrs since last cat entry	-4.098	**	-4.090	**	-4.090	**
	(0.1598)		(0.1598)		(0.1593)	
5-10 yrs since last cat entry	-4.125	**	-4.115	**	-4.115	**
	(0.1812)		(0.1811)		(0.1803)	
10+ yrs since last cat entry	-3.582	**	-3.538	**	-3.538	**
	(0.4003)		(0.3968)		(0.3968)	

** $p<0.05$ * $p<0.10$; results clustered on firm id

Figures

Figure 1. Knowledge space and market space

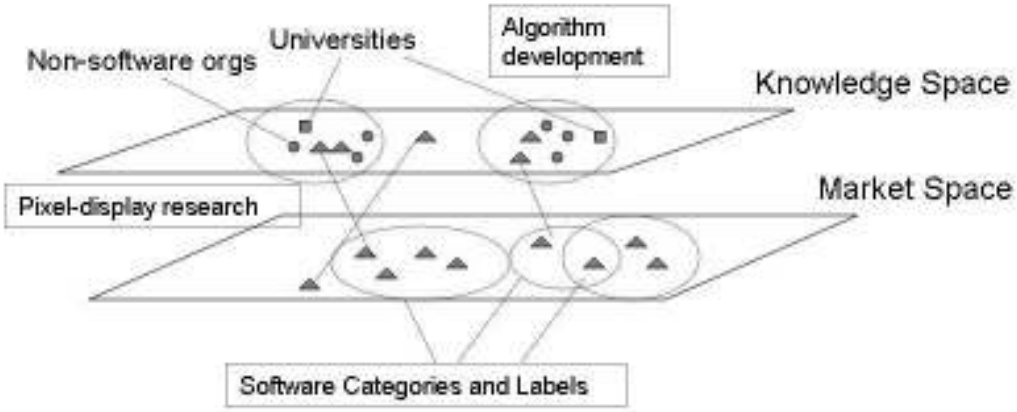


Figure 2. A simplified knowledge space with ten patents

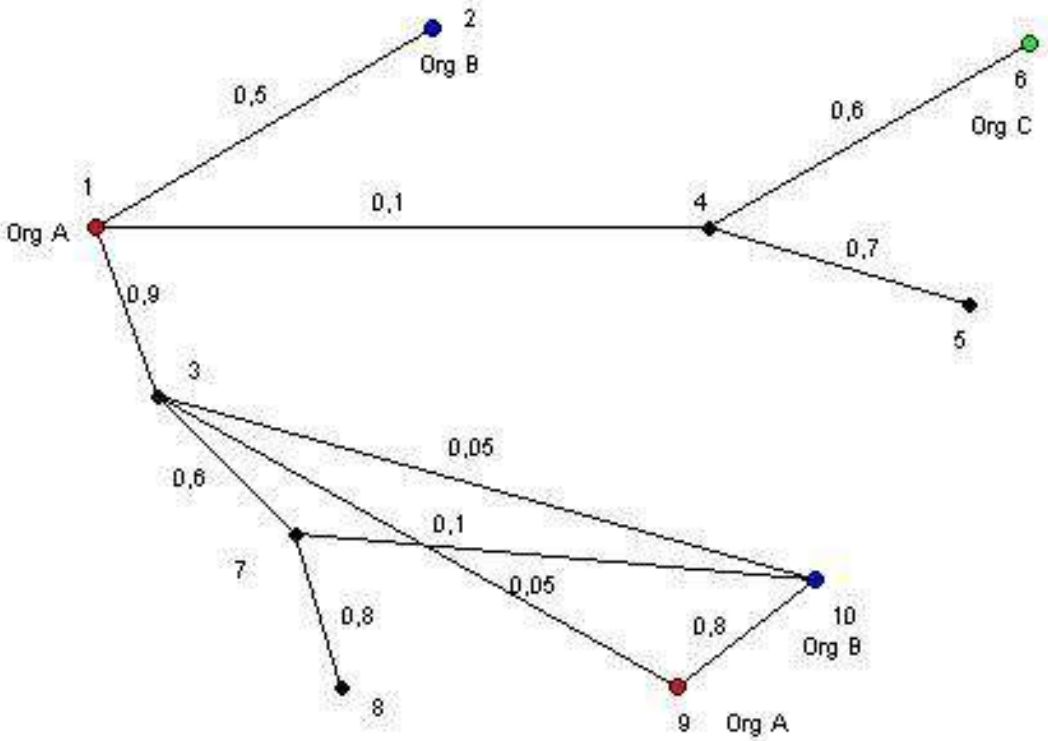
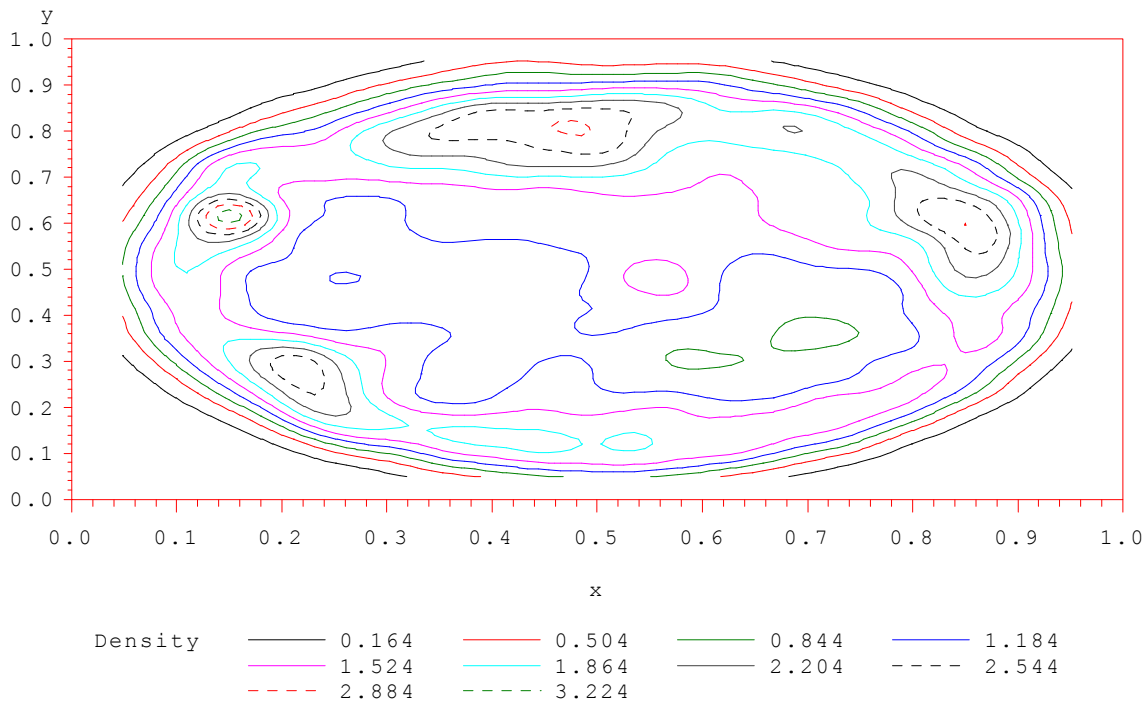
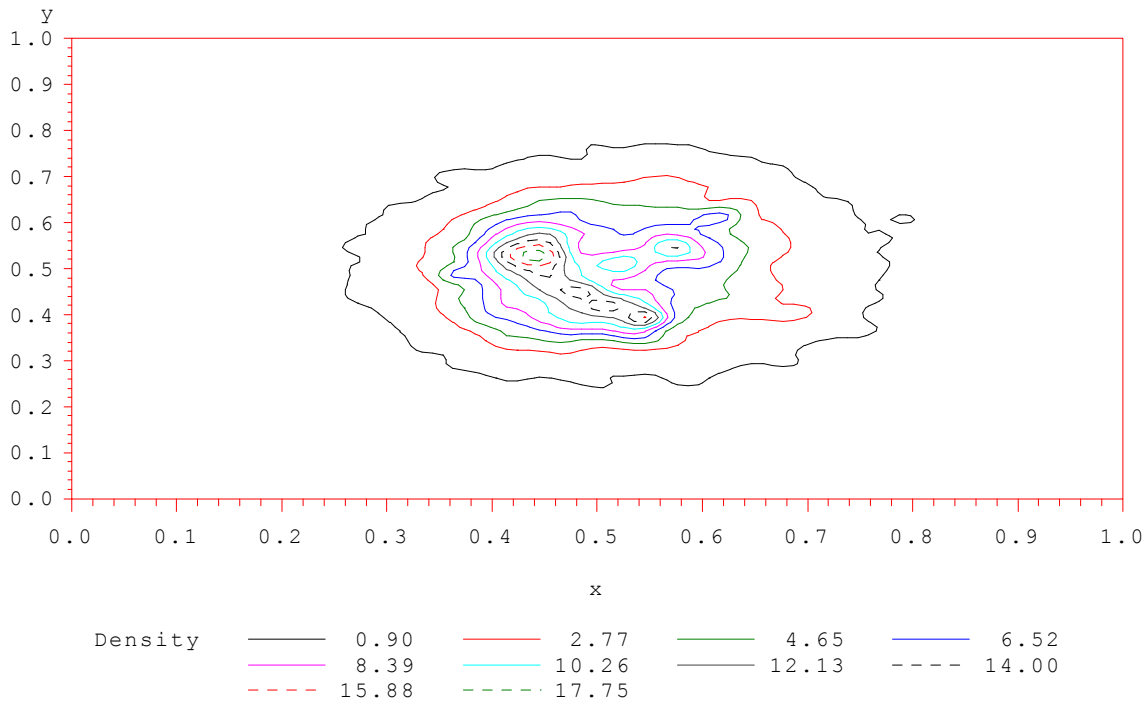


Figure 3. Knowledge space for selected years.

1992



1995



1998

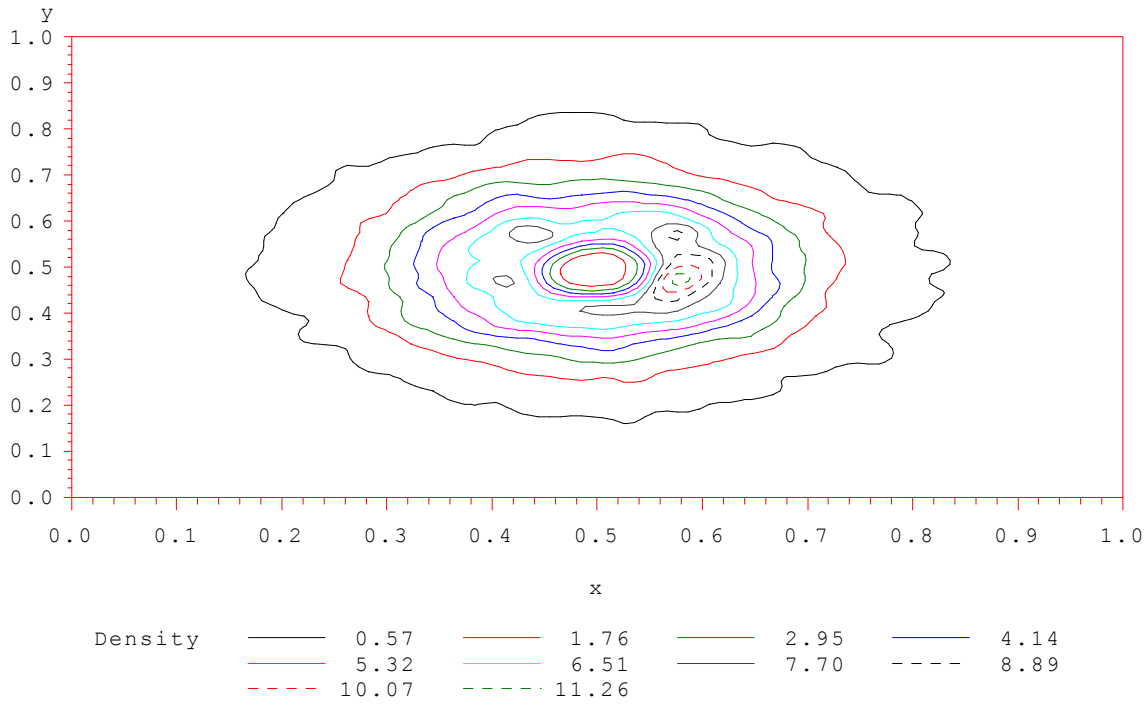


Figure 4. Example Classifications in Press Releases

Company	**Date**	**Description**
Citrix Systems	February 2000	Citrix Systems, Inc. is a global leader in application server software and services
Plasmon	August 2000	Plasmon, a leading manufacturer of automated data storage solutions, today announced its Diamond(TM) storage management software.
Watson General	May 1994	Watson General currently provides remote software monitoring systems.
Comergent Technologies	Sept 2002	Comergent Technologies(R) Inc., the leading provider of sell-side e-business software solutions
Accrue Software	October 1999	Accrue Software is a leading provider of e-business data collection and analysis software
ACP	July 2001	ACP provides enterprise web publishing and e-business solution
Broadvision	September 2000	Broadvision, a leading worldwide supplier of personalized e-business applications
Alliance	March 2001	Alliance offers the technical and business advantages of the Sybase Enterprise Portal with a wide range of e-business solutions, including content, e-commerce, and business process automation and analysis

Figure 5. Number of software organizations over time.

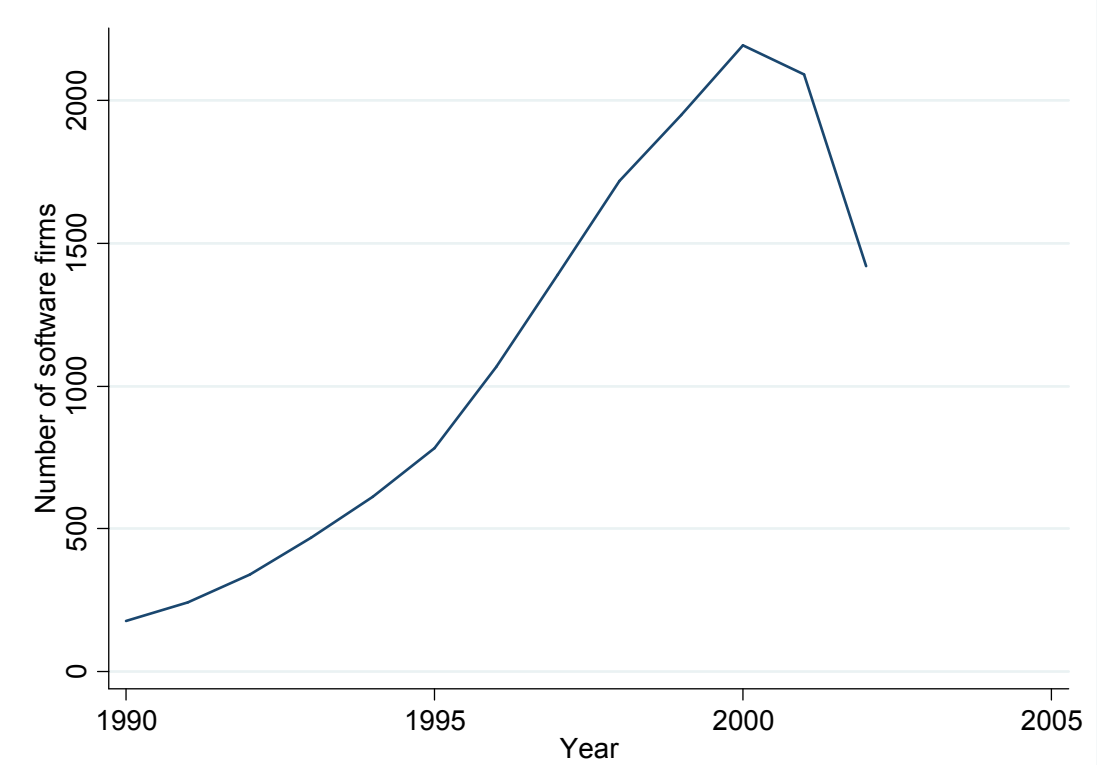


Figure 6. Number of categories or labels that classify software firms over time.

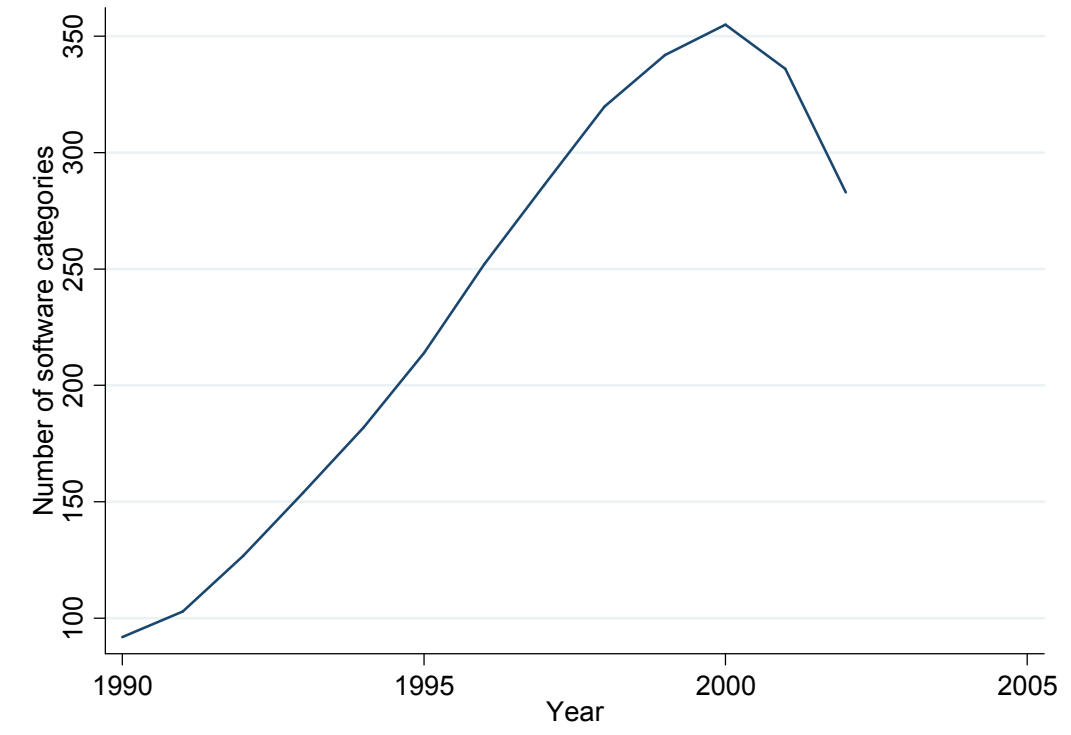


Figure 7. Frequency of fuzziness, by category-year, for labels and categories in the software industry.

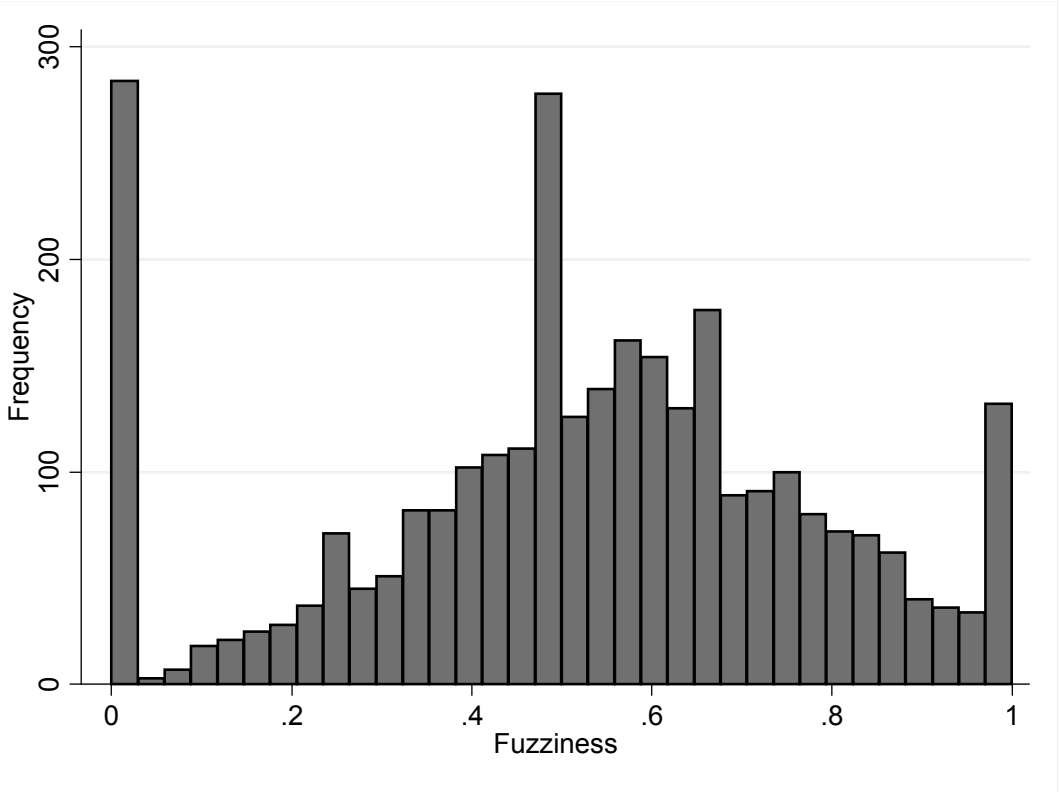


Figure 8. Frequency of lack of constraint, by category-year, for labels and categories in the software industry.

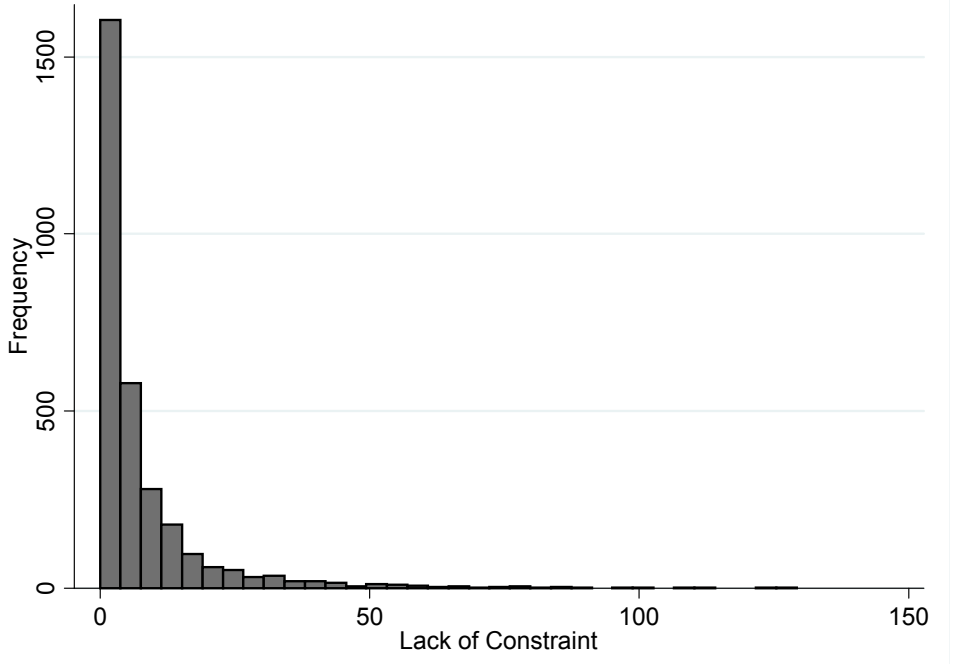


Figure 9. Relationship between category lack of constraint and fuzziness.

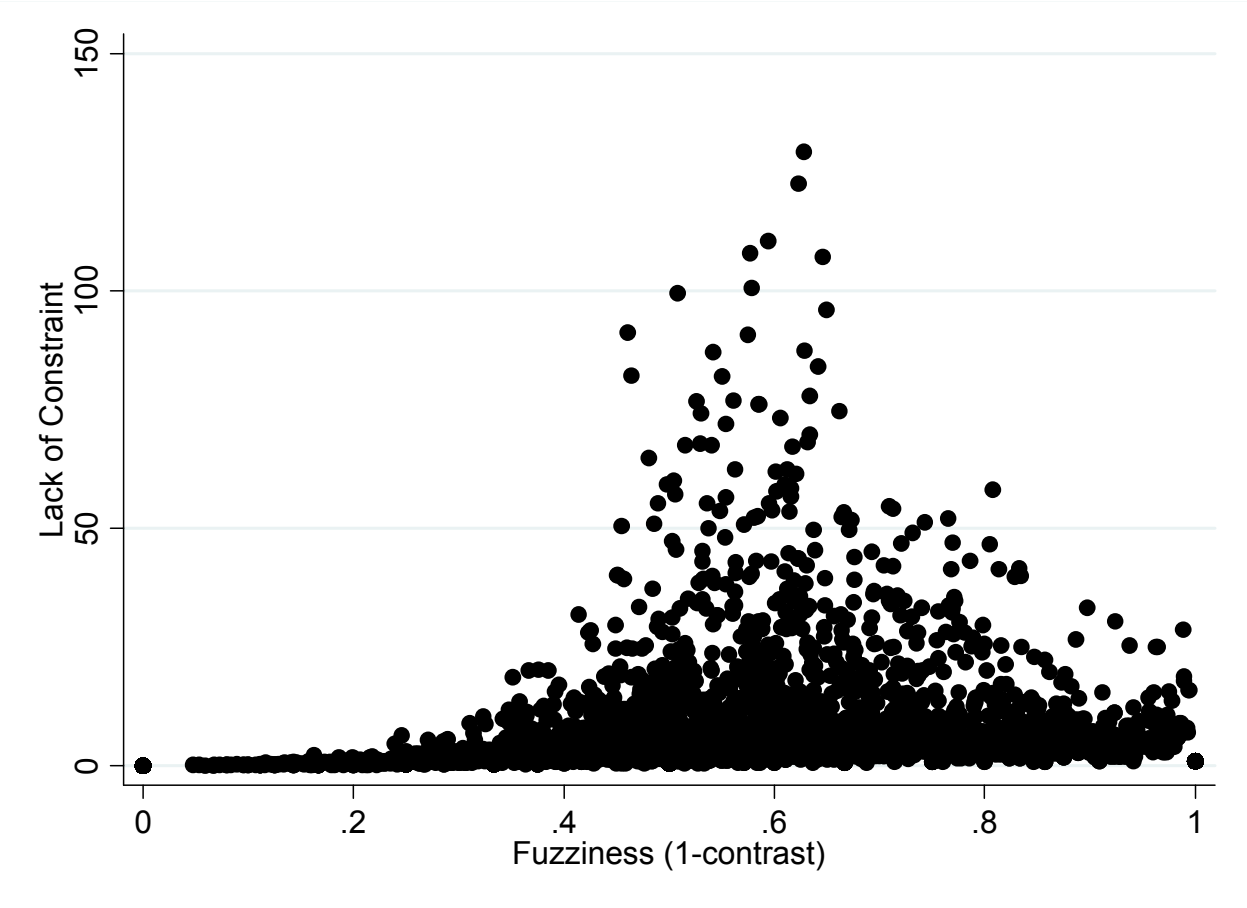
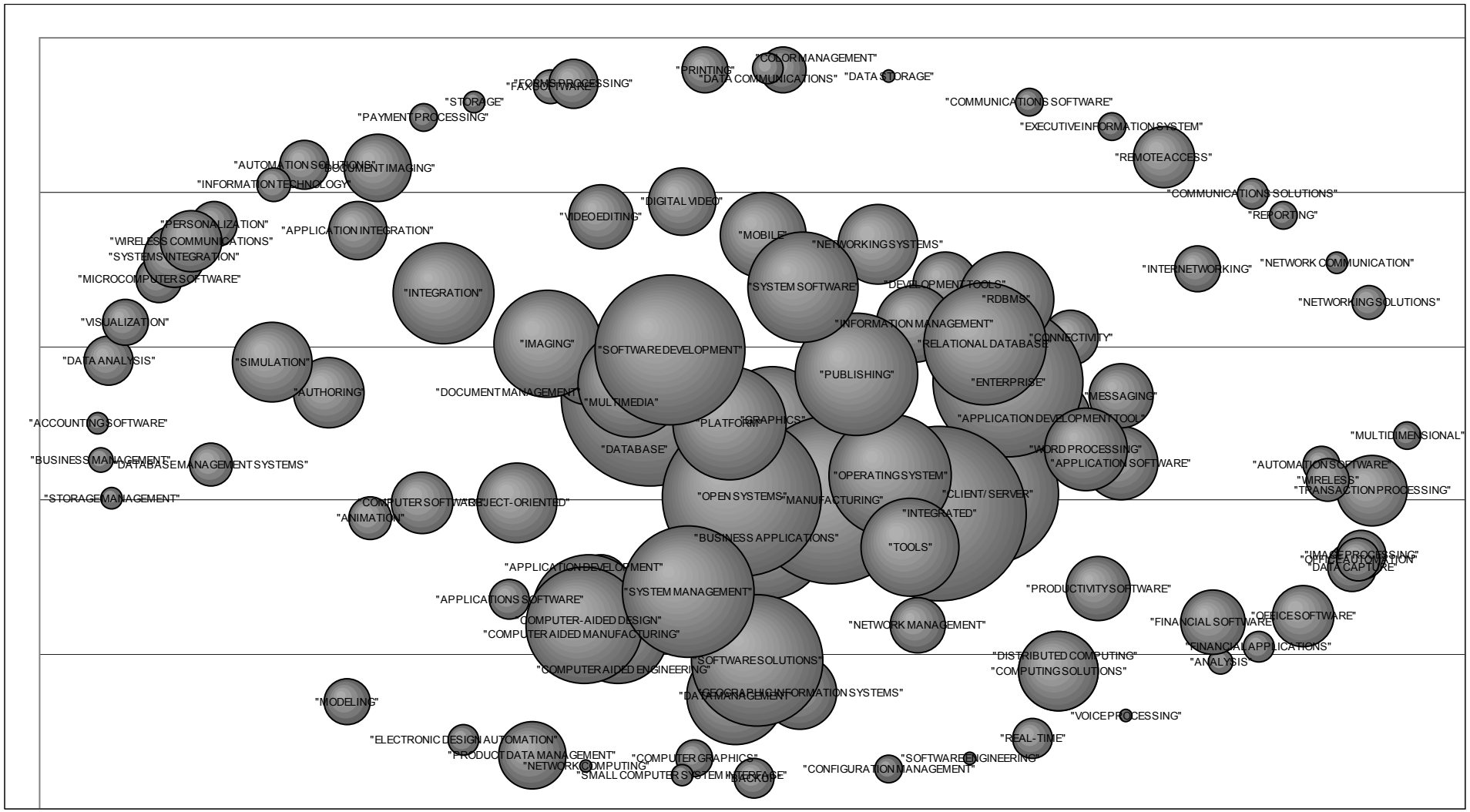
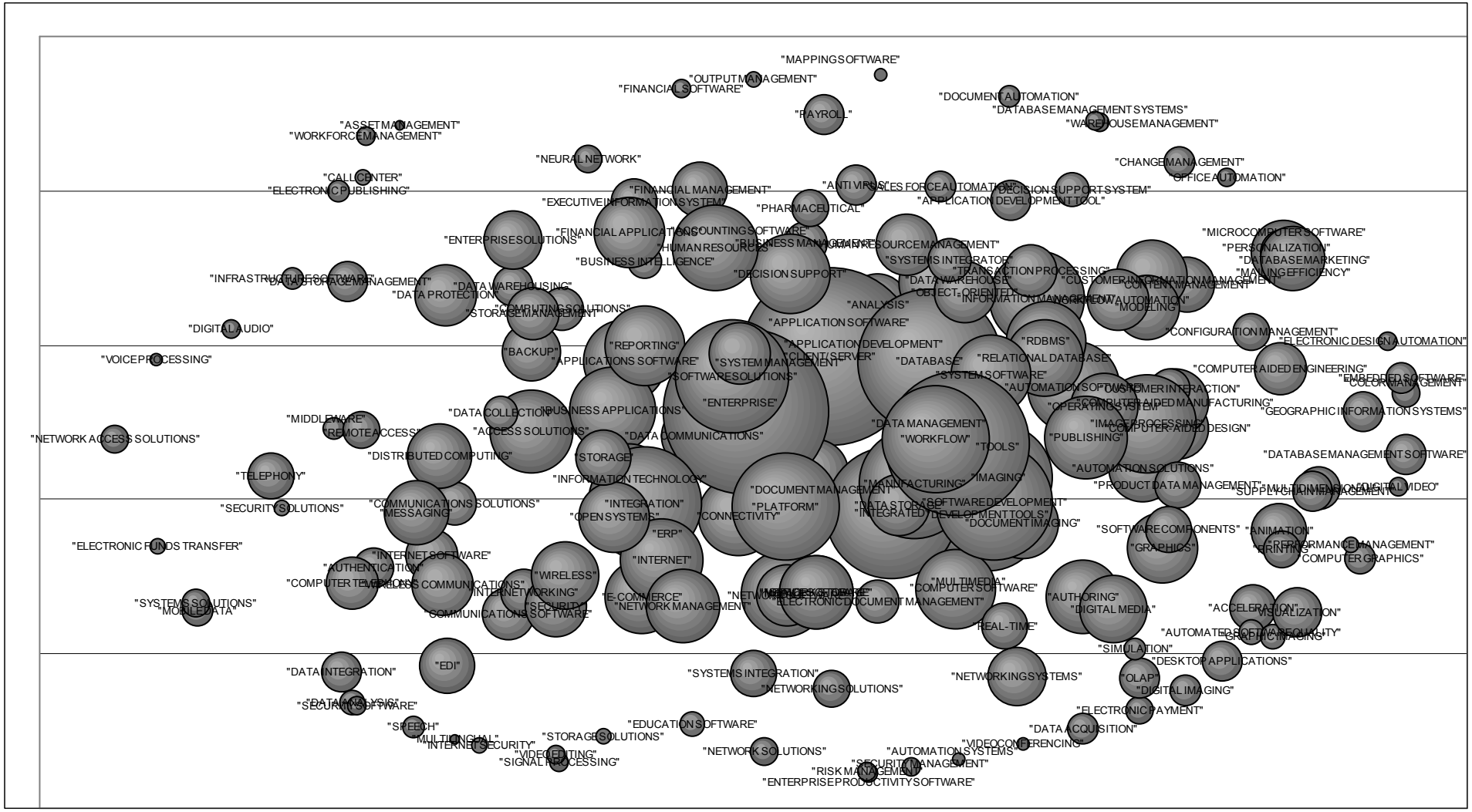


Figure 10. Categories and labels in the software industry over time, for selected years.

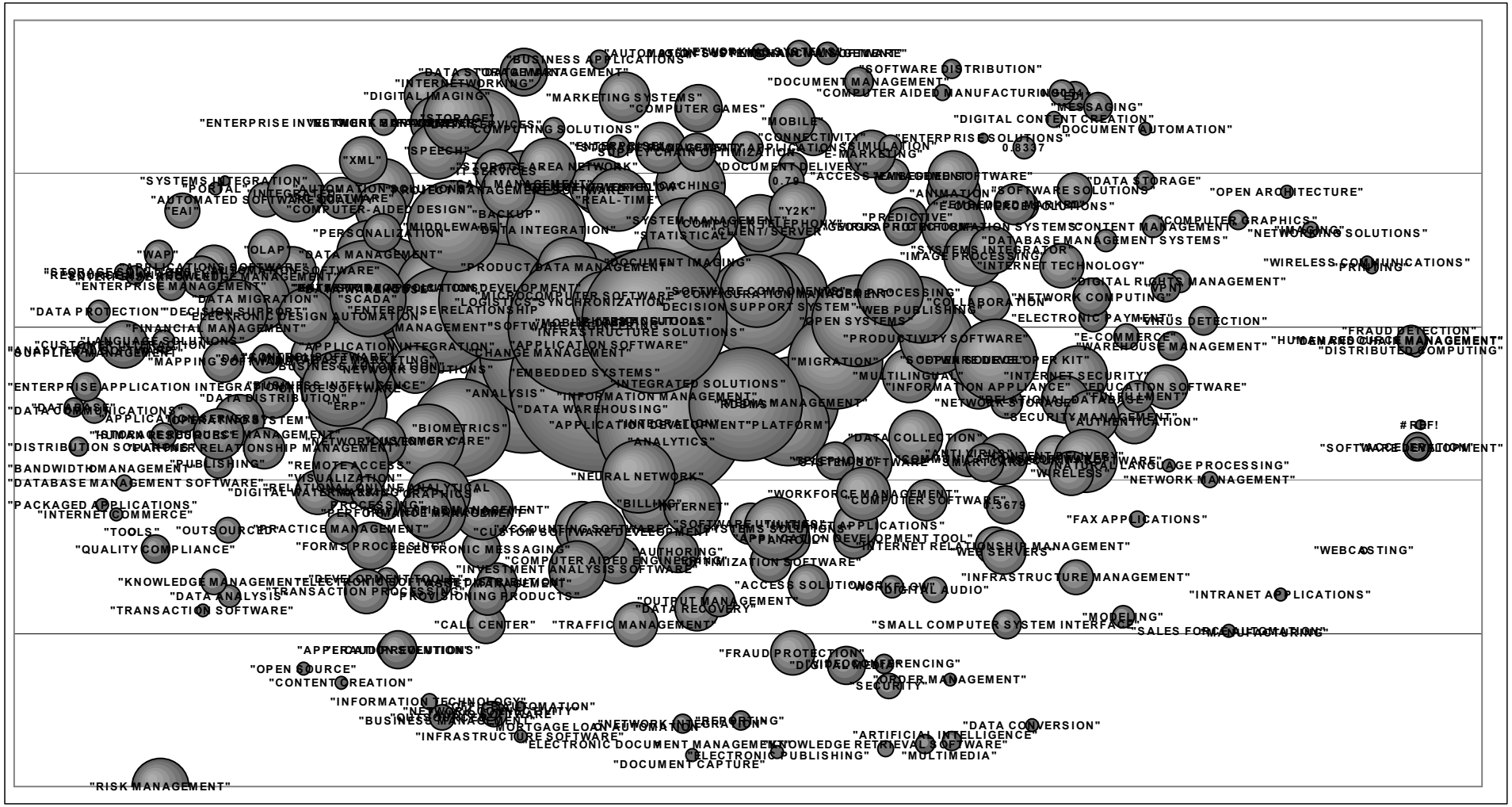
1992



1995


"MAPPINGSOFTWARE"
"OUTPUTMANAGEMENT"
"FINANCIALSOFTWARE"
"DOCUMENTAUTOMATION"
"DATABASEMANAGEMENTSYSTEMS"
"PAYROLL"
"WAREHOUSEMANAGEMENT"
"ASSETMANAGEMENT"
"WORKFORCEMANAGEMENT"
"NEURALNETWORK"
"CHANGEMANAGEMENT"
"CALLCENTER"
"OFFICEAUTOMATION"
"ELECTRONICPUBLISHING"
"ANTIVIRUS"
"SALESFORCEAUTOMATION"
"DECISIONSUPPORTSYSTEM"
"FINANCIALMANAGEMENT"
"EXECUTIVEINFORMATIONSYSTEMS"
"APPLICATIONDEVELOPMENTTOOL"
"PHARMACEUTICAL"
"MICROCOMPUTERSOFTWARE"
"ENTERPRISESOLUTIONS"
"FINANCIALAPPLICATIONS"
"ACCOUNTINGSOFTWARE"
"PERSONALIZATION"
"HUMANRESOURCES"
"BUSINESSMANAGEMENT"
"HUMANRESOURCEMANAGEMENT"
"DATABASEMARKETING"
"BUSINESSINTELLIGENCE"
"SYSTEMSINTEGRATOR"
"MAILINGEFFICIENCY"
"INFRASTRUCTURESOFTWARE"
"DATAWAREHOUSE"
"CONTENTMANAGEMENT"
"DATAWAREHOUSING"
"DATAPROTECTION"
"OBJECT-ORIENTED"
"MODELING"
"DIGITALAUDIO"
"COMPUTINGSOLUTIONS"
"STORAGEMANAGEMENT"
"ANALYSIS"
"APPLICATIONSOFTWARE"
"CONFIGURATIONMANAGEMENT"
"ELECTRONICDESIGNAUTOMATION"
"VOICEPROCESSING"
"BACKUP"
"REPORTING"
"APPLICATIONDEVELOPMENT"
"CLIENT/SERVER"
"RDBMS"
"APPLICATIONSSOFTWARE"
"SYSTEMMANAGEMENT"
"DATABASE"
"RELATIONALDATABASE"
"SOFTWARESOLUTIONS"
"SYSTEMSOFTWARE"
"COMPUTERAIDEDENGINEERING"
"EMBEDDEDSOFTWARE"
"COLORMANAGEMENT"
"ENTERPRISE"
"CUSTOMERINTERACTION"
"COMPUTERAIDEDMANUFACTURING"
"GEOGRAPHICINFORMATIONSYSTEMS"
"MIDDLEWARE"
"DATACOLLECTION"
"BUSINESSAPPLICATIONS"
"OPERATINGSYSTEM"
"REMOTEACCESS"
"ACCESSSOLUTIONS"
"DATACOMMUNICATIONS"
"DATAMANAGEMENT"
"IMAGEPROCESSING"
"NETWORKACCESSSOLUTIONS"
"WORKFLOW"
"COMPUTER-AIDEDDESIGN"
"PUBLISHING"
"TOOLS"
"DISTRIBUTEDCOMPUTING"
"STORAGE"
"DATABASEMANAGEMENTSOFTWARE"
"TELEPHONY"
"INFORMATIONTECHNOLOGY"
"AUTOMATIONSOLUTIONS"
"MANUFACTURING"
"IMAGING"
"PRODUCTDATAMANAGEMENT"
"DOCUMENTMANAGEMENT"
"SUPPLYCHAINMANAGEMENT"
"SECURITYSOLUTIONS"
"COMMUNICATIONSSOLUTIONS"
"INTEGRATION"
"PLATFORM"
"SOFTWAREDEVELOPMENT"
"MESSAGING"
"OPENSYSTEMS"
"CONNECTIVITY"
"INTEGRATEDDEVELOPMENTTOOLS"
"DOCUMENTIMAGING"
"SOFTWARECOMPONENTS"
"ANIMATION"
"ELECTRONICFUNDSTRANSFER"
"ERP"
"GRAPHICS"
"PERFORMANCEMANAGEMENT"
"INTERNETSOFTWARE"
"INTERNET"
"PRINTING"
"AUTHENTICATION"
"COMPUTERGRAPHICS"
"WIRELESS"
"COMPUTERTELEPHONY"
"WIRELESSCOMMUNICATIONS"
"MULTIMEDIA"
"INTERNETWORKING"
"COMPUTERSOFTWARE"
"SYSTEMSSOLUTIONS"
"E-COMMERCE"
"AUTHORING"
"ACCELERATION"
"MOBILEDATA"
"NETWORKMANAGEMENT"
"DIGITALMEDIA"
"VISUALIZATION"
"SECURITY"
"ELECTRONICDOCUMENTMANAGEMENT"
"COMMUNICATIONSSOFTWARE"
"REAL-TIME"
"AUTOMATEDSOFTWAREQUALITY"
"SIMULATION"
"DESKTOPAPPLICATIONS"
"EDI"
"SYSTEMSINTEGRATION"
"NETWORKINGSYSTEMS"
"OLAP"
"DATAINTEGRATION"
"NETWORKINGSOLUTIONS"
"DIGITALIMAGING"
"DATAANALYSIS"
"SECURITYSOFTWARE"
"ELECTRONICPAYMENT"
"SPEECH"
"EDUCATIONSOFTWARE"
"DATAACQUISITION"
"MULTIMEDIAL"
"STORAGESOLUTIONS"
"VIDEOCONFERENCING"
"INTERNETSECURITY"
"NETWORKSOLUTIONS"
"AUTOMATIONSYSTEMS"
"VIDEOEDITING"
"SECURITYMANAGEMENT"
"SIGNALPROCESSING"
"RISKMANAGEMENT"
"ENTERPRISEPRODUCTIVITYSOFTWARE"

1998


"BUSINESS APPLICATIONS"
"SOFTWARE DISTRIBUTION"
"DOCUMENT MANAGEMENT"
"INTERNETWORKING"
"DIGITAL IMAGING"
"MARKETING SYSTEMS"
"COMPUTER GAMES"
"MESSAGING"
"DIGITAL CONTENT CREATION"
"MOBILE"
"DOCUMENT AUTOMATION"
"ENTERPRISE SOLUTIONS"
"SPEECH"
"XML"
"STORAGE AREA NETWORK"
"E-MARKETING"
"SYSTEMS INTEGRATION"
"DOCUMENT DELIVERY"
"DATA STORAGE"
"ANIMATION"
"SOFTWARE SOLUTIONS"
"OPEN ARCHITECTURE"
"AUTOMATED SOFTWARE"
"EAI"
"COMPUTER-AIDED DESIGN"
"BACKUP"
"REAL-TIME"
"Y2K"
"PREDICTIVE"
"COMPUTER GRAPHICS"
"CONTENT MANAGEMENT"
"NETWORKING SOLUTIONS"
"PERSONALIZATION"
"DATA INTEGRATION"
"CLIENT/SERVER"
"DATABASE MANAGEMENT SYSTEMS"
"WAP"
"OLAP"
"DATA MANAGEMENT"
"STATISTICAL"
"SYSTEMS INTEGRATOR"
"IMAGE PROCESSING"
"INTERNET TECHNOLOGY"
"WIRELESS COMMUNICATIONS"
"PRODUCT DATA MANAGEMENT"
"DOCUMENT IMAGING"
"PRINTING"
"ENTERPRISE MANAGEMENT"
"SCADA"
"DIGITAL RIGHTS MANAGEMENT"
"DATA MIGRATION"
"DECISION SUPPORT SYSTEM"
"COLLABORATION"
"NETWORK COMPUTING"
"DATA PROTECTION"
"ELECTRONIC DESIGN AUTOMATION"
"OPEN SYSTEMS"
"ELECTRONIC PAYMENT"
"FINANCIAL MANAGEMENT"
"PRODUCTIVITY SOFTWARE"
"E-COMMERCE"
"WAREHOUSE MANAGEMENT"
"FRAUD DETECTION"
"APPLICATION INTEGRATION"
"APPLICATION SOFTWARE"
"DISTRIBUTED COMPUTING"
"MAPPING SOFTWARE"
"EMBEDDED SYSTEMS"
"MULTILINGUAL"
"INTERNET SECURITY"
"ENTERPRISE APPLICATION INTEGRATION"
"INTEGRATED SOLUTIONS"
"INFORMATION APPLIANCE"
"DATA DISTRIBUTION"
"ERP"
"ANALYSIS"
"INFORMATION MANAGEMENT"
"DATA WAREHOUSING"
"NETWORK STORAGE"
"DATABASE"
"SECURITY MANAGEMENT"
"AUTHENTICATION"
"BIOMETRICS"
"APPLICATION DEVELOPMENT PLATFORM"
"DATA COLLECTION"
"ANALYTICS"
"DISTRIBUTION SOFTWARE"
"PUBLISHING"
"NATURAL LANGUAGE PROCESSING"
"BANDWIDTH MANAGEMENT"
"WIRELESS"
"VISUALIZATION"
"NEURAL NETWORK"
"NETWORK MANAGEMENT"
"DATABASE MANAGEMENT SOFTWARE"
"WORKFORCE MANAGEMENT"
"PACKAGED APPLICATIONS"
"BILLING"
"INTERNET"
"COMPUTER SOFTWARE"
"INTERNET COMMERCE"
"FAX APPLICATIONS"
"TOOLS"
"OUTSOURCED"
"SOFTWARE UTILITIES"
"APPLICATIONS"
"QUALITY COMPLIANCE"
"FORMS PROCESSING"
"ELECTRONIC MESSAGING"
"AUTHORING"
"INTERNET RELATIONSHIP MANAGEMENT"
"WEB SERVERS"
"WEBCASTING"
"KNOWLEDGE MANAGEMENT"
"INVESTMENT ANALYSIS SOFTWARE"
"ACCESS SOLUTIONS"
"INFRASTRUCTURE MANAGEMENT"
"DATA ANALYSIS"
"PROVISIONING PRODUCTS"
"DIGITAL AUDIO"
"INTRANET APPLICATIONS"
"TRANSACTION SOFTWARE"
"OUTPUT MANAGEMENT"
"DATA RECOVERY"
"MODELING"
"CALL CENTER"
"TRAFFIC MANAGEMENT"
"SMALL COMPUTER SYSTEM INTERFACE"
"SALES FORCE AUTOMATION"
"FRAUD PROTECTION"
"OPEN SOURCE"
"CONTENT CREATION"
"ORDER MANAGEMENT"
"SECURITY"
"INFORMATION TECHNOLOGY"
"DATA CONVERSION"
"ARTIFICIAL INTELLIGENCE"
"ELECTRONIC PUBLISHING"
"MULTIMEDIA"
"DOCUMENT CAPTURE"
"RISK MANAGEMENT"

2001

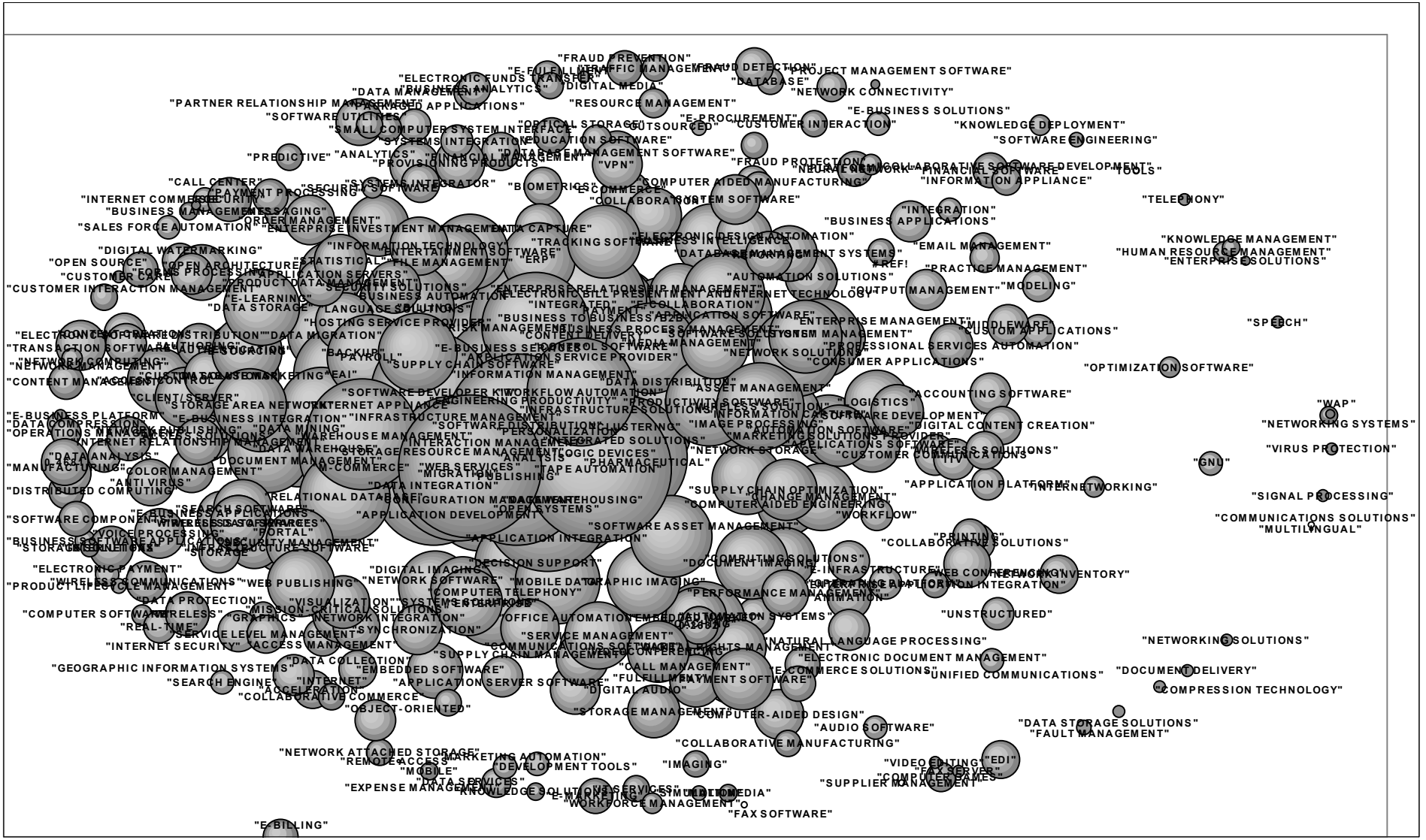

"FRAUD PREVENTION"
"FRAUD DETECTION"
"E-FULFILLMENT"
"TRAFFIC MANAGEMENT"
"PROJECT MANAGEMENT SOFTWARE"
"ELECTRONIC FUNDS TRANSFER"
"DATA MANAGEMENT"
"DIGITAL MEDIA"
"DATABASE"
"NETWORK CONNECTIVITY"
"PARTNER RELATIONSHIP MANAGEMENT"
"PACKAGED APPLICATIONS"
"RESOURCE MANAGEMENT"
"SOFTWARE UTILITIES"
"E-PROCUREMENT"
"E-BUSINESS SOLUTIONS"
"SMALL COMPUTER SYSTEM INTERFACE"
"OPTICAL STORAGE"
"OUTSOURCED"
"CUSTOMER INTERACTION"
"KNOWLEDGE DEPLOYMENT"
"SYSTEMS INTEGRATION"
"EDUCATION SOFTWARE"
"SOFTWARE ENGINEERING"
"PREDICTIVE"
"ANALYTICS"
"FINANCIAL MANAGEMENT"
"DATABASE MANAGEMENT SOFTWARE"
"FRAUD PROTECTION"
"NETWORK"
"COLLABORATIVE SOFTWARE DEVELOPMENT"
"PROVISIONING PRODUCTS"
"VPN"
"FINANCIAL SOFTWARE"
"TOOLS"
"CALL CENTER"
"SYSTEMS INTEGRATOR"
"BIOMETRICS"
"COMPUTER AIDED MANUFACTURING"
"INFORMATION APPLIANCE"
"INTERNET COMMERCE SECURITY"
"COLLABORATION SYSTEM SOFTWARE"
"TELEPHONY"
"BUSINESS MANAGEMENT"
"INTEGRATION"
"ORDER MANAGEMENT"
"BUSINESS APPLICATIONS"
"SALES FORCE AUTOMATION"
"ENTERPRISE INVESTMENT MANAGEMENT"
"DATA CAPTURE"
"ELECTRONIC DESIGN AUTOMATION"
"KNOWLEDGE MANAGEMENT"
"DIGITAL WATERMARKING"
"TRACKING SOFTWARE"
"INTELLIGENCE"
"EMAIL MANAGEMENT"
"HUMAN RESOURCE MANAGEMENT"
"OPEN SOURCE"
"INFORMATION TECHNOLOGY"
"DATABASE MANAGEMENT SYSTEMS"
"ENTERPRISE SOLUTIONS"
"STATISTICAL"
"FILE MANAGEMENT"
"ERP"
"#REF!"
"CUSTOMER CARE"
"AUTOMATION SOLUTIONS"
"PRACTICE MANAGEMENT"
"CUSTOMER INTERACTION MANAGEMENT"
"PRODUCT DATA MANAGEMENT"
"SECURITY SOLUTIONS"
"ENTERPRISE RELATIONSHIP MANAGEMENT"
"MODELING"
"E-LEARNING"
"BUSINESS AUTOMATION"
"ELECTRONIC BILL PRESENTMENT AND PAYMENT"
"INTEGRATED"
"E-COLLABORATION"
"OUTPUT MANAGEMENT"
"DATA STORAGE"
"LANGUAGE SOLUTIONS"
"INTERNET TECHNOLOGY"
"HOSTING SERVICE PROVIDER"
"BUSINESS TO BUSINESS (B2B)"
"ENTERPRISE MANAGEMENT"
"SPEECH"
"ELECTRONIC DISTRIBUTION"
"DATA MIGRATION"
"PAYMENT"
"APPLICATION SOFTWARE"
"MIDDLEWARE"
"TRANSACTION SOFTWARE"
"AUTHORING"
"BACKUP"
"E-BUSINESS SERVICES"
"MEDIA MANAGEMENT"
"CUSTOM APPLICATIONS"
"NETWORK COMPUTING"
"PATROL"
"APPLICATION SERVICE PROVIDER"
"NETWORK SOLUTIONS"
"PROFESSIONAL SERVICES AUTOMATION"
"NETWORK MANAGEMENT"
"SUPPLY CHAIN SOFTWARE"
"CONSUMER APPLICATIONS"
"OPTIMIZATION SOFTWARE"
"CONTENT MANAGEMENT"
"EAI"
"DATA DISTRIBUTION"
"CLIENT/SERVER"
"ASSET MANAGEMENT"
"ACCOUNTING SOFTWARE"
"STORAGE AREA NETWORK"
"INTERNET APPLIANCE"
"PRODUCTIVITY"
"LOGISTICS"
"WAP"
"E-BUSINESS PLATFORM"
"INFRASTRUCTURE SOLUTIONS"
"WIRELESS SOLUTIONS"
"INFORMATION CATALOG"
"DATA COMPRESSION"
"E-BUSINESS INTEGRATION"
"INFRASTRUCTURE MANAGEMENT"
"SOFTWARE DEVELOPMENT"
"CLUSTERING"
"IMAGE PROCESSING"
"DIGITAL CONTENT CREATION"
"NETWORKING SYSTEMS"
"OPERATIONS MANAGEMENT"
"DATA MINING"
"PERSONALIZATION"
"INFORMATION SOFTWARE"
"INTERNET RELATIONSHIP MANAGEMENT"
"WAREHOUSE MANAGEMENT"
"INTERACTION MANAGEMENT"
"INTEGRATED SOLUTIONS"
"MARKETING SOLUTIONS PROVIDER"
"APPLICATIONS SOFTWARE"
"VIRUS PROTECTION"
"DATA WAREHOUSE"
"STORAGE RESOURCE MANAGEMENT"
"LOGIC DEVICES"
"NETWORK STORAGE"
"WIRELESS SOLUTIONS"
"DATA ANALYSIS"
"ANALYSIS"
"CUSTOMER COMMUNICATIONS"
"GNU"
"MANUFACTURING"
"COLOR MANAGEMENT"
"M-COMMERCE"
"WEB SERVICES"
"PHARMACEUTICAL"
"TAPE AUTOMATION"
"ANTI VIRUS"
"PUBLISHING"
"APPLICATION PLATFORM"
"INTERNETWORKING"
"DISTRIBUTED COMPUTING"
"DATA INTEGRATION"
"SUPPLY CHAIN OPTIMIZATION"
"SIGNAL PROCESSING"
"RELATIONAL DATABASE"
"DATA WAREHOUSING"
"COMPUTER AIDED ENGINEERING"
"SOFTWARE COMPONENTS"
"OPEN SYSTEMS"
"APPLICATION DEVELOPMENT"
"WORKFLOW"
"COMMUNICATIONS SOLUTIONS"
"MULTILINGUAL"
"SOFTWARE ASSET MANAGEMENT"
"PRINTING"
"VOICE PROCESSING"
"PORTAL"
"APPLICATION INTEGRATION"
"COLLABORATIVE SOLUTIONS"
"STORAGE"
"DECISION SUPPORT"
"COMPUTING SOLUTIONS"
"ELECTRONIC PAYMENT"
"DIGITAL IMAGING"
"DOCUMENT IMAGING"
"E-INFRASTRUCTURE"
"WEB PUBLISHING"
"NETWORK SOFTWARE"
"MOBILE DATA"
"GRAPHIC IMAGING"
"WEB CONFERENCING"
"INVENTORY"
"PRODUCT LIFECYCLE MANAGEMENT"
"COMPUTER TELEPHONY"
"ENTERPRISE APPLICATION INTEGRATION"
"DATA PROTECTION"
"VISUALIZATION"
"PERFORMANCE MANAGEMENT"
"SYSTEM ARCHITECTURE"
"ANIMATION"
"COMPUTER SOFTWARE"
"WIRELESS"
"MISSION-CRITICAL SOLUTIONS"
"UNSTRUCTURED"
"GRAPHICS"
"NETWORK INTEGRATION"
"OFFICE AUTOMATION"
"EMBEDDED SYSTEMS"
"REAL-TIME"
"SYNCHRONIZATION"
"SERVICE MANAGEMENT"
"NATURAL LANGUAGE PROCESSING"
"NETWORKING SOLUTIONS"
"SERVICE LEVEL MANAGEMENT"
"INTERNET SECURITY"
"ACCESS MANAGEMENT"
"COMMUNICATIONS SOFTWARE"
"RIGHTS MANAGEMENT"
"SUPPLY CHAIN MANAGEMENT"
"ELECTRONIC DOCUMENT MANAGEMENT"
"DATA COLLECTION"
"GEOGRAPHIC INFORMATION SYSTEMS"
"EMBEDDED SOFTWARE"
"CALL MANAGEMENT"
"UNIFIED COMMUNICATIONS"
"DOCUMENT DELIVERY"
"SEARCH ENGINES"
"INTERNET"
"APPLICATION SERVER SOFTWARE"
"FULFILLMENT SOFTWARE"
"COMPRESSION TECHNOLOGY"
"ACCELERATION"
"COLLABORATIVE COMMERCE"
"DIGITAL AUDIO"
"OBJECT-ORIENTED"
"STORAGE MANAGEMENT"
"COMPUTER-AIDED DESIGN"
"AUDIO SOFTWARE"
"DATA STORAGE SOLUTIONS"
"FAULT MANAGEMENT"
"COLLABORATIVE MANUFACTURING"
"NETWORK ATTACHED STORAGE"
"MARKETING AUTOMATION"
"REMOTE ACCESS"
"DEVELOPMENT TOOLS"
"IMAGING"
"VIDEO EDITING"
"EDI"
"MOBILE"
"DATA SERVICES"
"SUPPLIER MANAGEMENT"
"EXPENSE MANAGEMENT"
"KNOWLEDGE SOLUTIONS"
"E-MARKETING"
"WORKFORCE MANAGEMENT"
"FAX SOFTWARE"
"E-BILLING"

Figure 11. Multiplier of the rate for firm creation of new categories