COMMUNITY ECOLOGY FOR INNOVATION CONCEPT: THE CASE OF CLOUD COMPUTING

Completed Research Paper

Jia Sun
College of Information Studies
University of Maryland, College Park
4105 Hornbake Building, South Wing
College Park, MD 20742
jsun@umd.edu

Ping Wang
College of Information Studies
University of Maryland, College Park
4105 Hornbake Building, South Wing
College Park, MD 20742
pwang@umd.edu

Abstract

Why do some concepts come to be highly popular, significantly reshaping the IT landscape, while others do not? We address this question by exploring the communities of organizations that underlie IT innovation concepts. In an illustrative investigation of the community for cloud computing, we integrate theories of organizational ecology and scale-free network and their associated methods. Substantial support of the theories and the effectiveness of the methods have led us to embark on a promising research program focused on IT innovation concepts and communities. This program affords several opportunities to break new grounds in IS research: (1) adding an ecological explanation to theories of IT innovations; (2) fertilizing new ground for IT innovation research; (3) stimulating research on scale-free networks associated with IT innovations; and (4) demonstrating the utility of computational discourse analysis.

Keywords: Information technology innovation, organizational ecology, scale-free network, innovation community, cloud computing, discourse analysis

Introduction

Big data, business analytics, cloud computing, and social media are innovation concepts very popular today in Information Technology (IT). These concepts have the potential to transform the ways we live, work, and play and thus it is important to understand the production, promulgation, and consumption of IT innovation concepts. Broadly, innovation concepts travel beyond the conventional product/service marketplace to spread in the marketplace for ideas, where analysts, consultants, journalists, and academics join vendors and adopters to collectively make sense of the concepts. Information Systems (IS) scholars have thoroughly researched the adoption and implementation of a wide array of IT innovations (Fichman 2004). Technology and Innovation Management (TIM) research has a long tradition of studying the development of technological innovations (including IT innovations) (Ruttan 2001). While numerous insights from IS and TIM have greatly enriched our understanding of the marketplace for new IT products and services, we know little about the marketplace for new IT concepts. *Why do some concepts come to be highly popular, significantly reshaping the IT landscape, while others do not?*

We address this fundamental question by exploring the communities of organizations that underlie IT innovation concepts. An innovation community is a set of organizations with interests in developing and/or utilizing the innovation. Innovation communities thrive and decay according to dynamic ecological processes and evolving social structures. To understand the ecology of such communities, we apply classic organizational ecology theory initially formulated in sociology (Hannan and Freeman 1977). To examine the structure of such communities, we employ the theories of scale-free networks originally discovered in the networks of hyperlinked websites (Barabási and Albert 1999). We integrate these theories and their associated methods in an illustrative investigation of the community for cloud computing, a key concept on today's IT innovation scene.

Substantial support of the theories and the effectiveness of the methods have led us to embark on a promising new research program focused on IT innovation concepts and communities. This program affords several opportunities to break new grounds in IS research: (1) adding an ecological explanation to theories of IT innovations; (2) fertilizing new ground for IT innovation research; (3) stimulating research on scale-free networks associated with IT innovations; and (4) demonstrating the utility of computational discourse analysis.

In what follows, we first review the main theories that inspire our study and develop the hypotheses. We then describe the methods and report the results from the empirical study. In closing, we discuss the implications of the results to IS research, making explicit comments on the breakthrough nature of a sustained research program on IT innovation concepts and communities.

Ecological and Network Theories of IT Innovation

IT Innovation Concept

An innovation is "an idea, practice, or object that is perceived as new by an individual or other unit of adoption" (Rogers 2003, p. 12). An IT innovation, accordingly, refers to the idea, practice, or object associated with a new information technology. Just as ideas are epistemologically different from physical practices or objects, an IT innovation exists in conceptual and/or material forms (Swanson and Ramiller 2004; Wang 2009). The *conceptual* form of an innovation is a set of ideas that describe the attributes, processes, and possible consequences of the innovation. For example, ideas underlying the cloud computing innovation may include the utility model for providing computing services and software as services delivered over the web. In contrast, the *material* form of an innovation refers to the existence of the innovation in the physical world. For instance, the material forms of cloud computing include the data centers, hardware, software, and computer and telecommunication networks that carry out various specific cloud computing services. Therefore, while the developers and adopters may directly interact with the materials associated with an IT innovation, they may join others, such as analysts, journalists, and academics, to discuss the innovation as a concept.

Theoretically, the notion of IT innovation concept may be considered to be a generic version of specialized concepts such as "technological frame" (Bijker 1987), "organizing vision" (Swanson and Ramiller 1997), and "management fashion" (Abrahamson and Fairchild 1999). The empirical examples of these concepts in IS research have been found, respectively, shaping the public discourse on how work is done over the Internet (Iacono and Kling 2001), guiding the evolution of an IT innovation targeted at small and medium enterprises (Currie 2004), and impacting the reputation and performance of *Fortune 500* companies innovating with popular new ITs (Wang 2010). Despite the important role innovation concepts play in the innovation processes and outcomes of numerous organizations of diverse types, the majority of empirical research has focused on the material form of IT innovations (Fichman 2004). Only a few empirical studies have examined innovations as concepts (e.g., Abrahamson and Fairchild 1999; Currie 2004; Wang 2009).

All previous research on the conceptual form of innovation seemed to share a common insight: While the material form of an innovation is often associated with a specific organization, the development, promulgation, and consumption of an innovation concept are not confined within the boundary of any organization, but require the work of a collective of organizational players. For example, it is true that Google provides a platform for cloud computing and the *Washington Post* adopted some cloud services. The concept of cloud computing, however, does not belong to any company. Any organization may develop, spread, and make sense of the cloud computing concept. Such collective endeavor is undertaken in a community of organizations interested in the innovation.

Table 1. Diverse Organizational Roles in the Cloud Computing Innovation Community

Role	Sample Organization	Sample Sentence in News Articles from ProQuest
Platform provider	Google and IBM	There's YouTube, IPTV, high-definition images and "cloud computing" – in which individuals and businesses use the centralized computing resources of Google and IBM data centers, instead of the local computing resources of their own PCs or office systems. (*Wall Street Journal*, 02/22/2008)
Application or service provider	Salesforce.com	Salesforce.com is now called a cloud application – after all, companies let it store their sales data, rather than running it on their own systems. (*New York Times*, 05/25/2008)
Adopter	General Services Administration	General Services Administration has announced that it will be the first federal agency to move e-mail to a cloud-based system, reflecting the government's push to adopt Web-based computing. (*Washington Post*, 12/03/2010)
Public investor	National Science Foundation	For example, the National Science Foundation, in addition to supporting the OpenFlow initiative, has financed the Global Environment for Network Innovations, or GENI. (*New York Times*, 05/22/2011)
Private investor	Intel Capital	Intel Capital wants to invest in companies that focus on education, cloud computing, information technology and related services. (*Wall Street Journal*, 05/13/2011)
Researcher	University of Virginia and Microsoft Research	Two researchers at the University of Virginia and four at Microsoft Research explored this possibility in a paper presented this year at the Usenix Workshop on Hot Topics in Cloud Computing. (*New York Times*, 11/27/2011)
Analyst	Forrester Research	Forrester Research analyst James Staten said the emphasis on data center systems will require computer makers to rethink their approach because they currently are structured to sell relatively small enterprise systems to a large number of corporate customers. (*Wall Street Journal*, 11/05/2008)
Consultant	McKinsey	The average server utilization in a data center, according to McKinsey, is 10 percent. That can be fairly easily increased to 18 percent, the consulting firm says, by adopting virtualization software. (*New York Times*, 04/20/2009)

Innovation Community

As previously noted, an innovation community is a set of organizations with interests in developing and/or utilizing the innovation (Wang and Ramiller 2009). Members of such a community are diverse, motivated by different interests and playing different roles. For instance, a quick scan of the news articles on cloud computing shows a list of roles and examples of organizations playing those roles in the cloud computing innovation community (Table 1).

Despite such diversity, organizations in an innovation community participate in developing and spreading the innovation concept as they make sense of the innovation. More importantly, each member's understanding of the innovation depends on the collective learning that takes place in the community (Wang and Ramiller 2009). Several empirical studies have examined such collective learning. Abrahamson and Fairchild (1999; 2001) used the Quality Circles innovation concept to illustrate competition among idea entrepreneurs (i.e., consultants, journalists, technical specialists, and scholars) in the marketplace for management knowledge. Wang and Ramiller (2009) depicted the community development of the Enterprise Resource Planning (ERP) concept as a cycle where community members contributed ideas to the meaning of the innovation concept and, in turn, the concept may guide members' next round of sense-making. More recently, Swanson (2012) summarized a series of innovation studies with the metaphor "IT innovation wave machine" to portray the institutional apparatus (mostly driven by vendors and consultants) that spins out waves of innovation concepts. With their intriguing findings, these pioneering studies invite more empirical research on the dynamic evolution and complex structure of the innovation community. As a next step to advance this line of inquiry, we propose ecology theory as a framework for understanding the relationships among organizations in an innovation community.

Ecology of IT Innovation Community

Over thirty years ago, organizational sociologists borrowed ecology theories from biology to conceptualize the evolution of organizations (Hannan and Freeman 1977). The ensuing paradigm of organizational ecology, as Baum and Amburgey (2002) reviewed, "aims to explain how social, economic and political conditions affect the relative abundance and diversity of organizations and to account for their changing composition over time" (p. 304). Although the general thesis of ecological theory may be applied to multiple levels of analysis, most application and adaptation of the theory has taken place at the population/industry level. Population ecology explains the evolution of established populations of organizations (such as the automobile industry) and emphasizes the factors that homogenize organizational forms and stabilize organizational populations. Astley (1985) criticized population ecology for failing to explain how populations originate in the first place, and thus favored an ecology theory at the community level. Community ecology "focuses on the rise and fall of populations as basic units of evolutionary change, simultaneously explaining forces that produce homogeneity and stability within populations and heterogeneity between them" (Astley 1985, p. 224).

As a prominent sociological theory, unlike other major theories such as institutional theory which has been applied in IS research productively for decades (see Currie 2011 for a recent review of institutional theory in IS), organizational ecology theory has only inspired sporadic studies in IS. For example, Chengalur-Smith and Sidorova (2003) drew from the population ecology literature to explore factors related to the survival of open source software projects. Nickerson and zur Muehlen (2006) developed an ecological approach to study how people and ideas move across institutions to create Internet standards. Wang et al. (2012) applied organizational ecology theory to understand the growth of Usenet newsgroups.

In our study of the evolution of IT innovation communities, we find ecological theory very useful, especially applied at the community, as opposed to population, level. This is because an innovation community consists of organizations in diverse populations/industries (as Table 1 illustrates). In particular, the density-dependent model widely adopted in organizational ecology research is illuminating because it rigorously models two ecological processes crucial to the development of innovation concepts: legitimation and competition.

According to organizational ecology, the evolution of a population manifests in its vital rates: the entry rate (at which organizations enter or are founded in the population) and exit rate (at which organizations fail or leave the population). The vital rates depend on legitimation of the population (to what extent is the population taken for granted) and competition among the organizations already in the population. The density-dependent model assumes that legitimation increases entry rates and decreases exit rates, while competition has the opposite effects. According to Hannan et al., (1995), "both legitimation and competition are assumed to be driven by the same observable variable: density, the number of organizations in a bounded organizational population" (p. 510).

Empirical support has been found for the density-dependent model in numerous studies of various industries in multiple countries (Hannan et al. 1995). However, it has rarely been applied to the community level, where different populations of organizations interact. Nevertheless, we found the legitimation and competition processes described in ecology theory resonate with the findings from studies of innovation community. For example, Wang and Swanson (2007) theorized the process by which organizations in diverse industries attempted to legitimize the innovation concept – Professional Services Automation. Abrahamson and Fairchild (2001) described the competition among knowledge industries for Quality Circles. They competed for both financial gain and attention – the currency of a knowledge economy. To test these theoretical arguments derived mostly from qualitative research, we raise two hypotheses to test the density-dependent model in an IT innovation community.

Hypothesis 1: Legitimation is positively associated with the entry rate of an IT innovation community.

Hypothesis 2: Competition is negatively associated with the entry rate of an IT innovation community.

Structure of IT Innovation Community

Density may be parsimonious to capture complex processes such as legitimation and competition. However, in IT, it is common to see a pair of innovation concepts and their associated communities of similar size in their emergent years. One concept takes off to become the "next big thing," whereas the other fades away quietly. This contrast implies that legitimation and competition, as indicated by density, may not be sufficient to explain the changing vitality of an innovation community. A further step might be to explore the structure of communities. To do so, we turn to contemporary network theory and we have found research on scale-free networks particularly interesting.

In studying how certain network structures facilitate or constrain certain network dynamical behaviors such as the spread of information, Barabási and Albert (1999) found that the World Wide Web exhibits the occurrence of highly connected vertices, a structure they called "scale-free." Simply defined, a scale-free network is the network whose nodes are not randomly or evenly connected, but includes a few "highly-connected" nodes. In scale-free networks, the distribution of node degree follows a power law in which most nodes have only a few connections and some have a large number of connections. In recent years, network research has uncovered a stunning range of physical and social networks that are scale-free (Barabási 2003).

These scale-free networks have been found extremely efficient in spreading things such as information, rumors, and viruses (Barabási 2003). For example, Bollt and ben-Avraham (2005) found that the scale-free network architecture results in a faster transit time between connected nodes, and the natural selection mechanism in scale-free networks also has a strong influence on the nature of diffusion. Duan et al. (2005) showed that using hub strategy can improve the contagion effects evidently. Dezső and Barabási (2002) discovered that by discriminating between the nodes, curing the highly connected nodes can restore a finite epidemic threshold and potentially eradicate a virus.

Considering the positive relationship between the scale-free network topology and the performance of the dynamic networks, we ask if highly-connected nodes in innovation community would help the community to grow.

Hypothesis 3: The degree of scale-freeness of an IT innovation community is positively associated with the entry rate of the community.

Methods

To test the hypotheses, we selected the innovation community for cloud computing, as an example of IT innovation concepts. Cloud computing, as defined by the U.S. National Institute of Standards and Technology (NIST), is "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" (Mell and Grance 2011, p. 2). This cloud model is composed of five essential characteristics (on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service), three service models (software as a service, platform as a service, and infrastructure as a service), and four deployment models (private, community, public, and hybrid cloud). It is interesting to note that this is NIST's 16th (and purportedly final) version of cloud computing definition. Even so, it is hardly the only version that everyone agrees with and numerous other definitions exist (Vaquero et al. 2009). Because of such diversity as reflected in the various definitions and multiple types of organizations in the cloud computing innovation community (Table 1) and the relatively notable size of the community, we selected cloud computing for our empirical study.

Data Collection

We identify organizations in the publicly available discourse on cloud computing. A discourse is "an interrelated set of texts, and the practices of their production, dissemination, and reception, that brings an object into being" (Phillips and Hardy 2002, p. 3). The object in our case is the cloud computing concept and the organizations that appear in the cloud computing discourse are the members of the cloud computing innovation community. Many sources of discourse exist. For this exploratory study, we focus on mainstream newspapers in the U.S. Specifically, we selected the ProQuest Newspapers database. The database archives full-text and full-image articles for the most respected national and regional newspapers such as *New York Times*, *Wall Street Journal*, *Washington Post*, *Los Angeles Times* and *Christian Science Monitor*.

In the ProQuest Newspapers database, we searched for the phrase "cloud computing" in the title, subject, abstract, and full text of news articles published during a ten-year period (2002 to 2011). The first article that mentioned cloud computing was published in *Wall Street Journal* in March 2007. In total, we collected 987 articles from the queries. Most of the news articles have approximately 5 to 30 paragraphs; a few of them have fewer than 5 or more than 100 paragraphs. Articles with more than 100 paragraphs are mostly company lists and thus we removed them. We extracted 19,212 paragraphs in total from the full text of the news articles containing cloud computing.

Data Processing

Identifying Organizations

We used Stanford NER, a Natural Language Processing (NLP) tool, to automatically identify the organization names from the paragraphs. Stanford NER is a Java implementation of a named entity recognizer. We wrote a PERL script to extract the organization names from the paragraphs.

Crowdsourcing organizational data cleaning

The precision of Stanford NER for organization identification was approximately 62% and thus the result we obtained had many errors. We relied on human judgment to correct the errors by employing CrowdFlower, a crowdsourcing service. Crowdsourcing is a process that involves outsourcing tasks to a distributed group of people. We uploaded the tasks of verifying the organizations extracted from news article data (2007-2011). For each task, we showed a sentence containing a possible organization's name and asked each CrowdFlower worker to tell us whether the name represents an organization by choosing "Yes," "No," or "Unknown." Here is an example of the tasks:

> *Is Accenture an organization based on the sentence below?*
>
> *Sample Sentence: He was senior executive and partner, communications and high technology strategy, at Accenture; and president, TataFone, at Tata Administrative Services.*

Each task was handled by multiple CrowdFlower workers. To ensure quality, 5% of the tasks were "gold standard" tasks for which we provided objective answers. This quality control mechanism not only prevented scanners from generating random or useless results, but also helped train new workers for better performance. Each task was handled by at least three trusted workers, whose answers to the "gold standard" tasks were closest to the answers we had provided. In total, 18 CrowdFlower workers completed

6,434 tasks in this study. We downloaded the results aggregated by CrowdFlower automatically. For each task, based on the workers' agreement, CrowdFlower provided an agreement score. We used the filter function in Excel to obtain a clean list of organizations whose agreement score of "Yes" was higher than 0.80, indicating a majority of CrowdFlower workers agreed that those are true organizations in the real world.

Calculating Co-occurrence of Organizations

Organizations in the cloud computing community are related in many ways and these relations form a network. To operationalize the network of organizations in the discourse, one way is to manually read each article and annotate the types of relations among organizations. Another way is to automatically detect the co-occurrences of organizations and treat the co-occurrences as network ties. The latter approach obviously loses the rich meanings of various organizational ties, but is scalable to the analysis of large collections of discourse data like ours (Kennedy 2008). So we selected the co-occurrence approach.

When two organizations co-occurred in the same paragraph in a news article, we assumed that there is a relation between the two organizations. To further validate the relation between each pair of co-occurring organizations, we used Fisher's exact test (p<=.05, one-tailed) to filter out the random co-occurrences. In probability theory, if two events are independent, the occurrence of one event makes it neither more nor less probable that the other occurs. Pearson's Chi-square test for independence has been widely applied to compare the independent groups (Larntz 1978). Fisher's exact test is more accurate than Chi-square test when the expected numbers are small (Larntz 1978).

Mapping and Visualizing Community Structure

After eliminating the random co-occurrences, we split the network edges (an edge link a pair of co-occurring organizations) by month from 2007 to 2011 according to the publication dates of the articles. Based on the monthly networks, we created the evolving organizational networks for cloud computing. These one-mode networks of organizations are neither directed nor weighted. We used NodeXL, an Excel add-on module, to visualize the temporal organization networks. We chose NodeXL because of its flexibility and versatility. In NodeXL, we grouped the network nodes into clusters using the clustering function embedded in NodeXL (Hansen et al. 2010).

Data Analysis

To test the hypotheses discussed above, we include variables representing legitimation, competition, and the scale-freeness measure of the organizational network in a general regression model to explain the changing entry rate month by month between October 2007 and November 2011.

Dependent Variable

Organizational entry rate is the dependent variable in our data analysis. An organization entered the network when it began to have at least one connection to another organization. Organization entry rate is the number of organizations that newly joined the network each month.

Independent Variables

We followed the classic density-dependent model to measure the two ecological processes (legitimation and competition) (Hannan et al. 1995). Legitimation is indicated by density, the number of organizations in the community. We used the quadratic function of density as the measure of competition.

Regarding network scale-freeness, Li et al. (2005) offered a potentially precise "scale-free metric." Briefly, let g be a graph with edge-set ε, and let the degree (number of edges) at a vertex i be d_i. The scale-free level can be defined as $S(g) = \sum_{(i,j)\in\varepsilon} d_i d_j$. This is maximized when high-degree nodes are connected to other high-degree nodes. The scale-freeness ratio is defined as $S(g) = \frac{s(g)}{s_{max}}$ where s_{max} is the maximum value of $s(h)$ for h in the set of all graphs with an identical degree distribution to g. A network with low $S(g)$ is "scale-rich;" and a network with $S(g)$ close to 1 is "scale-free." We implemented this metric in our study.

Control Variables

We would like to develop a model where the baseline entry rate changes freely over the historical period in response to changes in environmental conditions, avoiding the effects of environmental shifts on entry rate. So we included historical periods as control variables (Hannan et al. 1995). Our exploration of the data suggested that a model with five one-year periods is practical and thus we created four dummy variables for the five one-year periods as control variables.

Previous population ecology studies suggested that a surge of entries in the previous period indicates conditions favorable for the entries of new organizations (Carroll et al. 1993; Carroll and Swaminathan 1992). After a surge of new entries, since the pool of other potential entrants might be depleted, the effect of prior entries may weaken when there is a large number of recent entries. Therefore, the effect of prior entries is similar to that of density. Therefore we controlled for the number of organization entries in the previous month and for its quadratic term as well.

Analytical Models

Putting together the dependent, independent, and the one-year period control variables, we have model 1:

$$\lambda(t) = \beta_1 n_t + \beta_2 n_t^2 + \beta_3 sf_t + \sum_{i=1}^{4} \gamma_i y_i + \beta_0 \quad (1)$$

where $\lambda(t)$ denotes the organization entry rate of month t in the community; n_t denotes the number of organizations in the community just before time t (month t-1); sf_t is the scale-freeness measure of the organization network before time t (month t-1); y_i is the dummy variable for year i (1 corresponds to Year 2008 and 4 to Year 2011).

We included the organization entry rate in the previous month in the second model:

$$\lambda(t) = \beta_1 n_t + \beta_2 n_t^2 + \beta_3 sf_t + \beta_4 \lambda_{(t-1)} + \sum_{i=1}^{4} \gamma_i y_i + \beta_0 \quad (2)$$

In addition, we added the quadratic term of the organization entry rate in the previous month in Model 3:

$$\lambda(t) = \beta_1 n_t + \beta_2 n_t^2 + \beta_3 sf_t + \beta_4 \lambda_{(t-1)} + \beta_5 \lambda_{(t-1)}^2 + \sum_{i=1}^{4} \gamma_i y_i + \beta_0 \quad (3)$$

Results

Co-occurrences of Organizations

To illustrate the co-occurrences of organizations, Figure 1 shows a portion of the co-occurrence frequency matrix with each row or column representing an organization. The numbers on the diagonal are the numbers of paragraphs where the corresponding organizations appeared. For instance, Google was mentioned in 131 paragraphs extracted from the cloud computing news articles in 2008, while Sun Microsystems was only mentioned in 8 paragraphs. The numbers in the other cells represent the number of paragraphs containing the co-occurrences of the organization pairs. For instance, Google and Microsoft were co-mentioned in 48 paragraphs in 2008, whereas Dell and Google were not co-mentioned in 2008 at all.

To further understand the varied relations that the co-occurrences represent, we randomly selected 30 articles from this dataset. Our reading of these articles revealed five types of relations between organizations that co-occurred in these articles: competition, collaboration, merger and acquisition, adoption, and investment (Table 2). These categories are examples of the relations underlying the co-occurrences and the list in Table 2 is not exhaustive since a larger sample of the articles may reveal more relation types (e.g., the advisory relation between consultants and clients is a usual suspect). It is also useful to note that not all co-occurrences indicate meaningful relations between organizations; some organizational pairs simply appeared in the same paragraphs by coincidence.

	Google	Microsoft	Yahoo!	Dell	Hewlett-Packard	Sun Microsystems	…
Google	131						
Microsoft	48	188					
Yahoo!	28	35	79				
Dell	0	2	0	19			
Hewlett-Packard	1	0	1	3	10		
Sun Microsystems	1	2	1	1	1	8	
…							

Figure 1. Co-occurrences between Organizations in 2008

Table 2. Types of Relations among Co-occurring Organizations

Relation	Organization	Sample Sentence in News Articles from ProQuest
Competition	Microsoft and Google	Such is the potential for attracting users to Internet-based programs that Microsoft announced a 44.6 billion takeover bid for Yahoo!, the second-largest Web portal and Microsoft's best hope for battling Google in the world of cloud computing. (*The Baltimore Sun*, 02/08/2008)
Collaboration	Yahoo!, Computational Research Laboratories and Tata Sons	Yahoo! has struck an agreement with Computational Research Laboratories, a unit of Tata Sons Ltd., to pursue research on "cloud computing" technologies, in which programs are run from remote data centers instead of a local computer. (*New York Times*, 05/25/2008)
Merger and Acquisition	EMC and VMware	EMC acquired VMware in 2003 for 635 million and retained 86% of its stock following an eye-popping initial public offering last summer. (*Wall Street Journal*, 07/09/2008)
Adoption	Yahoo! and IBM	The corporate customers who have been trying out the I.B.M. systems include Yahoo! and other Internet companies. (*New York Times*, 04/23/2008)
Investment	Southeastern Asset Management and Sun Microsystems	But Sun's depressed stock - which hit its lowest level in 14 years Thursday - has attracted some big investors. Value investor Southeastern Asset Management disclosed last month that it holds about 20% of the company's shares. (*Wall Street Journal*, 11/15/2008)

Temporal Network Visualization

Figures 2 and 3 show the networks of organizations in 2007 and 2008. A layout with the entire network displays all connections in the network. However, when the network size increases, the network graph becomes dense and messy. Therefore, we visualize each network by collapsing small network clusters (identified in NodeXL) and expanding a few major clusters.

Figure 2 shows the collapsed network clusters (in the middle panel in the top row) and nine other clusters of organizations that appeared in the cloud computing community in 2007. One cluster consists of six universities: Stanford University, University of Maryland, University of Washington, University of California, Carnegie Mellon University, and Massachusetts Institute of Technology. This cluster stemmed from the formation of the Academic Cloud Computing Initiative (ACCI) in October 2007 as a multi-university project developed to enhance students' technical knowledge on cloud computing. Correspondingly, those six universities were involved in ACCI. Another cluster combines established organizations such Google, Yahoo!, AOL, and IBM; startups such as Zoho and Transmedia; and two market research firms. In 2007, the cloud computing innovation not only attracted the involvement of established organizations, but also offered web-based startups new opportunities to challenge the established firms.

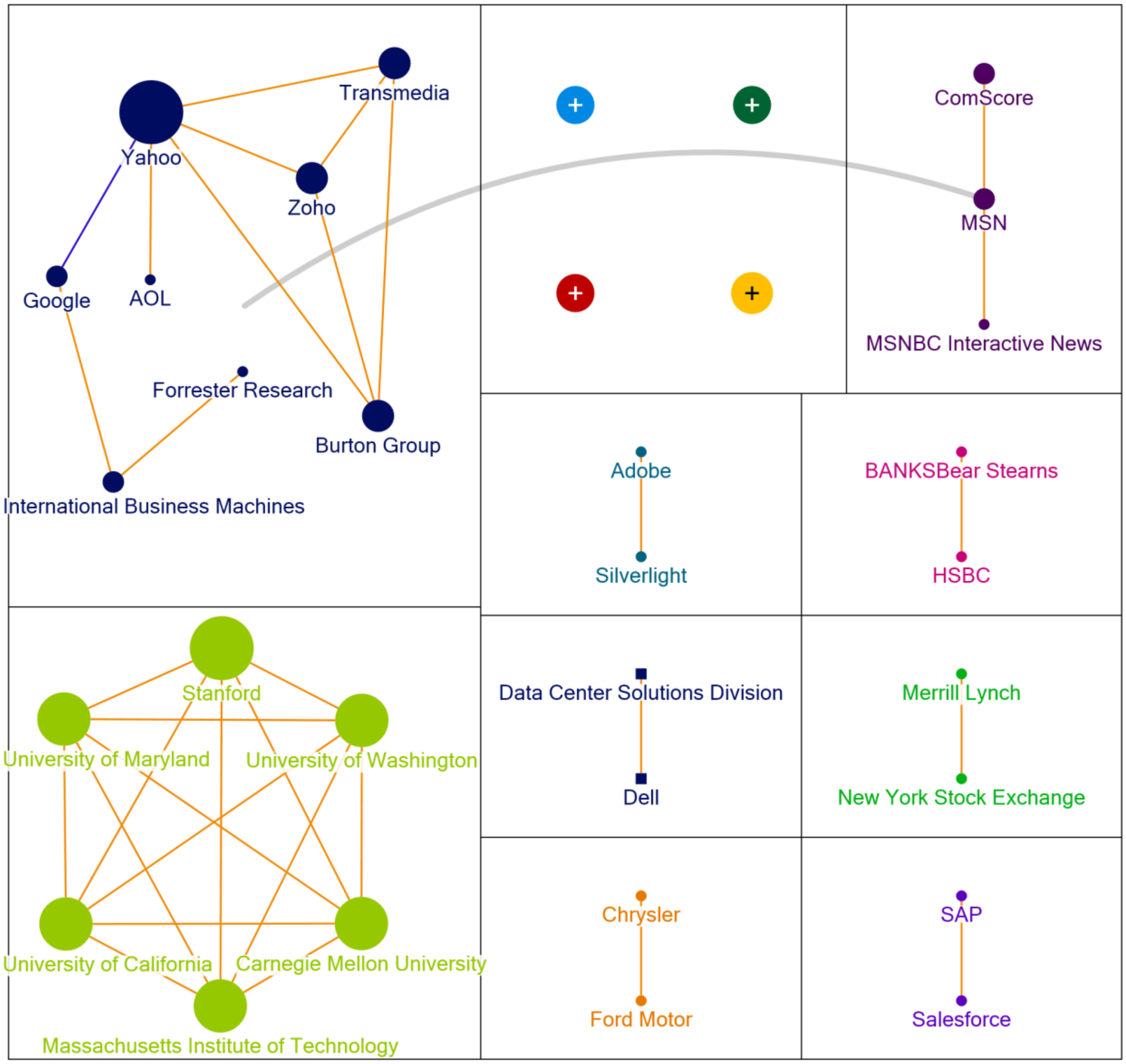

Figure 2. Organization Network Clusters of Cloud Computing in 2007

Figure 3 shows an overview of a large network with dozens of clusters (in the left panel) and zoomed-in views of five main clusters in 2008. From 2007 to 2008, as the size of the network increased from 40 to 214 nodes, the number of network clusters increased from 13 to 52. By filtering out the low degree edges, we were able to see major IT players in two different clusters in the cloud computing innovation community. Internet companies such as Google, Yahoo, Microsoft, Amazon, and Salesforce were in one cluster (the University of Wisconsin-Madison is an intriguing exception in that cluster), whereas computer hardware companies such as Dell, Hewlett-Packard, Intel, and Sun Microsystems were in another.

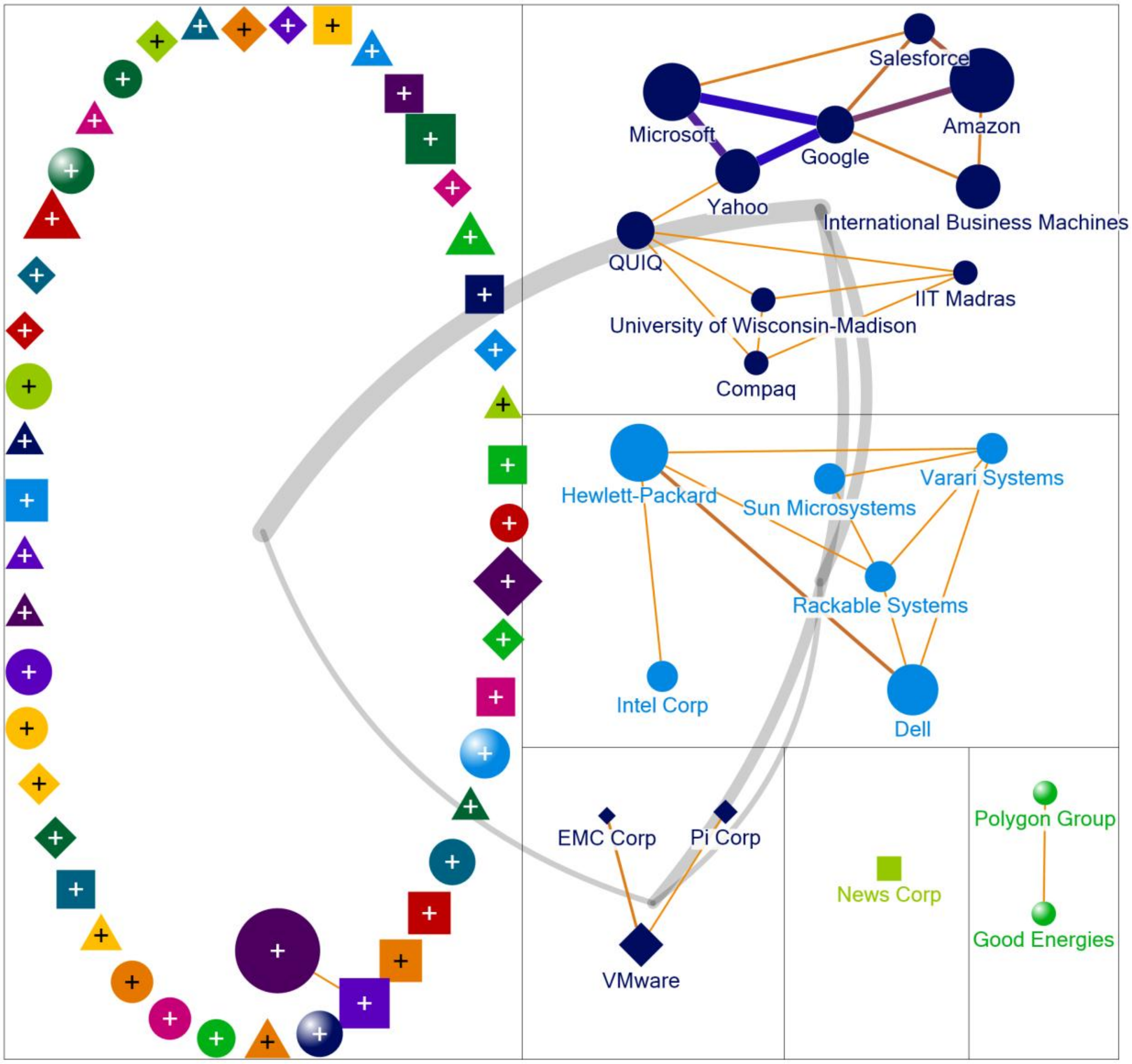

Figure 3. Organization Network Clusters of Cloud Computing in 2008

Regression Results

Figure 4 plots the dependent variable, the monthly organization entry rate (in grey), against the number of articles that mentioned cloud computing in the ProQuest Newspapers database (in black), an indicator of concept popularity in a few previous studies on innovation concepts (Abrahamson and Fairchild 1999; Wang 2010). Organization entry rate ranged from 0 to 105, and fluctuated more than the article count curve. The article count curve peaked in April 2011, and the organization entry rate peaked a month later. The high correlation (0.78) between the two variables suggests that both may be good indicators of the growth of the cloud computing innovation community.

Table 3 shows the descriptive statistics of the variables, not including the four one-year dummy variables. We present the descriptive statistics by year in order to show succinctly the evolution of the innovation community. Table 3 shows that on average, entry rate increased from 2007 to 2011, and reached an average of 61.90 in 2011. Density followed the same pattern. The scale-freeness of the network structure was declining from 2007 to 2011 as the density increased and community grew.

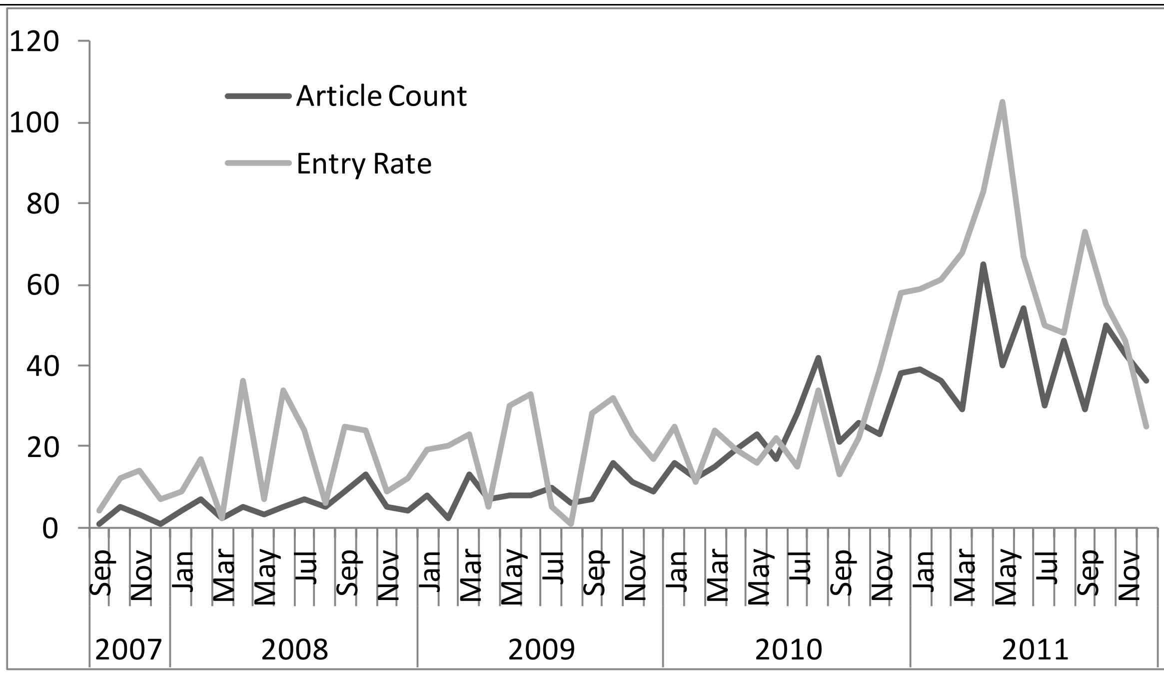

Figure 4. Concept and Entry Rate Trajectory for Cloud Computing in ProQuest Newspapers

Table 3. Descriptive Statistics of Variables in Hypotheses

Variables	2007		2008		2009		2010		2011	
	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.
Entry Rate	10.00	3.61	17.92	11.04	20.17	11.04	27.67	16.55	61.90	21.18
Density	29.67	10.69	148.08	69.09	368.00	70.53	615.67	83.90	1186.10	224.47
Density^2/1000	0.96	0.60	26.30	20.50	140.72	52.65	385.50	106.54	1452.62	524.53
Scale-freeness	0.96	0.06	0.66	0.06	0.55	0.02	0.50	0.01	0.49	0.01

Table 4 shows the results of regression based on the three analytical models described above. Models 1, 2, and 3 can explain 71.5%, 71.9%, 72.7% of the variance in organization entry rate, respectively. Because all hypotheses are directional, we used one-tailed tests. All three models show a significant, positive relationship between density and organization entry rate, suggesting a positive effect of legitimation on entry rate. Therefore, Hypothesis 1 is supported. All three models show a significant, negative relationship between organization entry rate and the quadratic term of density, suggesting a negative effect of competition on entry rate. Therefore, Hypothesis 2 is supported as well. Table 4 also shows a significant, positive effect of scale-freeness on entry rate in all the three models, and thus Hypothesis 3 is supported. None of the control variables has statistically significant relationship with the dependent variable.

Table 4. Results of Regression on Organization Entry Rate

Variables	Model 1		Model 2		Model 3	
	Coef.	S. E.	Coef.	S. E.	Coef.	S. E.
Independent Variables						
Density	0.229⁺	0.070	0.205⁺	0.077	0.198⁺	0.077
$Density^2/1000$	-0.120⁺	0.032	-0.109⁺	0.035	-0.104⁺	0.036
Scale-freeness	226.367⁺	91.039	205.301⁺	95.679	191.351⁺	96.131
Control Variables						
Prior Entry Rate			0.126	0.167	-0.16018	0.302
Prior Entry $Rate^2/1000$					3.36233	2.961
Period 2 (2008)	17.143	12.446	15.826	12.632	15.302	12.596
Period 3 (2009)	7.245	17.423	7.291	17.513	6.737	17.459
Period 4 (2010)	1.288	21.793	2.541	21.969	2.292	21.893
Period 5 (2011)	32.681	26.679	30.462	26.978	28.269	26.952
Multiple R-squared	0.715		0.719		0.727	
F-statistic (df)	15.03 (42)		13.09 (41)		11.86 (40)	

$^{+}p < .05$ (one-tailed test)

Discussion

Before we discuss the implications of these findings, we must acknowledge the limitations of the study.

Limitations

Most conspicuously, the empirical study is focused on the cloud computing innovation community, which is still developing. As we describe below, this is just the beginning of a long-term research program to investigate the concepts and communities for many IT innovations. Although this longitudinal study sheds light on the process by which one innovation comes to be increasingly popular, it cannot fully answer the research question why some innovations become popular while others do not. We are eager to compare the findings of this study with those of subsequent studies of other IT innovation concepts and communities. Second, the single source of discourse data we used in this study, the mainstream newspapers, however highly respected and representative, provides only a partial view of the cloud computing innovation community. Other data sources should be explored to overcome this limitation.

Third, our study illustrates that the variance/correlation approach is efficient in explaining the evolving vitality and popularity of an innovation community. However, the limitation of this approach in inferring causality is well known. To gain deeper understanding of the complex, dynamic causal relationships in the evolution of innovation communities, other approaches such as configurational methods (El Sawy et al. 2010) may complement what we have illustrated here. Fourth, just as in any other organizational ecology study, we used the density-dependent model to analyze the legitimation and competition processes in the innovation community. Both processes were measured by the same variable (density), except they were measured by the different functions of density. Neither legitimation nor competition was directly measured. This is a well-known limitation in all studies using the density-dependent model (Hannan et al. 1995). As we join others to develop more advanced measures for legitimation and competition, the benefits of applying and testing the density-dependent model on heterogeneous community for IT innovations outweigh the shortcomings of the model itself.

Implications to IS Research

We now discuss the implications of this study to IS research in four distinctive but related aspects.

Adding an ecological explanation to theories of IT innovations

In a seminal review, Fichman (2004) argued that IT innovation research may have converged on a "dominant paradigm" primarily focused on the adoption and implementation of IT systems in organizations. As it is increasingly difficult for the paradigm to inspire groundbreaking research, IT innovation researchers must go beyond the paradigm to explore new interesting directions. In this spirit, we have drawn on the distinction between the material and conceptual forms of IT innovations and presented a theoretical framework for studying IT innovation concepts (driving IT products and services) and innovation communities (encompassing the IT vendors and adopters, and many other organizations). At this supra-organizational, community level, we have added the ecology theory to existing explanations of how IT innovation concepts and communities emerge and evolve.

The ecological perspective presented here offers a quantitative, compelling, and parsimonious way to scientifically explain innovation popularity. Compared with the few pioneering qualitative studies of IT innovations at the community level (e.g., Currie 2004; Swanson 2012), this quantitative study has specified a set of clearly defined and measured variables, which helps replication and theory extension. Further, the fact that the model convincingly explains a large portion of the variance in organization entry rate suggests that the variables we have captured are good predictors of community dynamics. More impressively, this explanatory model relies on only a small set of factors and such parsimony makes it economical to pursue this line of research in the future. This scientific approach may augment experience- or expert-based methods for monitoring and understanding innovations (e.g., Gartner's Hype Cycles).

Fertilizing new ground for IT innovation research

From the ecological perspective, numerous new directions may be pursued in IT innovation research. As mentioned earlier, one is the innovation dimension, where more, multiple innovations' communities may be examined. Secondly, other sources of discourse data may be tapped into such as trade press, academic publications, blogsphere, and other social media where IT innovations live. By pooling the data from such diverse sources, researchers may be able to capture comprehensive and real-time snapshots of the innovation communities. Alternatively, analyzing the data from each source separately may help detect the nuances among the segments of an IT innovation community. Further, by expanding and shifting the examination windows, research may reveal intriguing temporal or historical patterns in IT innovations.

Stimulating research on scale-free innovation communities

Although we have found support for the theoretical argument that networks more scale-free would have higher performance, what is still unclear is the social processes underlying this network property. As IS researchers develop and evaluate physical and social networks, we have demonstrated that scale-freeness may be a sharp lens to add in the toolkit for understanding networks and communities for IT innovations. Many questions about the relationship between network structure and network performance remain. For example, how much theoretical advancement has scale-free network theory made from the classic insights on structural holes (Burt 1995), gate-keepers, and opinion leaders (Rogers 2003)?

Demonstrating the utility of computational discourse analysis

In this study, we have just had a taste of computational techniques such as NLP, automated text parsing, and crowdsourcing for data processing. The discourse data is massive, as we continue to explore other innovation concepts and communities, using other discourse sources, and in other historical periods. Our main lesson learned from this initial adventure is that humans and computers, because of their different fortes, should complement each other in order to efficiently and effectively analyze large collections of discourse data. Such division of labor and collaboration of skills are most wanted where off-the-shelf computing tools (such as Stanford NER in our case) are employed to generate a large amount of quick but noisy results.

Toward Sustained Research on Community Ecology for IT Innovation

In this illustrative analysis of organizations mentioned in the news articles about cloud computing, we applied organizational ecology to understand the legitimation of the innovation concept and the competition among organizations for collective attention. We also applied the theory of scale-free networks to examine the structure of the community of organizations involved in cloud computing in various ways. We found that legitimation and scale-free network structure were positively associated with the vital rate at which organizations join the community, whereas competition was negatively linked to community vitality.

As a professor registers for the webinar on big data, a consultant flies to the next city to work on her client's social media strategy, a publisher forecasts the sales figure for his next book on business analytics, and a market analyst releases her latest report on the frontiers of cloud computing, they are joining the communities that develop, disseminate, and consume these respective innovation concepts. What theories can IS researchers offer to explain the ebbs and flows of concepts in the booming and declining communities of interested and then uninterested people and organizations? What methods and tools can IS researchers offer to help everyone monitor, understand, and possibly shape the fast-changing innovation trends in IT? Our study is just a very modest start.

Acknowledgements

This study is a part of the STICK (Science & Technology Innovation Concept Knowledge-base) project, supported by the National Science Foundation's Science of Science and Innovation Policy program (Grant SBE-0915645). We thank other members of our research team, Ben Shneiderman, Yan Qu, Cody Dunne, Chen Huang, Phillip Lo, Liam Monahan, and Pengyi Zhang, for their help and suggestions.

References

Abrahamson, E., and Fairchild, G. 1999. "Management Fashion: Lifecycles, Triggers, and Collective Learning Processes," *Administrative Science Quarterly* (44:4), pp. 708-740.

Abrahamson, E., and Fairchild, G. 2001. "Knowledge Industries and Idea Entrepreneurs: New Dimensions of Innovative Products, Services, and Organizations," in *The Entrepreneurship Dynamic: Origins of Entrepreneurship and the Evolution of Industries,* C.B. Schoonhoven and E. Romanelli (eds.). Stanford, CA: Stanford University Press, pp. 147-177.

Astley, W.G. 1985. "The Two Ecologies: Population and Community Perspectives on Organizational Evolution," *Administrative Science Quarterly* (30:2), pp. 224-241.

Barabási, A.-L. 2003. *Linked: How Everything Is Connected to Everything Else and What It Means*. New York: Plume.

Barabási, A.-L., and Albert, R. 1999. "Emergence of Scaling in Random Networks," *Science* (286:5439), October 15, 1999, pp. 509-512.

Baum, J.A.C., and Amburgey, T.L. 2002. "Organization Ecology," in *The Blackwell Companion to Organizations,* J.A.C. Baum (ed.). Malden, MA: Blackwell, pp. 304-326.

Bijker, W.E. 1987. "The Social Construction of Bakelite: Toward a Theory of Invention," in *The Social Construction of Technological Systems,* W.E. Bijker, T.P. Hughes and T.J. Pinch (eds.). Cambridge, MA, London [England]: MIT Press, pp. 159-187.

Burt, R. 1995. *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Carroll, G.R., Preisendoerfer, P., Swaminathan, A., and Wiedenmayer, G. 1993. "Brewery and Brauerei: The Organizational Ecology of Brewing," *Organization Studies* (14:2), pp. 155-188.

Carroll, G.R., and Swaminathan, A. 1992. "The Organizational Ecology of Strategic Groups in the American Brewing Industry from 1975 to 1990," *Industrial and Corporate Change* (1:1), pp. 65-97.

Chengalur-Smith, S., and Sidorova, A. 2003. "Survival of Open-Source Projects: A Population Ecology Perspective," in *Proceedings of the 24th International Conference on Information Systems (ICIS),* Seattle, WA.

Currie, W.L. 2004. "The Organizing Vision of Application Service Provision: A Process-Oriented Analysis," *Information and Organization* (14:4), pp. 237-267.

Currie, W.L. 2011. "Institutional Theory of Information Technology," in *The Oxford Handbook of Management Information Systems: Critical Perspectives and New Directions,* R.D. Galliers and W.L. Currie (eds.). Oxford, New York: Oxford University Press, pp. 137-173.

Dezső, Z., and Barabási, A.-L. 2002. "Halting Viruses in Scale-Free Networks," *Physical Review E* (65:5), pp. 055103-1 - 055103-4.

Duan, W., Chen, Z., Liu, Z., and Jin, W. 2005. "Efficient Target Strategies for Contagion in Scale-Free Networks," *Physical Review E* (72:2), pp. 026133-1 - 026133-5.

El Sawy, O.A., Malhotra, A., Park, Y.K., and Pavlou, P.A. 2010. "Research Commentary—Seeking the Configurations of Digital Ecodynamics: It Takes Three to Tango," *Information Systems Research* (21:4), pp. 835-848.

Bollt, E.M. and ben-Avraham, D. 2005. "What Is Special About Diffusion on Scale-Free Nets?," *New Journal of Physics* (7:26), pp. 1-21.

Fichman, R.G. 2004. "Going Beyond the Dominant Paradigm for Information Technology Innovation Research: Emerging Concepts and Methods," *Journal of the Association for Information Systems* (5:8), pp. 314-355.

Hannan, M.T., Carroll, G.R., Dundon, E.A., and Torres, J.C. 1995. "Organizational Evolution in a Multinational Context: Entries of Automobile Manufacturers in Belgium, Britain, France, Germany, and Italy," *American Sociological Review* (60:4), p.. 509-528.

Hannan, M.T., and Freeman, J. 1977. "The Population Ecology of Organizations," *American Journal of Sociology* (82:5), pp. 929-964.

Hansen, D., Shneiderman, B., and Smith, M. 2010. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann.

Iacono, S., and Kling, R. 2001. "Computerization Movements: The Rise of the Internet and Distant Forms of Work," in *Information Technology and Organizational Transformation: History, Rhetoric, and Practice,* J. Yates and J. Van Maanen (eds.). Thousand Oaks, CA: Sage, pp. 93-136.

Kennedy, M.T. 2008. "Getting Counted: Markets, Media, and Reality," *American Sociological Review* (73:2), pp. 270-295.

Larntz, K. 1978. "Small-Sample Comparisons of Exact Levels for Chi-Squared Goodness-of-Fit Statistics," *Journal of the American Statistical Association* (73:362), pp. 253-263.

Li, L., Alderson, D., Tanaka, R., Doyle, J.C., and Willinger, W. 2005. "Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications (Extended Version)." Cornell University Library.

Mell, P., and Grance, T. 2011. "The NIST Definition of Cloud Computing," September, National Institute of Standards and Technology, Gaithersburg, MD, pp. 1-7.

Nickerson, J.V., and zur Muehlen, M. 2006. "The Ecology of Standard Processes: Insights from Internet Standard Making," *MIS Quarterly* (30), pp. 467-488.

Phillips, N., and Hardy, C. 2002. *Discourse Analysis: Investigating Processes of Social Construction*. Thousand Oaks CA: Sage Publications.

Rogers, E.M. 2003. *Diffusion of Innovations*, (5th ed.). New York, NY: Free Press.

Ruttan, V.W. 2001. *Technology, Growth, and Development: An Induced Innovation Perspective*. New York, Oxford: Oxford University Press.

Swanson, E.B. 2012. "The Manager's Guide to IT Innovation Waves," *MIT Sloan Management Review* (53:2), pp. 75-83.

Swanson, E.B., and Ramiller, N.C. 1997. "The Organizing Vision in Information Systems Innovation," *Organization Science* (8:5), pp. 458-474.

Swanson, E.B., and Ramiller, N.C. 2004. "Innovating Mindfully with Information Technology," *MIS Quarterly* (28:4), pp. 553-583.

Vaquero, L.M., Rodero-Merino, L., Caceres, J., and Lindner, M. 2009. "A Break in the Clouds: Towards a Cloud Definition," *SIGCOMM Computer Communication Review* (39:1), pp. 50-55.

Wang, P. 2009. "Popular Concepts Beyond Organizations: Exploring New Dimensions of Information Technology Innovations," *Journal of the Association for Information Systems* (10:1), pp. 1-30.

Wang, P. 2010. "Chasing the Hottest IT: Effects of Information Technology Fashion on Organizations," *MIS Quarterly* (34:1), pp. 63-85.

Wang, P., and Ramiller, N.C. 2009. "Community Learning in Information Technology Innovation," *MIS Quarterly* (33:4), pp. 709-734.

Wang, P., and Swanson, E.B. 2007. "Launching Professional Services Automation: Institutional Entrepreneurship for Information Technology Innovations," *Information and Organization* (17:2), pp. 59-88.

Wang, X., Butler, B.S., and Ren, Y. Forthcoming. "The Impact of Membership Overlap on Growth: An Ecological Competition View of Online Groups," *Organization Science*.