THE PROCESSES OF ECOSYSTEM EMERGENCE

July 2, 2014

Llewellyn D W Thomas
Imperial College Business School
Exhibition Road, London SW7 2AZ, UK
Telephone +44 (0)20 7594 8567
Email: llewellyn.thomas@imperial.ac.uk

Erkko Autio
Imperial College Business School
Exhibition Road, London SW7 2AZ, UK

ABSTRACT

The successful introduction of an ecosystem can rapidly undermine even apparent unassailable positions in seemingly matured and settled industries, however little is known about the processes through which an ecosystem emerges. Pioneering an institutional approach, we undertake a phasic analysis of the emergence processes of six digital service platform ecosystems. We demonstrate the uniqueness of each emergence process, and propose three phases of ecosystem emergence – initiation, momentum, and control. Path dependency and imprinting are then demonstrated through the finding of decreasing cross-case similarities with each successive phase.

Key words: ecosystem, platform, emergence, process, optimal matching
INTRODUCTION

As the examples of IBM in the 1980s and Nokia in the 2000s both demonstrate, the successful introduction of platform ecosystems can rapidly undermine apparent unassailable positions in seemingly matured and settled industries. With the continuous diffusion of ICT technologies across industry sectors and the associated spread in the use of ‘platform strategies’ (Gawer & Cusumano, 2002, 2008), managerial attention has remained on ‘business ecosystems’ or ‘innovation ecosystems’ from their introduction two decades ago (Moore, 1993). Indeed research attention seems to be increasing (see for instance Adner & Kapoor, 2010; Gulati, Puranam, & Tushman, 2012; Nambisan & Sawhney, 2011; Tiwana, Konynsksi, & Bush, 2010). To date, much of the ecosystem literature has focused on understanding the structure and dynamics of ecosystems, with research attention focused on issues such as value co-creation and value appropriation in ecosystem contexts (Adner & Kapoor, 2010; Gawer & Cusumano, 2002; Jacobides, Knudsen, & Augier, 2006).

However much less attention has been attached to the emergence of ecosystems. Although the early ecosystem literature proposes a four stage model of an ecosystem lifecycle (Moore, 1993) and more recently Gawer (2009) suggests an evolutionary typology of platform leadership, neither specifically examines the underlying processes of ecosystem emergence. The industrial economics literature offers insights into participant adoption strategies in multi-sided market contexts (Eisenmann, 2008; Eisenmann, Parker, & Van Alstyne, 2006; Hagiu & Eisenmann, 2007), and the alliance formation literature insights into the social context of prior alliances and interdependence (Gulati, 1995; Gulati & Gargiulo, 1999). In addition the behavioral strategy literature has indirectly considered ecosystem emergence strategies, suggesting alliance portfolios creation strategies (Ozcan & Eisenhardt, 2009), manipulation of organizational boundaries and market construction (Santos & Eisenhardt, 2009), and strategies to catalyze network formation (Hallen & Eisenhardt, 2012). Although these literatures provide valuable insight, they do not specifically consider the emergence processes of an ecosystem.

This lack of research on ecosystem emergence represents an important gap, given that many organizations, particularly in ICT sectors, today explicitly seek to create ecosystems and drive these to their advantage. As value creation processes become increasingly intertwined in today’s intertwined
Processes of ecosystem emergence

and highly specialized industrial landscape (Moore, 1996; Nalebuff & Brandenburger, 1996; Norman & Ramirez, 1994), it is important to start considering how ecosystems emerge in the first place, what activities and dynamics characterize ecosystem emergence and evolution, and under which conditions could ecosystem innovators stand a realistic prospect of achieving success (Ozcan & Eisenhardt, 2009). As such this paper seeks to understand: is there a generalized process of ecosystem emergence? For clarity, the focus of attention is not particularly on the activities of a particular actor, but of the more generalized processes which result in ecosystem emergence. We do not directly consider ecosystem creation, the actions and strategies through which a platform owner constructs an ecosystem.

This paper seeks to address this gap by empirically investigating the processes of ecosystem emergence and suggesting a three phase model of ecosystem emergence. Any investigation of ecosystem emergence necessarily needs to focus on a variety of different mechanisms. For instance, when an ecosystem has a platform that acts as its locus of coordination, then platform technological features have an important influence on ecosystem evolution (Gawer & Cusumano, 2002; Iansiti & Levien, 2004; Thomas, Autio, & Gann, 2014). In addition, the dependencies between the ecosystem participants are important influences on success and mobilization (Adner & Kapoor, 2010). Similarly, the variety of complementary asset providers and consumers that constitute the ecosystem results a consideration of membership, mobilization and support mechanisms (Gawer & Cusumano, 2002; Gulati et al., 2012; Iansiti & Levien, 2004; Wade, 1995). Finally cognitive mechanisms such as legitimacy, trust and meaning are also important to governance (Agerfalk & Fitzgerald, 2008; Gawer & Cusumano, 2002; Iansiti & Levien, 2004), as ecosystem participants work to develop the hierarchies and rules which govern their interactions.

Synthesizing these different mechanisms into a single approach, this paper pioneers an institutional approach to ecosystem emergence, integrating insights from dominant design theory, institutional entrepreneurship theory and social movement theory (Van De Ven & Hargrave, 2004). Dominant design theory provides the means to understand the social and institutional construction of the technological characteristics of an ecosystem (Kaplan & Tripsas, 2008; Tushman & Murmann, 1998; Tushman & Rosenkopf, 1992; Van De Ven & Garud, 1993), while social movement theory provides
the means to understand the collective action of ecosystem participants (Davis & McAdam, 2000; McAdam & Scott, 2005). Institutional entrepreneurship theory enables specific consideration of the actions of a platform owner in developing an ecosystem (Battilana, Leca, & Boxenbaum, 2009; DiMaggio, 1988; Garud, Jain, & Kumaraswamy, 2002).

The paper is structured as followed. The first section introduces an institutional approach to ecosystems and considers extant theories of the processes of ecosystem emergence. The second section introduces the research setting, detailing the six research cases and outlining the research methodology. The third section is an exposition of the results. The fourth section discusses the results in the context of our understanding of institution formation and change. The paper then concludes with contributions, limitations and directions for future research.

THEORY

The point of departure for such a research question is defining what is meant by an ecosystem. In management research, the term ‘ecosystem’ has been usually used to refer to a network of interconnected organizations that are linked to or operate around a focal firm or a platform (Adner, 2006; Adner & Kapoor, 2010; Iansiti & Levien, 2004; Moore, 1993, 1996; Teece, 2007). The difference with other network constructs in management research is that this network covers both production side and use side participants (Adner, 2012; Autio & Thomas, 2014; Iansiti & Levien, 2004). Following Autio and Thomas (2014), an ecosystem is defined as a network of interconnected organizations, organized around a focal firm or a platform, and incorporating both production and use side participants, and focusing on the development of new value through innovation.

Given this understanding of ecosystems, an ecosystem can be considered an institutional phenomenon. Structurally there is a strong resonance between the scope and participants of organizational fields and ecosystems, with both including suppliers, complementors, customers, competitors, universities, regulators, judiciary and standard setting bodies (DiMaggio & Powell, 1983; Iansiti & Levien, 2004; Scott, 2008; Teece, 2007). Thus similar to an ecosystem, an organizational field consists of the population of organizations operating in the same domain as indicated by the similarity of their services or products, also those other organizations that critically influence their performance, including exchange partners, competitors, funding sources and regulators (Scott, 2008). Both fields
Processes of ecosystem emergence

and ecosystems cut across traditional industries and focuses on the activities in which groups of organizations participate and their relationships to each other. Put differently, both ecosystems and institutional fields are the set of organizations that constitute a recognized area of life, characterized by structured network relations, and that share a set of institutions (Lawrence & Phillips, 2004). Hence an ecosystem is analogous to an institutional field in that it has its own institutional actors, logics and governance structures (Thomas & Autio, 2014b).

However the institutional nature of ecosystem is more profound than structural similarities, as the network of actors within a field or ecosystem are situated within a contingently stable alignment of material, organizational and discursive forces (Levy & Scully, 2007). Ecosystems and fields are the center of the common channels of dialogue and discussion around which markets, technologies, and issues arise (Adner, 2012; Beckert, 2010; Gawer & Phillips, 2013; Hoffman, 1999; Thomas & Autio, 2014b). More broadly, ecosystem and fields can both be considered arenas of “power relations” (Brint & Karabel, 1991; Gawer & Henderson, 2007; Teece, 2007). Iansiti and Levien (2004) have underlined the importance of legitimacy and trust in relation to ecosystem operations, and others have acknowledged the importance of openness, trust, tact and professionalism in developing an ecosystem (Agerfalk & Fitzgerald, 2008). The concept of ‘platform leadership’ of Gawer and Cusumano (2002) implies legitimacy through its focus on leadership – a platform owner cannot create or lead an ecosystem without being seen as legitimate by partners and other ecosystem participants (Olleros, 2008; Sivadas & Dwyer, 2000). Thus the rules that organize the ecosystem need to be legitimate in order for the successful operation of the ecosystem (Iansiti & Levien, 2004; Tiwana et al., 2010). Essentially, platform owners require social acceptability and credibility in the ecosystem to survive, beyond material resources, technical information and capabilities for production (Scott, Ruef, Mendel, & Caronna, 2000).

Theory suggests that there are a number of stages of ecosystem emergence. Moore (1993) has proposed a four stage model of the ecosystem lifecycle – birth, expansion, leadership and renewal – of which the first three are relevant when considering the emergence of a platform ecosystem. Hargrave and Van De Ven (2006) have suggested three phases of institutional technological innovation – emergence, development and implementation – as well as identifying the underlying process motors in
Suarez (2004) has suggested a five phase model of dominance design emergence - R&D build-up, technical feasibility, creating the market, decisive battle and post dominance – all of which are salient to understanding the emergence of an ecosystem. The notion of ‘institutionalization’ is a central concept in the organizational field literature. Meyer and Rowan (1977) defined institutionalization as the means by which social processes, obligations, and actualities take on a rule-like status in social thought and action. More specifically related to organizational fields, institutionalization can be understood as the process by which rules move from being abstractions to being constitutive of repeated patterns of interaction among field participants (Fligstein, 2001). From the institutional entrepreneurship literature, Greenwood, Suddaby, and Hinings (2002) have proposed a six stage model – precipitating jolts, deinstitutionalization, pre-institutionalization, theorization, diffusion and re-institutionalization – which is similar to the five stage model of Hinings, Greenwood, Reay, and Suddaby (2004). Within social movement theory, Davis and McAdam (2000) have proposed a linear model which emphasizes the identification of an opportunity, the gathering of resources to enable change, followed by collective action.

A synthesis of these emergence models suggests a first stage where there is the early technological development (Hargrave & Van De Ven, 2006; Moore, 1993; Suarez, 2004) as well as activities, such as sense-making and rule-making, that establish of the viability of the ecosystem and digital service (Greenwood et al., 2002; Suarez, 2004). There is also resource gathering, including the establishment of an organization structure for the personnel involved (Davis & McAdam, 2000) and the co-option of complementary asset providers and suppliers (Davis & McAdam, 2000; Moore, 1993; Suarez, 2004). As the ecosystem is nascent there is little competitor event or wider interest from society (Hargrave & Van De Ven, 2006; Moore, 1993; Suarez, 2004). This phase corresponds to the initial slow growth found by Agarwal and Bayus (2002) in the early stages of market evolution and sales take-off for new products.

A second stage is also suggested through the synthesis of these models. In this phase the ecosystem begins to grow rapidly (Agarwal & Bayus, 2002; Hargrave & Van De Ven, 2006; Moore, 1993; Suarez, 2004), as an installed base is created and network effects drive ecosystem growth (Sua-
Processes of ecosystem emergence

rez, 2004). Strategic maneuvering such as marketing and alliancing event is undertaken to drive adoption, as well as to provide the resources to scale up the ecosystem (Moore, 1993; Suarez, 2004). Increasing participant adoption, and marketing and discourse in wider society begins to legitimate the ecosystem and platform owner, complemented by processes of theorization and dissemination (Greenwood et al., 2002; Hargrave & Van De Ven, 2006). Competitive event is at its greatest here, as challengers also attempt to seize the opportunity (Agarwal & Bayus, 2002; Davis & McAdam, 2000; Hargrave & Van De Ven, 2006; Moore, 1993; Suarez, 2004).

A further stage is also suggested, where the ecosystem is now dominant and the competitive battles have been won (Greenwood et al., 2002; Moore, 1993; Suarez, 2004). Secure in the power structure it has created (Hinings et al., 2004) and its install base (Suarez, 2004), the ecosystem and platform owner have gained socio-political and cognitive legitimacy (Greenwood et al., 2002; Hinings et al., 2004). The focus of event moves from growth to control as activities move to interest satisfaction and value appropriation (Hinings et al., 2004; Moore, 1993; Suarez, 2004).

Taken together, these theories suggest that there are three stages of emergence; however there is little empirical work that validates the existence of these stages. Moreover there is little, if any, empirical work that considers the emergence processes of ecosystems. As such, our goal is to investigate if the emergence processes of ecosystems exhibit similarities across different ecosystems. By comparing the emergence processes of ecosystems across cases, we seek to provide for more systematic future research on ecosystem emergence processes, and also, provide rich insight that guides practitioner action in this domain. To create this contribution, we compare the emergence sequences of selected ecosystems using process methods.

METHODOLOGY

The research setting for the study are digital service platforms. Digital service platforms are an attractive setting for studying the creation of ecosystems as digital services are generally not necessary consumed directly or alone, but instead are consumed by other services or jointly consumed. This is due to the fact that many digital services are designed as incomplete systems (Garud, Jain, & Tuertscher, 2008), necessitating their inclusion in a network, or ecosystem, of other complementary services. It is through the ecosystem that the digital service platform provides the required utility to
the ultimate economic entity that consumes the service. This is enabled by “loosely coupled” integration methodologies, with “just-sufficient” protocols of interoperability and extensibility, which are a good design for widely distributed systems and networks (Orton & Weick, 1990). Digital service platforms also exhibit many of the characteristics of a high velocity environment, in that there are rapid changes in demand, competition, and technology (Bourgeois & Eisenhardt, 1988). In essence digital service platforms provide an ideal setting to observe the emergence processes as they happen over shorter timescales than in other environments. For clarity digital service platforms are defined as informational resources available for consumption where the consumers do not own the physical or informational infrastructure.

Six successful ecosystem emergence processes are investigated – Amazon, eBay, Facebook, Google, Salesforce, Wikipedia – which together illuminate the phenomenon of interest and cover a range of polar types (Pettigrew, 1990; Yin, 1984). In particular Amazon is an online retailer, eBay an online auction service, Facebook a social networking service, Google a search engine, Salesforce an online customer relationship management service, and Wikipedia an open source encyclopedia. See Table 1 for an overview of each Case.

An important differentiator for these digital service platforms is that each has a different business model, and in doing so they exemplify the main digital service business models. Put succinctly, Amazon has a retail based business model; eBay a business model that is brokerage based; Facebook has a business model that is predominantly based upon social driven advertising; Google’s business model is based on search-driven advertising; Salesforce.com has a subscription driven business model; and Wikipedia is open source, relying on its community and donations. A second differentiator is the types of participants within each platform ecosystem; Amazon is business to consumer service, while the services of eBay and Wikipedia link consumers to other consumers. Facebook and Google link businesses to consumers through advertising (and Facebook consumers to consumers for socially motivated reasons) and Salesforce is a business to business service. In addition these six cases vary by other important factors, including the effect of the dot-com crash, major crisis events, ecosys-
Processes of ecosystem emergence

tem participants, founder backgrounds and motivations, and early stage funding activities. Summariz-
ing, these six cases are heterogeneous within the context of digital services and potentially provide a level of generalizability beyond their context (Leonard-Barton, 1990).

Data was collected from a variety of archival sources, including critical histories authored by independent journalists, websites, press releases and news articles. For data sources that come from electronic sources (such as press releases, newsletters and corporate milestones), these were directly imported into a database. For data sources that required manual extraction of data, such as books, any action or incident that occurs was manually entered. The time frame ranged from the initial idea of the digital service, to the time when the success of the digital service and platform ecosystem are commonly acknowledged. Thus data collection often began before the digital service or company was launched, and finished a number of years later; for completeness an additional year was added after the commonly acknowledged success date (see Table 1). Following Garud and Van De Ven (1989), a discrete view of time was taken with the day as the temporal measurement unit. For incidents where an exact date could not be ascertained, an approximate date was utilized and noted.

From this basis a case narrative and timeline was compiled for each case in order to gain an overview of each. The data collection resulted in a sample of the incidents that constitute ecosystem emergence, not a full population, and hence the results necessarily needs to be understood in this context (Poole, Van De Ven, Dooley, & Holmes, 2000). In addition, due to the data sources utilized, there is most likely a bias towards the digital service platform in terms of the quantity of incidents and events have been recorded and collected, and as such the sequence data may represent the creation event as seen from the perspective of the platform owner. This is not considered to be an issue as the possible biases applies to all cases and any cross-case comparison would be valid.

To develop an understanding of ecosystem emergence requires statements about the temporal sequence of events that explain an observed stream of incidents (Poole et al., 2000), and it is important to distinguish between an incident as a raw datum, and an event as a theoretical construct (Abbott, 1984). Whereas an incident is an empirical observation, an event is not directly observed; it is a construct in a model that explains a pattern of incidents. For each event one can choose any number of incidents as indicators of the event, requiring a method called colligation (Abbott, 1984; McCullagh,
The first element of colligation is to understand the levels of analysis so as to tell a coherent story (Abbott, 1990b). Here relevant levels of analysis include the ecosystem, the digital service platform and the organization that provides that platform, and consistent with Garud and Van De Ven (1989) and theory, the wider environmental context. The next element of colligation consists of assembling the narrative by deciding its essential events at each of those levels and ensuring the same event granularity at each level (Abbott, 1984, 1990b). Here decision rules based around the nature of the incident were utilized, such as whether the incident was publicized in some way, named as critical or as a milestone, consisted in a change in the functionality or capability, or mentioned more than once from the data sources. In order to ensure similar granularity, where a group of incidents occurred over a number of days then they were gathered together into a single event. Finally the formal qualities of each event (Abbott, 1984) were also noted such as any duration and whether an event moves across levels. Table 1 details the number of events colligated for each Case.

To organize these events into a format conducive for process analysis, they were then coded using the ecosystem drivers as suggested by Thomas and Autio (2014a). Specifically, resource drivers are those concerned with the procurement and management of resources by the platform owner in an ecosystem. Technological drivers are those related to the design and provision of the underlying technologies in the platform ecosystem. Institutional drivers are those concerned with the institutional structure and operation of the ecosystem. Contextual drivers are important for ecosystem emergence, as wider technological, political and economic environment changes can influence the conditions in which the ecosystem emerges (Rao, Morrill, & Zald, 2000). See Figure 1 for code frequency across the six cases.

The codes were applied multi-functionally, meaning that multiple codes were applied to a single event, aligning with the concept of processes as multiple, intertwining narratives (Abbott, 1990a). For instance, a strategic partnership between a platform owner and a complementary asset provider which also involved a functional improvement to the digital service would be coded two times: once for the resource acquisition and once for the technological improvement.
Following the application of codes to the events, the data now consisted of six sequences of event codes, for example, a subset of the code sequence could appear as follows, R-T-C-T-R-I-C, with ‘R’ signifying a resource driver, ‘T’ a technology driver, ‘I’ an institutional driver and ‘C’ a context driver. Two phasic analysis techniques were applied to these coded sequences: optimal matching and direct inspection. Phasic analysis attempts to capture the overall coherency of development and change at a higher level than fine-grained micro-level structure (Poole et al., 2000). The fundamental assumption of phasic analysis is that the overall process can be described in units larger than the underlying events. Analysis was conducted using the R open source statistical software (R Core Team, 2012) with the TraMineR package for optimal matching (Studer, Ritschard, Gabadinho, & Müller, 2011).

Optimal matching is a technique that compares whether two sequences have similarities in the pattern of the codes, and requires that each sequence is represented by well-defined elements drawn from an alphabet of relatively small set of (repeating) event types (Poole et al., 2000). Optimal matching produces an index of the relative “distance” between any two sequences, and can be utilized on sequences of differing lengths. This index, known as the Levenshtein index, is the smallest possible cost of operations of insertion, substitution and deletion of sequence units required to align the two sequences. Put differently, this index represents the effort required to transform one sequence into another. As such the higher number returned in the index, the more different the sequences. A cost matrix determines the effort required for each insertion, substitution or deletion of a sequence element. We use a constant cost matrix; there is a value of one to each insertion, substitution and deletion. An alternative costing method, called TRATE, is based upon the probabilities of the transition of one code to another calculated from an analysis of the two sequences. This approach was not adopted as it results in differing costs of each insertion, substitution and deletion for each sequence comparison, making comparisons across differing sequence-pairs difficult to interpret.

The optimal matching algorithm always computes a distance measure for any pair of event sequences. However, because real-life event sequences are influenced by random events, it is very rare for any two event sequences to be perfectly matched. The challenge therefore remains to determine
the meaning of the distance measure. For example, if the optimal matching algorithm indicates a distance of 0.459 for two sequences, a procedure is required to determine whether this is indicative of meaningful difference or similarity between the two sequences. This is further complicated by the varying number of events within each case. For instance, the distance between Amazon-eBay is between 715 and 497 events, Amazon-Facebook is between 715 and 525 events, and eBay-Facebook is between 497 and 525 events. Given the influence of the number of insertions required to calculate the distance between each case-pair, the distances of each case-pair are not directly comparable. To explore the meaning of distance measures calculated from a dataset, clustering procedures are often used with optimal matching (Abbott, 1990b; Poole et al., 2000). Clustering allows the identification of a ‘most typical’ sequence by finding the sequence that minimizes some (possibly weighted) function of the distances to all other sequences (Abbott, 1990b). However, a clustering procedure is not possible here, as there are only six cases, and hence fifteen event sequence pairs in the dataset.

For this reason, we developed a randomization procedure to establish statistical benchmarks that could be used to compare the distance measures against.

1. For each case we first created an event set that was of equal length and contained the same frequency of codes as the actual event sequence (to use Amazon as an example, this would consist of a length of 715, of which 197 are resource codes, 179 technological codes, 220 institutional codes and 119 contextual codes). This is repeated for each case.

2. Using the ‘SAMPLE’ randomization algorithm in R and the respective event sets as a basis, we created 10,000 event sequences for each case where the ordering of events is randomized but the frequencies of the codes remain the same. In practice, this simply means that there was a different order in which the underlying events occurred. This creates a sample of the population of the code ordering permutations that could exist for each case.

3. For each case-pair comparison, we computed the distance of the 10,000 randomized event sequences of the first case to 10,000 randomized event sequences of the second. This procedure creates a database of 10,000 distance measures where 10,000 randomized event sequences of the first case sequence length and code frequency are compared against 10,000 randomized event sequences of the second case event sequence and code frequency. These 10,000 distances are a sample of the population of distances that would be created by calculating the distance between the populations of the code ordering permutations of each case-pair.

4. We then compared the actual distance measure computed for the actual sequences of the case-pair against the distance measure database. This enabled us to establish how the actual case-pair distance
Processes of ecosystem emergence

measured compared against the frequency distribution in the distance measure database. For example, if the mean distance was 0.5, and 95% of the random distances were within the range from 0.4 to 0.6 and the actual distance computed for the pair was 0.61, then this distance was equal to or greater than 95% of all distances in the database. Thus, the statistical significance of 0.11* indicates that the case-pair event sequences are statistically dissimilar with 95% confidence level (p < 0.05).

5. The meaning of this difference can be interpreted as indicating that the first case event sequence is substantially different to the second case event sequence. A distance measure of 0.39, where 0.39 is smaller than 95% of the values in the database would have indicated that the event sequence of the first case is statistically distinguishable and substantially similar to the second case.

A second technique utilized is direct inspection to determine if phases are suggested in the event sequences. Direct inspection, a qualitative technique, enables the determination of phases through interrogation of the case narratives and the manual discernment and sorting of pattern sequences into higher level sequences. In doing so direct inspection capitalizes on the background knowledge of the researcher and allows the modelling of sequence phases in a more discerning fashion than is possible by using algorithmic techniques (Poole et al., 2000). It is particularly useful when there are not many phases in the change sequence (Poole et al., 2000), something that the process models from ecosystems (see for instance, Moore, 1993), dominant design (see for instance, Hargrave & Van De Ven, 2006; Suarez, 2004), social movement theory (see for instance, Davis & McAdam, 2000) and institutional entrepreneurship (see for instance, Greenwood et al., 2002; Hinings et al., 2004) literatures suggest.

RESULTS

We now report the results of the comparison of the comparison of six empirical cases using phasic analysis. Figure 1 below details the three different phasic analyses that are reported below.

[Insert Figure 1 around here]

First detailed below are the results of the comparison of the full sequences for each case. This analysis provides insight into how similar the overall ecosystem emergence process is for each case. The second analysis considers the existence of phases within each case. Here the goal was to determine if there were sub-sequences within each case which could then be compared. The third analysis compares the phases across the cases.
**Analysis I: Complete sequence comparison**

This first analysis considers the similarity of the complete event sequences for all six cases, using optimal matching to determine the pairwise ‘distance’ of each case, as determined by the similarity of their complete event sequences. Fifteen case-pair distances are calculated, one for each pair: Amazon-eBay; Amazon-Facebook; Amazon-Google, and so on. Following this procedure there were 15 mean distances and 15 standard deviations – one each for each case-pair. These results are detailed in Table 2 below.

[Insert Table 2 around here]

The results indicate that the majority of the case-pairs are significantly different to each other, although with differing levels of significance. Amazon-Google, eBay-Salesforce, Amazon-Wikipedia, eBay-Wikipedia, Facebook-Wikipedia, Google-Wikipedia and Salesforce-Wikipedia are the most significantly different to each other (p<0.001). This means that the distance of the actual event sequences in each case-pair is greater than 99.9% of the 10,000 pairwise distances calculated from their randomly generated sequences. Amazon-Facebook, eBay-Goole, eBay-Facebook, Facebook-Goole and Google-Salesforce are also significantly different (p<0.01). This means that the distance of the actual event sequences in each case-pair is greater than 99% of the 10,000 pairwise distances calculated from their randomly generated sequences. Facebook-Salesforce is also almost significantly different, although less so than the others (p<0.05), with Amazon-eBay less so (p<0.1). This means that the distance of the actual event sequences between Salesforce and Google is greater than 95%, and Amazon and eBay greater than 90%, of the 10,000 pairwise distances calculated from their randomly generated sequences. The remaining case-pairs, eBay-Facebook and Amazon-Salesforce are not significantly different to each other. This means that their distances are similar to their respective random means. As the differences are positive, this suggests that the event sequence of each case-pair tends towards being different to each other.

**Analysis II: Three phases of emergence**
This second analysis considers whether there are sub-sequences (here called ‘phases’) within each case. Although the complete sequence analysis above indicates that the complete event sequences of each case are significantly different, sub-sequences within each complete event sequence are possible, as each case may have experienced the same sub-sequences, just in different ways.

To investigate whether there are sub-sequences within the complete event sequences, two phasic analyses are applied. The first, direct inspection, is retroductive, combining both inductive and deductive approaches (Poole et al., 2000). Insights and models from the existing theories of ecosystem and organizational field emergence as detailed in the theory section are above are applied to the case narratives and then adjusted in view of what is workable and informative. This retroductive approach provides a theoretical basis for the identification of phases and is consistent with the observation of Poole et al. (2000) that initial theoretical models can be used as ‘sensitizing constructs’ for conducting exploratory research when the study of change processes are at an embryonic stage.

The second analysis compares the (dis)similarity of a phase with other phases within a case, utilizing optimal matching to determine the pairwise ‘distance’ between each phase, as determined by the similarity of the event sequence within each phase. This analysis provides a validity check on the phase identification by determining if the phases in each case are different or similar to each other.

The characteristics of each phase are now discussed in three subsequent sub-sections in the context of existing ecosystem and institutional formation theories, including dominant design, social movement theory and institutional entrepreneurship, as well as the case narratives. For Phase II and Phase III, the key transition events for each case are also provided, providing further evidence of the existence of each phase.

**Phase I – Initiation**

Pre-existing environmental conditions meant that the emergence of an ecosystem was possible in each of the six cases (c.f. Hoffman, 1999). For Amazon, eBay, Salesforce and Wikipedia, the internet, specifically technology and increasing broadband provision, provided the ‘jolt’ that enabled the founders to successfully launch an ecosystem (Battilana et al., 2009; Greenwood et al., 2002; Hinings...
et al., 2004). For Facebook and Google, the poor performance of existing services (Friendster and Yahoo! respectively) led to efficiency mismatches that enabled the conditions for both services to launch (Seo & Creed, 2002). Attributes of the founders of each ecosystem, such as their network location and embeddedness, also suggested readiness for emergence (Battilana et al., 2009; Greenwood & Suddaby, 2006). For each case the founders were technically minded and cognizant of the opportunities that the internet provided. These conditions enabled the founders to identify new opportunities, gather resources, and introduce technically viable innovations (Greenwood et al., 2002; Hinings et al., 2004).

In this stage the early ecosystem participants cooperatively work together to co-opt key customers and partners (Davis & McAdam, 2000; Moore, 1993; Suarez, 2004). The cooperation of key customers and partners was more important for some cases than others. For example, Google’s strong relationship with Search Engine Watch, an important web-based commentator, and their constant attention to the maintenance of that relationship ensured a source of legitimacy. Similarly Amazon’s good relationship with the major book wholesalers ensured the success of the nascent digital service.

Each case cooperatively developed their nascent ecosystems. Each digital service was first developed in isolation, and access (and by implication membership) was then progressively released to beta testers (normally family and friends) until the digital service was released to general access. An exception here is eBay which opened up immediately to general access. However, uniquely among the other cases where there was significant event within the ecosystem from the beginning, eBay was seeded and tested by the founder himself before event started in earnest within the ecosystem. In all cases, once there were a small number of active participants within an ecosystem, the rich interaction between beta testers and the platform owner resulted in the co-production of the functionality of the digital service, and the definition of the logics that drove co-creation of value. This intense interaction addressed the technical potentialities for each ecosystem, as the feasibility of the digital service was de facto tested by early participants (Suarez, 2004). For those cases that faced (some little) competition during this phase, such as eBay, Facebook and Wikipedia, each competitor proposed a differing service and ecosystem design, leading to design competition (Hariharan & Prahalad, 1991). In parallel
with this, a set of regulative mechanisms was being created that enabled ecosystem emergence (Hoffman, 1999). For instance, very early on in their operation Amazon posted a ‘Bill of Rights’ for shoppers, and eBay published its ‘Ethical Guidelines’. Similarly, Wikipedia participants were releasing a series of policies that enabled the heterogeneous users to be able to create an encyclopedia. These regulations led to increasing returns so that participation in the ecosystem became viable, highlighting the role of incentives as a motivating force in ecosystem emergence (Hoffman, 1999; Scott, 2008).

Very early on in some ecosystems these early regulative institutions were complemented by the emergence of a normative institution (as first noted by Hoffman, 1999). As examples, on both eBay and Wikipedia, which had handed over much governance to ecosystem participants, particular behaviors of ecosystem participants were soon viewed as either constructive or to be discouraged. In the case of Wikipedia for instance, the behavior of ‘trolls’ \(^1\) was actively discouraged. On Facebook ecosystem participants very quickly developed idiosyncratic behaviors that gained normative significance. These were often beyond the regulative (and technical) control of the service, such as ‘friending competitions’ and the use of a photo to broadcast political affiliation. For those ecosystems which launched early in the history of the Internet, such as Amazon and eBay, much of this sense-making was as much cognitive as technical, as the potential of the internet, and hence the potential capabilities of the service, was not well understood at this time.

Summarizing, based upon the insights gained from theory and complemented through the particular histories of each case, Phase I consists of the initial idea and digital service development, resource gathering and early operation. With pre-existing conditions providing a fertile setting for the digital services to emerge, and following the progressive opening of the service to wider groups of participants, intense sense-making resulted in a set of technological attributes for the digital service and agreed-upon regulative, normative, and cognitive structures. Together these shaped the subsequent ecosystem emergence and created the necessary conditions for collective action (Hensmans, 2003; Hoffman, 1999).

---

\(^1\) A person who sows discord on the Internet by starting arguments or upsetting people (Wikipedia, retrieved on 20/07/2013).
Phase II – Momentum

In Phase II the ecosystem begins to grow rapidly, due the improving quality of the digital service and competitors’ entry (Agarwal & Bayus, 2002). Understanding of the ecosystem spreads rapidly as there is diffusion and dissemination of the ecosystems purpose and logic to potential participants (Greenwood et al., 2002; Hinings et al., 2004; Tolbert & Zucker, 1996). The accumulation of ecosystem participants generates a momentum that harnesses their distributed inputs that drives further adoption ecosystem through the ongoing regulation, evaluation, design, production and use of the digital service (Garud & Karnoe, 2003).

The case narratives provide insight into the key transition events that initiated this phase. For Amazon, Phase II commenced in late 1995 with the founder raising approximately $1 million from local Seattle investors. Following the investment round, the strategy became ‘Get Big Fast’, an explicit change from incremental organic growth to market domination, and business development staff were hired, an advertising agency brought on board, and an aggressive advertising campaign launched. For eBay, Phase II commenced at the beginning of 1997 with increasing numbers of participants, with over 330,000 completed auctions in the first quarter alone, giving rise to the moniker ‘the Great eBay Flood’. As a result, the service was unable to scale to meet demand and was continually crashing; despite this the number of participants continued to pour in. For Facebook, Phase II started with rapid growth in users after the summer 2004 college recess, aided by an automated process to add universities, as well as a VC investment round. To outflank emerging competitors, a ‘surround’ strategy was introduced, consisting of opening up access to the service to students at all the local universities where another social network had begun to take root. For Google, Phase II began with VC funding in June 1999, which resulted in extensive news coverage. This fuelled user adoption as the ‘Google’ brand became familiar to a wider audience beyond the core group of devotees. For Salesforce, Phase II commenced in February 2000 with high profile guerrilla marketing at the main competitor’s annual conference, as well as their own uniquely branded launch party. Together these led to a large influx of customer interest and a media firestorm, further inflamed when Siebel responded publically. For Wikipedia, Phase II commenced with escalating mainstream media attention...
Processes of ecosystem emergence

(such as the New York Times), rapidly increasing the number of contributors and the arrival of specialized tools and services to assist in editing.

During this stage the ecosystem opens up a large market by scaling up with both suppliers and partners to achieve maximum market coverage (Moore, 1993). Put differently, the nascent market that was created in the first phase is rapidly expanded in this phase through coalitions and outreach such as marketing (Suarez, 2004). In each case during this phase increasing network effects drove market size, to such an extent that eBay, Salesforce and Wikipedia experienced serious technological failures due to the overloading of their service. In these three cases the service outage and the resulting furor within the ecosystem almost destroyed confidence in (and the legitimacy of) the emerging ecosystem. Although Amazon, Facebook and Google also experienced service failures due to the explosive growth each was experiencing, in these cases they did not have the same effect on the legitimacy of the ecosystem. Of particular note here is Facebook which had an explicit policy of slowing growth through the limited opening of the service to new ecosystem participants to ensure service stability.

In all cases there was ongoing technical development of the digital service platform. However, the main functional dimensions of the platform had already been agreed by this time and were being implemented. Instead, as suggested by social movement theory and institutional entrepreneurship theory, there was increasing introduction of regulative mechanisms to coordinate activities, as well as rising participant commitments as normative aspects developed (Hoffman, 1999; Scott, 2008). In addition, cognitive mechanisms of coordination also began to emerge, through discourse between participants and the platform owner, receipt of awards from commentators, as well as high levels of press and societal interest that established the ecosystem and digital service in society. These provided validation and objectification of the ecosystem and service not only to the participants, but also for wider society (Lampel & Meyer, 2008; Rao, 1994; Scott, 2008). Processes of theorization, legitimation and dissemination led to the commitment of the participants, as well as to the establishment of a power structure (Greenwood et al., 2002; Hinings et al., 2004). However these processes were contested (c.f. Benford & Snow, 2000). For eBay and Wikipedia, where the governance of the ecosystem had been handed over to participants, any change in regulative or normative structures was often met with a large push back from the ecosystem participants. For those cases where governance remained more
under the control of the platform owner, the nature of the contestation varied. For instance, often the
digital service platform’s own tools were used against it when ecosystem participants protested. In the
case of Facebook, technological design changes to the service sometimes led to protests which were
collectively organized through platform functionality, such as groups. For those cases that remained
tightly controlled, such as Google and Salesforce, the contestation often originated outside the ecosys-
tem, but implicitly supported by ecosystem participants.

This process of socio-cognitive construction was aided by the aggressive use of marketing and
public relations, and technical tools which further raised the profile of the ecosystems. In particular, as
each ecosystem was finding themselves in a favorable position with respect to the emerging systemic
power relations, they acted to mobilize a bias in their favor to conquer the field (Hensmans, 2003).

Much of this framing event (Benford & Snow, 2000; Snow & McAdam, 2000) was undertaken with
the goal of actor mobilization and collective action (Battilana et al., 2009; Seo & Creed, 2002). For
instance, both Amazon and eBay, with their more consumer oriented ecosystems, focused marketing
attention on advertising that corresponded to festive dates within the consumer calendar (such as Fa-
thers’ Day and the December holiday season). Google celebrated festive dates through its Doodles. In
doing so they were framing themselves as part of the wider US institutional and societal environment.

Amazon and Google enabled their service to be deployed onto third party websites, thus not only driv-
ing more traffic to their ecosystems, but also advertising the service that they were offering to poten-
tial participants who had not yet discovered their service. In this manner they emphasized their tech-
nical efficiency aspects. Salesforce, with its business to business orientation, undertook aggressive
guerrilla marketing against the market dominant players, such as Seibel, emphasizing its ‘otherness’
(Snow & Benford, 1988; Snow, Rochford, Worden, & Benford, 1986). For Facebook, Google and
Wikipedia, marketing and public relations event focused not so much on driving participants to the
site but mitigating scandals, such as privacy concerns, that periodically erupted (Rao et al., 2000).

This phase is also characterized by the main competitive battles as the ecosystem strives to
dominate its challengers (Davis & McAdam, 2000; Moore, 1993; Suarez, 2004). Although much of
the dominant design literature emphasizes that this competitive battle is for ownership of the techno-
logical standards that underpin a service (Hariharan & Prahalad, 1991), in the example of the cases at
hand this was not such a concern as the main internet standards had been established. Instead, the main competitive thrust was to drive participant adoption so as to reach a tipping point and allow network effects to largely guarantee dominance; this was particularly so for those cases – eBay, Facebook, Google and Wikipedia – that relied on network effects for their lock-in. Thus much of the framing and market outreach activities were directed towards participant adoption. In addition to the actions of the platform owner, collectively the ecosystem participants also began to contest the domain vis-à-vis the competition (Davis & McAdam, 2000). For example, when Amazon received a poor review for the service that it offered, the offending media outlet was bombarded with complaints, and a wider recognition of the value of the digital service emerged from the independent (un-orchestrated) actions of the ecosystem participants. Similarly in the case of eBay, when a competitor placed a stooge onto the eBay boards to promote a competitive digital auction service, ecosystem members responded aggressively to remove it. For Salesforce, the success of their roadshows generated such high levels of goodwill towards the digital service and ecosystem that individual participants, alone or collectively, would evangelize the ecosystem independently of any guidance from Salesforce themselves. Indeed this event was so prevalent that marketing activities were established to help support it. Similarly for Wikipedia, individuals and groups of individuals would collectively act to maintain the coherence and stability of the service during large influxes of new ecosystem participants.

Summarizing, based upon the insights gained from theory and complemented through the particular histories of each case, in Phase II the ecosystem begins to grow rapidly, driven by increasing numbers of participants (drive by positive network effects), aggressive marketing, much press and societal interest, as well as competitor event. The market takes-off, and understanding of the ecosystem spreads rapidly as there is diffusion and dissemination of its value logic.

**Phase III – Control**

In Phase III, the ecosystem has now emerged and has socio-political and cognitive legitimacy (Greenwood et al., 2002; Hinings et al., 2004). Power relations stabilize and become more or less taken for granted and the platform owner is freer to act (Hensmans, 2003). Throughout this period regulative, normative and cognitive structures continue to develop, although these tend to become
self-reinforcing, enhancing the dominant position of the platform owner in the ecosystem and the ecosystem itself (DiMaggio & Powell, 1983; Hoffman, 1999). This powerful position is reinforced through the installed base and network effects working to their advantage, assuring market dominance (Suarez, 2004). With the platform owner in such a powerful position, the nature of competition shifts from market momentum to competition for profits (Hariharan & Prahalad, 1991). This reflects the reality that value appropriation processes would not be able to occur unless the platform owner was in a powerful position vis-à-vis the other ecosystem participants. With the ecosystem established as the undisputed leader focus now shifts to closer control of the activities of ecosystem participants and value appropriation.

The case narratives provide insight into the transition events that initiated this phase, and which further support the identification of this phase. For Amazon, already the undoubted market leader, Phase III began in early 2000 with an explicit shift to profitability and value appropriation. Staff were made redundant, unprofitable lines of business closed down, and functions and programs of the digital service changed to drive profitability. For eBay, Phase III was ushered in mid-2000 with the resolution of the Million Auction March, a (successful) participant protest to move one million transactions to competing auction services. Resolution meant that ecosystem participants now recognized eBay’s tensions to operate as a listed company, so conflict between the ecosystem and platform owner declined. As a consequence, the functionality of the digital service platform was enhanced with the goal to increase revenues, such as the introduction of ‘Buy-It-Now’. In addition, wider society and the press now viewed eBay as a neutral venue. For Facebook, Phase III began in 2009 as control was finally consolidated over the apps available through the platform, as well as formal acknowledgement of the wide ranging privacy concerns. A particular action here was opening up the terms of use to a vote for the users. In addition, internally operations had coalesced around a particular logic of value co-creation, and advertising and other value extraction techniques were now mature and continually tweaked to enhance revenue. For Google, Phase III began in 2004 as the service became the undisputed leader and the remaining competition (such as Microsoft and Yahoo!) were unable to maintain parity. Socially the service was now embedded in society, although there was an increasing wariness as to privacy. The shift to value appropriation was underlined with the winning of a court case which
allowed trademarked terms to be used as keywords and a subsequent change in advertising policies further entrenched the focus on value appropriation. For Salesforce, Phase III began in 2006 with the rectification of the technical issues that had caused serious service delivery, and launch of trust.salesforce.com, a service that publically detailed the status of the service delivery. Press interest began to wane as the software as a service model was accepted as viable. In addition, there was a shift to value appropriation with the hiring of staff to focus on revenue collection, and release of monetization strategy documents. For Wikipedia, Phase III commenced at the end of 2005 with the media crisis of the Seigenthaler incident, who wrote an op-ed piece in USA Today, entitled ‘A False Wikipedia Biography’. This crisis led to an imposition of tighter editorial control, which in contrast to previous attempts, was welcomed by the ecosystem tired of the previous scandals and edit wars. In addition the Wikipedia Foundation formalized its fundraising process and technologically the service was stabilized.

As well as the importance of maintaining strong bargaining power in relation to other participants in the ecosystem, the platform owner needs to provide a compelling vision for the future (Gawer & Phillips, 2013; Moore, 1993; Sharapov, Thomas, & Autio, 2013). In this manner the platform owner in this phase corresponds to the ‘platform leader’ of Gawer and Cusumano (2002) and the ‘keystone’ of Iansiti and Levien (2004). For instance, Gawer and Henderson (2007) and Gawer and Phillips (2013) emphasize that a platform owner must use a number of mechanisms at a either firm level or ecosystem level in order to ensure ecosystem dominance. At the firm level, such mechanisms include organizational structure and processes that signal that the future viability of the ecosystem and the intention of the platform owner. As examples, Amazon underwent a wide ranging restructure to enable profitability; eBay instigated specialist teams to signal its intentions to collaborate with the ecosystem; and Google appointed senior executives to global communications and outreach roles. Other mechanisms involve discourse, such as explicitly framing a strategy for the future. As examples, eBay publically defended its actions regarding antiques auctions in the face of criticism and outlined its future strategy for the first time, Salesforce publically announced its monetization strategy, and Wikipedia issued a series of press releases detailing how the service was to be run into the future.
Taken together, these strategies highlight how a dominant platform owner in an ecosystem can maintain its position within the ecosystem, and also signal its viability and success beyond the ecosystem.

Summarizing, based upon the insights gained from theory and complemented through the particular histories of each case, in Phase III, the focus of event moves from expansion to control and value appropriation. Power relations stabilize and become more or less taken for granted, technological change is driven by maintenance and value extraction reasons. Regulative, normative and cognitive structures continue to develop, becoming self-reinforcing, enhancing the dominant position of the platform owner in the ecosystem and the ecosystem itself, and the ability of the platform owner to control and extract value.

**Optimal matching validation**

In the analysis above, three phases have been identified through direct inspection. The goal of this section is to prove that each phase within each case is dissimilar to the others, thus validating the existence of each phase. To do so the event sequences of each phase are compared through optimal matching to determine a pairwise ‘distance’ between each phase within a case, as determined by the (dis)similarity of their phase sub-sequences. In this analysis the phase event sequences for each case are compared with the other phase event sequences within that case. Eighteen phase-pair distances are calculated: for Amazon, one for Phase I-II, one for Phase II-III, and one for Phase I-III; for eBay, one for Phase I-II, one for Phase II-III, and one for Phase I-III; and so on. Following this procedure there were 18 mean distances and 18 standard deviations – one each for each phase-pair (3) of each case (6). These results are detailed in Table 4 below.

For all phases in all cases except Amazon and eBay, the sign of the variance is positive for all phase comparisons, providing evidence that each of the phases is not similar to the others within a case. For Amazon, Facebook, Google, Salesforce and Wikipedia, some phases are significantly different. Phase I-III of Salesforce and Phase II-III of Wikipedia are the most significantly different (p<0.001). This means that the distance between the phase-pairs is greater than 99.9% of the 10,000 random phase-pairs with the same code frequency. Similarly Phase I-III of Google and Phase I-III of Wikipedia are also significantly different (p< 0.01). This means that the distance between the two
phases is greater than 99% of the 10,000 random phase-pairs with the same code frequency. Phase II-III of Amazon, Phase I-III of Facebook and Phase II-III of Google are also significantly different (p<0.05). This means that the distance between the phase-pairs is greater than 95% of the 10,000 random phase-pairs with the same code frequency. The remainder of the phase-pairs are not significantly different or similar, further indicating that within each case there is no similarity between the phases.

eBay has positive variances from random mean for Phase I-II and Phase II-III, however the variance from the random mean for Phase I-III is negative but not significant. As the distance from Phase I to Phase II is positive, and Phase II to Phase III is positive, these results still suggest that the sequences within each of the three phases differ and hence validate the phase identification. Amazon also has a non-significant negative variance for Phase I-II, however the size of the variance is small (0.001). This suggests that Amazon may have begun Phase II earlier than their explicit strategy of “Get Big Fast” would suggest; perhaps network effects had already begun to drive the momentum phase.

**Analysis III: Phase sequence comparison**

Analysis II above has suggested that there are three phases of ecosystem emergence that occur during the development of an ecosystem. The purpose of this section is to apply optimal matching to determine how similar or different each phase is across cases. In this analysis the event sequences of a phase are compared between the six cases. Forty five phase-case-pair distances are calculated: fifteen for Phase I, fifteen for Phase II and fifteen for Phase III. Alternatively expressed, three pairwise distances are calculated for Amazon-eBay (one for each phase), three for Amazon-Facebook, and so on. Following this procedure there were 45 mean distances and 45 standard deviations – one for each phase of each case-pair. Table 4 below presents the results.

[Insert Table 4 around here]

For Phase I the results indicate that, in contrast to the complete sequence comparison of Analysis I, a large majority of the phase-case-pairs are no different to the random mean and the significant differences between sequences have disappeared. This indicates that the difference between each phase-case-pair is no different than the average of 10,000 phase-case-pairs calculated from random sequences with the same code frequency. Furthermore, the majority of these phase-case-pair comparisons now have a negative variance from their random mean. This implies that each phase-case-pair is
tending towards a similar event sequence. This finding is emphasized through the significant similarity of the Amazon-Google event sequences for this phase (p<0.01), where the distance of the event sequences for the phase-case-pair, relative to one another, is less than 90% of the 10,000 pairwise distances calculated from the randomly generated sequences. In contrast, the phase-case-pair of Salesforce-Wikipedia is significantly different (p<0.01), where the distances of the event sequences for the phase-case-pair, relative to one another, is higher than 95% of the 10,000 pairwise distances calculated from the randomly generated sequences. Similarity the phase-case-pairs of Amazon-eBay, eBay-Google and Google-Wikipedia are significantly positive, but only at a low level of significance (p<0.1). In general, the disappearance of the significant differences between each case-pair, and the negative variance from the random mean, suggest there is some similarity between the phases across the cases.

In contrast, in Phase II the similarity that was observed in Phase I has mostly disappeared and instead there is now predominantly dissimilarity between the phase-case-pairs. The variances for all the phase-case-pairs, except Amazon-eBay and Google-Salesforce, are positive from their respective random means with eight significant, suggesting that these phase-case-pairs are either dissimilar or tending towards being dissimilar. Amazon-Google and Facebook-Google are significantly different (p<0.01) with eBay-Salesforce less so (p<0.1). Strikingly all phase-case-pair comparisons with Wikipedia are highly significantly different (p<0.001). This means that the distance of the event sequences for each phase-case-pair, relative to one another, are higher than 99.9% of the 10,000 pairwise distances in the randomly generated sequences. The negative variances of Amazon-eBay and Google-Salesforce are insignificantly negative. These results suggest that there are no longer any similarities in the underlying processes during emergence, and that, at best, any two ecosystem emergence process is no more similar than any two random sequences.

Phase III appears to maintain the dissimilarities between the sequences for all cases.² In particular, the variance from the respective random mean for the vast majority of phase-case-pairs is posi-

² Caution should be exercised in attempting to extract too much insight as the results are right censored due to continuing control activities past the end date that are not included in the analysis.
 Processes of ecosystem emergence

tive, with nine significant. This suggests that the sub-sequence for each phase-case-pair is either dis-
similar or tending towards being dissimilar. In contrast to the previous phase, where there was strong
significant differences for all phase-case-pairs with Wikipedia and weaker significant differences be-
tween the other phase-case-pairs, the levels of significance are much reduced for Wikipedia although
it remains significantly different to all other cases in this phase. For instance, Amazon-Wikipedia and
Google-Wikipedia are now the only phase-case-pairs that are highly significantly different (p<0.001).
Facebook-Wikipedia and Salesforce-Wikipedia remain significantly different although less so
(p<0.05). Google-Salesforce now exhibits a high significant difference (p<0.001), followed by Amaz-
on-Facebook (p<0.01), Facebook-Salesforce (p<0.05) and eBay-Salesforce (p<0.1). The exceptions
are Amazon-eBay and Amazon-Salesforce which has a non-significant negative variance from its ran-
dom mean, suggesting that the respective sub-sequences within this phase tend to similarity. These
results suggest that significant differences remain between each of the sub-sequences, but with re-
duced levels of significance.

Taking the three phases together, this analysis suggests that ecosystems begin with fairly simi-
lar processes of resource acquisition and technological development (Phase I) which then have unique
momentum processes of ecosystem growth, competition, and press and societal interest (Phase II). In
Phase III each case settles into their own unique processes of value appropriation and ecosystem con-
trol. Figure 2 below plots the variance of the distance of each phase-case-pair from their random
means across phases.

[Insert Figure 2 around here]

In this figure the x-axis consists of the three phases, and the y-axis the variance of each phase-
case-pair from its respective random mean. An additional measure, the average of the phase-case-pair
variances from their respective random means for each phase, is also plotted. This figure suggests that
the differences in the event sequences increase from one phase to the next. In Phase I the sub-se-
quencies of each phase-case-pair tend to similarity. In Phase II the sub-sequences of each phase-case-
pair tend to dissimilarity, and in Phase III there is a greater tendency to dissimilarity, despite de-
creased levels of significance.
DISCUSSION

The above analysis indicates that the processes of ecosystems creation are unique to the ecosystem – each ecosystem emerges in its own manner. However the analysis above also suggests that despite the unique sequence of emergence, there are three common phases – Initiation, Momentum and Control. The Initiation phase consists of the initial idea and digital service development, resource gathering and early operation. In the Momentum phase, the ecosystem begins to grow rapidly, with investment, increasing numbers of participants (driven by positive network effects), aggressive marketing, much press and societal interest, as well as competitor event. In the Control phase, the focus of event moves from growth to control and value appropriation, as the ecosystem is established as the undisputed leader. A further finding is that there is similarity in the structure of the Initiation activities between differing ecosystems, however as the ecosystem emerges through the Momentum and Control phases, the early differences within the ecosystem result in significant process differences.

The finding that the complete emergence process of an ecosystem is unique has a range of support within the institutional formation literature. For instance, theorists have noted that the institutional logics and power relationships within a field are unique and that they emerge uniquely (Fligstein, 2001; Scott, 2008). Schilling (2009) and Thomas et al. (2014) suggest that there are multiple pathways to the emergence of an ecosystem when the firm is able to adjust the level of ‘openness’ of the technological design. Others have noted that firm level capabilities, such as financial resource acquisition, alliance formation, personnel recruitment, and product design, emerge in different ways although appear similar once created (Eisenhardt & Martin, 2000).

The finding of three phases provide substantive empirical support for the ecosystem emergence model of Moore (1993) and also support models of dominant design emergence (Suarez, 2004) and the sequential process models of institutional entrepreneurship (Greenwood et al., 2002; Hinings et al., 2004). Moreover these findings also support the suggestions of an evolutionary dynamic in both dominant design and institutional entrepreneurship theory. In particular, institutional emergence has been suggested to be driven by a variance, selection, retention mechanism drives (Barley & Tolbert, 1997; Van De Ven & Garud, 1994). This particular dynamic was observed in the cases where there
was strong community development and identity, such as eBay and Wikipedia. The findings also support the institutional models that argue that institutionalization and collaboration are interdependent. For instance, Phillips, Lawrence, and Hardy (2000) suggest that organizational fields provide the rules and resources upon which collaboration is constructed, and collaboration provides a context for the on-going processes of structuration that sustain the organizational fields of the participants.

However, the finding of increasing dissimilarity across successive phases provides evidence counter to what would be suggested by institutional theory. Given that in the latter phase the ecosystems are moving towards control, value appropriation and optimization, it would be reasonable to consider that wider institutional concerns, particularly regulatory pressures, could also exert a homogenizing influence on activities in later phases (DiMaggio & Powell, 1983). For instance, as the majority of cases experienced an IPO (Facebook and Wikipedia excepted), the resource activities required for the preparation and execution of the IPO, and particular the subsequent requirements for reporting, could impact both the frequency and pattern of resource, institutional and contextual events. Similarly, case responses to societal concerns such as privacy (Facebook and Google), viability (Amazon and Salesforce) and trust and fraud (eBay and Wikipedia), as well as engagement with regulators and other institutions, could also influence the both the frequency and patterns of institutional and contextual events. However, Figure 2 above demonstrates that there is no homogenizing effect and instead value appropriation and control activities tend to further differentiate the underlying ecosystem event sequences.

An alternative explanation for the increasing dissimilarity across phases between the emergence sequences of ecosystems is imprinting and path dependence. Organizations are imprinted by the forces present at the time of their creation, leading to them to assume specific characteristics that tend to be carried forward (Stinchcombe, 1965). For instance, dominant design theorists have outlined the formative influence that pre-existing technologies have on the emergence of a new design (Anderson & Tushman, 1990; Murmann & Frenken, 2006). Furthermore, differences in technologies, beliefs and other founding conditions can lead to differing attributes of an ecosystem (Scott, 1983). Social movement theorists have also emphasized that the form of a social movement bears the imprint of the specific opportunity that gives it life (McAdam, McCarthy, & Zald, 1996). For example, each of the
founders came from different backgrounds which shaped the subsequent processes of emergence. Bezos and Benioff (Amazon and Salesforce) both came from corporate backgrounds, and Page and Brin and Zuckerberg (Google and Facebook) were students, and these different starting positions influenced the service development process and resource acquisition of the respective ecosystems. Similarly the developer background of Omidyar and Wales (eBay and Wikipedia) meant that the hacker ethos and ideas of open participation were integral to the early development of eBay and Wikipedia. As a consequence, the conditions in which each platform ecosystem began influenced the constituent events and dynamics that drove the emergence of the ecosystem.

Path dependence occurs when the end state depends on the particular sequence of unfolding events (Arthur, 1989; David, 1985; Garud & Karnoe, 2001). The decisions which are taken throughout the emergence of an ecosystem also shape its final state, as each provides the raw material that subsequent decisions are made (David, 1985). In particular, the early decisions of founders, as well as the early influence of customers and other stakeholders, can have a strong influence on the structure and dynamics of an ecosystem due to the amplification effect of structuration processes (Lawrence & Phillips, 2004). Similarly, processes such as breakthrough, translation and bricolage that shape the ecosystem mean that an emergence sequence will be idiosyncratic to the decisions made (Campbell, 2005; Garud & Karnoe, 2003). For instance, each case faced challenges – technological failure (eBay and Salesforce), internal conflict within the ecosystem (eBay, Facebook and Wikipedia), and intense societal pressure (Google and Facebook) – each of which was addressed uniquely according to the demands and interests of the service, ecosystem and founders. How each case addressed these challenges influenced the subsequent actions that were taken. For instance, Facebook’s habit of initially ignoring the demands or concerns of ecosystem participants, responding arrogantly, and then eventually backing down, led to feature retractions, ongoing conflict with both ecosystem participants and society, and a subsequent requirement for specialized resources and technologies that was not shared with other cases. Thus the decisions taken both enable and constrain the activities of the ecosystem, as it became embedded in paths that it tried to shape and which in turn shaped it (Garud & Karnoe, 2003). As a consequence the relationship between a platform owner, the ecosystem and society is the
product its own distinct history, reflecting the assimilation of values and demands over time and the residues of past processes (Astley & Van De Ven, 1983; Scott, 1983).

CONCLUSION

To conclude, the emergence processes of six digital service platform ecosystems were empirically analyzed using phasic analysis to determine the similarities and differences between each creation sequence. Three phases of ecosystem emergence were suggested – Initiation, Momentum, and Control – where cross-case similarities decrease from one phase to the next as the ecosystem emerges. In particular the cross-case analysis emphasizes the effect of imprinting and path dependency on ecosystem emergence.

A key contribution of this research is to the ecosystems literature. This is one of the first studies to explicitly consider the processes of ecosystem emergence, and has provided empirical support for the existing model of Moore (1993). In addition the research pioneers an institutional approach to ecosystems, acknowledging the many institutional factors that lie in extant research and developing them into a coherent analysis. A further contribution is to the prescriptive literature. At present the managerial literature has focused extensively on the importance of interdependence (Adner, 2006) and platform ecosystem management strategies (Cusumano & Gawer, 2002). In addition the prescriptive literature that does address platform ecosystem emergence has taken a piecemeal approach, such as the steps for getting the different sides of a multi-sided market on board (Hagiu & Eisenmann, 2007), or differing strategies to achieve platform leadership (Gawer & Cusumano, 2008). As such the identification of the three phases of emergence, as well as the increasing divergence in processes that results from imprinting and path dependency, not only provide a more holistic approach, but also indicate that following best practices in ecosystem contexts may not be recommended.

There are a number of limitations of this research. As the empirical research considered ecosystems in digital service platform contexts, these results may not be generalizable beyond internet-based businesses. However within the digital service context these results are salient, as they encompass the six main business models. Further, with the continuous diffusion of ICT technologies across industry sectors and the associated spread in the use of platform strategies, it is feasible that generalizability
Processes of ecosystem emergence

may increase. A further limitation is that the success bias in the results may have painted a biased picture of ecosystem emergence. A further limitation concerns the methodology: the data collection techniques utilized mean that only a sample of the possible events were identified and as such the analyzed sequences are incomplete. A more technical limitation concerns the fact that each event is treated equally in importance, meaning that minor events have the same effect on the sequence as one which was game-changing.

In conclusion, we have found that the processes of ecosystem emergence are unique to each ecosystem, and that each successful ecosystem emergence has three common phases. There are path dependency and imprinting effects suggested by the decreasing similarity of the emergence processes of each cases in successive phases. It is hoped that these findings will inspire researchers to further investigate how ecosystems emerge.
REFERENCES


Processes of ecosystem emergence


<table>
<thead>
<tr>
<th>Year founded</th>
<th>Amazon</th>
<th>eBay</th>
<th>Facebook</th>
<th>Google</th>
<th>Salesforce</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Founder</td>
<td>Jeff Bezos</td>
<td>Pierre Omidyar</td>
<td>Mark Zuckerberg</td>
<td>Sergei Brin &amp; Larry Page</td>
<td>Marc Benioff</td>
<td>Jimmy Wales</td>
</tr>
<tr>
<td>Founder back-ground</td>
<td>Hedge Fund executive</td>
<td>Software engineer</td>
<td>Undergraduate student</td>
<td>Doctoral students</td>
<td>Software executive</td>
<td>Bank trader</td>
</tr>
<tr>
<td>Founding ideal</td>
<td>Commercial success</td>
<td>Create the perfect market</td>
<td>Make the world more open</td>
<td>Make all the world’s information freely accessible and available</td>
<td>Free the business from enterprise software</td>
<td>Free access to the world’s knowledge</td>
</tr>
<tr>
<td>Competitors at founding</td>
<td>Books.com; clbooks.com</td>
<td>None</td>
<td>MySpace.com; Friendster.com; Yahoo!; Altavista; Excite</td>
<td>None</td>
<td>None</td>
<td>Encyclopedias</td>
</tr>
<tr>
<td>Initial Funding</td>
<td>Founder &amp; Family</td>
<td>Founder</td>
<td>Founders</td>
<td>Founder</td>
<td>Founder &amp; angel investors</td>
<td>Founder</td>
</tr>
<tr>
<td>Year became profitable</td>
<td>2001</td>
<td>1995</td>
<td>TBD</td>
<td>TBD</td>
<td>2002</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>Business Model</td>
<td>Retail</td>
<td>Brokerage</td>
<td>Social advertising</td>
<td>Search advertising</td>
<td>Subscription</td>
<td>Open source</td>
</tr>
<tr>
<td>Marketing Approach</td>
<td>Word of mouth</td>
<td>Word of mouth</td>
<td>Word of mouth</td>
<td>Word of mouth</td>
<td>Guerrilla</td>
<td>Word of mouth</td>
</tr>
<tr>
<td>Major Crisis Event</td>
<td>Dotcom crash</td>
<td>Technology failure</td>
<td>Privacy</td>
<td>Privacy</td>
<td>Technology failure</td>
<td>Quality</td>
</tr>
<tr>
<td>Effect of dotcom crash</td>
<td>Major; potential for bankruptcy</td>
<td>Minor; slump in listings</td>
<td>Not Applicable</td>
<td>None</td>
<td>Mid; cash-flow issue solved through fee re-</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Ecosystem participants</td>
<td>Customer, wholesaler suppliers, distributors, partner websites, associate websites, marketplace vendors</td>
<td>Customer, escrow services, insurance services, information services, software, packaging materials, collection</td>
<td>User, app developers, advertisers, Connect enabled websites, ‘Like’ enabled websites</td>
<td>Customer, advertisers, advertising affiliates, search affiliates, app developers</td>
<td>Customer, consultants, solution developers, AppExchange developers, software partnerships, publishing</td>
<td>Users, editors, bot developers, add-on developers</td>
</tr>
</tbody>
</table>
### TABLE 2 – Code Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Amazon</th>
<th>eBay</th>
<th>Facebook</th>
<th>Google</th>
<th>Salesforce</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resource</strong></td>
<td>27.55%</td>
<td>25.25%</td>
<td>18.29%</td>
<td>22.45%</td>
<td>15.18%</td>
<td>7.05%</td>
</tr>
<tr>
<td></td>
<td>(197)</td>
<td>(126)</td>
<td>(182)</td>
<td>(134)</td>
<td>(114)</td>
<td>(52)</td>
</tr>
<tr>
<td><strong>Technological</strong></td>
<td>25.03%</td>
<td>18.31%</td>
<td>34.67%</td>
<td>33.33%</td>
<td>11.98%</td>
<td>29.54%</td>
</tr>
<tr>
<td></td>
<td>(179)</td>
<td>(91)</td>
<td>(182)</td>
<td>(199)</td>
<td>(90)</td>
<td>(218)</td>
</tr>
<tr>
<td><strong>Institutional</strong></td>
<td>30.77%</td>
<td>43.06%</td>
<td>23.62%</td>
<td>22.61%</td>
<td>52.60%</td>
<td>40.51%</td>
</tr>
<tr>
<td></td>
<td>(220)</td>
<td>(214)</td>
<td>(124)</td>
<td>(135)</td>
<td>(395)</td>
<td>(299)</td>
</tr>
<tr>
<td><strong>Contextual</strong></td>
<td>16.64%</td>
<td>13.28%</td>
<td>23.43%</td>
<td>21.61%</td>
<td>20.24%</td>
<td>22.90%</td>
</tr>
<tr>
<td></td>
<td>(119)</td>
<td>(66)</td>
<td>(123)</td>
<td>(129)</td>
<td>(152)</td>
<td>(169)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>(715)</td>
<td>(497)</td>
<td>(525)</td>
<td>(597)</td>
<td>(751)</td>
<td>(738)</td>
</tr>
</tbody>
</table>

Note: The number in parentheses represents actual code count.

### TABLE 3 – Optimal Matching of Complete Case Sequences

<table>
<thead>
<tr>
<th></th>
<th>Amazon</th>
<th>eBay</th>
<th>Facebook</th>
<th>Google</th>
<th>Salesforce</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amazon</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>eBay</strong></td>
<td>0.019 +</td>
<td>0.000</td>
<td>0.013</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.643)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>Facebook</strong></td>
<td>0.029 **</td>
<td>0.013</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.691)</td>
<td>(0.789</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>Google</strong></td>
<td>0.044 ***</td>
<td>0.037 **</td>
<td>0.036 **</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.719)</td>
<td>(0.760)</td>
<td>(0.700)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>Salesforce</strong></td>
<td>0.008</td>
<td>0.043 ***</td>
<td>0.022 *</td>
<td>0.030 **</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.767)</td>
<td>(0.676)</td>
<td>(0.778)</td>
<td>(0.815)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>Wikipedia</strong></td>
<td>0.108 ***</td>
<td>0.051 ***</td>
<td>0.074 ***</td>
<td>0.082 ***</td>
<td>0.069 ***</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.855)</td>
<td>(0.717)</td>
<td>(0.738)</td>
<td>(0.785)</td>
<td>(0.768)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001; numbers in parentheses are actual distances.
## TABLE 4 – Internal Phase Optimal Matching

<table>
<thead>
<tr>
<th></th>
<th>Amazon</th>
<th>eBay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phase I</td>
<td>Phase II</td>
</tr>
<tr>
<td>Phase I</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Phase II</td>
<td>-0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Phase III</td>
<td>0.011 *</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.678)</td>
</tr>
<tr>
<td>Facebook</td>
<td>Phase I</td>
<td>Phase II</td>
</tr>
<tr>
<td>Phase I</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Phase II</td>
<td>0.007</td>
<td>0.000</td>
</tr>
<tr>
<td>Phase III</td>
<td>0.075 *</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.691)</td>
</tr>
<tr>
<td>Salesforce</td>
<td>Phase I</td>
<td>Phase II</td>
</tr>
<tr>
<td>Phase I</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Phase II</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Phase III</td>
<td>0.269 ***</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.815)</td>
</tr>
</tbody>
</table>

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001; numbers in parentheses are actual distances.
TABLE 5 – Cross-Case Optimal Matching

Phase I optimal matching cross case comparison

<table>
<thead>
<tr>
<th></th>
<th>Amazon</th>
<th>eBay</th>
<th>Facebook</th>
<th>Google</th>
<th>Salesforce</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eBay</td>
<td>0.081 +</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>0.009</td>
<td>0.061</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google</td>
<td>0.141 **</td>
<td>0.102 +</td>
<td>-0.005</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salesforce</td>
<td>0.069</td>
<td>0.074</td>
<td>-0.068</td>
<td>0.003</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Wikipedia</td>
<td>0.012</td>
<td>0.074</td>
<td>0.050</td>
<td>0.095 +</td>
<td>0.142 **</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001; numbers in parentheses are actual distances.

Phase II optimal matching cross case comparison

<table>
<thead>
<tr>
<th></th>
<th>Amazon</th>
<th>eBay</th>
<th>Facebook</th>
<th>Google</th>
<th>Salesforce</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eBay</td>
<td>-0.008</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>0.009</td>
<td>0.008</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google</td>
<td>0.050 **</td>
<td>0.012</td>
<td>0.056 **</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salesforce</td>
<td>0.005</td>
<td>0.025 +</td>
<td>0.011</td>
<td>-0.014</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Wikipedia</td>
<td>0.110 ***</td>
<td>0.074 ***</td>
<td>0.073 ***</td>
<td>0.075 ***</td>
<td>0.065 ***</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001; numbers in parentheses are actual distances.

Phase III optimal matching cross case comparison

<table>
<thead>
<tr>
<th></th>
<th>Amazon</th>
<th>eBay</th>
<th>Facebook</th>
<th>Google</th>
<th>Salesforce</th>
<th>Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eBay</td>
<td>-0.001</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>0.046 *</td>
<td>0.064</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google</td>
<td>0.009</td>
<td>0.058</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salesforce</td>
<td>-0.021</td>
<td>0.095 +</td>
<td>0.126 *</td>
<td>0.225 ***</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Wikipedia</td>
<td>0.167 ***</td>
<td>0.049</td>
<td>0.072 *</td>
<td>0.165 ***</td>
<td>0.064 *</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001; numbers in parentheses are actual distances.
FIGURE 1 – Outline of analyses conducted

Analysis I
Complete Sequence Comparison
Goal: Determine how similar (or different) the overall ecosystem emergence process is across cases
Methods: Optimal Matching

Analysis II
Sub-Sequence Investigation
Goal: Determine if there are sub-sequences ("phases") within each case which could then be compared
Methods: Direct Inspection

Analysis III
Sub-Sequence Comparison
Goal: Determine if there are similarities (or differences) in each sub-sequence across cases
Methods: Optimal Matching

FIGURE 2 – Phase-case-pair variances from random mean distance