

# Niche Formation in the Mashup Ecosystem

Michael Weiss<sup>1</sup>, Solange Sari<sup>1</sup>, and Nadia Noori<sup>1</sup>,

<sup>1</sup> Institute for Technology Entrepreneurship and Commercialization, Carleton University,  
1125 Colonel By Dr, Ottawa, ON K1S 5B6, Canada  
[michael.weiss@carleton.ca](mailto:michael.weiss@carleton.ca), [ssari@connect.carleton.ca](mailto:ssari@connect.carleton.ca), [nmoori@connect.carleton.ca](mailto:nmoori@connect.carleton.ca)

**Abstract.** Mashups are constructed within a complex ecosystem of data providers who offer open APIs to users, users who combine APIs into mashups, and platforms like the ProgrammableWeb, or Mashape that facilitate the construction and publication of mashups. In this paper, we argue that the evolution of the mashup ecosystem can be explained in terms of ecosystem niches anchored around hub or keystone APIs. To demonstrate the formation of niches in the mashup ecosystem, we model groups of related mashups as species, and reconstruct the evolution of mashup species through phylogenetic analysis.

**Keywords:** Mashups, Recombinant Innovation, Ecosystems, Keystones, Growth, Evolution, Speciation, Niche Formation.

## 1 Introduction

Mashups are situational applications that combine services provided by third parties through open APIs, as well as user-owned data sources [1]. The creation of mashups is supported by a complex ecosystem of interconnected data providers, mashup platforms, and users [2, 3]. In our own previous work we have examined the structure and evolution of the mashup ecosystem [3], and mashup speciation [4].

Our goal in this paper is to explain the evolution of the mashup ecosystem through the lens of the speciation. The paper provides evidence of the formation of niches within the mashup ecosystem that are anchored around hub or keystone APIs, and offers techniques for analyzing niche formation based on phylogenetics.

In the following, we first review related work on recombinant innovation, ecosystems, and technology evolution. We then describe our research method, and report on our findings on niche formation in the mashup ecosystem. We conclude the paper with a discussion of our findings and areas for future work.

## 2 Related Work

### 2.1 Recombinant Innovation

Innovation can be described as a process of recombination, i.e. the construction of new ideas from existing ones [5]. The notion of recombinant innovation is closely linked to that of modularity, which allows the creation of new products by mixing and matching components [6]. Imitation is one of the primary means of innovation [7]. When developers are creating new mashups, they often start with another mashup as a

“blueprint” for their own mashups [4]. Simulation models confirm that mashup development is largely the result of a copying process [6].

## **2.2 Ecosystems**

In an ecosystem, value is co-created by ecosystem members who both collaborate and compete [9]. Research on the mashup ecosystem has found that the distribution of API use follows a power law, implying that the ecosystem has a small number of hub APIs that provide the base functionality for a large number of complementors [3]. Hubs naturally emerge in ecosystems [9]. These hubs provide the stable common assets for the mashup ecosystem. Co-creation of new functionality in the mashup ecosystem is anchored around those common assets.

As observed by [10] for innovation ecosystems, these hubs can be grouped into multiple tiers of keystones. The success of an ecosystem requires providing access to information on the innovation architecture, participating in standardization efforts, as well as investing in the providers of complements [11]. These activities, performed by a focal company, facilitate cumulative innovation. An example is Google’s ecosystem [12]. At its core is Google’s vast computing infrastructure that enables Google to leverage third-party innovation while maintaining architectural control.

## **2.3 Technology Evolution**

Adner & Levinthal [13] study the emergence of new technologies through the lens of biological speciation. They define speciation as the separation of one evolving population from its antecedent population. Speciation allows populations to follow different evolutionary paths. There are two processes at work: adaption (when technology becomes adapted to the needs of a particular niche), and resource abundance (how many resources are available in a niche to sustain the innovation).

Based on mechanisms of speciation and extinction, Weiss & Sari [4] describe an evolutionary model that generate clusters of mashups, that is, niches in the mashup ecosystem, and estimate the diversification of the mashup ecosystem over time. The model represents a mashup as an individual of an evolutionary species. They reconstruct the evolution of mashups through phylogenetic analysis.

# **3 Research Method**

## **3.1 Data collection**

The data for our study was collected from of the ProgrammableWeb ([www.programmableweb.com](http://www.programmableweb.com)), a repository of open APIs and mashups. There are other websites that provide similar services, such as Mashape ([www.mashape.com](http://www.mashape.com)). However, the ProgrammableWeb provides the most comprehensive collection. It should be noted, though, that the ProgrammableWeb only lists publicly accessible mashups, and internally used enterprise mashups are not listed.

The extracted data was used to produce datasets for the population of APIs and mashups in the mashup ecosystem. The API dataset included the name, publication date and category of each API, and the mashup dataset included mashup name,

publication date, tags, and APIs used. The sampling period was 04/09/2005 (inception of the mashup ecosystem) to 22/01/2013, and it includes 2656 days. Over this time period, a total of 8245 APIs (of which 1186 APIs were used in at least one mashup) and 6868 mashups were published in the repository.

### 3.2 Data analysis

To identify hub APIs we compute the contributions of each API to mashups and rank by the number of mashups they contribute to. We then determine the set of APIs that is responsible for 1/3 of the contributions to mashups (this cutoff is chosen according to Bradford’s law [14]). This provides a set of candidate hub APIs to be examined more closely by constructing phylogenetic trees in the next stage of the analysis.

To assess the relative impact that hub APIs have on the mashup ecosystem over time, we also compute their cumulative contributions. These curves will have the typical S-shape of an adoption cycle [15]. The inflection points in the S-curves mark events of significant interest to understanding the evolution of the ecosystem.

Finally, we reconstruct the evolution of the mashup ecosystem by constructing a phylogenetic tree of mashup species. A phylogenetic tree captures the evolutionary relationships between species of mashups. The tree was estimated using the neighbor-joining method [16], as implemented in the ape library (ape.mpl.ird.fr) in the statistics package R (www.r-project.org). A mashup species is a group of similar mashups.

Similar mashups will appear in related branches of the tree. The similarity of two mashups can be computed as the overlap in their APIs using the Jaccard index [4]. Each mashup can be represented as a set of APIs. For example, given two mashups  $m1 = \{\text{Google Maps, Flickr}\}$  and  $m2 = \{\text{Flickr, Amazon eCommerce}\}$ , the similarity is  $1/3 = 0.33$ , as both mashups share Flickr, and the total number of elements is 3.

## 4 Findings

### 4.1 Growth of hub APIs

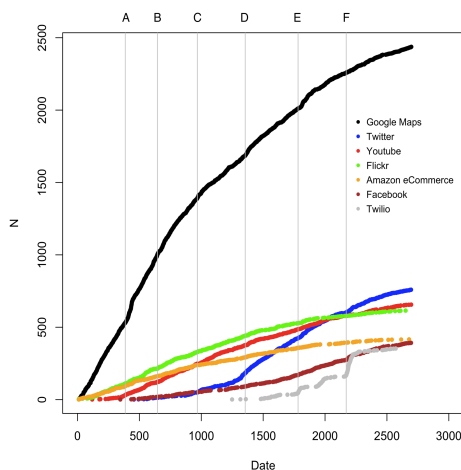
Table 1 lists the candidate hub APIs and their contributions together with their date of introduction and category assigned to them on submission.

**Table 1.** Hub APIs and their contributions to mashups

Core API	Contribution	Date published	Category
Google Maps	2437	2005-12-05	Mapping
Twitter	759	2006-12-08	Social
YouTube	656	2006-02-08	Video
Flickr	615	2005-09-04	Photos
Amazon eCommerce	416	2006-04-04	Shopping
Facebook	392	2006-08-16	Social
Twilio	353	2009-01-10	Telephony

The graph in Fig. 2 shows the cumulative contribution of each API. Initially, adoption of an API is low. This is followed by a period of steep growth, and subsequent saturation. Some of the curves (eg Google Maps) only show the steep

growth and subsequent saturation portions of the S-curve. Here, we can assume that the early stages of adoption precede the creation of the ProgrammableWeb. In other cases (eg Twilio), the whole adoption cycle is captured within the graph. The growth stage is when an API will make its greatest impact on the ecosystem. These are periods where one would expect “bursts of innovation” [13] driven by this API.



**Fig. 2.** Contributions of hub APIs over time. Date is the number of days since inception of the mashup ecosystem. N is the number of mashups an API contributes to. Vertical lines marked with capital letters indicate the cumulative total number of mashups in 1000 increments.

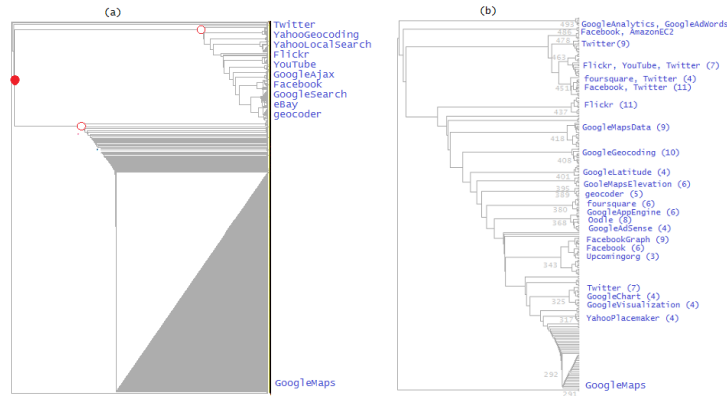
## 4.2 Niche formation

Expecting that niches are anchored around hub APIs, we constructed phylogenetic trees centered on those APIs to identify characteristics of the niches. In Fig. 2, we indicated each cumulative 1000 mashup increment by a vertical line to allow cross-referencing between the evolution of hub APIs and the APIs in each niche.

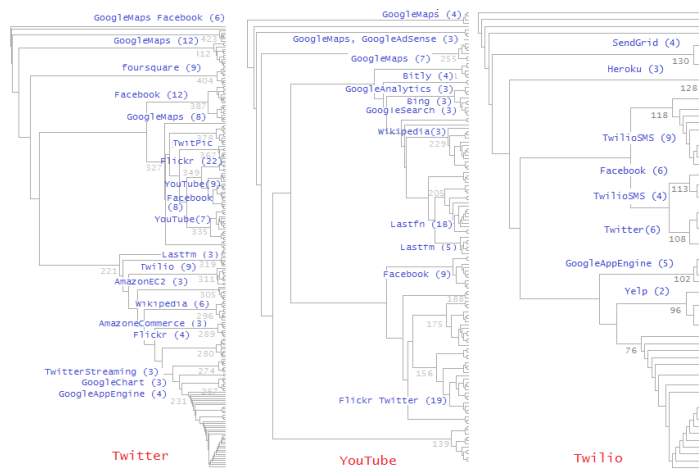
As we examine these trees, we observe that the impact of hub APIs varies with time. API dominance and complementarity of APIs are some of the interesting observations we can make. For instance, in Fig. 3(a) we can observe the initial dominance of Google Maps, as represented by a cluster of mashups that only use Google Maps. Later, as shown in Fig. 3(b), the clusters become more evenly distributed, as we there are more cluster with APIs that complement GoogleMaps, such as Twitter and YouTube, or other APIs by Google, such as GoogleSearch.

One way to understand the impact of hub APIs on the evolution of the mashup ecosystem is to align growth stages in their S-curves (see Fig. 2) with the phylogenetic trees for the corresponding time window. Fig. 4 offers a more detailed perspective of each of the APIs complementing Google Maps past the 5000 mashups (E) mark. It shows the phylogenetic trees for Twitter, YouTube, and Twilio. Each of these APIs creates a niche within the mashup ecosystem, where it drives the evolution of this niche as its hub API. A similar analysis can be conducted within each of those

niches. We can identify sub-niches such as the niche anchored around Facebook in the Twitter niche (4a), and Last.fm in the YouTube niche (4b).



**Fig. 3.** Phylogenetic trees comparing Google Maps API evolution (a) before and (b) after 1727 days. This date correspond to 5000 mashups (marked with an E in Fig. 2).



**Fig. 4.** Phylogenetic trees of the Twitter, YouTube, and Twilio niches after 1727 days.

## Conclusion

In conclusion, we find that the evolution of the mashup ecosystem can be explained in terms of ecosystem niches anchored around hub or keystone APIs. Those are APIs that have a significant impact on the evolution of the ecosystem. To help study niche

formation we developed a technique based on phylogenetic analysis. This technique involves creating phylogenetic trees for specific time windows when particular APIs are dominant. Furthermore, we observed the formation of niches within niches.

The results of our research are, however, far from final. We are still at the beginning of our understanding of how the mashup ecosystem evolves. One venue for future research should be around governance strategies for hub API providers. For example, how can hub API providers encourage the creation of complementary APIs that strengthen their niche. Another venue to explore is the creation of a new generation of mashup directories, for example, a tool that allows developers to browse a “tree of life” of mashups and to discover new opportunities for mashups. Such a tool could also be used by providers to learn about emerging needs for new APIs.

## References

1. Daniel, F., Matera, M., & Weiss, M. (2011), Next in mashup development: User-created apps on the web, *IT Pro*, Sep/Oct, 22-28
2. Yu, S., & Woodard, C. (2008), Innovation in the programmable web: Characterizing the mashup ecosystem, *Mashups 2008, ICSOC 2008 Workshops*, Springer, 136-147.
3. Weiss, M., & Gangadharan, G.R. (2010), Modeling the Mashup Ecosystem: Structure and Growth, *R&D Management*, 40(1), 40-49.
4. Weiss, M., & Sari, S. (2011), Diversity in the mashup ecosystem, *International Conference on Web Information Systems and Technologies*.
5. Hargadon, A. (2002), Brokering knowledge: linking learning and innovation. *Research in Organizational Behavior*, 24, 41–85.
6. Ethiraj, S., & Levinthal, D. (2004), Modularity and innovation in complex systems. *Management Science*, 50, 2, 159–173.
7. Bentley, A., Earls, M., & O'Brian, M. (2011), *I'll Have What She's Having: Mapping Social Behavior*, MIT Press.
8. Sari, S., & Weiss, M. (2010), Evolution of the mashup ecosystem by copying, *Workshop on Web APIs and Services Mashups*, ACM, 11.
9. Thomas, L., & Autio, E. (2012), Modeling the ecosystem: A meta-synthesis of ecosystem and related literatures, *DRUID 2012*.
10. Hagel, J., Brown, J. S., & Davison, L. (2008), Shaping strategy in a world of constant disruption, *Harvard Business Review*, October, 81-89.
11. West, J. (2006), Does appropriability enable or retard open innovation. In: Chesbrough, H., Vanhaverbeke, W., & West, J. (eds), *Open Innovation: Researching a New Paradigm*. Oxford, Oxford University Press, 109–133.
12. Iyer, B., & Davenport, T.H. (2008), Reverse engineering google's innovation machine. *Harvard Business Review*, 86, 4, 58–69.
13. Adner, R., & Levinthal, D. (2002), The emergence of emerging technology, *California Management Review*, 45(1), 50-66.
14. Bradford, S. (1950), *Documentation*, Washington, DC, Public Affairs Press.
15. Rogers, E. (1983), *Diffusion of Innovation*, Free Press.
16. Gascuel, O. (1997), BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data, *Molecular Biology and Evolution*, 14, 685-695.