



# Data correction and evolution analysis of the ProgrammableWeb service ecosystem<sup>☆</sup>

Mingyi Liu<sup>a</sup>, Zhiying Tu<sup>b</sup>, Yeqi Zhu<sup>a</sup>, Xiaofei Xu<sup>a</sup>, Zhongjie Wang<sup>a,\*</sup>, Quan Z. Sheng<sup>c</sup>

<sup>a</sup> Faculty of Computing, Harbin Institute of Technology, Harbin, China

<sup>b</sup> School of Computer Science and Technology, Harbin Institute of Technology, Weihai, China

<sup>c</sup> Department of Computing, Macquarie University, Sydney, NSW 2109, Australia

## ARTICLE INFO

### Article history:

Received 3 February 2021

Received in revised form 19 August 2021

Accepted 20 August 2021

Available online 27 August 2021

### Keywords:

Service ecosystem

Evolution analysis

ProgrammableWeb

Dynamic network model

APIs

Mashups

## ABSTRACT

The evolution analysis on Web service ecosystems has become a critical problem as the frequency of service changes on the Internet increases rapidly. Developers need to understand these evolution patterns to assist in their decision-making on service selection. *ProgrammableWeb* is a popular Web service ecosystem on which several evolution analyses have been conducted in the literature. However, the existing studies have ignored the quality issues of the *ProgrammableWeb* dataset and the issue of service obsolescence. In this study, we first report the quality issues identified in the *ProgrammableWeb* dataset from our empirical study. Then, we propose a novel method to correct the relevant evolution analysis data by estimating the life cycle of application programming interfaces (APIs) and mashups. We also reveal how to use three different dynamic network models in the service ecosystem evolution analysis based on the corrected *ProgrammableWeb* dataset. Our experimental experience iterates the quality issues of the original *ProgrammableWeb* and highlights several research opportunities.

© 2021 Elsevier Inc. All rights reserved.

## 1. Introduction

With the development of Web 2.0 and the wide adoption of service-oriented architecture (SOA), many services now expose their features in the form of application programming interfaces (APIs). Multiple APIs can be easily composed into an application, also called *mashups* (Ngu et al., 2010), that creates and delivers unique new value to customers. This growing phenomenon is called the *API economy* (Brown et al., 2014). As domain barriers have opened and cross-border cooperation and cross-border integration have become more common, the promotion of the API economy in the Internet era has gradually weakened the concept of the traditional domain. These represent substantial changes to the traditional Internet ecosystem, leading to the evolution of service ecosystems.

The API economy has enabled businesses in regard to cross-border integration and innovation, creating an increasing number of new applications. Moreover, as users experience these new applications, they may discover new needs, which not only further accelerates the innovation process but also intensifies market competition. Service providers need to be sensitive to the changes of user needs and preferences and constantly bring

forth new services. Therefore, studying the evolution of service ecosystems is important because it can offer insights and significant benefits from different perspectives. From a business perspective, evolution analysis assists decision-making by helping service providers and market regulators understand the evolutionary patterns of a service ecosystem, thereby guiding sustainable and healthy service/service ecosystem development. For example, through evolution analysis, service providers can learn the collaboration strategies of their competitors and discover popular market evolution trends, allowing them to adjust their business strategies for fleeting innovation opportunities and thereby maintaining or enhancing the competitiveness of their services. From a technical perspective, evolution analysis can mine interpretable prior knowledge from data to facilitate other downstream tasks, such as service recommendation, service discovery, and service composition, thereby accelerating the pace of service development (Bouguettaya et al., 2017; Sheng et al., 2014).

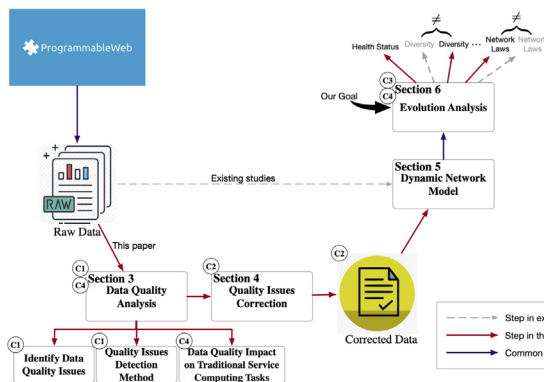
*ProgrammableWeb*<sup>1</sup> is the largest online API store platform, and it collects a large number of third-party APIs and mashups. Every day, new APIs/mashups emerge, existing APIs perish, and different APIs dynamically cooperate to create new mashups. Thus, it is a typical Web service ecosystem. In addition, *ProgrammableWeb* serves as a standard research dataset in the field of service computing. As a typical representative of the real Internet ecosystem, it has been employed to support many service

<sup>☆</sup> Editor: Fabio Palomba.

\* Corresponding author.

E-mail address: [rainy@hit.edu.cn](mailto:rainy@hit.edu.cn) (Z. Wang).

<sup>1</sup> <https://www.programmableweb.com/>.



**Fig. 1.** Overview of this paper relative to existing studies in the context of evolution analysis.

science studies, especially in the fields of service recommendation (Ma et al., 2021; Botangen et al., 2020; Kalaï et al., 2018), service discovery (Adeleye et al., 2019; Xu et al., 2017), service evolution analysis (Tian et al., 2017), and quality of service (QoS) prediction (Chen et al., 2020, 2017). As of January 9, 2021, *ProgrammableWeb* collected 23,881 APIs and 7973 mashups, with detailed information such as date of creation, category, profile, and active status.

**Motivation:** A number of studies have focused on the evolution analyses of the *ProgrammableWeb* service ecosystem from a variety of perspectives. The analysis results of all the existing studies show a positive and optimistic attitude towards the healthy development of the *ProgrammableWeb* service ecosystem. Some studies have also provided a generative model of the service ecosystem network to guide service recommendation and service discovery. These studies (Wittern et al., 2014; Zhong et al., 2014; Huang et al., 2013; Pan and Chai, 2018) are usually explicitly or implicitly based on two assumptions: (i) the *ProgrammableWeb* dataset is high quality and credible, and (ii) the obsolescence of APIs and mashups and the impact of such demise can be ignored. However, these assumptions require more careful data processing because the *ProgrammableWeb* dataset has certain flaws. For example, quality issues in labeling, especially regarding the active availability status and obsolescence times of APIs and mashups, are important. Ignoring the demise of APIs and mashups makes it more difficult to assess the true health status of the service ecosystem, promotes blindly optimistic conclusions, leads to inaccurate evolutionary patterns, and damages downstream tasks.

Fig. 1 shows the overview of this paper relative to existing studies in the context of evolution analysis. Existing studies directly use the raw data crawled from the *ProgrammableWeb* for dynamic network modeling followed by evolutionary analysis. However, in this paper, we first identify and correct the erroneous information related to evolution in the *ProgrammableWeb* dataset, specifically for the active state and time-related attributes. Then, we discuss how to use dynamic networks to model the service ecosystem. After establishing a dynamic network model, we analyze the evolution of the service ecosystem using the network properties and network visualization approaches. Finally, we discuss the problems that exist in the service ecosystem and their impacts on other service tasks.

The main contributions and innovations of this study are as follows:

- C1 We identify data quality issues, including data incompleteness, data errors and data noise, in the commonly used *ProgrammableWeb* dataset by exploiting statistical methods, automated network request testing, and manual inspections.

- C2 We correct the active status of APIs and mashups in the *ProgrammableWeb* dataset based on the results of automated network request testing and propose a normal distribution-based method to estimate the active time of APIs and mashups. We have released the new *ProgrammableWeb* dataset with the active status and active time corrected for other researchers to conduct related studies.<sup>2</sup>

- C3 We conduct an evolution analysis based on the corrected *ProgrammableWeb* dataset, and the results show that the original dataset is not suitable for service ecosystem research. The assumption of perfect data quality misleads statistical results of the data and affects the choice of algorithms. If these factors are not considered, the results of the existing research works based on the old dataset should be revisited, such as various conclusions about the current health status, diversity, and network laws of the *ProgrammableWeb* service ecosystem.

- C4 During the evolution analysis, we analyze the potential impact of the data quality issues of the original *ProgrammableWeb* dataset from both business and technical viewpoint, and provide some suggestions for addressing or avoiding these quality problems. We also highlight some opportunities for future research using the new *programmableWeb* dataset.

The remainder of this paper is organized as follows. Section 2 discusses the related works. Section 3 presents various data quality issues related to the evolutionary analysis of the *ProgrammableWeb* dataset. Section 4 provides a data restoration method based on probability estimation. Section 5 clarifies how to use dynamic networks to model service ecosystems. Section 6 explains the methods, reports the results of the evolution analysis of the service ecosystem, and discusses the impacts of other quality issues of the original *ProgrammableWeb* dataset on traditional service computing tasks and the new challenges and opportunities that the corrected *ProgrammableWeb* dataset introduces for traditional tasks. Finally, Section 8 offers some concluding remarks.

## 2. Related works

In recent years, the evolution analysis of service ecosystems, such as the *ProgrammableWeb*, has been studied extensively. The existing studies focus on individual service state changes and the changes to the service network topology that are intended to help developers select and integrate appropriate services into their own applications. The ultimate goal of these studies is to provide prior knowledge to help solve various traditional service computing problems, such as service recommendation and service label prediction (Yao et al., 2015).

For example, Wang et al. (2009) discovered user behavior patterns in mashup communities by studying the network and clustering properties of the *ProgrammableWeb* service ecosystems. The authors argued that mashup communities possess scale-free properties and frequently used APIs attract large numbers of users. Weiss and Gangadharan (2010) examined the structure of the mashup ecosystem and its growth over time and concluded that (i) the distribution of mashups over APIs follows a power law and that (ii) the complexity of mashups continues to increase. Huang et al. (2014) proposed a three-phase network prediction approach (NPA) to study both usage patterns and the evolution traces of the entire *ProgrammableWeb* service ecosystem for evolution-aware service recommendation. The authors

<sup>2</sup> The dataset is available on GitHub: <https://github.com/HIT-ICES/Corrected-ProgrammableWeb-dataset>.

in Zhong et al. (2014) considered the evolution of service ecosystems for service recommendation, extracted service evolution patterns by exploiting latent Dirichlet allocation (LDA) and time series predictions, and then used these patterns to guide service recommendation. Lyu et al. (2014) proposed a three-level view model for Web service ecosystem visualization. Such visualization is valuable not only for understanding the ecosystem but also for providing support to service consumers, helping them to discover appropriate services. Bai et al. (2017) developed a tailored topic model to mine effective representations of service ecosystems that addresses both service evolution and information sparsity, while Gan et al. (2019) proposed a novel approach for service multilabel recommendation using deep neural networks.

In addition to exploring the Web service ecosystem represented by the *ProgrammableWeb*, the evolution of other types of service ecosystems have also been explored in recent studies. For instance, the work in Sampaio et al. (2017) focuses on microservice ecosystems. Aggregating structural, deployment, and run-time information of an evolving microservice system into one model provides actionable insights to help developers manage service upgrades, architectural evolution, and changing deployment trade-offs. Wang et al. (2018a) also focused on microservice ecosystems by proposing a distributed knowledge-based evolution model (DKEM) that can discover stable evolution patterns and automatically explore new and more stable cooperation among services. In addition to these technical-level service ecosystems, some researchers have turned their interests to the evolution of business-level service ecosystems. A very recent work in Liu et al. (2020), Liu et al. (2020) proposed a novel multilayer network-based service ecosystem model (MSEM) that can be constructed automatically by mining massive textual datasets from the Internet. Via the introduction of the concept of service events, MSEM can not only explore evolutionary patterns but also determine the driving factors of the evolution.

### 3. Data quality issues

We collected a total of 23,678 APIs and 7766 mashups from the *ProgrammableWeb* website (including data in the deathpool). After checking and testing these data, we identified three serious data quality issues related to prior evolutionary analyses using the original *ProgrammableWeb* dataset. These issues are described in detail in the rest of this section.

#### 3.1. Untrustworthy death time

*ProgrammableWeb* places the deprecated APIs and mashups into a collection called the “deathpool”. The time at which a deprecated API or mashup enters the deathpool is marked and this time is often referred to as the “death time” of the API or the mashup.

We conducted statistical studies on the death times of APIs and mashups in the deathpool provided by *ProgrammableWeb* and the results are shown in Figs. 2 and 3. The results indicate that the death times of most APIs and mashups in the deathpool are clustered at specific periods (e.g., April 2014). Interestingly, some APIs/mashups are created even later than the creation time; for example, *Iron Mountain Policy Center*<sup>3</sup> was marked as deprecated in April 2014 but was actually submitted on January 21, 2020. This might be the result of the automatic processing at a certain time because the death time labels are obviously unreasonable and are therefore unsuitable for Web service evolution studies. In terms of distribution, the distribution of death times between 2018 and 2020 is more in line with the prior knowledge of

experts. We also compiled statistics on the survival duration of APIs that died during this period, as shown in Fig. 7, and found no data in the results that violate common sense, such as a service that dies before its birth. To further confirm the reliability of this portion of the data, we randomly selected 20 APIs and investigated their real death times. We found that the average death time error was 50 days, which is an acceptable result. Therefore, we believe that recent data (those with death times from 2018 to 2020) are reliable estimations of API/mashup death times. This aspect will be discussed in more detail in Section 4.1.

#### 3.2. Incorrect API status

We tested the status of APIs that was marked as available in two ways:

- (1) We detected whether the descriptive text of the API explicitly indicates its available status. When the text indicates “no longer available”, we consider the API’s status to be dead (or obsolete).
- (2) For the rest of the APIs, we tested the results of accessing their API endpoints (URL addresses) under strict conditions to prevent available APIs from being marked as dead APIs when it is determined that the API is not available. We conclude that an API is regarded as obsolete only when its URL is not reachable or when the network access request returns a 404 status code.

After using automatic network request tests to check<sup>4</sup> the availability status of all APIs and manually analyzing their descriptive texts, we summarize three different death patterns of unavailable APIs, as illustrated in Fig. 4:

- *Death*: The API no longer provides its services in any form; i.e., it has entirely and permanently perished from the ecosystem.
- *Transfer*: The original API is no longer available, but all its functions were transferred to another API that continues to provide services. This pattern usually occurs when an API is acquired or the name is changed.
- *Split*: The original API is no longer available; its services have been split into several independent APIs, and each independent API provides a portion of the original API’s services.

Table 1 shows the statistical results of the available status of the APIs on *ProgrammableWeb*. From the table, we can see that only approximately 44.7% of the APIs are truly active and that only 14.8% of the unavailable APIs are correctly marked as unavailable (including APIs in the deathpool). In addition, 85.2% of the unavailable APIs are incorrectly marked as active. The proportions of different death patterns are listed in Table 1. Most of the unavailable APIs (over 99.9%) are caused by the death of the APIs. The number of unavailable APIs caused by *transfer* and *split* is relatively small, 17 and 7, respectively. However, the APIs that are unavailable due to *split* and *transfer* have a substantial impact on the API service ecosystem. For example, the most commonly used APIs, such as *Google Map*, *Facebook*, and *Twitter*, have all split. The statistical results in Fig. 6 also support this point.

<sup>4</sup> All the network request tests were conducted on the Google Cloud platform. We repeated the tests three times on different dates and in different periods and manually reviewed 100 randomly selected results.

<sup>3</sup> <https://www.programmableweb.com/api/iron-mountain-policy-center>.

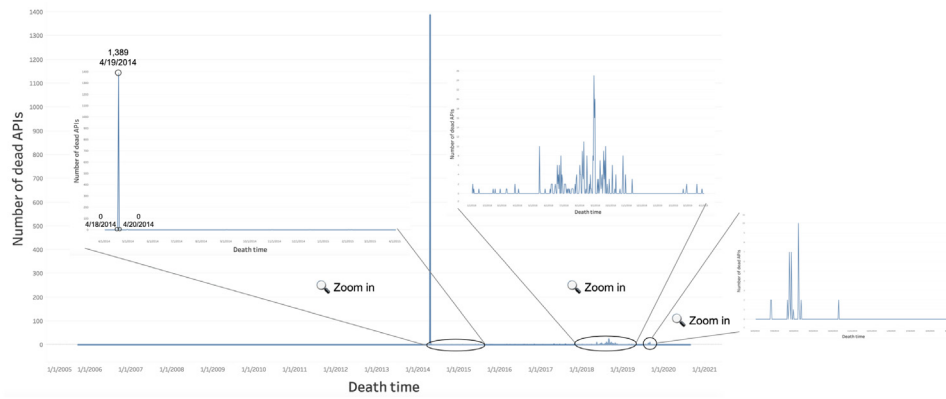


Fig. 2. The death time distribution of the APIs in the deathpool.

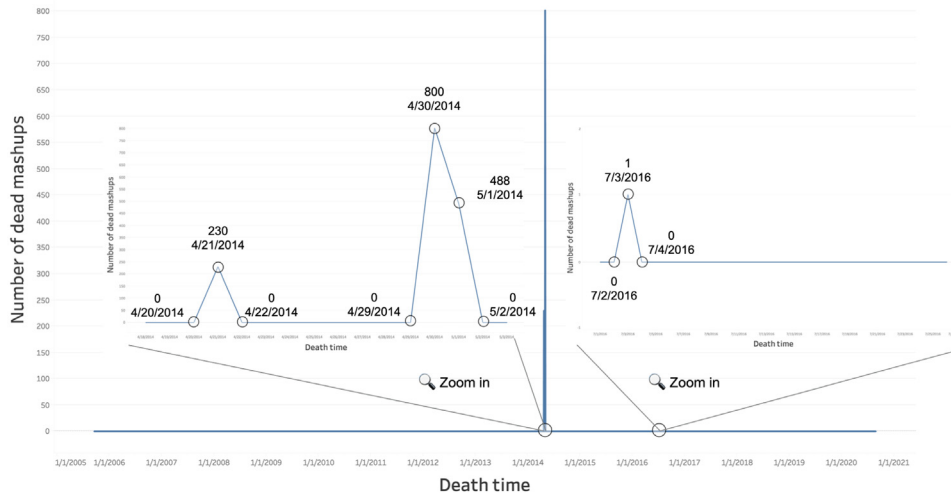


Fig. 3. The death time distribution of the mashups in the deathpool.

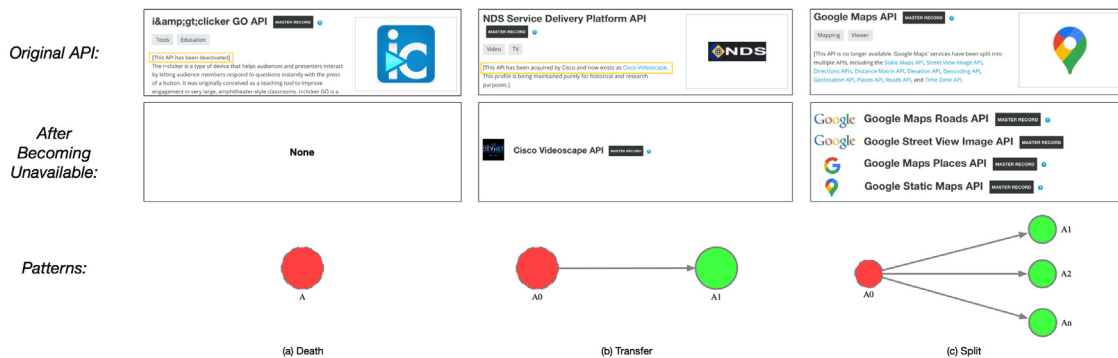


Fig. 4. Three different API death patterns: the red circles represent unavailable APIs, while the green circles represent available APIs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 3.3. Incorrect mashup information

The data quality issues in mashups are primarily twofold: (a) *incorrect activity status* and (b) *incorrect API composition information*.

An incorrect activity status means that mashups cannot properly provide the services they describe but are still marked as available. We invited 5 coders who are all graduate students working on service computing topics, to perform manual analysis. For each mashup, two coders are randomly selected for manual checking. Coders were required to report whether content in the

homepage of this mashup is inconsistent with the description in the page of this mashup on *ProgrammableWeb*. For those mashups that did not match, they were asked to check the history information at WebArchive<sup>5</sup> and then summarized the reasons for the unavailability. When two coders come to different conclusions on one mashup, this mashup will be voted by the remaining three coders. Finally, we use Inter-Rater Reliability (IRR)[28] to measure the level of agreement. During the manual checking, we did not find any cases that different coders have different

<sup>5</sup> <http://web.archive.org>.



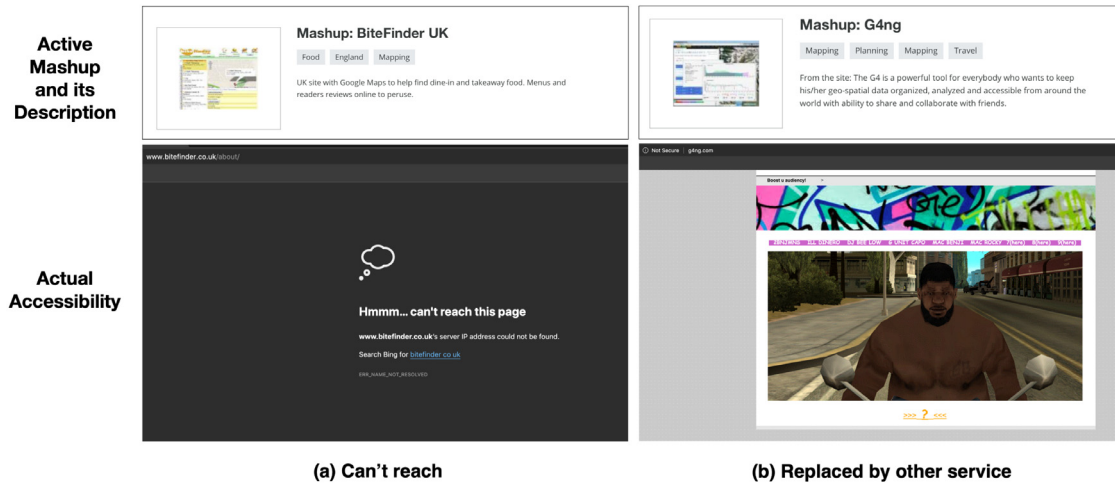


Fig. 5. Two different mashup death patterns.

Table 1

The statistical results of the available status of the APIs on *ProgrammableWeb*. All the APIs in the deathpool are marked as **dead**. \* denotes the overall percentage, \*\* denotes the row percentage.

| Truth       |          | Labeled         |                 |
|-------------|----------|-----------------|-----------------|
|             |          | Available       | Unavailable     |
| Available   |          | 10,534 (44.7%*) | –               |
| Unavailable | Dead     | 11,087          |                 |
|             | Split    | 7               | 1,934 (14.8%**) |
|             | Transfer | 17              |                 |

conclusions on one mashup, and this is mainly because of the relevant professional background of all five coders and the clear criteria for their judgment. Therefore, the IRR is 100% agreement between coders on 100% samples, which is sufficient agreement among multiple coders. Based on the provided report, we conclude that mashups are unavailable due to two different patterns: *unreachable* and *replaced*. Fig. 5 illustrates these two patterns with examples. Based on the above two patterns, the remaining mashups are determined to be unavailable if they satisfy any of the following conditions:

- *Unreachable*: The homepage is not accessible or does not exist (404).
- *Replaced*: The homepage's HTML source code does not contain information about the mashup.

We manually checked 530 mashups and this number is greater than 366 (the minimum number of mashups for statistical representation with a 95% confidence level and 5 intervals). The 530 mashups were selected based on the following steps:

Step 1 330 mashups were manually checked at random, including 130 available mashups and 200 unavailable mashups. The purpose of this step is to build mashup status detection algorithm. We stopped when we checked 200 unavailable mashups because (1) for these unavailable mashups, coders not only need to determine their availability, but more importantly, to identify exact unavailable time of these mashups by searching on <http://web.archive.org>, which is an extremely time-consuming task. (2) during the examination, we found that after 50 unavailable mashups were found, no new reasons for mashup unavailability emerged, so we believe that 200 unavailable mashups are sufficient for summarizing the patterns of mashups unavailability and providing sufficient a priori knowledge for the algorithm to be constructed.

Step 2 200 mashups were randomly selected from mashups marked as *unavailable* by the algorithm for manual check. The purpose of this step is to check the validity of the rule-based algorithm we have constructed. We chose 200 to be consistent with the number of 200 unavailable mashups in the previous step.

The test results show that the true numbers of available and unavailable mashups are 2,489 (32.0%) and 5,277 (68.0%), respectively, while the numbers given by *ProgrammableWeb* are 6,247 (80.4%) and 1,519 (19.6%), respectively (including mashups in the deathpool).

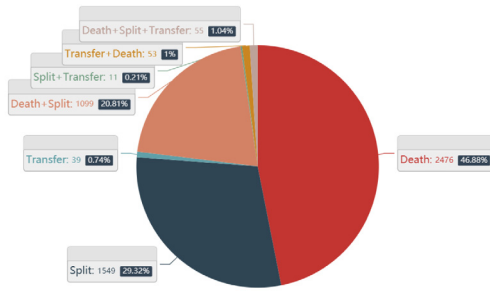
Incorrect API composition information refers to mashups that are still available but are marked as invoking some unavailable APIs. We found 4,942 mashups that invoked unavailable APIs, and 1,934 that were still available. Fig. 6 shows mashups invoking the different types of unavailable APIs, where each sector is labeled with the number and percentage of mashups using a particular type of unavailable API. For example, in Fig. 6(b), the number of available mashups invoking only *split* APIs is 2,174, which is 56.9% of the number of available mashups that invoke unavailable APIs. The figure provides some insights that have been overlooked in the past studies:

1. APIs that are widely used and provide rich services tend to split and subsequently provide more refined services.
2. The importance of different APIs in mashups is different. Some APIs do not participate in a mashup or could be replaced by other APIs in the mashup. Thus, even if these APIs are abandoned, the mashup can still guarantee the availability of services.

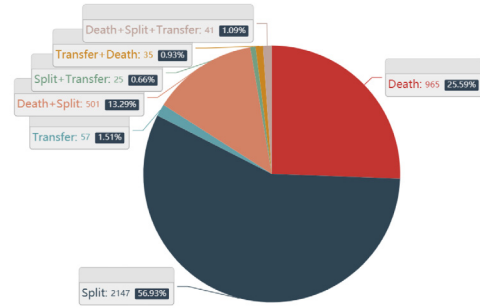
### 3.4. Other quality issues

This section discusses several other data quality issues in the original *ProgrammableWeb* dataset. These issues are not relevant to the evolutionary analysis but are relevant to other tasks such as API tag prediction. This section also covers some quality issues that cannot be corrected without external data or manpower. It should be noted that these data quality issues do not affect the data correction methods in Section 4: (1) these issues are independent of the quality issues raised in Section 3.1, Section 3.2 and Section 3.3; (2) these issues are evolutionary-independent.

The descriptive texts of APIs and mashups contain considerable noise, and the information in the texts is sparse, which hinders tasks that require semantic support, such as service tag



(a) Ratio of all mashups that invoke different types of unavailable APIs



(b) Ratio of available mashups that invoke different types of unavailable APIs

**Fig. 6.** Ratio of mashups that invoke different types of unavailable APIs.

prediction (Gan et al., 2019) and service recommendation (Ma et al., 2021). Bai et al. (2017) also reported this problem and tried to use service representation-latent Dirichlet allocation (SR-LDA) to mine API/mashup topic words to alleviate the noise and sparsity problems. However, based on the results of their study, there is still much room for improvement. The issue of the quality of the description text will not have an impact on the research questions or methods in this paper, as it is only private property of API/mashup that is independent of time and API/mashup status.

Toy mashups refer to the mashups that do not offer practical application value to users. Such mashups are usually the result of developers practicing or are created just for fun, such as *Website-Grader*<sup>6</sup> and *Trump's Circle of Death*.<sup>7</sup> Generally, these toy mashups should be treated as noisy data and filtered out. However, much manual annotation is required to accurately identify these toy mashups. Therefore, this issue is usually overlooked in research. Nevertheless, we note that toy mashup noise has a potential impact for algorithms operating on the entire dataset and that the extent of this influence depends on the application scenario. Fortunately, the toy mashups have no effect on most of the research questions in this paper because these research questions use only part of the data and automatically filter out the toy mashups. The toy mashups do have a small impact on RQ1 in Section 6, but they do not affect the conclusion drawn for the question.

A mashup is formed by combining various APIs in a certain sequence. Unfortunately, no relevant information exists regarding the combination sequence in the original *ProgrammableWeb* dataset. Moreover, the original sequence is almost impossible to

automatically recover based on information stored in the original *ProgrammableWeb*. Therefore, we make a compromise and focus on the co-occurrences of APIs in mashups.

### 3.5. Impact of data quality issues

The implications of the abovementioned data quality issues for service computing tasks, except for the evolution analysis, are discussed in this section; implications for the evolution analysis are explained in detail after the answers to each of the research questions in Section 6.

Incorrect API/mashup availability status will significantly affect the reliability of the results generated by application scenarios, such as service recommendation, service selection, service combination, etc., that need to satisfy user requirements. Since they will recommend services that have been unavailable for a long time, this obviously does not meet the needs of realistic conditions.

Incorrect death time and availability status can be a major threat to the correctness of various time-aware methods. The original *ProgrammableWeb* dataset is presented as an incremental dynamic network in which only new nodes (APIs, mashups) and edges (invoke relations) are added. However, the fact that the API/mashup fails at different times makes the data structure of an incremental dynamic network unrealistic, and the nodes and edges in this network can exit the network. This causes many operations that could be performed on an incremental dynamic network to be inoperable or costly in terms of time and space.

Incorrect API composition information, the lack of API combination sequence information, and toy mashups will affect the reliability of the results of tasks such as service composition, service discovery and service selection because these tasks rely on the statistical results of known API compositions, and these data quality issues lead to unreliable and missing API composition profiles.

## 4. Data correction

In this section, the data quality issues mentioned in Section 3 are addressed to better support the evolution analysis of the service ecosystem. Before expounding on the details of the method, we wish to clarify the following points:

- Our method is not perfect in the sense that the data correction results are not totally accurate; they are a probability estimate.
- Because this paper focuses on service ecosystem evolution, our method aims at the whole service ecosystem; in other words, we can guarantee the rationality of the overall distribution of the data but cannot guarantee the accuracy of individual samples.
- To balance the algorithm's complexity, human cost, and other factors, we ignored some unimportant factors and made several reasonable assumptions.

### 4.1. Estimated death time

It is not feasible to directly estimate the death times of APIs and mashups. Therefore, we calculated the death times by estimating survival days. The statistical results of the longevity of APIs placed in the deathpool from 2018 to 2020 are shown in Fig. 7.

Based on Fig. 7, we assume that the longevity of APIs and mashups ( $x_1, x_2, \dots, x_n$ ) follows a normal distribution  $\mathcal{N}(\mu, \sigma^2)$ , which is one of the most common and applicable distributions in the real world. Then, the key problem is how to learn the approximate values of parameters  $\mu$  and  $\sigma^2$  based on the sample

<sup>6</sup> <https://www.programmableweb.com/mashup/website-gradercom>.

<sup>7</sup> <https://www.programmableweb.com/mashup/trumps-circle-death>.

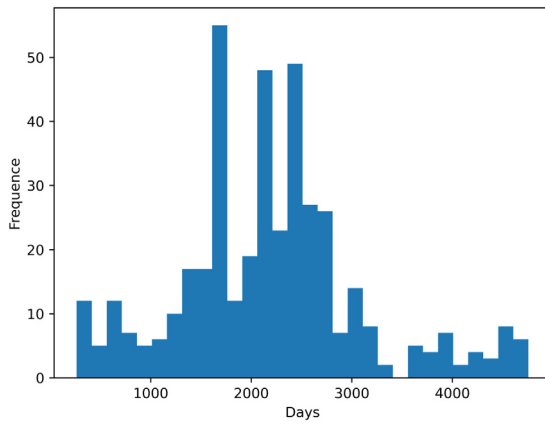


Fig. 7. The distribution of survival days for APIs placed in the deathpool from 2018 to 2020.

$(x_1, x_2, \dots, x_n)$ . The standard approach for addressing this problem is to use the maximum likelihood method, which requires maximization of the log-likelihood function:

$$\ln \mathcal{L}(\mu, \sigma^2) = \sum_{i=1}^n \ln f(x_i | \mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \quad (1)$$

where  $f(x)$  is the probability density function (PDF). Taking the derivatives of  $\mu$  and  $\sigma^2$  and solving the resulting system of first-order conditions yields the maximum likelihood estimates:

$$\hat{\mu} = \bar{x} \equiv \frac{1}{n} \sum_{i=1}^n x_i, \quad (2)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (3)$$

To ensure that the distribution estimated using the longevity of APIs placed in the deathpool from 2018 to 2020 is representative, we first randomly selected 200 unavailable APIs/mashups and manually checked their survival days

$(x'_1, x'_2, \dots, x'_n)$  through the Web snapshots provided by *Wayback Machine*<sup>8</sup> to represent the overall distribution. Then, we estimate the values of parameters  $\hat{\mu}'$  and  $\hat{\sigma}'^2$  based on the sample  $(x'_1, x'_2, \dots, x'_n)$  using Eqs. (2) and (3). Finally, we use the Z-statistic to test whether the two distributions are the same:

$$Z = \frac{\hat{\mu} - \hat{\mu}'}{\sqrt{\hat{\sigma}^2 + \hat{\sigma}'^2}} \quad (4)$$

Normally, in qualitative terms:

- If the Z-statistic is less than 2, the two distributions are the same.
- If the Z-statistic is between 2.0 and 2.5, the two distributions are marginally different.
- If the Z-statistic is between 2.5 and 3.0, the two distributions are significantly different.
- If the Z-statistic is more than 3.0, the two distributions are highly significantly different.

The Z-statistic for these two distributions is 0.385, which is much less than 2. Therefore, we can state that the distribution estimated using the longevity of APIs is the same as the distribution of the whole dataset, which means that the data in Fig. 6 are representative.

For each obsolete API or mashup  $x$ , we generate a longevity value,  $d_x$ , from the normal distribution  $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$ . The unavailable time  $end_x$  can be calculated as follows:

$$end_x = \begin{cases} start_x + d_x & \text{if } start_x + d_x \leq \beta_x \\ random(start_x, \beta_x) & \text{otherwise,} \end{cases} \quad (5)$$

where  $start_x$  is the creation time of  $x$ ;  $\beta_x$  denotes the earliest time that  $x$  is confirmed to be unavailable. In this paper, we take the value as the date of the first network request test (September 10, 2020). We also introduce a random distribution  $random(\cdot)$  to prevent  $end_x$  from being later than  $\beta_x$ . In particular, with the *transferred/split* APIs, we do not have to use such a method to estimate the unavailable time. Suppose an API  $x$  has *split/transferred* into a new API set  $\{x_1, x_2, \dots, x_m\}$ . The unavailable time of  $x$  is:

$$end_x = \max(start_{x_1}, start_{x_2}, \dots, start_{x_m}). \quad (6)$$

#### 4.2. Correct API composition for mashups

We use  $M_t = \{a_1, a_2, \dots, a_k\}$  to represent an available mashup at time  $t$ , where  $a$  denotes an API included in the mashup. Suppose at time  $t + 1$  the mashup  $M_{t+1}$  is still available, but API  $a_i$  is no longer available. The representation of  $M_{t+1}$  depends on the pattern of  $a_i$  being unavailable, as shown in Eq. (7):

$$M_{t+1} = \begin{cases} M_t - \{a_i\} & \text{if death} \\ M_t - \{a_i\} + \{a'_i\} & \text{if transfer, } a_i \rightarrow a'_i \\ M_t - \{a_i\} + \{a_{i,1}, \dots, a_{i,n}\} & \text{if split, } a_i \rightarrow \{a_{i,1}, \dots, a_{i,n}\} \end{cases} \quad (7)$$

The handling of *death* and *transfer* cases is easy to understand. For *split* APIs, we made a reasonable assumption due to imperfections in the information. We assumed that all the APIs  $a_{i,j}$  generated by splitting will be used by the mashup to ensure complete functionality despite knowing that only one subset would be used in reality. For example, mashup *Mosoto*<sup>9</sup> was composed of *Box*<sup>10</sup> and *Facebook*<sup>11</sup> services when it was first created. Thus, it can be denoted as  $\{Box, Facebook\}$ . Then, something happened on *Facebook* that caused it to split into *Facebook Ads*,<sup>12</sup> *Facebook Atlas*,<sup>13</sup> *Facebook Graph*<sup>14</sup> and *Facebook Marketing*.<sup>15</sup> Subsequently, the representation of *Mosoto* is  $\{Box, Facebook Ads, Facebook Atlas, Facebook Graph, and Facebook Marketing\}$ . Next, when *Facebook Ads* and *Facebook Atlas* are deprecated, the *Mosoto* representation changes again to  $\{Box, Facebook Graph, and Facebook Marketing\}$ .

### 5. Service ecosystem dynamic network model

The service ecosystem is a continuously evolving complex network system consisting of service entities and interactions between them. This system can be naturally modeled as three different dynamic networks in which the structure changes over time depending on different scenarios: (i) the Mashup-API network (M-A), (ii) the API-API network (A-A), and (iii) the category-category network (C-C).

<sup>9</sup> <https://www.programmableweb.com/mashup/mosoto>.

<sup>10</sup> <https://www.programmableweb.com/api/box>.

<sup>11</sup> <https://www.programmableweb.com/api/facebook>.

<sup>12</sup> <https://www.programmableweb.com/api/facebook-ads>.

<sup>13</sup> <https://www.programmableweb.com/api/facebook-atlas>.

<sup>14</sup> <https://www.programmableweb.com/api/facebook-graph>.

<sup>15</sup> <https://www.programmableweb.com/api/facebook-marketing>.

<sup>8</sup> <http://web.archive.org>.

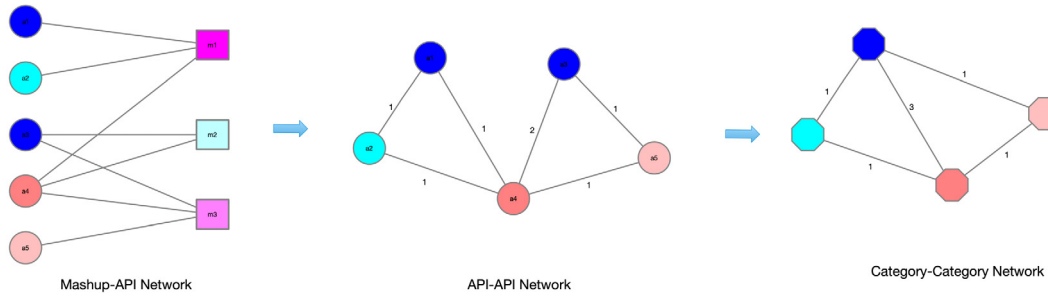


Fig. 8. Conversion between the three dynamic network types.

**Definition 1.** The Mashup-API network (M-A) is a dynamic bipartite graph  $G_{MA} = \{A, M, E_{MA}\}$ , where  $A$  refers to the APIs,  $M$  refers to the mashups,  $E_{MA} = \{(u, v, start, end) | u \in M, v \in A\}$  denotes the invoking relations between mashups and APIs, and  $start$  and  $end$  represent the duration of each relation. An API/mashup  $x \in A \cup M$  can be denoted as  $x = \{start, end, c\}$ , where  $c$  represents the primary category to which  $x$  belongs.  $G_{MA}^t = \{A^t, M^t, E_{MA}^t\}$  is used to represent a snapshot corresponding to the M-A at time  $t$ .

**Definition 2.** The API-API network (A-A) is a homogeneous network serial snapshot, denoted as  $G_{AA} = \{G_{AA}^1, G_{AA}^2, \dots, G_{AA}^n\}$ .  $G_{AA}^t = \{A^t, E_{AA}^t\}$  is the A-A snapshot at time  $t$  generated by a Mashup-API network snapshot  $G_{MA}^t$ .  $E_{AA}^t = \{(u, v, w) | u, v \in A^t\}$  represents the API concurrence relations, and  $(u, v, w)$  indicates that API  $u$  and API  $v$  are invoked together in an available mashup  $w$  at time  $t$ .

**Definition 3.** The category-category network (C-C) is a hypergraph serial snapshot, denoted as  $G_{CC} = \{G_{CC}^1, G_{CC}^2, \dots, G_{CC}^n\}$ .  $G_{CC}^t = \{C^t, E_{CC}^t\}$  is the C-C snapshot at time  $t$  generated by an API-API network snapshot  $G_{AA}^t$ , where  $C$  is a set of categories to which the APIs belong.  $E_{CC}^t = \{(u, v, w) | u, v \in C\}$  indicates the number of times that an API of category  $u$  and an API of category  $v$  are invoked together in an available mashup  $w$  at time  $t$ .

Fig. 8 shows the conversion process between the three dynamic network types. The square in the figure represents a mashup, the circle represents an API, and the color of the shape denotes the primary category. The link in the figure denotes the invoking/concurrence relations. These three different dynamic networks reflect the evolution of the *ProgrammableWeb* service ecosystem from different levels, which will be discussed in detail in the next section.

## 6. Service ecosystem evolution analysis

This section analyzes the service ecosystem evolution from the perspective of complex networks. The subsequent analysis in this section is organized by answering six research questions and identifying research challenges and opportunities based on the analysis. The title of each research question consists of two parts. The first part concerns which dynamic network model is selected to support the research question, and the second part explains the research question. As shown in Fig. 9, we have constructed 6 research questions at three levels: the service ecosystem as a whole (RQ1, RQ2, RQ3), the local structure of the service ecosystem (RQ3, RQ4) and service individuals in the service ecosystem (RQ5, RQ6). In particular, in RQ1, RQ2, RQ4 and RQ5, we found different conclusions from existing studies. Furthermore, in RQ3 and RQ6 we found that the current *ProgrammableWeb* service ecosystem cannot satisfy the complex service market requirement.

It is worth mentioning that although the method discussed in Section 4 cannot guarantee that an API/mashup individual is

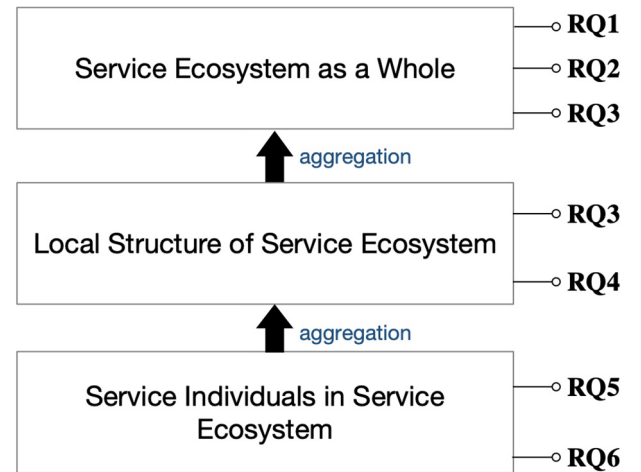


Fig. 9. The link between six research questions.

completely correct, it does guarantee the accuracy of the overall distribution of the *ProgrammableWeb* service ecosystem. We repeated the experiments several times independently and found that the results have no influence on the conclusions of the evolutionary analysis in this paper.

### RQ1: [M-A] What is the health status of the *ProgrammableWeb* service ecosystem?

The number of available APIs and mashups is the most intuitive indicator of the health and prosperity of a service ecosystem. Fig. 10 shows how the number of available APIs and mashups changes over time. The blue line (i.e., No death) denotes that APIs and mashups are always available; the green line (i.e., Deathpool) denotes using the deathpool information provided by the original *ProgrammableWeb* to determine whether an API or mashup is available; and the red line (i.e., Corrected) denotes the results of our corrected *ProgrammableWeb* dataset.

To the best of our knowledge, all the existing studies are based on the first two scenarios (No death or Deathpool), and consequently, they provide “optimistic” estimates regarding the health and prosperity of the *ProgrammableWeb* service ecosystem (Tian et al., 2017). However, our corrected data show that the reality might be much worse.

Based on the change curve of the number of available APIs and mashups, there is no evidence to support continuous “explosive” growth, which has often been mentioned in the existing studies. The results show that after 2014, the *ProgrammableWeb* service ecosystem actually began to decline. Although the number of new APIs remains stable year over year, the rate of API obsolescence has continued to accelerate, especially after 2018. While the addition of new mashups has stagnated since 2014, the rate of obsolescence has remained steady. We argue that the reason for



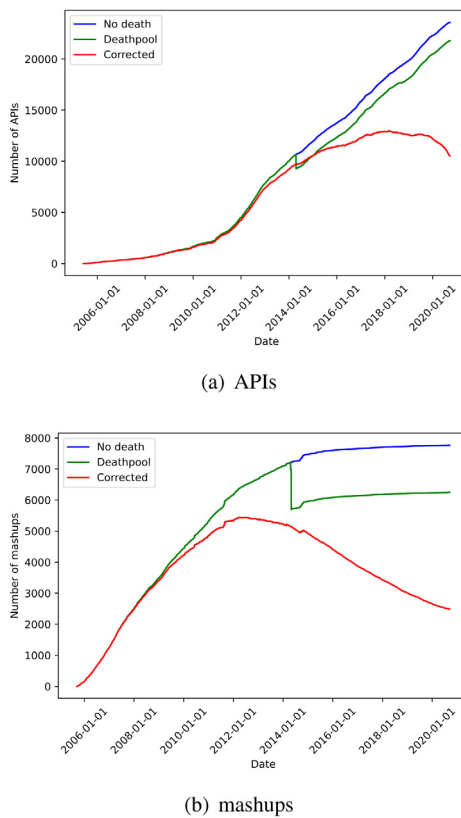


Fig. 10. The numbers of available APIs and mashups change over time (daily).

this phenomenon is that the popularity of the mobile Internet has led to an increasing shift of Web mashups to mobile apps (Wang et al., 2017), which is also supported by online statistics.<sup>16</sup>

In summary, the changes in the number of mashups and APIs in the *ProgrammableWeb* service ecosystem indicate that **the ecosystem is in poor health and could become extinct unless necessary steps are taken to remedy the situation.**

#### RQ2: [A-A] Does the degree distribution of the service ecosystem network comply with the power law?

In almost all studies on service ecosystem network evolution, *degree distribution* is a core element. Many studies (Adeleye et al., 2018, 2019; Lyu et al., 2014; Zhou and Wang, 2019) have concluded that the degree distribution of the A-A network complies with the power law, and a generative model of the A-A network is built to support downstream tasks. However, based on the theoretical analysis and our experimental results, we believe that this conclusion may be misleading due to insufficient statistics and inappropriate evaluation indicators.

Fig. 11 presents the results of the degree distribution changes over time. The figure shows that the degree distribution is quite consistent with the power law,<sup>17</sup> especially on the corrected *ProgrammableWeb* dataset. The existing studies usually adopt the *p-value* as evidence that the degree distribution of the *ProgrammableWeb* service ecosystem network conforms to the power law. We first followed (Adeleye et al., 2018) and conducted a *goodness-of-fit* test based on the Kolmogorov–Smirnov (KS) distance to calculate a *p-value* to quantitatively measure the plausibility of the power law. The results are shown in Fig. 12.

<sup>16</sup> <https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/>.

<sup>17</sup> If consistent, the overall trend should be close to a straight line, not a curve.

We can see from Fig. 12 that the *p-value* fluctuates substantially. Thus, we argue that the results do not support that the power law is a plausible hypothesis for the data:

- A *p-value* depends upon both the magnitude of association and the precision of the estimate (the sample size), while a smaller sample size tends to result in a larger *p-value*. *p-values* are typically used to reach a conclusion of “significant” or “not significant” based on whether the *p-value* is larger than a threshold. Consequently, *p-values* are more similar to qualitative measurements than quantitative measurements. For a detailed discussion we refer readers to Clauset et al. (2009).
- *P-values* are computed based on the assumption that the null hypothesis is true. A *p-value* represents the probability that the data deviates from the null hypothesis by a certain amount. Consequently, a *p-value* measures the compatibility of the data with the null hypothesis—not the probability that the null hypothesis is correct. Therefore, it is unreasonable to compare *p-values* associated with different null hypotheses as in Adeleye et al. (2018).

**To summarize, we argue that the current statistics and measurement methods are insufficient, inappropriate, or have difficulty supporting the hypothesis that the degree distribution complies with the power law.** Downstream tasks such as service recommendation and service discovery based on the existing findings may risk incorrect hypotheses and degrade the effectiveness of the method.

#### RQ3: [C-C] What are the changes in diversity and popularity of different categories in the service ecosystem?

Diversity reflects the vitality of an ecosystem and is also positively related to the richness of the services that a service ecosystem can provide (Huang et al., 2013; Kochan et al., 2003). Understanding the changing trends of popular service categories in a service ecosystem can provide developers with suggestions when providing new services. The C–C network naturally reflects the diversity and the popularity of various categories in the service ecosystem. The number of nodes indicates the diversity of service categories, and the locations (generated by visual layout) and sizes of nodes (number of aggregated APIs) can be used to indicate popularity.

Fig. 14 depicts the diversity of the *ProgrammableWeb* service ecosystem changes over time by counting the number of nodes in the C–C network, which is generated from the corresponding A–A network after removing isolated APIs. **The red curve shows that the diversity of the *ProgrammableWeb* service ecosystem first increased, then entered a plateau, and began to decline rapidly after 2017.**

The evolution of the C–C network is visualized using *Gephi*<sup>18</sup> and shown in Fig. 13, which also indicates the changes in the popularity of different categories. In particular, *Mapping* has been a popular category from the beginning. Although the node size has undergone a relative decrease, it still captures the center place, which means that this category is often used with other categories (we all use such services everywhere). For example, when searching for a restaurant on a map, users also obtain additional information, such as customers' reviews and discounts.

*eCommerce* also appeared from the beginning, but it did not undergo rapid development until recent years. However, its popularity has remained consistent. In recent years, *eCommerce* nodes have become one of the largest node clusters in the graph; that is, *eCommerce* is included as part of other categories' services more than ever before. Therefore, when creating a new mashup

<sup>18</sup> An open graph viz platform: <https://gephi.org>.

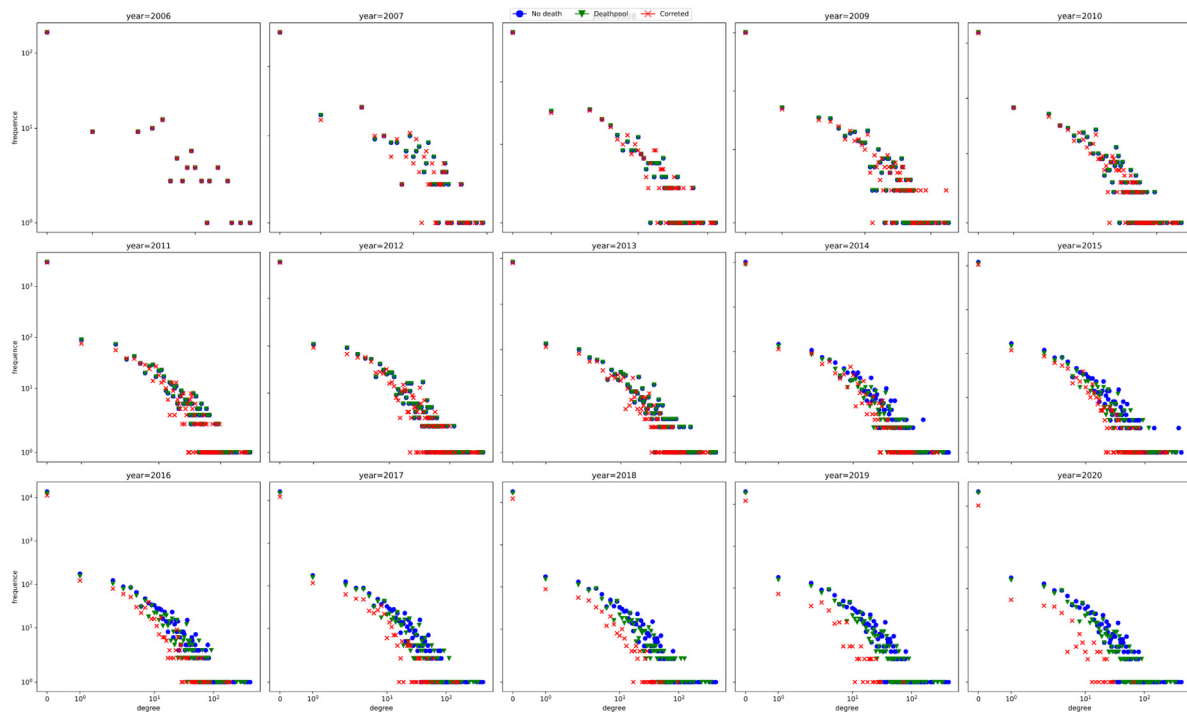


Fig. 11. The degree distribution changes over time (yearly).

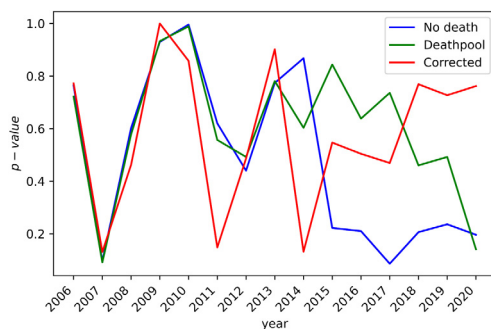


Fig. 12. The  $p$ -value for the power law changes over time (yearly).

to provide a new service, the mashup may achieve a higher probability of recommendation if it includes eCommerce services.

Unlike the categories mentioned above, the *search* node was initially very large and then gradually decreased. It also moved to the edge of the graph. This implies that using complete search services is not as necessary as it was before. Instead, services may implement their own searches rather than relying on search services provided by others.

Additionally, Fig. 13 also reflects the slow update and incomplete inclusion of APIs and mashups in the current *ProgrammableWeb* dataset. For example, the APIs of speech recognition, natural language processing (NLP), image processing and other cognitive services, which have been promoted by large companies such as *Microsoft*, *Google*, *Amazon*, and *Baidu* in recent years, have not been included.

#### RQ4: [A-A] What are the changes in the number and size of connected components over time?

The connected components form a subgraph in which any two vertices are connected to each other by paths that are connected to no additional vertices in the subgraph. Connected components are an important feature of the A-A network. On the one hand,

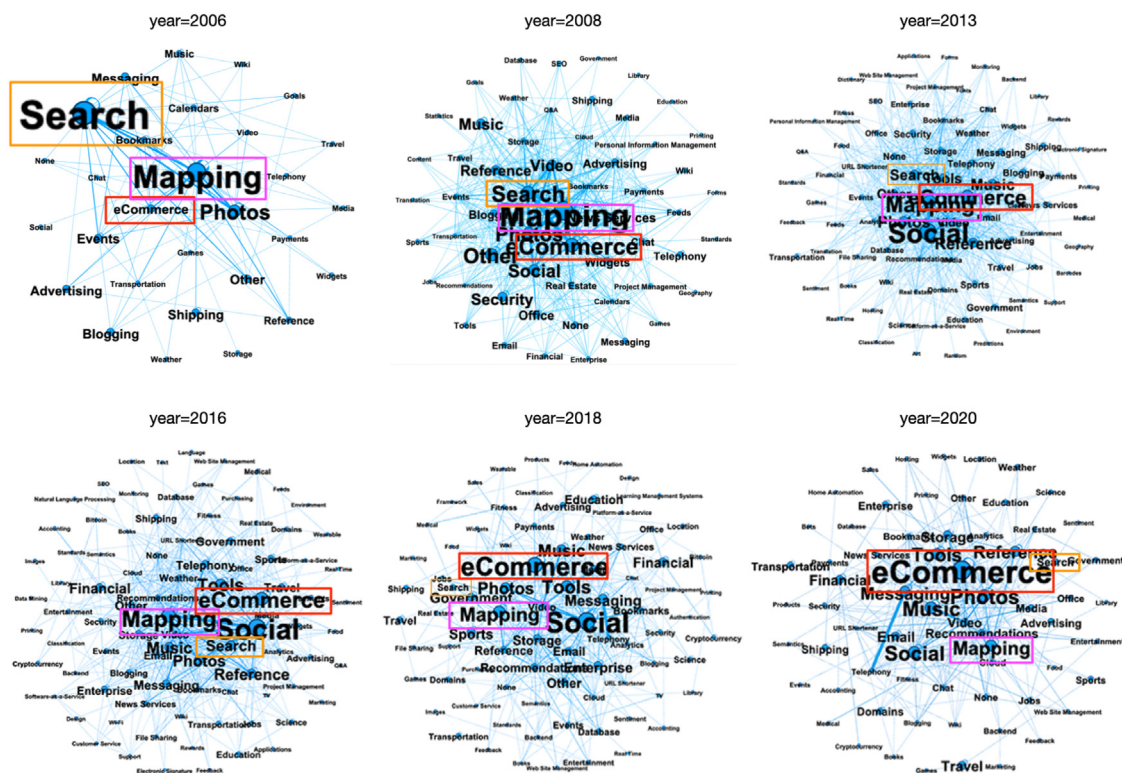
connected components can reflect the development of the service ecosystem via their number and size. Intuitively, we know that if the entire A-A network is a large connected component, more combinations are available, making it more likely that rich applications (mashups) will be created. On the other hand, many service recommendation and service discovery algorithms (Zhong et al., 2014; Gao et al., 2016; Adeleye et al., 2018) are essentially performed on connected components that reach a certain scale.

Fig. 15 shows how the number of connected components and the size of the largest connected components has changed over time. The total number of connected components first grew, then stabilized, and then began to decline sharply after 2018, but the number of components greater than 4 remained stable. The size of the largest connected component has continued to shrink since 2013; by 2020, it was below 200. The figure also shows that many isolated connected components exist in the A-A network and that most of the connected components are very small (less than 5). From a data science perspective, most of the data (isolated small components) provide poor information when performing tasks such as data-driven service recommendation, service discovery, and service composition. The size of the largest connected component is also quite small, which introduces the small data problem and greatly limits the ability to apply advanced machine learning methods (Zhang et al., 2019, 2020; Goyal and Ferrara, 2018; Wang et al., 2018b) to service computing.

#### RQ5: [A-A] Are mashups that use frequently co-occurring APIs more likely to survive?

The existing service recommendation methods (Ma et al., 2021; Gao and Wu, 2017) essentially involve statistical analysis and processing of the frequency of co-occurring APIs. The more frequently an API pair appears, the more likely it is to be recommended to developers. This naturally raises the question of whether frequently co-occurring APIs truly represent a good combination. More directly, is the survival rate of mashups that use frequently co-occurring APIs higher?

Fig. 16 demonstrates the changes of mashups' average survival rates over the co-occurring frequencies of APIs. The survival rate

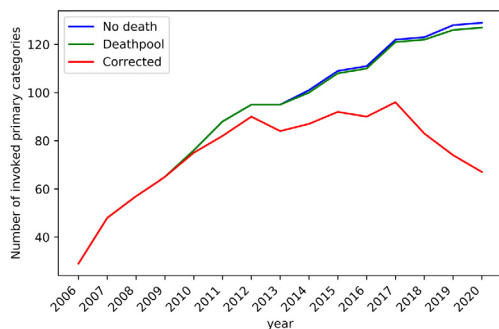


**Fig. 13.** Visualization of the evolution of the category-category network (yearly).

Table 2

Top 5 most frequently co-occurring API pairs and 5 low-frequency but high-survival-rate API pairs.

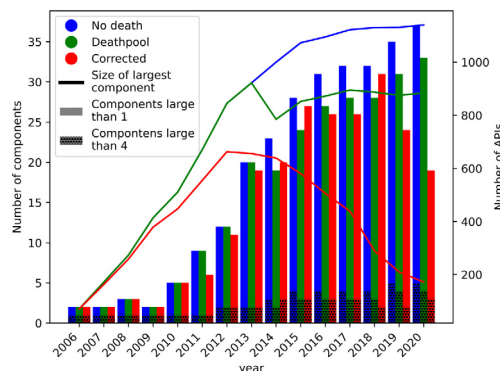
| API1                      | API2                  | Active use | Total use | Survival rate | Avg days |
|---------------------------|-----------------------|------------|-----------|---------------|----------|
| /api/twitter              | /api/google-maps      | 63         | 185       | 0.34          | 1,405.09 |
| /api/youtube              | /api/flickr           | 36         | 184       | 0.2           | 1,589.16 |
| /api/twitter              | /api/facebook         | 64         | 179       | 0.36          | 1,414.26 |
| /api/google-maps          | /api/flickr           | 47         | 169       | 0.28          | 1,608.62 |
| /api/google-maps          | /api/youtube          | 48         | 157       | 0.31          | 1,596.5  |
| /api/google-visualization | /api/google-maps      | 8          | 10        | 0.8           | 769.0    |
| /api/facebook             | /api/instagram-graph  | 9          | 13        | 0.69          | 1,094.75 |
| /api/instagram-graph      | /api/twitter          | 12         | 18        | 0.67          | 1,212.17 |
| /api/google-maps          | /api/google-geocoding | 21         | 33        | 0.64          | 1,117.75 |
| /api/google-maps          | /api/zillow           | 7          | 11        | 0.64          | 1,070.25 |



**Fig. 14.** The number of invoked primary categories (diversity) changes over time (yearly).

reaches the highest level when the frequency is between 40 and 60 and then decreases, which clearly indicates that using high-frequency pairs does not mean a tendency to survive.

Table 2 lists some co-occurring APIs as examples: the first five are the most frequently co-occurring API pairs, and the last five are low-frequency (e.g., less than or approximately 30 times) but



**Fig. 15.** The number of connected components and the size of the largest connected component (yearly).

high-survival-rate API pairs. Clearly, the mashups that invoke the frequently co-occurring API pairs do not have a higher survival rate. The survival rates of the five API pairs with the highest usage frequency are all below 40%, and those that use (*YouTube*



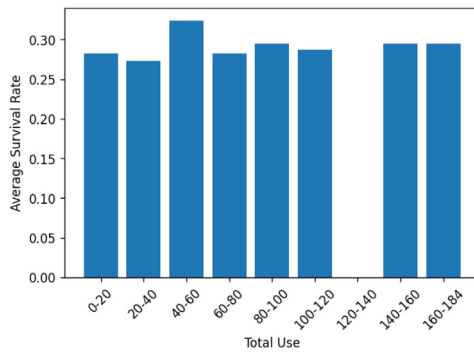


Fig. 16. Survival rate distribution of mashups that use co-occurring APIs.

and Flickr) are as low as 20%. However, the survival rate of some API combinations used at low frequencies exceeds 60%, and the highest reach 80%. Another interesting observation is that among the high-survival API pairs, in most cases, the two APIs are from the same developer, or one of the API categories is *Social* (e.g., Twitter, Facebook, or Instagram Graph).

In conclusion, **mashups that use frequently co-occurring API combinations do not mean that they are more likely to survive**. This phenomenon has brought new challenges to service recommendation and service composition, such as how to consider factors such as survival rate and developer relationships in recommendation or composition.

#### RQ6: [M-A] Are new mashup sizes becoming larger over time?

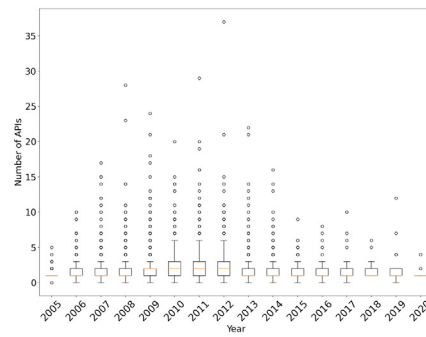
With the rapid development of the Internet, user requirements have become increasingly complex. It has become a trend for service providers to connect more external applications to their own platforms to satisfy these increasingly complex user requirements (Wu et al., 2015; Xu et al., 2015; Weiss and Gangadharan, 2010). For example, WeChat and Alipay achieved this by *Mini-program*; Microsoft released *Office*<sup>19</sup> to enable users to access Word, PowerPoint and Excel in one app. Therefore, a large size mashup tends to have more complete functionalities.

We wanted to know whether the abovementioned evolutionary trend is also occurring in Web application-based mashups, that is, whether mashups have changed to meet the increasingly complex requirements of users.

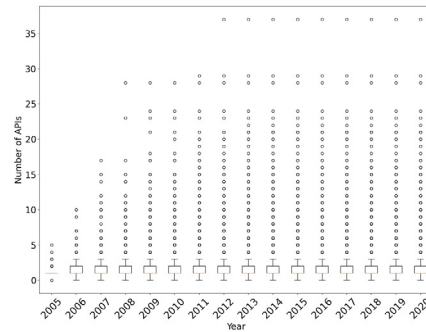
A mashup's size can be denoted by the number of APIs it invokes. Fig. 15 shows a box chart of the number of APIs invoked by newly submitted mashups each year. From the figure, we can see that most mashups provide only a simple service and the number of APIs they invoke is less than 3. The overall size of newly submitted mashups changed very little between 2006 and 2020: the average numbers of APIs they invoke remains steady between 1.2 and 2.5. The boxplot in Fig. 17(b) shows the changes in size among all mashups. Apparently, no large change has occurred since 2012 in any part of this boxplot. Another interesting phenomenon is that the size of new mashups gradually increased each year prior to 2012. The most complex mashup added in 2012 was composed of 36 APIs, which is a fairly dramatic number. However, after 2012, the complexity of newly submitted mashups began to decline.

**The size of mashups in recent years has not increased**, which may imply that Web-based mashups cannot adapt well to the increasingly complex requirements of users. The popularity of the mobile Internet, the lack of open APIs, and the existence of commercial barriers may be other reasons for the decline in Web-based mashups.

<sup>19</sup> <https://play.google.com/store/apps/details?id=com.microsoft.office.officehubrow>.



(a) The complexity of new mashups changes over time.



(b) The complexity of all mashups changes over time.

Fig. 17. The complexity of mashups changes over time (yearly).

## 7. Threats to validity

In this section, we will discuss the threats to validity in detail. There are three main threats: **An Alternative Method**, **A More Appropriate Probability Distribution**, **Time Errors** and **Accuracy of Rule-based Techniques**.

**An Alternative Method.** There are two categories of methods to perform the estimation of death time:

1. One is the overall estimation based on the probability distribution represented by the method proposed in Section 4.1. This kind of method is independent of the information in the specific API/mashup and can only guarantee a reasonable overall sample distribution but cannot provide a precise estimation.
2. Another method that we did not adopt is to construct a regression model and use it to predict the number of days each unavailable API/mashup survives. Ideally, this is the best method because it ensures both a reasonable overall distribution and individual accuracy. Unfortunately, this method is not practicable on the *ProgrammableWeb* dataset. First, the amount of data is too small, which hinders deep learning-based regression models. Second, the information provided in the *ProgrammableWeb* dataset makes it difficult to identify features related to survival days, which hinders the feature engineering-based machine learning regression models.

As we stated at the beginning of Section 4, our proposed method is not perfect and can only guarantee a reasonable overall data distribution, but this is sufficient in the scenario of evolutionary analysis. When we have more labeled data and richer information, it will be better to construct a regression model, which



will produce more accurate results. Therefore, it is important to emphasize that the method proposed in this paper is a relatively reasonable and feasible way to address the evolutionary analysis scenario under the current data conditions.

**A More Appropriate Probability Distribution.** In Section 4.1, we assumed that the longevity of APIs/mashups follows a normal distribution and confirmed that the overall distribution and the data used for estimation were identically distributed by the Z-test. However, this does not indicate that the normal distribution is the best distribution. As more labeled data are provided, we may find that the survival days of the service follow a more complex distribution so that we can replace the normal distribution now used for estimation with a more accurate probability distribution. It should be noted that under the condition of limited observation data, it is not suitable to choose a complex distribution—it is more robust to utilize a simple distribution such as the normal distribution.

**Time Errors.** In Section 4.1, we select 20 APIs to check their dead times and found an error of 50 days. This error may slightly affect some of the descriptions in some of the research questions; for example, the turning points in RQ1 may be offset, but these do not affect the final conclusions of the research questions in this paper. We have also repeated the experiment several times to mitigate this threat.

**Rule-based Techniques.** We use rule-based techniques when checking the availability of APIs and Mashups. Due to the limitation of rule-based techniques, there may be a small number of false positive and false negative samples, which may result in slight variations in specific values, such as those in Table 1. To deal with this issue, we have taken repeated tests and manual sampling to verify the validity of these techniques. More manual checking and improving the rules-based technology can further reduce the risks associated with the rules-based technology.

## 8. Conclusion

In this paper, we analyzed the evolution of the *ProgrammableWeb* service ecosystem from a dynamic network perspective. We first summarized the quality issues in the original *ProgrammableWeb* dataset and analyzed the negative effects of these quality issues on traditional service computing tasks. Then, we proposed novel methods to correct the evolution-related data quality issues, including API/mashup availability status, API/mashup death time, and mashup composition. Finally, we conducted a set of extensive experimental analyses on the corrected *ProgrammableWeb* dataset, and the most intriguing finding is that the development of the *ProgrammableWeb* service ecosystem is less optimistic than that reported by the existing studies. In fact, we discovered considerable evidence that the *ProgrammableWeb* is declining. Our empirical analysis has identified a number of research challenges and opportunities. To ensure that the *ProgrammableWeb* community continues to flourish, we encourage researchers, developers, and managers to address these challenges with innovative solutions.

## CRediT authorship contribution statement

**Mingyi Liu:** Conceptualization, Methodology, Software, Writing – original draft, Visualization. **Zhiying Tu:** Conceptualization, Writing – original draft, Resources. **Yeqi Zhu:** Software, Visualization, Investigation. **Xiaofei Xu:** Funding acquisition. **Zhongjie Wang:** Conceptualization, Writing – original draft, Project administration, Funding acquisition. **Quan Z. Sheng:** Writing – original draft.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The research in this paper is partially supported by the National Key Research and Development Program of China (No 2018YFB1402500) and the National Natural Science Foundation of China (61772155, 61832004, 61802089, 61832014).

## References

- Adeleye, O., Yu, J., Yongchareon, S., Han, Y., 2018. Constructing and evaluating an evolving web-API network for service discovery. In: Proceedings of the 16th International Conference on Service-Oriented Computing, pp. 603–617.
- Adeleye, O., Yu, J., Yongchareon, S., Sheng, Q.Z., Yang, L.H., 2019. A fitness-based evolving network for web-APIs Discovery. In: Proceedings of the Australasian Computer Science Week Multiconference, pp. 1–10.
- Bai, B., Fan, Y., Tan, W., Zhang, J., 2017. SR-LDA: Mining effective representations for generating service ecosystem knowledge maps. In: 2017 IEEE International Conference on Services Computing (SCC). IEEE, pp. 124–131.
- Botangen, K.A., Yu, J., Sheng, Q.Z., Han, Y., Yongchareon, S., 2020. Geographic-aware collaborative filtering for web service recommendation. Expert Syst. Appl. 151, 113347.
- Bouguettaya, A., Singh, M.P., Huhns, M.N., Sheng, Q.Z., Dong, H., Yu, Q., Neiat, A.G., Mistry, S., Benattallah, B., Medjahed, B., Ouzzani, M., Casati, F., Liu, X., Wang, H., Georgakopoulos, D., Chen, L., Nepal, S., Malik, Z., Erradi, A., Wang, Y., Blake, M.B., Dustdar, S., Leymann, F., Papazoglou, M.P., 2017. A service computing manifesto: the next 10 years. Commun. ACM 60 (4), 64–72.
- Brown, A., Fishenden, J., Thompson, M., 2014. API economy, ecosystems and engagement models. In: Digitizing Government: Understanding and Implementing New Digital Business Models. Palgrave Macmillan UK, London, pp. 225–236.
- Chen, Z., Shen, L., Li, F., 2017. Exploiting web service geographical neighborhood for collaborative QoS prediction. Future Gener. Comput. Syst. 68, 248–259.
- Chen, Z., Sun, Y., You, D., Li, F., Shen, L., 2020. An accurate and efficient web service QoS prediction model with wide-range awareness. Future Gener. Comput. Syst. 108, 275–292.
- Clauset, A., Shalizi, C.R., Newman, M.E., 2009. Power-law distributions in empirical data. SIAM Rev. 51 (4), 661–703.
- Gan, Y., Xiang, Y., Zou, G., Miao, H., Zhang, B., 2019. Multi-label recommendation of web services with the combination of deep neural networks. In: Proceedings of International Conference on Collaborative Computing: Networking, Applications and Worksharing. Springer, pp. 133–150.
- Gao, W., Chen, L., Wu, J., Bouguettaya, A., 2016. Joint modeling users, services, mashups, and topics for service recommendation. In: Proceedings of 2016 IEEE International Conference on Web Services (ICWS). IEEE, pp. 260–267.
- Gao, W., Wu, J., 2017. A novel framework for service set recommendation in mashup creation. In: Proceedings of 2017 IEEE International Conference on Web Services (ICWS). IEEE, pp. 65–72.
- Goyal, P., Ferrara, E., 2018. Graph embedding techniques, applications, and performance: A survey. Knowl.-Based Syst. 151, 78–94.
- Huang, K., Fan, Y., Tan, W., 2014. Recommendation in an evolving service ecosystem based on network prediction. IEEE Trans. Autom. Sci. Eng. 11 (3), 906–920.
- Huang, P.-Y., Liu, H.-Y., Chen, C.-H., Cheng, P.-J., 2013. The impact of social diversity and dynamic influence propagation for identifying influencers in social networks. In: Proceedings of 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). 1, IEEE, pp. 410–416.
- Kalaï, A., Zayani, C.A., Amous, I., Abdelghani, W., Sèdes, F., 2018. Social collaborative service recommendation approach based on user's trust and domain-specific expertise. Future Gener. Comput. Syst. 80, 355–367.
- Kochan, T., Bezrukova, K., Ely, R., Jackson, S., Joshi, A., Jehn, K., Leonard, J., Levine, D., Thomas, D., 2003. The effects of diversity on business performance: Report of the diversity research network. Hum. Resour. Manag. 42 (1), 3–21.
- Liu, M., Tu, Z., Wang, J., Wang, Z., 2020. A novel multi-layer network model for service ecosystems. In: Proceedings of 2020 International Conference on Service Science (ICSS), pp. 23–30.
- Liu, M., Tu, Z., Xu, X., Wang, Z., 2020. A data-driven approach for constructing multilayer network-based service ecosystem models. arXiv preprint arXiv: 2004.10383.
- Lyu, S., Liu, J., Tang, M., Kang, G., Cao, B., Duan, Y., 2014. Three-level views of the web service network: An empirical study based on ProgrammableWeb. In: Proceedings of 2014 IEEE International Congress on Big Data. IEEE, pp. 374–381.

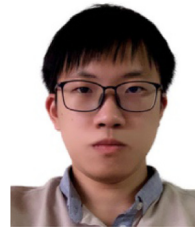
- Ma, Y., Geng, X., Wang, J., 2021. A deep neural network with multiplex interactions for cold-start service recommendation. *IEEE Trans. Eng. Manage.* 68 (1), 105–119.
- Ngu, A.H., Carlson, M.P., Sheng, Q.Z., Paik, H., 2010. Semantic-based mashup of composite applications. *IEEE Trans. Serv. Comput.* 3 (1), 2–15.
- Pan, W., Chai, C., 2018. Structure-aware mashup service clustering for cloud-based internet of things using genetic algorithm based clustering algorithm. *Future Gener. Comput. Syst.* 87, 267–277.
- Sampaio, A.R., Kadiyala, H., Hu, B., Steinbacher, J., Erwin, T., Rosa, N., Beschastnikh, I., Rubin, J., 2017. Supporting microservice evolution. In: *Proceedings of 2017 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, pp. 539–543.
- Sheng, Q.Z., Qiao, X., Vasilakos, A.V., Szabo, C., Bourne, S., Xu, X., 2014. Web services composition: A decade's overview. *Inform. Sci.* 280, 218–238.
- Tian, Y., Kochhar, P.S., Lo, D., 2017. An exploratory study of functionality and learning resources of web APIs on ProgrammableWeb. In: *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering*, pp. 202–207.
- Wang, J., Chen, H., Zhang, Y., 2009. Mining user behavior pattern in mashup community. In: *Proceedings of 2009 IEEE International Conference on Information Reuse & Integration*. IEEE, pp. 126–131.
- Wang, X., Feng, Z., Chen, S., Huang, K., 2018a. DKEM: A distributed knowledge based evolution model for service ecosystem. In: *Proceedings of 2018 IEEE International Conference on Web Services (ICWS)*. IEEE, pp. 1–8.
- Wang, H., Wang, J., Wang, J., Zhao, M., Zhang, W., Zhang, F., Xie, X., Guo, M., 2018. Graphgan: Graph Representation Learning with Generative Adversarial Nets. In: *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Wang, S., Wang, Z., Xu, X., Sheng, Q.Z., 2017. App update patterns: How developers act on user reviews in mobile app stores. In: *Proceedings of the 15th International Conference on Service-Oriented Computing (ICSOC)*, pp. 125–141, Malaga, Spain.
- Weiss, M., Gangadharan, G., 2010. Modeling the mashup ecosystem: Structure and growth. *R&D Manage.* 40 (1), 40–49.
- Wittern, E., Laredo, J., Vukovic, M., Muthusamy, V., Slominski, A., 2014. A graph-based data model for API ecosystem insights. In: *Proceedings of 2014 IEEE International Conference on Web Services*. IEEE, pp. 41–48.
- Wu, Z., Yin, J., Deng, S., Wu, J., Li, Y., Chen, L., 2015. Modern service industry and crossover services: Development and trends in China. *IEEE Trans. Serv. Comput.* 9 (5), 664–671.
- Xu, X., Liu, Z., Wang, Z., Sheng, Q.Z., Yu, J., Wang, X., 2017. S-ABC: A paradigm of service domain-oriented artificial bee colony algorithms for service selection and composition. *Future Gener. Comput. Syst.* 68, 304–319.
- Xu, X., Sheng, Q.Z., Zhang, L.-J., Fan, Y., Dustdar, S., 2015. From big data to big service. *Computer* 48 (7), 80–83.
- Yao, L., Sheng, Q., Ngu, A., Yu, J., Segev, A., 2015. Unified collaborative and content-based web service recommendation. *IEEE Trans. Serv. Comput.* 8 (3), 453–466.
- Zhang, Z., Cui, P., Zhu, W., 2020. Deep learning on graphs: A survey. *IEEE Trans. Knowl. Data Eng.*
- Zhang, S., Tong, H., Xu, J., Maciejewski, R., 2019. Graph convolutional networks: A comprehensive review. *Computational Soc. Netw.* 6 (1), 11.
- Zhong, Y., Fan, Y., Huang, K., Tan, W., Zhang, J., 2014. Time-aware service recommendation for mashup creation. *IEEE Trans. Serv. Comput.* 8 (3), 356–368.
- Zhou, S., Wang, Y., 2019. Service ranking in service networks using parameters in complex networks: A comparative study. *Cluster Comput.* 22 (2), 2921–2930.



**Mingyi Liu** received his B.S. degree from the School of Computer Science and Technology, Harbin Institute of Technology in 2018. He is currently pursuing the Ph.D. degree in software engineering at Harbin Institute of Technology (HIT), China. His research interests include service ecosystem models, service evolution analysis, data mining and knowledge graphs.



**Zhiying Tu** is an associate professor of School of Computer Science and Technology at Harbin Institute of Technology (HIT). He holds a Ph.D. degree in Computer Integrated Manufacturing (Productique) from the University of Bordeaux. Since 2013, He began to work at HIT. His research interest is Service Computing, Enterprise Interoperability, and Cognitive Computing. He has 20 publications as edited books and proceedings, refereed book chapters, and refereed technical papers in journals and conferences. He is the member of IEEE Computer Society, and CCF China.



**Yeqi Zhu** is an undergraduate from Faculty of Computing, Harbin Institute of Technology (HIT). He will start studying for a master's degree from 2021. His research interests include knowledge graphs and service recommendations.



of the IEEE and ACM.

**Xiaofei Xu** is a professor at Faculty of Computing, and vice president of Harbin Institute of Technology. He received the Ph.D. degree in computer science from Harbin Institute of Technology in 1988. His research interests include enterprise intelligent computing, services computing, Internet of services, and data mining. He is the associate chair of IFIP TC5 WG5.8, chair of INTEROP-VLab China Pole, fellow of China Computer Federation (CCF), and the vice director of the technical committee of service computing of CCF. He is the author of more than 300 publications. He is member



**Zhongjie Wang** is a professor at Faculty of Computing, Harbin Institute of Technology (HIT). He received the Ph.D. degree in computer science from Harbin Institute of Technology in 2006. His research interests include services computing, mobile and social networking services, and software architecture. He is the author of more than 40 publications. He is a member of the IEEE.



Outstanding Research Contribution (2012), and Microsoft Fellowship (2003).

**Quan Z. Sheng** is a professor and Head of Department of Computing at Macquarie University, Sydney, Australia. His research interests include service oriented computing, distributed computing, Internet computing, and Internet of Things. He holds a Ph.D. degree in computer science from the University of New South Wales (UNSW) and did his post-doc as a research scientist at CSIRO ICT Centre. He has more than 360 publications. Prof Quan Z. Sheng is the recipient of AMiner Most Influential Scholar Award on IoT (2019), ARC Future Fellowship (2014), Chris Wallace Award for