

雷文全,唐明董,张祥平,等.面向 Mashup 的 Web API 生态网络结构分析[J].湖南科技大学学报(自然科学版),2019,34(4):97-103.doi:10.13582/j.cnki.1672-9102.2019.04.014

Lei W Q, Tang M D, Zhang X P, et al. A Network Analysis of the Mashup-oriented Web API Ecosystem[J]. Journal of Hunan University of Science and Technology ( Natural Science Edition ), 2019, 34( 4 ): 97- 103. doi: 10. 13582/ j. cnki. 1672- 9102. 2019. 04. 014

# 面向 Mashup 的 Web API 生态网络结构分析

雷文全,唐明董\*,张祥平,夏艳敏

(湖南科技大学 计算机科学与工程学院,湖南 湘潭 411201)

**摘要:**随着 API 经济的发展,越来越多的公司通过 Web API 开放自己的数据和服务,以实现商业增值.与此同时,通过调用和合成不同 Web API 的 Mashup 应用开发模式也逐渐流行,形成了面向 Mashup 的 Web API 生态系统.然而,目前对这种 Web API 生态系统的认识还不够深入,对 API 关系网络和 API 提供者关系网络缺少分析.本文利用 ProgrammableWeb.com 上的真实数据对 Web API 生态网络结构进行了分析,提出了挖掘 Web API 协作网络以及 Web API 提供者合作网络的方法,在此基础上分析了这些网络的基本结构特征,如网络尺寸、平均度、聚集系数、同配/异配性等.分析结果不仅能加深对 Web API 生态系统的认识,还可以为 Web API 的有效发现提供支持.

**关键词:**Web API; 网络分析; Mashup; 结构特征; Web API 提供者

**中图分类号:**TP399 **文献标志码:**A **文章编号:**1672-9102(2019)04-0097-07

## A Network Analysis of the Mashup-oriented Web API Ecosystem

Lei Wenquan, Tang Mingdong, Zhang Xiangping, Xia Yanmin

(School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China)

**Abstract:** With the advent of API economy, there was an increasing number of companies opening their data and services to the public by publishing Web APIs, aiming at gaining extra economic benefits. Meanwhile, Mashup applications which were made by invoking or combining different Web APIs become more popular than ever, and consequently raising up the Mashup-oriented ecological system. However, so far there still lacked of deep understanding of the Web API ecological networks, e.g., the Web API collaboration and API provider collaboration networks. Base on the real data from ProgrammableWeb.com, an approach to constructing collaboration network of Web APIs and API providers was proposed, and the structural features of these networks were analyzed, including network size, average degree, degree distribution, aggregate coefficient and so on. The results can not only help understanding the Web API ecosystem, but also provide support for effective Web API discovering.

**Keywords:** Web API; network analysis; Mashups; structural features; Web API provider

随着互联网的发展和 Web 服务相关技术<sup>[1-3]</sup>的进步,越来越多的软件服务开始出现在互联网上,这些服务提供了应用程序访问接口(API)以供外部调用.开发者可以通过调用或合成网络上的开放服务或 API,快速开发各类应用.由于开放服务或 API 所带来的增值效应,包括 Google, Baidu 在内的越来越多的

收稿日期:2018-06-10

基金项目:国家自然科学基金资助项目(61572186)

\*通信作者,E-mail:mdtang@126.com

公司均发布了 Web API,提供各种信息和数据服务.越来越多的 Web API 出现在网上,将不同功能的 Web API 和互联网资源通过 Mashup 方式合成,可以快速开发 Web 应用<sup>[4-5]</sup>.随着 Mashup 的流行,Web API 生态系统<sup>[6]</sup>得到了迅速发展.

面向 Mashup 的 Web API 生态系统是由 Web API, Web API 提供者, Mashup 等以及它们之间存在的关联关系(引用、合作或竞争等)构成的开放复杂系统<sup>[7-10]</sup>.图 1 为一个 Web API 生态系统示例.图 1 中显示了 Mashup 应用、应用开发者(Developer)、Web API 和 Web API 提供者(Provider)以及它们之间的关系.例如,100 Most Powerful Celebrities 和 Mileage Calculator 为 Mashup 应用;Yahoo Geocoding, YouTube, Google Calendar 和 Google Maps 为 Web API; yahoo.com 为服务提供者; koolade 和 mibazaar 为应用开发者.显然, Yahoo Geocoding 和 YouTube 构成协作关系, yahoo.com 和 google.com 构成合作关系.

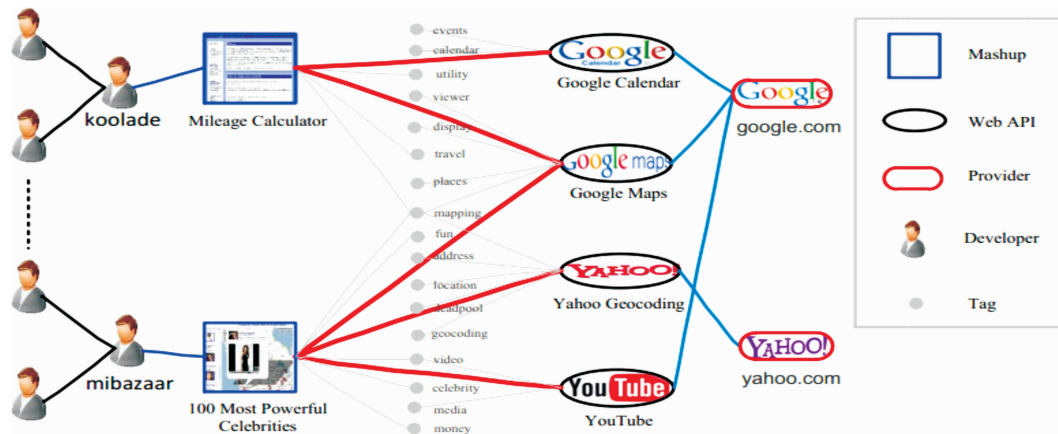


图 1 Web API 生态系统示例

为了认识和理解 Web API 生态系统,已有一些学者开展了相关的研究<sup>[11-16]</sup>.他们分析了 Web API 生态系统的规模和成分,也发现了 API 生态系统中的小世界(Small-World)、幂律(Power-Law)等结构特征.例如,Cherifi 等<sup>[17]</sup>通过比较用语法分析和语义分析 2 种方式抽取服务依赖关系得到的服务网络的结构特征,发现它们都具有小世界特征和社区化特征;Huang 等<sup>[18]</sup>利用 Programmable.com 上的真实数据构建 Mashup 网络并对其结构属性进行分析,发现了幂律、异配性(Disassortative)等重要结构特征.

然而,目前的工作虽然分析了 Web API 生态网络的一些统计特征,但是对 Web API 生态网络的认识还远不够全面和深入.本文使用 ProgrammableWeb.com<sup>[19]</sup>上的真实数据构建出 Web API 协作关系网络和 Web API 提供者网络.利用 Gephi 工具将这些网络可视化,揭示了关系较强的节点对,并统计了这些网络的基本结构特征,如网络规模、平均度、聚集系数和同配/异配性等,验证了 Web API 生态网络的小世界、幂律等网络特性.

## 1 Web API 数据集

本文使用的数据集来自于 ProgrammableWeb.com,该网站是目前最受欢迎的 Web API 门户网站.我们获取了 Web API 和 Mashup 的相关属性.Web API 包含的属性主要有 name(名称),Tags(标签)和 URL(地址); Mashup 涉及的属性主要有 name, Tags 和它使用的 API.表 1 和表 2 显示了 Web API 和 Mashup 的主要属性.

表 1 Web API 的基本属性举例

API name	Tags	Category	URL	Protocol
Shiny Ads	advertising	Advertising	http://ShinyAds.com	REST
Websnapt	image	Security	http://www.	JavaScript
	security		Websnapt.com/code/	

表 2 Mashup 的基本属性举例

Mashup name	Tags	APIs
Earth Sandwich	mapping, science	Google Maps
AuctionPixie	auction, shopping	eBay, hostip.info

ProgrammableWeb.com 数据集的统计结果如表 3 所示.从表 3 中可以得出:数据集 ProgrammableWeb.com 中 Mashup 的总数量为 6 970, Web API 的总数量为 9 135, 而 9 135 个 Web API 中,目前只有 1 194 个 Web API 被用来构建 Mashup 应用,因此多数 Web API 的使用率比较低,意味着 Web API 市场尚处于开始阶段,仍有很大的发展和应用空间.

表 3 ProgrammableWeb.com 数据集的统计结果

Mashup 总数量	6 970
Web API 总数量	9 135
每个 Mashup 平均包含 Web API 的个数	1.31
Mashup 中使用的 Web API 总数量	1 194
Web API 使用的 Tag 总数量	1 727
每个 Web API 平均使用 Tag 的个数	3.275
Web API 提供商总数量	7 242
每个提供商平均发布 Web API 的个数	1.261

## 2 Web API 生态网络关系挖掘

从 Web API 生态系统中可以挖掘 2 种典型关系,即 Web API 协作关系和 Web API 提供者合作关系.简单地说,Web API 协作关系指 2 个 API 共同出现在至少 1 个 Mashup 应用中,即被同一个 Mashup 所组合;Web API 提供者合作关系是指在它们拥有的 Web API 之间存在协作关系.

### 2.1 Web API 协作关系网络

定义 1 (Web API 协作关系网络):若 2 个 Web API 出现在相同的 Mashup 中,称它们具有协作关系,Web API 协作关系网络表示全部 Web API 与 Web API 之间的协作关系.这种关系可以采用  $d \times d$  的矩阵  $A = [a_{ij}]_{d \times d}$  表示,矩阵中元素的值  $a_{ij}$  等 Web API<sub>*i*</sub> 与 Web API<sub>*j*</sub> 的协作关系强度,即共同参与过的 Mashup 应用个数.

图 2 给出了根据定义推导出 Web API 协作网络以及计算边权的过程示意图.例如,Web API<sub>1</sub> 在 Mashup<sub>1</sub>, Mashup<sub>2</sub> 和 Mashup<sub>3</sub> 中出现过,Web API<sub>2</sub> 在 Mashup<sub>1</sub> 和 Mashup<sub>2</sub> 中出现过;由于 Web API<sub>1</sub> 和 Web API<sub>2</sub> 在 Mashup<sub>1</sub> 和 Mashup<sub>2</sub> 这 2 个 Mashup 中共同出现过,所以得到边权  $(API_1, API_2) = 2$ .

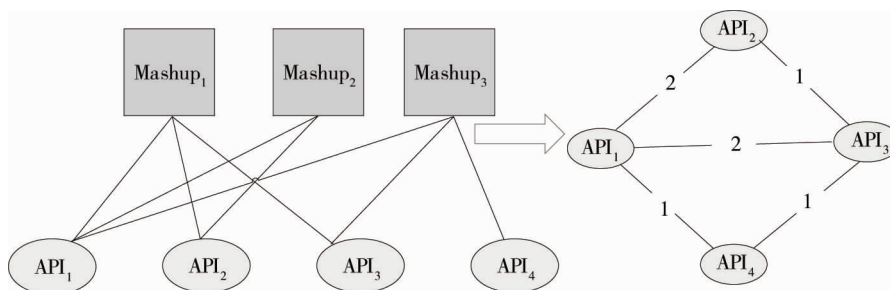


图 2 Web API 协作关系网络举例

### 2.2 Web API 提供者合作关系网络

定义 2 (Web API 提供者合作关系网络):若 2 个 Web API 提供者之间存在协作关系的 Web API,则称它们具有合作关系,Web API 提供者合作关系网络表示全部 Provider 与 Provider 之间的合作关系.这种关系可以采用  $d \times d$  的矩阵  $C = [c_{ij}]_{d \times d}$  表示,矩阵中元素的值  $c_{ij}$  等于 Provider<sub>*i*</sub> 与 Provider<sub>*j*</sub> 中具有协作关系

的Web API对数之和比上  $Provider_i$  与  $Provider_j$  的数量之积.

根据定义 2,我们可以得出 Web API 提供者合作关系边权值的计算公式

$$C(i,j) = \frac{\sum A(i,j)}{P(i)P(j)}. \tag{1}$$

式中:  $C(i,j)$  为 Web API 提供者合作关系网络边权值;  $\sum A(i,j)$  为  $Provider_i$  与  $Provider_j$  中具有协作关系的 Web API 对数之和;  $P(i)$  为  $Provider_i$  中的 Web API 个数;  $P(j)$  为  $Provider_j$  中的 Web API 个数.

根据定义 2 及式(1),由 Web API 协作关系网络推导出 Web API 提供者合作关系网络边权值的示意图如图 3 所示. $Pro_1$  提供了 Web API<sub>1</sub> 这 1 个 Web API,  $Pro_2$  提供了 Web API<sub>2</sub> 和 Web API<sub>4</sub> 这 2 个 Web API;根据 Web API 协作关系网络的定义,我们可以得出:Web API<sub>1</sub> 与 Web API<sub>2</sub> 的协作关系边权值为 2,Web API<sub>1</sub> 与 Web API<sub>4</sub> 协作关系边权值为 1,故  $Pro_1$  与  $Pro_2$  中共现了 3 对 Web API;所以边权  $(Pro_1, Pro_2) = 3/2 = 1.5$ .

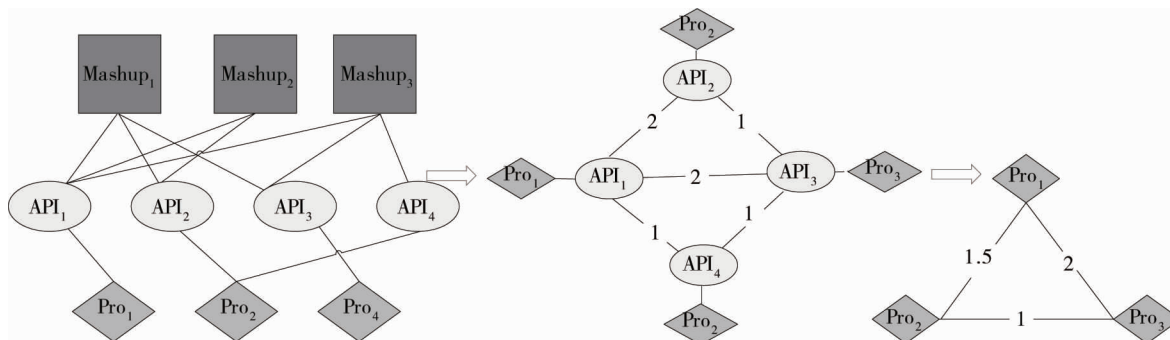


图 3 Web API 提供者合作关系网络举例

### 3 Web API 生态网络的结构分析

#### 3.1 Web API 协作关系网络分析

在得到 Web API 协作关系网络后,将连接度最大的前 10 个 Web API 罗列出来了,结果如表 4 所示.

将构建好的 Web API 协作关系网络用 Gephi 工具可视化,为了能够方便地展示该网络图,将度数小于 61 的节点和与之相连接的边全部都过滤掉,用剩下的 54 个节点和 959 条边构建出了一个简单的网络图,如图 4 所示.

表 4 度值最大的前 10 个 Web API

Web API	度值
Yahoo Shopping	414
Buildasearch	382
Netflix	356
Windows Sidebar Gadgets	327
Alexa Web Information Service	314
Factual	195
Commission Junction	182
Bloglines	162
Twitter	148
Amazon Product Advertising	143



图 4 Web API 协作关系网络可视化

统计其网络属性,得到表 5.从表 5 中可以看到:Web API 协作关系网络的网络直径很小,并且平均聚集系数很高.从这 2 个数据中可以分析得出:Web API 协作关系网络具有“小世界”网络属性.

图 5 描述了对数坐标系下的 Web API 协作关系网络中度分布的拟合图,其中横坐标为节点度值  $K$

的对数,纵坐标为度值为  $K$  的节点数目  $N$  的对数.从图 5 中可以看出:Web API 协作关系网络的度分布成幂律分布,这表明 Web API 协作关系网络是“无标度”网络.

表 5 Web API 协作关系网络属性

网络属性	属性值
节点数	9 161
边数	8 560
平均度	1.869
平均加权度	7.984
网络直径	7
平均聚集系数	0.745
同配系数	0.287

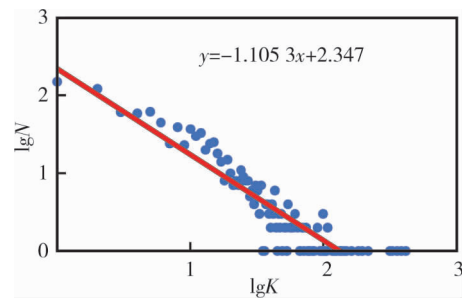


图 5 Web API 协作关系网络度分布

对 Web API 协作关系网络的关系最强节点对进行统计,并将协作关系最强的 10 对 Web API 罗列了出来,结果如表 6 所示.

表 6 协作关系最强的 10 对 Web API

序号	Web API <sub>1</sub>	Web API <sub>2</sub>	协作次数
1	Yahoo Shopping	Buildasearch	342
2	Windows Sidebar Gadgets	Netflix	340
3	Windows Sidebar Gadgets	Yahoo Shopping	322
4	Google Maps	Zillow	298
5	Yahoo Shopping	Netflix	296
6	Alexa Web Information Service	Buildasearch	288
7	Factual	Netflix	212
8	Buildasearch	Netflix	210
9	Windows Sidebar Gadgets	Buildasearch	196
10	Alexa Web Information Service	Yahoo Shopping	192

### 3.2 Web API 提供者合作关系网络分析

在得到 Web API 提供商合作关系网络后,将度值最大的前 10 个 Web API 提供商罗列出来了,结果如表 7 所示.

表 7 度值最大的前 10 个 Web API 提供商

提供者	度值
google.com	519
yahoo.com	261
flickr.com	229
facebook.com	218
twitter.com	205
amazon.com	161
last.fm	107
delicious.com	92
foursquare.com	85
x.com	84

将构建好的 Web API 提供者合作关系网络用 Gephi 工具可视化,为了能够方便地展示该网络图,我们将度值小于 30 的节点及与之相连接的边过滤掉,用剩下的 53 个节点和 662 条边构建了一个简单的网络图,如图 6 所示.



图6 Web API提供者合作关系网络可视化

统计其网络属性,得到表8.从表8中可以看到:该网络的网络直径很小,平均聚集系数很高,我们可以分析出Web API提供者合作关系网络具有“小世界”网络模型属性.

图7描述了在对数坐标系下的Web API提供商合作关系网络中度分布的拟合图,其中横坐标为节点度值K的对数,纵坐标为度值为K的节点数目N的对数.从图7中可以看出:Web API提供商合作关系网络的度分布成幂律分布,这表明Web API提供商合作关系网络是“无标度”网络.

表8 Web API提供者合作关系网络属性

网络属性	属性值
节点数	7 242
边数	4 398
平均度	1.215
平均加权重	0.595
网络直径	6
平均聚集系数	0.797
同配系数	0.494

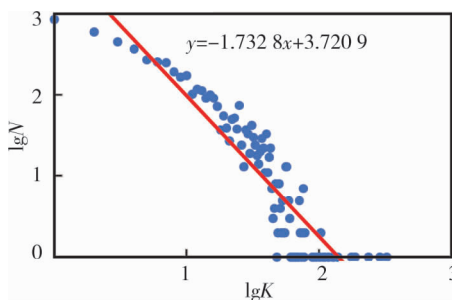


图7 Web API提供商合作关系网络度分布

对Web API提供者合作关系网络的关系最强节点对进行统计.表9给出了合作关系最强的10对Web API提供者.

表9 合作关系最强的10对Web API提供者

序号	Provider <sub>1</sub>	Provider <sub>2</sub>	强度
1	google.com	yahoo.com	253
2	amazon.com	google.com	138
3	facebook.com	google.com	108
4	google.com	twitter.com	68
5	amazon.com	yahoo.com	52
6	strikeiron.com	google.com	46
7	microsoft.com	google.com	37
8	google.com	x.com	35
9	bing.com	google.com	34
10	google.com	github.com	32

## 4 结论

1) 本文利用 Programmable.com 上的真实数据构建出了Web API生态网络并将其可视化,统计分析了这些网络的规模、平均度、度分布、聚集系数、同配/异配性等基本结构特性,验证了Web API协作关系网

络具有的“小世界”“幂律”网络特性.

2)提出了 Web API 提供者合作关系网络的概念、模型及构建方法,发现 API 提供者合作关系网络也具有“小世界”“幂律”网络特性.

3)分析结果不仅有助于认识和理解 Web API 生态网络内部结构,通过挖掘和分析 API 以及 API 提供者之间的关系还可以为 Web API 的高效发现提供支持.

#### 参考文献:

- [1] 陈世展,冯志勇,王辉.服务关系及其在面向服务计算中的应用[J].计算机学报,2010,33(11):2068-2083.
- [2] Feng Z, Lan B, Zhang Z, et al. A study of semantic Web services network[J].The Computer Journal, 2015, 58(6): 1293-1305.
- [3] 王忠杰,徐飞,徐晓飞.支持大规模个性化功能需求的服务网络构建[J].软件学报,2014,25(6):1180-1195.
- [4] Wang G, Liu J X, Cao B Q, et al. Mashup service classification and recommendation based on similarity computing[C]//CGC, 2012: 621-628.
- [5] Li H C, Liu J X, Cao B Q, et al. Integrating tag, topic, co-occurrence, and popularity to recommend Web APIs for Mashup creation[C]//IEEE International Conference on Services Computing. IEEE, 2017:84-91.
- [6] Barros A P, Dumas M. The rise of web service ecosystem[J]. IT professional, 2006, 8(5):31-37.
- [7] Endress N B. Semantic Mashups[M]. Heidelberg: Springer, 2013:1-51.
- [8] 吕赛霞.Web2.0 服务生态系统的挖掘与分析[D].湘潭:湖南科技大学,2015.
- [9] Wittern E, Laredo J, Vukovic M. A graph-based data model for API ecosystem insights[C]//ICWS, 2014: 41-48.
- [10] Gu Z, Xu B, Li J. Service data correlation modeling and its application in data-driven service composition[J]. IEEE Transaction on Services Computing, 2010, 3(4):279-291.
- [11] Chen X, Zheng Z, Liu X, et al. Personalized qos-aware Web service recommendation and visualization[J]. IEEE Transactions on Services Computing, 2013, 6(1): 35-47.
- [12] Yu S, Woodard C J. Innovation in the programmable Web: Characterizing the mashup ecosystem[C]//The 6th International Conference on Service Oriented Computing. Springer, Sydney, Australia, 2009:136-147.
- [13] Wu J, Chen L, Zheng Z, et al. Clustering Web services to facilitate service discovery[J]. Knowledge and information systems, 2014, 38(1):207-229.
- [14] Zhong Y, Fan Y, Huang K, et al. Time-aware service recommendation for Mashup creation in an evolving service ecosystem [C]//ICWS, 2014:25-32.
- [15] Tang M D, Xie F F, Cao B Q, et al. Exploring Web services from a network perspective using multi-level views[J].Journal of Web Engineering, 2016, 15(5/6):501-520.
- [16] 刘建勋,石敏,周栋,等.基于主题模型的 Mashup 标签推荐方法[J].计算机学报,2017,40(2):520-534.
- [17] Cherifi C, Labatut V, Santucci J F. Web services dependency networks analysis[C]//International Conference on New Media and Interactivity. Istanbul, TR, 2010.
- [18] Huang K, Fan Y, Tan W. An empirical study of programmable web: a network analysis on a service-mashup system[C]//IEEE 19th International Conference on Web Services (ICWS), 2012: 552-559.
- [19] Du V A. 9000 APIs: Mobile Gets Serious[EB/OL]. [2016-06-30]. <https://www.programmableweb.com/news/9000-apis-mobile-gets-serious/2013/04/30>.