

Constructing a Global Social Service Network for Better Quality of Web Service Discovery

Wuhui Chen, Incheon Paik, *Member, IEEE*, and Patrick C.K Hung, *Member, IEEE*

Abstract—Web services have had a tremendous impact on the Web for supporting a distributed service-based economy on a global scale. However, despite the outstanding progress, their uptake on a Web scale has been significantly less than initially anticipated. The isolation of services and the lack of social relationships among related services have been identified as reasons for the poor uptake. In this paper, we propose connecting the isolated service islands into a *global social service network* to enhance the services' sociability on a global scale. First, we propose *linked social service-specific principles* based on *linked data principles* for publishing services on the open Web as *linked social services*; then, we suggest a new framework for constructing the *global social service network* following *linked social service-specific principles* based on complex network theories. Next, an approach is proposed to enable the exploitation of the *global social service network*, providing Linked Social Services as a Service. Finally, experimental results show that our approach can solve the quality of service discovery problem, improving both the service discovering time and the success rate by exploring service-to-service based on the *global social service network*.

Index Terms—global social service network, Linked data principles, Linked social service, Link-as-you-go

1 INTRODUCTION

WEB services were expected to have a tremendous impact on the Web, as a potential solution for supporting a distributed service-based economy on a global scale. However, despite outstanding progress, uptake on the Web has been significantly less than initially anticipated [1], [2]. On the one hand, the number of services available on the Web is far less than expected. For example, one of the largest indexes of publicly available Web services currently provides 28,500 Web services with their corresponding documentation. While there are trillions of Web pages available on the Web, the number of publicly available Web services in one large composition service system is not greater than 4000, which is very small [3]. On the other hand, a large body of research about Web service discovery has been devoted to keyword- or semantic-based discovery to improve the quantity and quality of service-matching performance. Nevertheless, from a recent survey [4], most services published on the Web have not been used, and only a few of the services on the Web have been discovered, composed or invoked. These meager results handicap the service environment, resulting in a vicious circle of creation, publication, location, and composition of services in the computer industry. From investigation of several technological perspectives of Web services, the main reasons can be inferred as the following.

First, all the approaches based on current service descriptions, such as WSDL, Web APIs or Ontology Web Language for Services (OWL-S), only consider services as

isolated functional islands with no links to related services: unfortunately, this hampers service discovery and composition. Nowadays, services published in UDDI or on the Web based on current service description approaches know only about themselves, but not about the peers that they would like to work with in composition or that they would compete against for service selection [5]. These isolated service islands mean that service discovery is confronted with the following issues. One of them is that most approaches to service discovery lack consideration of interactions with the service consumers, so the usability threshold for service consumers is still high. Service consumers cannot discover services by following links that interest them, as they do when exploring Web pages. Another issue is that in most cases, service consumers are not limited to using a single service, but want to locate multiple services that can work together. However, guiding service consumers to discover services, starting from a service at hand and extending to peer services, which can be combined into more complex functions, is still a challenging issue because current services are isolated.

Second, services are considered only in terms of their own functional and nonfunctional properties; and the service's social activities, defined as engaging in significant social interaction with peer services via network models, are ignored. Web services are intended to be composed with related services in mind, and their functionality and nonfunctionality are interdependent, controlling services' social behavior [6]. Therefore, it is important to consider services' social activities for satisfying users, such as learning service's past social interactions for future use. For example, *PaymentService* from provider A was composed together with *ReservationService* from provider B successfully and more

- Wuhui Chen and Incheon Paik are with the School of Computer Science and Engineering University of Aizu, Fukushima, Japan. E-mail: chenwuhui21@gmail.com, paikic@u-aizu.ac.jp.
- Patrick C.K Hung is with the Business and Information Technology University of Ontario Institute of Technology (UOIT), Canada. E-mail: patrick.hung@uoit.ca.

Manuscript received (insert date of submission if desired). Please note that all acknowledgments should be placed at the end of the paper, before the bibliography.

frequently than *PaymentService* from provider C even though they have similar functionality and Quality of Service (QoS). Thus, when *ReservationService* from provider B is located, *PaymentService* from provider A has a higher probability of satisfying the user than *PaymentService* from provider C. To enhance service's sociability for improving the quality of service discovery, however, no approach provides a network model on a Web scale for supporting service's social activities. As well, other issues for the lack of success of Web service such as trustworthiness between service consumer and provider, service security, and business transactions are outside the scope of this work.

This paper proposes a methodology to drive an innovation from service islands to a *global social service network* to support service's social activities. In the *global social service network*, services are interlinked to related services from different sources functionally across the Web and in turn external services may link to them functionally using social link. To connect isolated service islands, *linked social service* is proposed to connect service islands into the *global social service network* with social link by following *linked social service-specific principles*. To develop the service's sociability for improving the quality of service discovery, the *global social service network* is constructed considering not only the service's own functional and nonfunctional detail but also the service's past social interactions and popularity, to provide a network model having properties that reflect the social reality of supporting the service's social activities. Our contributions may be summarized as follows:

1. Linked social service-specific principles based on linked data principles has been proposed for publishing isolated services as linked social service on the open web.
2. A new platform has been proposed for constructing global social service network to connect isolated service islands for supporting service's social activities.
3. An effective service discovery approach called *link-as-you-go* has been proposed to enable exploring service-to-service based on global social service network.

The remainder of this paper is structured as follows. In Section 2 discusses the background and related work. In Section 3, we propose *linked social service-specific principles* based on *linked data principles*. In Section 4, we present an algorithm to recommend social services using the quality of social link and propose a novel platform to construct a *global social service network* and analyze it using complex network theories. In Section 5, an approach is proposed to enable exploitation of the *global social service network*, providing Linked Social Service as a Service. Finally, evaluations of the *global social service network* and service discovery are provided in Section 6.

2 BACKGROUND AND RELATED WORK

We consider the two issues of functionality and QoS, and service's sociability; and, additionally, our basic motivation: *linked data* and its application to services.

2.1 Functionality and QoS

Service's functionality is a set of functional properties that

represent the description of the service tasks in terms of operation signatures, such as operation names and input/output schema; while QoS is a set of nonfunctional properties that encompass performance characteristics, such as execution cost, time, reliability, and security [7]. Web services, Semantic Web Service (SWS) and Web APIs are service models that consider only the functionality and QoS of service but not the service's social activities [5]. On the one hand, the stack of SWS and Web service technologies has brought a considerable level of complexity and yet suffers from the fact that most existing service discovery approaches are still in their infancy [8] [9] [10]. Current discovery techniques are registry-based, such as UDDI; however, UDDI was not adopted widely enough [11]. One of the main reasons for the lack of success of UDDI is the isolation of the services without social relationships among related services in these registries [12]. On the other hand, Web APIs are generally described using plain, unstructured HTML, except for a few that use the XML-based format WADL [13]; Web APIs are also isolated data source islands instead of comprising a single global data space because no hyperlinks can be set between data items in different Web APIs.

2.2 Service Sociability

Service's sociability is the skill, tendency or property of being sociable or social, and of interacting well with related services, which is supported by the network models we refer to here as service social networks [6]. The issue with service's sociability is how best to capture the way Web services interact via service social networks, and to know with whom they have worked in the past and with whom they might work in the future. A service social network is constructed to reflect services' social reality, describe the mutual consciousness of mutual agreement about a social situation and to support the services' future social activities. Therefore, by connecting distributed services into one single service social network, we can capitalize on users' willingness to interact, share, collaborate, and make recommendations for improving the quality of service discovery and service composition.

As a consequence, some approaches have been proposed to use service social networks to enable GPS-like support service discovery and service composition [5], [12], [14], [15]. However, all these approaches struggle with constructing a single service social network dynamically, as they do when constructing the World Wide Web. This is hampering services' sociability. These approaches do not consider interlinking Web services on the open Web into one single service social network to enhance services' sociability, as they interlink two persons when they are friends in a social network or link two actors who acted in the same movie in an actor collaboration graph. In this paper, we propose a methodology to interlink distributed services into a *global social service network* with social link, using the quality of social link considering the service's social activities and popularity to provide a global network model to support services' social activities.

2.3 Linked Data: Spawning New Wave of Services

Linked data technology provides the benefits of machine-understandable data from the Web and improved data discovery by using links among data items. The semantic annotation is done using a lightweight ontology so that economic and efficient annotation and deployment of *linked data* are enabled. The linked data [16] are managed using *linked data principles* [17] that suggest: 1) using URIs for things; 2) using HTTP URIs when people search for data; 3) using standards (RDF, SPARQL) for looking up a URI; and 4) including links to other URIs for rich discovery. Therefore, Pedrinaci et al. [18] propose *linked services* by following the *linked data principles*: services are published as *linked services* of data on the Web. However, services in the approach are still isolated, so *linked services* failed to provide a proper mechanism to interlink distributed services into a global service space for service discovery and composition.

3 REVOLUTION: FROM SERVICE ISLANDS TO LINKED SOCIAL SERVICES

To connect distributed service islands into a service social network, we first propose *linked social service-specific principles* that provide a basic recipe for publishing and interlinking services into service social networks. Then we suggest *linked social service*, which considers not only the functionality and QoS of service but also the service's sociability.

3.1 Linked Social Service-specific Principles

Before any significant uptake of services can take place on the Web, proper mechanisms for creating, publishing and discovering services must be in place. In this respect, our previous review of the state of the art discussed in Section 2 shows that:

- *Linked data* provides best practices for publishing and interlinking structured data on the Web [18].
- Service's sociability is extremely significant for providing a better quality of service discovery and composition [5], [6], [12].
- Connecting distributed service islands into a single service social network provides a network model for service's social activities [12].

This review inspires us to construct service social networks based on *linked data principles* to enhance service's sociability for providing a better quality of service discovery. Here we propose *linked social service-specific principles* that include four items to provide a basic recipe for publishing and connecting services into a service social network.

1. Services are published on the open Web following *linked data principles*.
2. Services are built upon the Web of data.
3. Services shall link to related services or be linked from related services functionally, using social link.
4. Interlink services must reflect service social reality.

The first service-specific principle advocates publishing services as *linked data* on the Web of data following *linked data principles*. Therefore, service

descriptions are comprised of structured data about reusable functionality on the Web. Their inputs and outputs, their functionality, their nonfunctional properties, their social information and knowledge are described in terms of lightweight RDF(S) vocabularies, so that applications can easily discover and process their descriptions just as retrieving Linked data.

The second principle advocates building services upon the Web of data because the Web of data can be used as background knowledge [19] to provide suitable ontologies that can be used, extended, and combined to create domain ontologies for annotating services.

The third principle advocates using social links to connect services functionally so they can be discovered and composed effectively by following these social links. For example, TicketInfoService, ReserveTicketService and PayTicketService can be linked together into a single service social network. Note that social link is a subclass of an RDF link in the Web of data, which indicates the service's social relationship between services.

Finally, to develop service's sociability for social activities, the fourth principle advocates considering the service's past social interactions and the service's popularity for interlinking services into a network model to reflect service social reality. If two services worked together frequently in the past then they are more likely to work together in the future.

By following *linked social service-specific principles*, services can be published in the service social network as *linked social services*. *Linked social services* are built upon the Web of data and link to related services and are linked by related services using social links into a service social network for social activities. Thus, *linked social service* is a set of functionalities published as *linked data* on the Web of data following *linked social service-specific principles*, interlinking to related services from different sources across the Web and in turn being linked to from external services functionally with social link.

4 GLOBAL SERVICE SPACE: GLOBAL SOCIAL SERVICE NETWORK

Linked social service is proposed to overcome the current service model by interlinking isolated service islands into a *global social service network* to develop the service's sociability. In this section, we will suggest a new framework for constructing a *global social service network* following *linked social service-specific principles*.

Definition 1 (Global Social Service Network). A *global social service network* is a service social network for service's social activities; its structure is a directed graph $G = \langle V, E \rangle$ on the Web, where:

- V represents a set of nodes, with each node being a *linked social service*; and
- E represents a set of directed edges, with each edge corresponding to social link.

A *global social service network* connects cross-domain distributed services together by social link. On the one hand, just as RDF links in the Web of data connect distributed data into a single global data space, *linked*

social service enables social links to be formed between services in different service sources, and therefore connects these services into the *global social service network* on the Web; this enhances the service's sociability and collaboration for service discovery and composition. On the other hand, the *global social service network* describes service societies' features such as social relations and social states, and provides a basis for inferring, planning, and coordinating social activities, thereby reflecting service social reality to capitalize on users' willingness to interact, share, collaborate, and make recommendations for improving the quality of service discovery.

4.1 Motivating Scenario

Suppose that there are three service providers from different domains who want to publish their services on the Web. The following information describes the services.

For service provider A from the medical domain:

- **W1:** Given symptom information, *DiagnosticService* returns disease information.
- **W2:** Given disease information, *getRelatedClinicInformationService* returns address type and clinic information, which is a complex data type that includes *clinicCity* and *clinicState*.
- **W3:** Given disease information, *getClinicInfoService* returns address type, *clinicCity* and *clinicState*.

For service provider B from the geography domain:

- **W4:** Given address type, *clinicCity* and *clinicState*, *getAddressInfoService* returns *zipCode* and *clinicAddress*.
- **W5:** Given *diseaseInfo* and clinic information, *getClinicAddressInfoService* returns *zipCode* and *clinicAddress*.
- **W6:** Given address type and clinic information, *getAddressInformationService* returns *zipCode* and *clinicAddress*.
- **W7:** Given two addresses, such as *clinicAddress* and *hotelAddress*, *findDirectionService* returns *mapImages* and *drivingDirection*.
- **W8:** Given *mapImages*, *getWeatherService* returns weather information.

For service provider C from the trip domain:

- **W9:** Given *zipCode*, *getTourService* returns *tourAddress*.
- **W10:** Given *zipCode*, *getHotelInformationService* returns *hotelAddress*.
- **W11:** Given *clinicAddress*, *getRestaurantService* returns *restaurantAddress*.
- **W12:** Given *hotelAddress*, *getHotelInfoService* returns *hotelInformation*.

4.2 Pattern of Social Link

To connect distributed services into a *global social service network*, social links are formed between isolated services. However, rather than simply connecting services, we use the pattern of social link to make typed statements that link arbitrary services. Here we define the pattern of social link that represents the functional relationships between resource service and target services according to service data correlations, which are data mappings

between the input/output attributes of services [20].

- T_n is a set of target services on the open Web, where n is the number of services.
- R is a resource service that refers to a known service.
- S_n, S'_n are sets of services that refer to T_n or R .
- e, e' are data elements. Each element is an addressable subcomponent of a data entity.
- D_n, D'_n are data entities that refer to input or output of a service with arbitrary data type. Each data entity is a set of data elements, such as $D = \{e_1, e_2, \dots\}$.
- f is a transformation that maps D_n to D'_n .
- $S_n \rightarrow D_n$ is an output relation: S_n produces D_n .
- $D_n \rightarrow D'_n$ is a data correlation: D'_n is a direct subset of D_n .
- $D'_n \rightarrow S'_n$ is an input relation: S'_n consumes D'_n .

Based on these concepts, we define the pattern of social link as follows.

Definition 2 (Pattern of Social Link). A pattern of social link is a triple (R_O, R_L, R_I) , where:

- $R_O = \{(S_n, D_n) \mid S_n \rightarrow D_n\}$.
- $R_L = \{(D_n, D'_n) \mid D_n \rightarrow D'_n\}$.
- $R_I = \{(D'_n, S'_n) \mid D'_n \rightarrow S'_n\}$.

To make typed statements to link peer services that can be worked together for service compositions, *Peer social link* is proposed to connect services that can be combined to provide a more complex service. *Peer social link* can be illustrated by the following rules, including sequential, parallel and conditional routing.

1) *Sequential Incoming-SocialLink* $L(\leftarrow)$: a Sequential Incoming-SocialLink $L(\leftarrow)$ is an $R \times T$ relation as shown in Fig. 1 (a), which can be denoted as follows:

$L(R \leftarrow T)$ if and only if (iff)

$$\forall e' \in R.D', \exists e \in T.D: f(e) = e'$$

Example 1 ($L(\leftarrow)$). Suppose there are two services W1 as T and W2 as R, as defined in the scenario of Section 4.1. All e from W2's D can be perfectly generated from W1's D' . So W1 and W2 can be linked using $L(W2 \leftarrow W1)$.

2) *Parallel Incoming-SocialLink* $L(< \oplus)$: a Parallel Incoming-SocialLink $L(< \oplus)$ is an $R \times T_n$ relation ($1 < i < n$) as shown in Fig. 1 (b). T_i, D_i are merged for $R.D'$, which can be denoted as follows ($1 < k < l < n$):

$L(R < \oplus T_n)$ iff

$$\forall e' \in R.D', \exists e \in \bigcup_1^n T_i.D_i: f(e) = e' \wedge T_k.D_k \cap T_l.D_l = \emptyset$$

3) *Choice Incoming-SocialLink* $L(< | |)$: a Choice Incoming-SocialLink $L(< | |)$ is an $R \times T_n$ relation ($1 < i < n$) as shown in Fig. 1 (c) that holds if:

$L(R < | | T_n)$ iff

$$\forall e' \in R.D', \exists e \in T_i.D_i: f(e) = e'$$

4) *Sequential Outgoing-SocialLink* $L(\rightarrow)$: a Sequential Outgoing-SocialLink $L(\rightarrow)$ is an $R \times T$ relation as shown in Fig. 1 (d) that holds if:

$L(R \rightarrow T)$ iff

$$\forall e' \in T.D', \exists e \in R.D: f(e) = e'$$

5) *Parallel Outgoing-SocialLink* $L(\oplus >)$: a Parallel Outgoing-SocialLink $L(\oplus >)$ is an $R \times T_n$ relation ($1 < i <$

n) as shown in Fig. 1 (e), which can be denoted as follows ($1 < k < l < n$):

$$L(R \oplus > T_n) \text{ iff}$$

$$\forall e \in R.D, \exists e' \in \bigcap_l T_l.D_l : f(e') = e \wedge T_k.D_k \cap T_l.D_l = \emptyset$$

6) *Choice Outgoing-SocialLink* $L(| >)$: a Choice Outgoing-SocialLink $L(| >)$ is an $R \times T_n$ relation ($1 < i < n$) as shown in Fig. 1 (f) that holds if:

$$L(R | > T_n) \text{ iff}$$

$$\forall e' \in T_i.D'_i, \exists e \in R.D : f(e) = e'$$

Further, to make typed statements that link services that perform a specific common function, *Cluster social link* is proposed to connect services offering similar functionalities. A *Cluster social link* $L(=)$: is an $R \times T$ relation that holds if:

$$L(R = T) \text{ iff}$$

$$\forall R.D, \exists T.D : R.D = T.D \wedge \forall R.D', \exists T.D' : R.D' = T.D'$$

4.3 Quality of Social Link

We consider four generic quality criteria for social link that can be denoted by $L(R, T_n)$: 1) Dependency Satisfaction Rate; 2) QoS Preference; 3) Sociability Preference; and 4) Preferential Service Connectivity.

4.3.1 Dependency Satisfaction Rate

To form social links for connecting distributed services, it is important to consider the functional relationships between R and T_n according to service data correlations, which are data mappings between the input/output attributes of services, to ensure quality of the social link.

Definition 3 (Dependency Satisfaction Rate). Given R and T , the Dependency Satisfaction Rate $Q_{DSR}(R, T)$ measures the degree of service data correlation between R and T for social link, and can be denoted as follows:

$$Q_{DSR}(R, T) = \frac{2 \times |T.D \cap R.D'|}{|T.D| + |R.D'|} \quad (1)$$

$Q_{DSR}(R, T)$ estimates the proportion of the common input/output attributes of R and T that are well specified, to ensure a correct data flow between R and T . A higher value indicates better quality of the social link. As defined in Section 4.2, D and D' are data entities that refer to input and output of a service and have arbitrary data type. Each data entity is a set of data elements. Suppose that $T.D = \{e_1, e_2, \dots\}$ and $R.D' = \{e'_1, e'_2, \dots\}$, then $Q_{DSR}(R, T)$ can be denoted in more detail as follows:

$$Q_{DSR}(R, T) = \frac{2 \times |\{e_1, e_2, \dots\} \cap \{e'_1, e'_2, \dots\}|}{|\{e_1, e_2, \dots\}| + |\{e'_1, e'_2, \dots\}|}$$

Example 2 ($Q_{DSR}(R, T)$). Suppose that $W2$ as R and target services $W5$ and $W6$ exist and are defined in the scenario of Section 4.1. According to equation (1), $Q_{DSR}(W2, W6)$ is calculated as:

$$\frac{2 \times |\{\text{addressType, clinicCity, clinicState}\} \cap \{\text{addressType, clinicCity, clinicState}\}|}{|\{\text{addressType, clinicCity, clinicState}\}| + |\{\text{addressType, clinicCity, clinicState}\}|} = 1.$$

And $Q_{DSR}(W2, W5)$ is computed as follows:

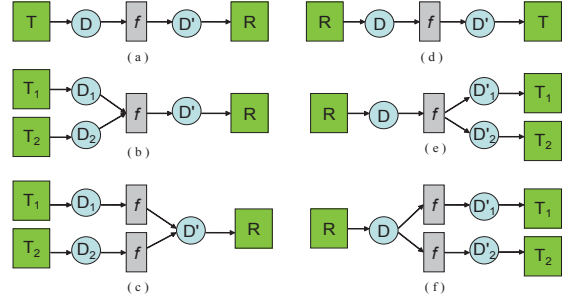


Fig. 1. Peer social link patterns.

$$\frac{2 \times |\{\text{addressType, clinicCity, clinicState}\} \cap \{\text{clinicCity, clinicState, diseaseInfo}\}|}{|\{\text{addressType, clinicCity, clinicState}\}| + |\{\text{clinicCity, clinicState, diseaseInfo}\}|} = 4/6.$$

Through the $Q_{DSR}(R, T)$, we can know that $L(W2, W6)$ has better quality than $L(W2, W5)$.

4.3.2 QoS Preference

When several target services have similar values of $Q_{DSR}(R, T)$ for a given R , their QoS properties such as price, availability, reliability and reputation become important to ensure the quality of the social link. Suppose that QoS_{max} and QoS_{min} are respectively the maximal and minimal QoS values of the target services; then the QoS Preference can be defined as follows.

Definition 4 (QoS Preference). Given R and a set of target services T_n , the QoS preference $Q_{QoS}(R, T)$ provides a measure for the degree of QoS satisfaction between R and T , and can be denoted as follows:

$$Q_{QoS}(R, T) = \frac{QoS_T - QoS_{min}}{QoS_{max} - QoS_{min}} \quad (2)$$

$Q_{QoS}(R, T)$ estimates the performance of T for discriminating among target services in terms of QoS to ensure the quality of the social link. Higher values indicate better quality of the social link.

Example 3 ($Q_{QoS}(R, T)$). Suppose there are three target services, $W4$ with QoS value 0.8, $W5$ with QoS value 0.6, and $W6$ with QoS value 0.9, and $W2$ as R , as defined in the scenario of Section 4.1. Using equation (1), we can calculate that both $Q_{DSR}(W2, W4)$ and $Q_{DSR}(W2, W6)$ are equal to 1. Using equation (2), $Q_{QoS}(W2, W4)$ and $Q_{QoS}(W2, W6)$ can be calculated as:

$$Q_{QoS}(W2, W6) = \frac{0.9 - 0.6}{0.9 - 0.6} = 1.$$

$Q_{QoS}(W2, W4)$ can similarly be calculated as:

$$Q_{QoS}(W2, W4) = \frac{0.8 - 0.6}{0.9 - 0.6} = \frac{2}{3}.$$

Thus, $Q_{QoS}(R, T)$, indicates that $L(W2, W6)$ is better quality than $L(W2, W4)$.

4.3.3 Sociability Preference

In most cases, Web services are not used alone, but are composed with peer services to accomplish a complex task. For examples, A scientific workflow precisely describes a multistep procedure to streamline a composition of tasks and the dataflow among them [21],

[22]; and an e-business workflow choreographs several business transactions simultaneously [23]. Interestingly, we found that among these workflows, if service X is contained in one workflow, then the probability of finding service Y in the same workflow is high. In other words, from learning service's past social interactions, we can find some usage patterns to know with whom the service has worked in the past and with whom it would prefer to work in the future [14], [15]. To illustrate the usage pattern, we use service association rules, which can be defined as:

$$X \Rightarrow Y \text{ where } X, Y \subset I \text{ and } X \cap Y = \phi.$$

Here, we treat services as items and workflows as transactions. Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of n binary services called *items*, and let $T = \{T_1, T_2, \dots, T_m\}$ be a set of transactions (service invocations) in a service invocation database. Each transaction in a database D has a unique transaction ID and contains a subset of the items in I .

Service association rules mean that if the workflow contains X , then the workflow has a good chance of containing Y . Here, we calculate support and confidence to find a dominating rule from the database as follows.

First, the support $\text{supp}(X)$ of an itemset X is defined as the proportion of transactions in the dataset that contain the itemset. For example, in Table 1, the itemset $\{S_1, S_4\}$ has a support of $3/5 = 0.6$, because it occurs in 60% of all transactions (3 out of 5).

Second, the confidence of a service association rule is defined as $\text{conf}(X \Rightarrow Y) = \text{supp}(X \cup Y) / \text{supp}(X)$. For example, the service association rule $\{S_1\} \Rightarrow \{S_4\}$ has a confidence of $0.6/0.8 = 0.75$ in the database in Table 1, which means that for 75% of the transactions containing S_1 , the service association rule is correct.

Table 1

Database with four items (services) and five transactions

Transaction ID	S_1	S_2	S_3	S_4
1	1	1	0	1
2	1	0	0	1
3	1	1	1	0
4	1	0	1	1
5	0	1	0	1

To know with whom service has worked in the past and with whom it would prefer to work in the future, we use the term Sociability Preference so that we can measure how often service T appears in transactions that contain R .

Definition 5 (Sociability Preference). Given R and T , the Sociability Preference $Q_{SP}(R, T)$ provides a measure for how often service T appears in transactions that contain R for the quality of the social link. The confidence of finding T when R is already known can be denoted as follows:

$$Q_{SP}(R, T) = \text{conf}(R \Rightarrow T) = \frac{\text{supp}(R \cup T)}{\text{supp}(R)} \quad (3)$$

$Q_{SP}(R, T)$ means that if the workflow contains R , then the workflow is likely to contain T . Higher values indicate

better quality of the social link.

4.3.4 Preferential Service Connectivity

A particularly important quantity in the *global social service network* is the shortest path between two services, d , defined as the smallest number of social links that must be followed to navigate from one service to the other for service discovery. To reduce the diameter of a *global social service network* and form a small-world network, services attach preferentially to services that are already well linked [24]. Most real networks exhibit preferential connectivity, meaning a new node links with higher probability to nodes that already have a large number of links. To incorporate preferential attachment, we use the term Preferential Service Connectivity.

Definition 6 (Preferential Service Connectivity). Given R and T , the Preferential Service Connectivity $Q_{PSC}(R, T)$ provides one measure for the probability that R will be connected to T . It depends on the connectivity K_i (K_i is the number of links in node i) of T and can be denoted as follows:

$$Q_{PSC}(R, T) = \frac{K_i}{\sum_j K_j} \quad (4)$$

$Q_{PSC}(R, T)$ means that the likelihood of linking to a target service depends on the service node's degree. Higher values indicate better quality of the social link.

Example 4 ($Q_{PSC}(R, T)$). Suppose there are three target services: W_4 with connectivity value 8, W_5 with connectivity value 6, and W_6 with connectivity value 10; and W_2 as R , where these services are defined in the scenario of Section 4.1. Using equation (4), $Q_{PSC}(W_2, W_4)$ can be calculated as:

$$Q_{PSC}(W_2, W_4) = \frac{8}{24} = \frac{1}{3}$$

Similarly, $Q_{PSC}(W_2, W_6)$ can be calculated as:

$$Q_{PSC}(W_2, W_6) = \frac{10}{24} = 5/12$$

Thus, $Q_{PSC}(R, T)$, indicates that $L(W_2, W_6)$ is better quality than $L(W_2, W_4)$.

4.3.5 Combining the Qualities

Given the above quality criteria, here we combine them to evaluate the quality of the social link using quality aggregation rules. The rules for aggregating quality values for any social link are provided in Table 2. Using the aggregation rules, the quality vector of social link can be defined by the quality of social link $Q(R, T_n)$.

Definition 7 (Quality of Social Link). Given R and a set of target services T_n , the quality of social link $Q(R, T_n)$ provides one measure for the quality of the links between R and T_n for social link, and can be denoted as follows:

$$Q(R, T_n) = \langle Q_{DSR}(R, T_n), Q_{QoS}(R, T_n), Q_{SP}(R, T_n), Q_{PSC}(R, T_n) \rangle \quad (5)$$

Equation (5) means that the quality of social link between R and T_n depends on the four quality criteria. Higher

values indicate better quality of social link. First, selecting links with the best functional quality $Q_{DSR}(R, T_n)$ will ensure easy end-to-end integration between services by minimizing semantic and syntactic mediators, and by providing seamless deployment and execution of computed compositions. The choice of such criteria should be dominant to reduce the cost of data mediation. Second, selecting links including services with the best nonfunctional quality values will ensure quality of social link in aspects understandable by most users, such as price and response time. Such criteria should be dominant when data shared by services are not so heterogeneous (e.g., services aligning their data at description time where most of their exchanged data match perfectly), or when all data mediators are known at design time [25]. Third, selecting links including services with the best $Q_{SP}(R, T)$ will ensure the quality of social link by learning with whom the services have worked in the past and with whom they would prefer to work in the future. The choice of such criteria should be dominant to reflect the service social reality: services that interacted frequently in the past have a high probability of being linked with social link. Finally, selecting links including services with the best $Q_{PSC}(R, T)$ will ensure the quality of social link by linking to well-known, popular services with high connectivity so that they have a high probability of being recruited by the well-known services.

4.4 Construction of Global Social Service Network

By *linked social service-specific principles*, services can be published as *linked social services* and added to a *global social service network* with social link, which allows services to be discovered and composed effectively.

Definition 8 (Global Social Service Network Construction). *Given a set of services $\{S_n\}$, Global Social Service Network Construction is a process of publishing services $\{S_n\}$ as linked social services in a global social service network $G = \langle V, E \rangle$ by following linked social service-specific principles, such that:*

$$\forall S_i \in \{S\}. \exists S_j \in G.V : L(S_i, S_j)$$

The definition means that new services are added to a *global social service network* with social link by following *linked social service-specific principles*. To construct a *global social service network*, we first introduce a social link recommendation algorithm to recommend social services using quality of social link; then, an algorithm for constructing a *global social service network* will be

introduced; finally, we will analyze the social links in the *global social service network* using complex network theories.

4.4.1 Recommending Social Services Using Quality of Social Link

To publish services as *linked social services*, it is important to find social services to link together functionally. To ensure the quality of social relationships between R and T, we developed the social link recommendation algorithm to recommend social services using quality of social link as defined above. First, for each pattern of social link we recommend social services as social link candidates based on $Q_{DSR}(R, T_n)$ to ensure functionality. Second, we rank all the social link candidates with different patterns of social link using the quality of social link to ensure the quality of social services. Given that the Incoming-SocialLink model recommendation is similar to the outgoing-SocialLink model, here we only show the Incoming-SocialLink model recommendation as Algorithm 1.

The algorithm adopts two steps. In step 1, it selects a set of social service candidates for R according to the Q_{DSR} , which has a threshold λ to ensure the functionality quality of social services. First, the *Cluster social link* candidates are recommended according to Q_{DSR} (lines 2-5); then service sets that have strong functionality relationships with the R are selected (lines 6-13); finally, based on the service sets, it classifies the social service candidates according to their patterns of social link including $L(<| |)$, $L(\leftarrow)$, and $L(<\oplus)$ (lines 15-30). In step 2, it ranks and selects m social services with higher qualities of social link to ensure the quality of social service (lines 32-35).

The time complexity of the algorithm is determined by two time components: (1) the amount of time needed to find the maximal $Q(R, T_n)$ among the nodes in V ; (2) the amount of time for the calculation of $Q(R, T_n)$ of each node. The two components can be represented by the number of iterations of the until loop times and the complexity of those steps in each iteration, respectively. The number of iterations is decided by the complexity of global social service network, which can be denoted as $X = |V|$. In each iteration, the amount of time needed for the calculation of the value of $Q(R, T_n)$ which including four criterions is a constant, denoted by A. For X iteration, the complexity of the until loop of the algorithm is $A \times X$, so the time complexity of the algorithm is: $A \times |V|$.

Table 2
Quality aggregation rules for social link

Pattern of social Link	Quality Criterion			
	$Q_{DSR}(R, T_n)$	$Q_{QoS}(R, T_n)$	$Q_{SP}(R, T_n)$	$Q_{PSC}(R, T_n)$
$L(R \leftarrow T)$ or $L(R \rightarrow T)$	$Q_{DSR}(R, T)$	$Q_{QoS}(R, T)$	$Q_{SP}(R, T)$	$Q_{PSC}(R, T)$
$L(R < \oplus T_n)$ or $L(R \oplus > T_n)$	$\sum_1^n Q_{DSR}(R, T_i)$	$\sum_1^n Q_{QoS}(R, T_i)$	$\frac{1}{n} \times \sum_1^n Q_{SP}(R, T_i)$	$\frac{1}{n} \times \sum_1^n Q_{PSC}(R, T_i)$
$L(R < T_n)$ or $L(R > T_n)$	$\text{Min}\{Q_{DSR}(R, T_i)\}$	$\text{Max}\{Q_{QoS}(R, T_i)\}$	$\frac{1}{n} \times \sum_1^n Q_{SP}(R, T_i)$	$\frac{1}{n} \times \sum_1^n Q_{PSC}(R, T_i)$
$L(R = T)$	$Q_{DSR}(R, T)$	$Q_{QoS}(R, T)$	$Q_{SP}(R, T)$	$Q_{PSC}(R, T)$

Algorithm 1: Recommend Social Services**Input:** $G \langle V, E \rangle$, threshold λ , resource service S , $Q(R, T_n)$, m **Output:** A set of m social services $\{S_m\}$ **Variables:** service has SocialLink $S.L$, input of service $S.I$, output of service $S.O$, service set $\{S_1\}$, $\{S_2\}$, $\{S_{ss}\}$

```

2. For each  $S_i \in G.V$  do
2.   if  $(Q_{DSR}(S.I, S_i.I) > \lambda \ \&\& \ Q_{DSR}(S.O, S_i.O) > \lambda)$  then
3.      $S_i.L := L(=)$ ;
4.      $\{S_{ss}\} := \{S_{ss}\} \cup S_i$ ;
5.   end
6.   else if  $(Q_{DSR}(S.I, S_i.O) > \lambda \ || \ S_i.O \supset S.I)$  then
7.      $S_i.L := L(\leftarrow)$ ;
8.      $\{S_1\} = \{S_1\} \cup S_i$ ;
9.   end else
10.  else if  $(S_i.O \subset S.I)$  then
11.     $S_i.L := L(< \oplus)$ ;
12.     $\{S_2\} = \{S_2\} \cup S_i$ ;
13.  end else
14. End
15. For each  $S_i \in G.V$  do
16.  if  $(\exists \max\{S_n\} \subset \{S_1\}, \forall S_k \in \{S_n\}: Q_{DSR}(S_k, S_j) > \lambda)$  then
17.     $S_n.L = L(< |)$ ;
18.     $\{S_1\} = \{S_1\} - \{S_n\}$ ;
19.     $\{S_{ss}\} = \{S_{ss}\} \cup \{S_n\}$ ;
20.  end
21. End
22.  $\{S_{ss}\} = \{S_{ss}\} \cup \{S_1\}$ ;
23. while  $\{S_2\} \neq \Phi$  do
24.  if  $(\exists \min\{S_n\} \subset \{S_2\}, Q_{DSR}(S, S_n) > \lambda)$  then
25.     $S_n.L = L(< \oplus)$ ;
26.     $\{S_2\} = \{S_2\} - \{S_n\}$ ;
27.     $\{S_{ss}\} = \{S_{ss}\} \cup \{S_n\}$ ;
28.  end
29.  else break;
30. End else
31. end
32. if  $(\{S_{ss}\} \neq \Phi)$  then
33.   $\{S_m\} = \text{rankingSocialServiceQuality}(\{S_{ss}\}, m)$ 
34.  return  $\{S_m\}$ .
35. end

```

a *global social service network* are updated to reflect service social reality according to service's past social interactions using the quality of social link. Some "old" links may be rewritten by some "new" social links with higher quality of social link.

To incorporate the growing character of the network, starting with a small number (m_0) of vertices, at every time step we add a new vertex with $m \leq m_0$ edges that link the new vertex to m different vertices already present in the network. To incorporate preferential attachment, we assume that the probability Π_i that a new vertex will be connected to vertex i depends on the connectivity k_i of that vertex, such that $\Pi_i = k_i / \sum_j k_j$. To incorporate the competitive aspect, we assign a fitness parameter η_i so that at every time step a new service W is added to the *global social service network* with a fitness η_i that depends on Q_{DSR} , Q_{QoS} and Q_{SP} . To incorporate the adaptation aspect to reflect service's social reality, we rewrite "old" social links having a low quality of social link and add a new social link with a high quality of social link throughout the lifetime of the network. Based on the previous analysis, we also quantify the quality of social link as:

$$Q(R, T_n) = \frac{\eta_i k_i}{\sum_j \eta_j k_j}, \quad (6)$$

where k_i is the degree of node i , and η_i is a fitness parameter, as each node has an intrinsic ability to compete for edges at the expense of other nodes. η_i can be calculated as:

$$\eta_i = (W_{DSR} Q^2_{DSR}(R, T_n) + W_{QoS} Q^2_{QoS}(R, T_n) + W_{SP} Q^2_{SP}(R, T_n)).$$

Here, $W_{DSR} + W_{QoS} + W_{SP} = 1$. Higher values indicate that node i is better at competition.

Considering the issues discussed above, we use an algorithm to build our network that depends on four parameters: m_0 (the initial number of nodes); m (the number of links added or rewired at every step of the algorithm); p (the probability of adding links); and q (the probability of edge rewiring). The procedure starts with m_0 nodes already present in the network and performs one of the following three actions at every step.

1. With probability p , $m \leq m_0$, new links are added. The starting point of the link is chosen uniformly, and the end point of the new link is chosen using algorithm 1 according to the quality of social link of equation (6). The process is repeated m times.
2. With probability q , m edges are rewired. For this purpose, node i is chosen at random, and its link L_{ij} with lowest quality of social link is chosen. The link L_{ij} is deleted and another node z is selected using algorithm 1 according to the quality of social link of equation (6), and the new link L_{iz} is added.
3. With probability $1 - p - q$, a new node with m links is added. These new links connect the new node to m other nodes chosen using algorithm 1 according to the quality of social link of equation (6).

Once the desired N nodes are obtained, the algorithm

4.4.2 Constructing a Global Social Service Network

To construct a *global social service network* for better service discovery, we must consider four generic aspects of the *global social service network*. First is a growth aspect. A *global social service network* is open, formed by the continual addition of new services; thus, the number of vertices, N , increases throughout the lifetime of the network, just as the www grows exponentially in time by the addition of new Web pages and the research literature constantly grows because of the publication of new papers. Second is a Preferential Service Connectivity aspect. In a *global social service network*, the probability that two vertices are connected is not random and uniform, but exhibits Preferential Service Connectivity: there is a higher probability of linking to a vertex that already has a large number of connections. Third is a competitive aspect, as each node has an intrinsic ability to compete for edges at the expense of other nodes. For example, a higher $Q_{DSR}(R, T)$ between R and T makes the T more competitive. Fourth is an adaptation aspect. Social links in

stops. Actions 1 and 2 are designed to satisfy the adaptation aspect of the network, and action 3 can successfully achieve the growth aspect. The quality of social link with fitness η_i satisfies the competitive and preferential attachment aspect well.

4.4.3 Social Link of Global Social Service Network

According to equation (6), the probability Π_i that a new node will connect to a node i already present in a *global social service network* depends on the connectivity k_i and on the fitness η_i of that node, such that:

$$\prod_i = \mathcal{Q}(R, T_n) = \frac{n_i k_i}{\sum_j k_j n_j}. \quad (7)$$

This generalized preferential attachment [26], [27] ensures in the simplest possible way that fitness and connectivity jointly determine the rate at which new links are added to a given node. To address the scaling properties of this model we use a continuum theory [28], allowing us to predict the connectivity distribution, which is the probability that a node has k social links. A node i will increase its connectivity k_i at a rate that is proportional to the probability that a new node will attach to it, giving:

$$\frac{\partial k_i}{\partial t} = m \frac{n_i k_i}{\sum_j k_j n_j}. \quad (8)$$

The factor m appears because each new node adds m social links to the network. To solve (8), we assume that similarly to the scale-free model [29], the time evolution of k_i follows a power law, but there is multiscaling in the network, i.e., the dynamic exponent depends on the fitness η_i :

$$k_{n_i}(t, t_0) = \left(\frac{t}{t_0} \right)^{\beta(n_i)}, \quad (9)$$

where t_0 is the time at which node i was born. The dynamic exponent $\beta(n)$ is bounded, i.e., $0 < \beta(n) < 1$, because a node always increases the number of social links over time ($\beta(n) > 0$), and $k_i(t)$ cannot increase faster than t ($\beta(n) < 1$). We first calculate the mean of the sum $\sum_j (\eta_j k_j)$ over all possible realizations of the quenched noise $\{\eta\}$. Because each node is born at a different time, t_0 , the sum over j can be written as an integral over t_0 :

$$\begin{aligned} \left\langle \sum_j n_j k_j \right\rangle &= \int dn \rho(n) n \int_1^t dt_0 k_n(t, t_0) \\ &= \int dn n \rho(n) m \frac{(t - t^{\beta(n)})}{1 - \beta(n)}. \end{aligned} \quad (10)$$

Because $\beta(n) < 1$, in the $t \rightarrow \infty$ limit, $t^{\beta(n)}$ can be neglected compared with t , and thus we obtain:

$$\left\langle \sum_j n_j k_j \right\rangle \stackrel{t \rightarrow \infty}{=} C m t (1 + O(t^{-\varepsilon})), \quad (11)$$

where

$$\begin{aligned} \varepsilon &= (1 - \max_n \beta(n)) > 0, \\ C &= \int dn \rho(n) \frac{n}{1 - \beta(n)}. \end{aligned} \quad (12)$$

Using equation (11), and the notation $k_\eta = k_{n_\eta}(t, t_0)$, the

dynamic equation (8) can be written as:

$$\frac{\partial k_n}{\partial t} = \frac{n k_n}{C t}, \quad (13)$$

which has a solution of form (9), given that

$$\beta(n) = \frac{n}{C}, \quad (14)$$

thereby confirming the self-consistent nature of assumption (9). To complete the calculation, we must determine C from equation (12) after substituting n/C for $\beta(n)$:

$$1 = \int_0^{n_{\max}} dn \rho(n) \frac{1}{\frac{C}{n} - 1}, \quad (15)$$

where n_{\max} is the maximum possible fitness in the network. Apparently, (15) is a singular integral. However, because $\beta(n) = n/C < 1$ for every value of n , we have $C > n_{\max}$. Thus, the integration limit never reaches the singularity. Note also that, because $\sum_j \eta_j k_j \leq (n_{\max}) \sum_j k_j = 2mt(n_{\max})$, we have, using equation (11), that $C \leq 2(n_{\max})$.

Finally, we can calculate the connectivity distribution $P(k)$, which gives the probability that a node has k social links. If there is a single dynamic exponent, β , the connectivity distribution follows the power law $P(k) \sim k^{-r}$, where the connectivity exponent is given by $r = 1/\beta + 1$. However, here we have a spectrum of dynamic exponents $\beta(\eta)$; thus, $P(k)$ is given by a weighted sum over different power laws. To find $P(k)$ we must calculate the cumulative probability that for a certain node $k_\eta(t) > k$:

$$P(k_n(t) > k) = P\left(t_0 < t \left(\frac{m}{k}\right)^{C/n}\right) = t \left(\frac{m}{k}\right)^{\frac{C}{n}}. \quad (16)$$

Thus, the connectivity distribution is given by the integral:

$$P(k) = \int_0^{n_{\max}} dn \frac{\partial P(k_n(t) > k)}{\partial t} \propto \int dn \rho(n) \frac{C}{n} \left(\frac{m}{k}\right)^{\frac{C}{n}+1}. \quad (17)$$

5 LINKED SOCIAL SERVICE AS A SERVICE

In this section, we propose a novel approach for discovering services on the *global social service network*, providing Linked Social Services as a Service (LSSaaS). Our approach, called *link-as-you-go*, allows users to start browsing with one service and then navigate along links to related services by following social links in the *global social service network*. This is similar to exploring the Web of data by following RDF links, or exploring for friends in a social network, and illustrates the service's social activities, thereby improving the usability threshold for service consumers.

5.1 Link-as-you-go: Going Deeper into the Global Social Service Network Following Social Links

Consider a scenario in which a man feels sick. He would first want to find out what illness he is suffering from, and then find which clinic can best deal with the illness. He must also find information about hotels and

restaurants near the clinic because his health care may take a couple of days, and finally, he must know how to travel to the hotels and restaurants from the clinic.

No single service can satisfy all aspects of this scenario. To fulfill this request, multiple services must be discovered and combined. The set of services required to fulfill his request are as follows. First, he needs *DiagnosticService* (W1) to obtain a possible diagnosis based on input of his symptoms. Next, he requires *getRelatedClinicInformationService* (W2) to obtain information on which clinic can deal with his health problem, and *getAddressInformationService* (W6) to obtain more detailed address information. Next, he requires *getHotelInformationService* (W10) and *getRestaurantService* (W11) to obtain information about hotels and restaurants. Finally, having obtained the required addresses, *findDirectionService* (W7) is required to obtain a map.

Traditional service discovery approaches, such as keyword-similarity-based discovery or input/output ontology concept-based discovery, lack interaction with service consumers. Service consumers cannot participate in the discovery process, which is a "black box." Moreover, services are currently published as isolated service islands so that they can only know about themselves rather than about their peers. This makes it difficult for current service discovery approaches to guide service consumers to discover services from the service at hand to peer services that can be composed. To discover services that can effectively fulfill a complex request, we first define the service discovery problem in the *global social service network* using social links.

Definition 9 (Web Service Discovery Problem). Suppose that a request λ has initial input parameters $\lambda.I$ and desired output parameters $\lambda.O$. The Web service discovery problem in global social service network is defined as finding a subnetwork which contains a set of services $\{w\}$, such that:

- 1) $\forall w_i \in \{w\} : w_i.I \subseteq \lambda.I$;
- 2) $\forall i \in \lambda.O, \exists w_i \in \{w\} : i \in w_i.O$;
- 3) $\forall w_i \in \{w\}, \exists w_j \in \{w\} : L(w_i, w_j)$.

Just as we use *linked data* to reuse data, discover data from relevant datasets and integrate data from large numbers of formerly unknown data sources, social link can be used for service discovery. Our novel approach, *Link-as-you-go*, allows users to start browsing in one service and then navigate along social links into related services to explore service-to-service, so that users can explore the *global social service network* more deeply on the open Web. *Link-as-you-go* opens the black box for service discovery so that a service consumer can discover services by following links of interest in the same manner as exploring Web pages or finding new friends through already known friends in Facebook. The social links of the *linked social service* can be denoted as follows:

$$SL = \{L_1, L_2, \dots, L_n\}, 1 \leq n \leq N,$$

where SL denotes a set of social links, and L_1, L_2, \dots, L_n indicate that *linked social services* have N social links,

where

$$N = \int_0^{\infty} P(k)dk \propto \int_0^{\infty} dk \int_0^{\infty} dn \rho(n) \frac{C}{n} \left(\frac{m}{k}\right)^{\frac{C}{n}+1}. \quad (18)$$

Thus, to explore a *global social service network* by following these links T times, the number of services potentially explored (SE) can be denoted as follows:

$$SE = N^T. \quad (19)$$

Therefore, by using the *link-as-you-go* approach, services can be discovered effectively by following the social links. The following is a solution for the previous scenario: assume now that one is in the position of W6, which is *GetAddressInformationService* as shown in Fig. 2 (a). Here, W6 links to peer services that can be combined together, such as W2, W10 and W11. One can explore from W6 to W10 by following links, as shown in Fig. 2 (b), after which one can explore services from W10 to more services, such as W7 and W11. In this manner, one can easily explore the *global social service network* more deeply to discover the services needed.

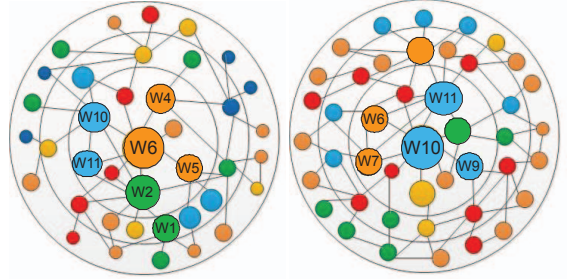


Fig. 2. (a): Linking from W6.

(b): Linking from W10.

6 IMPLEMENTATION AND EVALUATION

We analyzed the performance of our approach by:

- Observing the properties of the *global social service network* in Section 6.2.1
- Evaluating the effectiveness of LSSaaS in terms of service discovering time in Section 6.2.2
- Evaluating the quality of LSSaaS in terms of success rate in Section 6.2.3
- Observing the evolution of the *global social service network* over $Q_{sp}(R, T)$ and $Q_{psc}(R, T)$, by varying the number of transaction items and the parameter m , which is the number of links added and/or rewired at every step of the construction procedure in Section 6.2.4

6.1 Context of Experimentation

6.1.1 Implementation Details

Service-Of-A-Service (SOAS) is a platform to construct *global social service network*; it also enables exploitation of the *global social service network*, providing LSSaaS. SOAS currently has about 2000 services registered from the OWL-S test collection and the SAWSDL test collection, about 3000 services from the owls-mx_2_0 test collection [9], around 5000 services indexed by Seekda.com, and about 500 services from ProgrammableWeb.com. SOAS has five main modules: firstly, the *Crawler* module crawls

distributed services from the Web of documents and receives the service publication by service providers, who simply follow *linked social service-specific principles*. Secondly, the *Translator* module receives isolated services based on service descriptions such as those of OWL-S, WSMO or WSDL, and translates them into RDF descriptions based on the SOAS model described in our previous work [30], which provides a minimal and common conceptual model and uses social links to connect service islands. DBpedia [31], a common vocabulary on the Web of data, provides lightweight ontologies for service semantic annotation. Thirdly, the *Recommender* module recommends social services for isolated services using the recommendation algorithm 1 described in Section 4.4.1 based on the quality of social link. Fourthly, the *Global Social Service Network* module adds services into the *global social service network* with social links using the construction procedure described in Section 4.4.2. Sesame [32] is used as an RDF repository for storage, inferencing and querying of services. Finally, To enable exploitation of the *global social service network*, the *link-as-you-go* module offers service consumers an *link-as-you-go* approach to explore service-to-service to go deeper into the *global social service network*.

6.1.2 Experimental Setup

For state-of-the-art approaches, we use two strategies, one of which is based only on syntactic matching between services without considering any ontology reasoning. The other uses semantic matching between services. In this experiment, we compared our approach with that of Seekda.com and ProgrammableWeb.com for strategy one, and of Iserve [18] and OWLS-MX [9] for strategy two. Throughout the experiments, we used the following two evaluation metrics.

1. T (service discovering time): measures how much time a service consumer takes to find the services needed, in seconds. This is a measure of the effectiveness of the service discovery approach.
2. Success rate: defined as the probability that a user satisfies the first five services recommended by the proposed framework. It measures the quality of service discovery.

Smaller values of T indicate better solutions. The success rate, ranging from 0.0 to 1.0, indicates the adaptability of the proposed framework because it approaches 1.0 as the proposed framework adapts to the user's preferences.

To evaluate T and the success rate, 50 participants, of whom 15 were general users, 20 were IT professionals, and 15 were Web service professionals, tried to discover services including 12 specific services defined in the scenario of Section 4.1 to accomplish the task described in the scenario of Section 5.1, and recorded the time they took and the satisfaction values for the recommended services. In addition, 50 participants have joined in our experiment to explore services based on global social service network for three months to prepare transaction items. Finally, we got 60,000 transaction items from their usage histories as experiment datasets for $Q_{SP}(R, T)$. We set the value of the other variables in the construction

procedure as follows: about 10,000 services ($N = 10,000$) were added to the *global social service network*, $m_0 = m = 4$, $p = q = 0.3$, and $W_{DSR} = 0.4$, $W_{QoS} = 0.2$, $W_{SP} = 0.4$. Our experiments were conducted using an Intel(R) Core2 CPU, 2.4 GHz, 4 GB RAM, and Java 2 Enterprise Edition V1.7.0.

6.2 Evaluation

6.2.1 Properties of the Global Social Service Network

We evaluated the properties of the *global social service network* by:

- observing the shortest path between two services in the *global social service network*, which is defined as the smallest number of social links that must be followed to navigate from one service to the other
- observing the social link distribution on the *global social service network* to show the number of social links for each *linked social service*.

Fig. 3 is another indicator of the density of the *global social service network*. We compute the components by starting with an arbitrary service. The figure shows the distance of any service in the main cluster from this starting object. Almost all of the objects have a distance between 5 and 9 from the starting object and are thus within a short distance from the starting object, indicating that the *global social service network* forms a small-world network. Fig. 3 must be interpreted cautiously, because it depends on the (randomly selected) starting object.

In Fig. 4, the X-axis represents the number of social links and the Y-axis represents the connectivity distribution $P(k)$. Fig. 4 shows that the tail of the distribution follows $P(k) \propto k^{-r}$ with $r = 1.975$ (where r is a parameter whose value is typically in the range $2 < r < 3$). Even though r does not exceed 2, the probabilities are highly skewed so that their shape is similar to the Zipf distribution.

6.2.2 Effectiveness of LSSaaS

To show that our *link-as-you-go* approach can effectively improve service discovery, we evaluated the effectiveness of LSSaaS by observing the total service discovering time

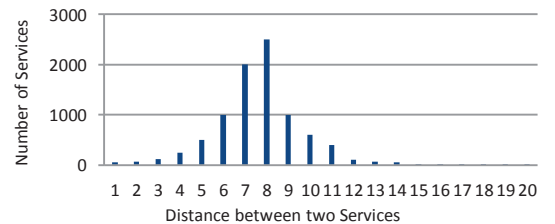


Fig. 3. Numbers of services at given distances from an origin service.

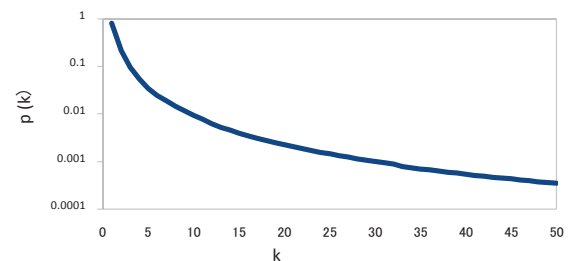


Fig. 4. Probabilities of occurrence of number of social links.

while changing the number of desired services and the participant group, including the general user group, IT professional group, and Web services professional group. We also compared our approach with state-of-the-art approaches in terms of service discovering time.

The first experiment evaluated the effectiveness of our approach by changing the number of desired services from 1 to 12. In most cases, service consumers are not limited to discovering a single service, but want to locate multiple services. As the number of desired services increases, the service discovering time is expected to increase. Fig. 5 shows that this was the case, but our approach performs far better than other approaches, especially for discovering large numbers of desired services. This occurs because once the participants find one service, then other services can be discovered by following the social links.

In the second experiment, we evaluated the effectiveness of our approach by observing the total service discovering time for 10 desired services over the participant groups. In most cases, to discover services, users are expected to specify their requests with associated service and semantic information, which is not realistic if they do not have the appropriate knowledge. Table 3 shows that general users could discover services they wanted more easily using our approach than with other approaches. The results indicate that the current use threshold of the existing service discovery approaches is still high and our approach can improve the usage threshold effectively by linking one service to others, just as exploring friends in Facebook.

Table 3
Service discovering time by different groups.

Approaches	Participant group		
	General users	IT professional	Web service professional
Seekda.com	1100 s	721 s	650 s
ProgrammableWeb	1950 s	1205 s	980 s
Iserve	1020 s	605 s	550 s
OWLS-MX	950 s	520 s	495 s
SOAS	310 s	295 s	205 s

6.2.3 Quality of LSSaaS

To show that our *link-as-you-go* approach can effectively guarantee the quality of service discovery, we evaluated the quality of LSSaaS by observing the success rate while changing the number of users and the number of services, and compared our approach with state-of-the-art approaches in terms of success rate.

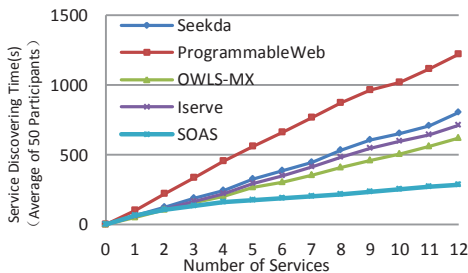


Fig. 5. T based on different approaches.

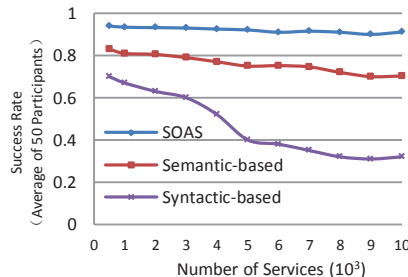


Fig. 6. Success rate versus number of services.

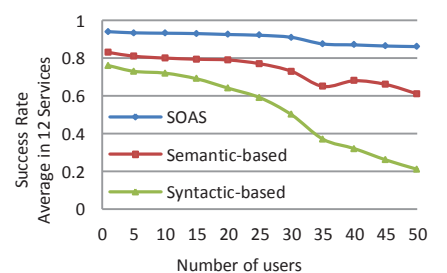


Fig. 7. Success rate versus number of users.

The first experiment evaluated the quality of service discovery by observing the success rate when the number of services changed from 1000 to 10,000. As the number of services increases, the number of service candidates for each query will increase, and we can expect that the success rate will decrease. Fig. 6 shows that as the number of services increases, the success rate decreases. However, in comparing with other approaches, our approach maintained a higher success rate, even when the number of services reached 10,000, which indicates that our approach is more scalable and more suitable for large service sets. This is because global social service network successfully reflects the service's social reality by learning the service's past social interactions and by adding new links and rewiring links based on the quality of social link.

The second experiment evaluated the quality of service discovery by observing the success rate while changing the number of users from 1 to 50. Note that in this experiment, different users had different user preferences. Thus, as the number of users increases, it is expected that the success rate will decrease. Fig. 7 shows the success rate according to the number of users. The results show that as the number of users increases, the success rate decreases. However, compared with other approaches, our approach maintains a higher success rate as the number of users increases. This is because service's past social interactions are collected to learn user preferences so that our approach can adapt well to user's preferences.

6.2.4 Evolution of the Global Social Service Network

We evaluated two factors that change the shape of the *global social service network*, the parameter m and services' past social interactions. We also evaluated how the network's shape affects the service discovery in terms of service discovering time and success rate. To measure how much the social links in the *global social service network* are changed by service's past social interactions, we used the evaluation metric $k(t)$, which can be calculated as:

$$k(t) = \frac{N}{\sum_j k_j}$$

Here, N is the total number of new social links added and/or rewired in the *global social service network*; $\sum_j k_j$ is the total number of social links in the *global social service network*; and $k(t)$ is a measure of the impact of service's past social interactions on the *global social service network*.

In the first experiment, we evaluated the impact of $Q_{sp}(R, T)$ on the *global social service network* by varying the number of transaction items from 10,000 to 60,000. Note

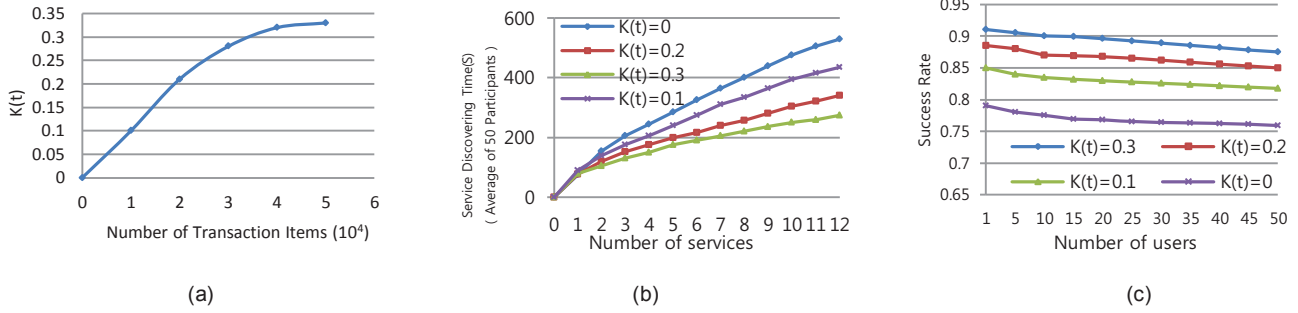


Fig. 8. Effect of number of transactions on the global social service network.

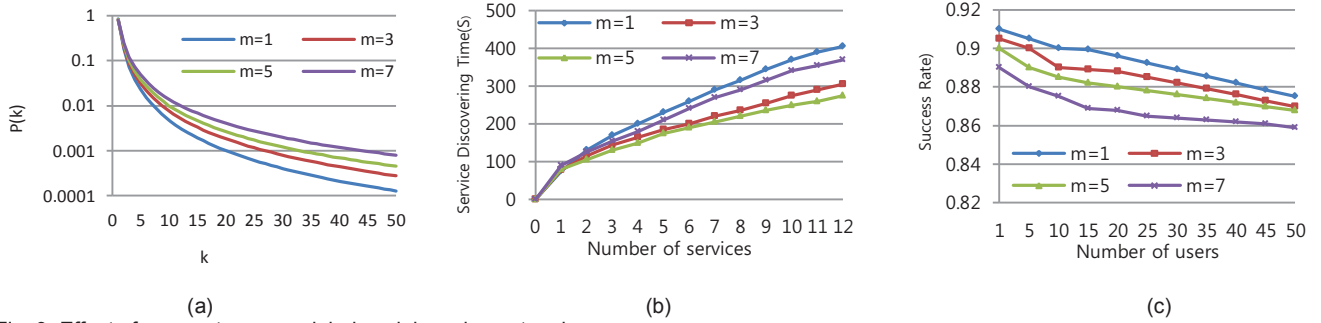


Fig. 9. Effect of parameter m on global social service network.

that we collected transaction items from usage histories. Fig. 8(a) shows $k(t)$ according to the number of transaction items. The results show that as the number of transaction items increases, $k(t)$ increases; when the number of transaction items reaches 50,000, $k(t)$ increases more slowly.

Second, to show the impact of $k(t)$ on *link-as-you-go* in terms of service discovering time, we varied the value of $k(t)$ from 0.1 to 0.3 by changing the number of transaction items and observed the service discovering time according to the number of desired services. Fig. 8(b) shows that the service discovering time decreases as $k(t)$ increases; higher values indicate better effectiveness. This occurs because users' past usage histories are learned to reflect the service's social reality and to satisfy the users' requirements.

Third, to show the impact of $k(t)$ on *link-as-you-go* in terms of success rate, we varied the value of $k(t)$ from 0.1 to 0.3 while changing the number of transaction items and observing the success rate according to the number of users. Fig. 8(c) shows that the success rate was higher with higher $k(t)$ values. This occurs because some successful service invocations and compositions are learned from the services' past social interactions and are applied to recommend social services using the quality of social link.

In the second experiment, we first evaluated the impact of the parameter m on the global social service network by varying the value of m from 1 to 7. Recall that m represents the number of links added and/or rewired at every step of the construction algorithm. As m increases, the number of social links for each node will increase. Fig. 9(a) shows the probability $P(k)$ that a vertex in the network interacts with k other vertices. The results show that as m increases, the probability for vertices with a higher degree is higher and the average degree for all

vertices in the network increases. Second, to show the impact of degree distribution on *link-as-you-go* in terms of service discovering time, we varied the value of m from 1 to 7 and observed the service discovering time according to the number of desired services. Fig. 9(b) shows that the performance of service discovering time was best when $m = 3$. Our experimental results show that as m increases, users find it more difficult to select desired services when they are provided with more service candidates. Third, to show the impact of m on *link-as-you-go* in terms of success rate, we varied the value of m from 1 to 7 and observed the success rate according to the number of users. Fig. 9(c) shows that the success rate decreases as m increases. However, the success rate is still high enough to guarantee the quality of service discovery.

7 CONCLUSIONS

In this paper, to improve the quality of service discovery, we have proposed a methodology to drive an innovation from isolated service islands to *linked social service*. Thus, services can link to and be linked by related services functionally on the Web into a *global social service network*, enabling exploration from service to service. The experimental results show that our approach can solve the quality of service discovery problem, improving not only the service discovering time but also the success rate by exploring service-to-service based on the *global social service network*. In our observation, one limitation of our approach could not involve user's feedback such as positive feedback and negative feedback, we will invest a trust model to incorporate the effect of feedback to better ensure the quality of social link in future work. Another limitation of our model can be described as lack of social influence analysis. To model the social influence on global social service network and identify the representative nodes to better explore global social service network is

another direction for our future work. Another area of investigation is to develop some approaches to clarify service requirements for the service-based economy at the global scale. Our approach can therefore impel service providers to publish their services on the Web into *global social service network*, and motivate service consumers to use services from the *global social service network*.

REFERENCES

[1] J. Davies, J. Domingue, C. Pedrinaci, D. Fensel, R. Gonzalez-Cabero, M. Potter, M. Richardson, and S. Stincic, "Towards the Open Service Web," *BT Technology Journal*, vol. 26, no. 2, 2009.

[2] C. Petrie, "Practical Web Services," *IEEE Internet Computing*, vol. 13, no. 6, pp. 94–96, 2009.

[3] E. Al-Masri and Q. H. Mahmoud, "Investigating Web Services on the World Wide Web," *Proc. 17th IEEE Int'l World Wide Web Conf.*, pp. 795–804, 2008.

[4] W. Jiang, D. Lee, and S. Hu, "Large-Scale Longitudinal Analysis of SOAP-Based and RESTful Web Services," *Proc. 19th IEEE Int'l Web Service Conf.*, pp. 218–225, 2012.

[5] Z. Maamar, H. Hacid and M. N. Huhns, "Why Web Services Need Social Networks," *IEEE Internet Computing*, vol. 15, no. 2, pp. 90–94, 2011.

[6] F. Y. Wang, D. Zeng, K. M. Carley, W. Mao, "Social Computing: From Social Informatics to Social Intelligence," *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 79–83, 2007.

[7] F. De Paoli, M. Palmonari, M. Comerio, and A. Maurino, "A Meta-model for Non-functional Property Descriptions of Web Services," *Proc. 6th IEEE Int'l Web Service Conf., Proc.*, Beijing, China, pp. 393–400, 2008.

[8] Y. Han, S. Chen, and Z. Feng, "Optimizing Service Composition Network from Social Network Analysis and User Historical Composite Services," *AAAI Technical Report SS-12-04: Intelligent Web Services Meet Social Computing*, pp. 39–45, 2012.

[9] M. Klusch, B. Fries, and K. Sycara, "Automated semantic web service discovery with OWLS-MX," *Proc. 15th IEEE Int'l Autonomous agents and multiagent systems Conf.*, pp. 915–922, 2006.

[10] A. Brogi, S. Corfini, and R. Popescu, "Semantics-Based Composition-Oriented Discovery of Web Services," *ACM Trans. Internet Technology*, vol. 8, no. 4, pp. 1–39, 2008.

[11] T. Pilioura and A. Tsalgatidou, "Unified Publication and Discovery of Semantic Web Services," *ACM Trans. Web*, vol. 3, no. 3, pp. 1–44, 2009.

[12] Z. Maamar, N. Faci, L. Krug Wives, Y. Badr, P. Bispo Santos, and J. Palazzo, M. de Oliveira, "Using Social Networks for Web Services Discovery," *IEEE Internet Computing*, vol. 15, no. 4, pp. 48–54, 2011.

[13] M. Hadley, "Web Application Description Language," *Member submission, W3C*, August 2009.

[14] J. Zhang, W. Tan, J. Alexander, I. Foster, and R. Madduri, "Recommend-As-You-Go: A Novel Approach Supporting Services-Oriented Scientific Workflow Reuse," *Proc. 8th IEEE Int'l Services Computing Conf.*, pp. 48–55, Jul. 2011.

[15] W. Tan, J. Zhang, R. Madduri, I. Foster, D. D. Roure and C. Goble, "ServiceMap: Providing Map and GPS Assistance to Service Composition in Bioinformatics," *Proc. 8th IEEE Int'l Services Computing Conf.*, pp. 632–639, Jul. 2011.

[16] C. Bizer, T. Heath, T. B. Lee, "Linked Data – the Story So Far," *J. Semantic Web Inf.*, vol. 5, no. 3, pp. 1–22, 2009.

[17] C. Bizer, T. Heath, T. B. Lee, "Linked Data: Principles and State of the Art," *17th Int'l World Wide Web Conf.*, 2008.

[18] C. Pedrinaci, and J. Domingue, "Toward the Next Wave of Services: Linked Services for the Web of Data," *J. Universal Computer Science*,

vol. 16, no. 13, pp. 1694–1719, 2010.

[19] M. Aquin, E. Motta, M. Sabou, S. Angeletou, L. Gridinoc, V. Lopez, and D. Guidi. "Toward a New Generation of Semantic Web Applications," *IEEE Intelligent Systems*, vol. 23, no. 3, pp. 20–28, 2008.

[20] Z. Gu, B. Xu, and J. Li, "Service Data Correlation Modeling and Its Application in Data-Driven Service Composition," *IEEE Trans. Services Computing*, vol. 3, no. 4, pp. 279–291, 2010.

[21] I. J. Taylor, *Workflows for e-Science: Scientific Workflows for Grids*. Springer, pp. 428–449, 2007.

[22] W. Tan, J. Zhang, and I. Foster, "Network Analysis of Scientific Workflows: A Gateway to Reuse," *IEEE Computer*, vol. 43, no. 9, pp. 54–61, Sept, 2010.

[23] I. Paik, W. Chen, Semantic Pattern Templates for Automating Business Choreography On Web Services, *International Journal of Business Process Integration and Management*, Vol. 6, no. 2, Dec, 2012.

[24] K. S. Nature, R. Albert, and H. Jeong, "Diameter of the World-Wide Web," *Nature*, vol. 401, pp. 398–399, 1999.

[25] F. Lecue and N. Mehandjiev, "Seeking Quality of Web Service Composition in a Semantic Dimension," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 6, pp. 942–959, 2011.

[26] A. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*. vol. 286, pp. 509–512, 1999.

[27] G. Bianconi and A.-L. Barabási, "Bose–Einstein Condensation in Complex Networks," *Phys. Rev. Lett.*, vol. 86, no. 24, pp. 5632–5635, Jun. 2001.

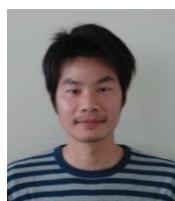
[28] R. Albert and A. Barabási, "Topology of Evolving Networks: Local Events and Universality," *Phys. Rev. Lett.*, vol. 85, pp. 5234–5237, 2000.

[29] G. Bianconi, "Competition and Multiscaling in Evolving Networks," *Europhys. Lett.*, vol. 54, pp. 436–442, 2001.

[30] W. Chen and I. Paik, "Improving Efficiency of Service Discovery Using Linked Data-based Service Publication," *Inf Syst Front*, DOI 10.1007/s10796-012-9381-x, Sep. 2012.

[31] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. "DBpedia: A Nucleus for a Web of Open Data," *Proc. 6th IEEE Int'l Semantic Web Conf.*, pp. 722–735. November 2008.

[32] J. Broekstra, A. Kampman and F. V. Harmelen, "Sesame: An Architecture for Storing and Querying RDF Data and Schema Information," *Spinning the Semantic Web*, pp. 197–222, 2003.



Wuhui Chen received the bachelor's degree in 2008 from Software College Northeast University, China. He received the master's degree in 2011 from School of Computer Science and Engineering, University of Aizu, Japan. Now he is a Ph.D student in School of Computer Science and Engineering, University of Aizu, Japan. His research interests include Service Computing, Cloud Computing, Mobile Service, and Social Computing.



Incheon Paik received the M.E. and Ph.D. degrees in Electronics Engineering from Korea University in 1987 and 1992, respectively. Now he is an associate professor in the University of Aizu, Japan. Research interests include Semantic Web, Web Services and Their Composition, Web Data Mining, Awareness Computing, Security for e-Business, and Agents on Semantic Web. He served several conferences as a program chair and program committee member for numerous international conferences. He is a member of ACM, IEEE, and IPSJ.



Patrick C. K. Hung is an Associate Professor at the Faculty of Business and Information Technology in University of Ontario Institute of Technology (UOIT) in Canada and an Adjunct Professor at the State Key Laboratory of Software Engineering at Wuhan University in China. He is a founding committee member of the IEEE International Conference of Web Services, IEEE International Conference on Services Computing, and IEEE Congress on Services. He is an associate editor of the IEEE Transactions on Services Computing, International Journal of Web Services Research and International Journal of Business Process and Integration Management.