

# ReputationNet: a Reputation Engine to Enhance ServiceMap by Recommending Trusted Services

Jinhui Yao<sup>1,2</sup>, Wei Tan<sup>3</sup>, Surya Nepal<sup>2</sup>, Shiping Chen<sup>1,2</sup>,  
Jia Zhang<sup>4</sup>, David De Roure<sup>5</sup>, Carole Goble<sup>6</sup>,

<sup>1</sup>*School of Electrical and Information Engineering, University of Sydney, Australia*

<sup>2</sup>*Information Engineering Laboratory, CSIRO ICT Centre, Australia*

<sup>3</sup>*IBM T.J. Watson Research Center, USA*

<sup>4</sup>*Department of Computer Science, Northern Illinois University, USA*

<sup>5</sup>*Oxford e-Research Centre, Oxford University, UK*

<sup>6</sup>*School of Computer Science, University of Manchester, UK*

*jin.yao@sydney.edu.au, wtan@us.ibm.com, surya.nepal@csiro.au, shiping.chen@csiro.au,*

*jiazhang@cs.niu.edu, david.deroure@oerc.ox.ac.uk, carole.goble@manchester.ac.uk*

**Abstract**—The concept of Service Oriented Architecture (SOA) enables flexible and dynamic collaborations among different service providers. Backed up by SOA, scientific workflows can bring together various scientific computing tools and resources all offered as services to answer complex research questions. However, studies conducted on myExperiment show that although the sharing of service-based capabilities opens a gateway to resource reuse, in practice, the degree of reuse is very low. This motivates us to propose ServiceMap to provide navigation support through the network of services in building scientific workflows. In this paper, we propose an extension of ServiceMap, i.e., ReputationNet that incorporates the reputation of services/workflows and their publishers to reinforce its capability in terms of service and workflow recommendations. We develop a novel model of the reputation aspects of the services/workflows, and we propose heuristic algorithms to provide service recommendations based on reputations. Experiments have been conducted with workflows on myExperiment to evaluate the effectiveness and validity of the ReputationNet approach for service recommendations.

**Index Terms**—service orientated architecture; scientific workflow; reputation; service composition

## I. INTRODUCTION

The concept of Service Oriented Architecture (SOA) enables flexible and dynamic collaborations among different service providers. A service can either be used in a standalone manner or be composed with other services to form a value-added workflow [25]. Backed up by SOA, scientific workflows can bring together various scientific computing tools and resources, which can be offered as services to answer complex research questions [27]. In this context, workflow models describe the relationship of individual computational components and their input and output data in a declarative way. A workflow engine is responsible for enforcing the predefined invocation sequences among collaborative services. In astronomy, for example, scientists are using workflows to generate science-grade mosaics of the sky [20], and examine the structure of galaxies [30]. In bioinformatics, workflows can help understand the underpinnings of complex diseases [23][24].

In the scientific research, there is a growing trend that researchers publish workflows so that their experimental routines can either be used directly or adapted and repurposed. As a result of this on-going effort, several domain-specific online workflow repositories have evolved in recent years, including the UK-based myExperiment project [13] ([www.myexperiment.org](http://www.myexperiment.org)) comprising of more than 1,700 life-science workflows. The advent of these online repositories makes it possible to assess the state of art of scientific workflow and promote the reuse of the best practices. However, studies [28] conducted on myExperiment indicate that although the sharing of service-based capabilities opens a gateway to resource reuse, in practice, such reuse is very low. In response, the ServiceMap framework [29] has been developed. The ServiceMap framework employs association rule mining and matrix-based searching algorithms aiming to provide a GPS-like support to: 1) help domain scientists better understand various usage patterns of the existing services; and 2) provide a system level support to recommend possible services and their compositions.

ServiceMap provides a framework for navigating through the network of services when building scientific workflows. To utilize this network to find desirable services or service compositions for composing workflows, one needs to choose services or service compositions against other similar alternative ones. Therefore, it is essential to provide guidelines for selecting the best services/service compositions based on observable information about the services/workflows and their publishers. In this paper, we use trust on service/workflow and its publisher, as perceived by consumers, for selecting the best alternative. We extend the ServiceMap framework and generate the ReputationNet by incorporating trust of both services/workflows and their publishers into the service network, to reinforce the capability of ServiceMap in terms of service and workflow recommendations. Our contributions are three-fold: 1) we develop a novel modeling approach to evaluate and present the reputation aspect of the services, workflows and the workflow authors; 2) we propose heuristic algorithms to provide service recommendations based on reputations; and 3)

we conduct experiments using workflows on myExperiment to evaluate the effectiveness of the reputation-based recommendations. Note that although our project focuses on studying workflows on social platforms such as myExperiment, these approaches can be applied to other domains.

## II. BACKGROUND AND RESEARCH MOTIVATION

Our ServiceMap framework addresses two questions that domain scientists frequently ask when exploiting external Web services in building a scientific workflow:

- Q1: *Given the services I plan to use, what are the other services often used together with them, by other scientists?*
- Q2: *Given two or more services I want to use together, can I find an operation chain, which is already used by others, to connect them?*

In an attempt to provide the answers, all the myExperiment workflows were analyzed. Two disjoint networks (graphs) were produced: 1) an undirected workflow-service network, describing the association rules among the services (Q1); and 2) a directed operation network to search for the reachable links between given services (Q2).

To utilize ServiceMap for service recommendations, intuitive enough, one may need to find which services have the strongest association with the target service and they can be recommended with higher confidence; and one may need to find which service composition is more likely to be better than the others when there are multiple reachable links between given services. However, this task is challenging. An empirical study on the two networks in the ServiceMap, revealed very limited reusability of the services and the service compositions among the workflows. Thus, the likelihood of a service or a service composition being used in multiple workflows is low. Certainly, this issue brings significant difficulties for service recommendation.

Two examples are shown in Figure 1. The directed graph in Figure 1a shows a case where three different paths (service compositions) are found linking service *a* (*WSDbfetch.wsd*) and *b* (*WSClustalW2.wsd*) are found. The undirected graph in Figure 1b shows a case that service *c* (*WSWUBlast.wsd*) is found to have associations with four other services (i.e. the services used together in the same workflow). The two cases are derived by ServiceMap using real workflows stored on myExperiment. The numbers on the directed/undirected links indicate the number of occurrences of the link, e.g., there are three workflows in which service composition  $S_a \rightarrow S_1$  is used. However, based solely on numbers of occurrences, identifying the most optimal path or association is difficult. Considering the number of total workflows available (more than 1700 as by February 2012), none of the paths or associations has the number of occurrences dominantly more significant than the others. Meanwhile, many other factors may have impact. Firstly, for instance, the credentials of the authors of the workflows may suggest the credibility of the service composition or association, e.g. if *path 2* is the only one designed by an expert in the field, although it includes smaller number of occurrences of the links compared to the

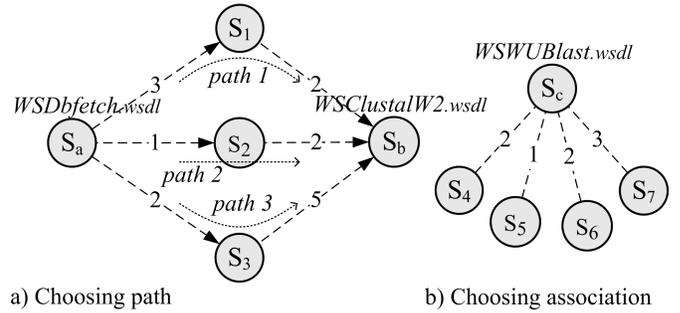


Figure 1. Motivation examples

alternatives, *path 2* should be deemed more credible. Another factor to consider in evaluations can be the popularity of the designed workflows, as the adoption by the community could be an indicator of their reliability and usefulness.

While the community of myExperiment is developing rapidly, the sparseness discovered in the empirical study and the lack of important factors reveal the necessity to employ alternative methods to complement the network analysis of ServiceMap, in order to deliver accurate and effective recommendations. This motivates the development of ReputationNet to consider the reputation aspects of the workflows and their designers to reinforce ServiceMap.

## III. REPUTATION BASED SERVICE RECOMMENDATION

We define *trust* as the belief that a user has regarding the intention and capability of a service/workflow to behave as expected. We use *reputation* as a mechanism of establishing the belief about the services and their publishers' ability to deliver, through collective perception of the users/workflows that have interacted with the service in the past. This mechanism has been successfully applied in Internet marketplaces such as eBay, and Amazon as well as Web services using the concept of reputation [19]. The notion underpinning the reputation-based trust models is to capture consumers' perception of the consumed service and use it to evaluate the reputation of the service [7]. A social platform like myExperiment provides rich data that can be analyzed to synthesize useful reputation insights.

Figure 2 shows the overall approach of ReputationNet. We download all the workflows from myExperiment to build the ServiceMap, which constructs both an undirected network capturing the associations between the services, and a directed network capturing the service compositions. The method to build ServiceMap is reported in [29]. In the meanwhile we download other reputation information about the workflows and their designers, (e.g. view and download times of the workflow, credits received by the designers, the friends the designer is connected to, etc.) to compute their reputation scores. Combining the network structure of the services, the service associations as well as the service Compositions with the reputation scores, we form the ReputationNet. ReputationNet enables us, for instance, to identify the most reputable association (an extension to Q1) that exists between a given service and others related to it. Similarly, for two or more

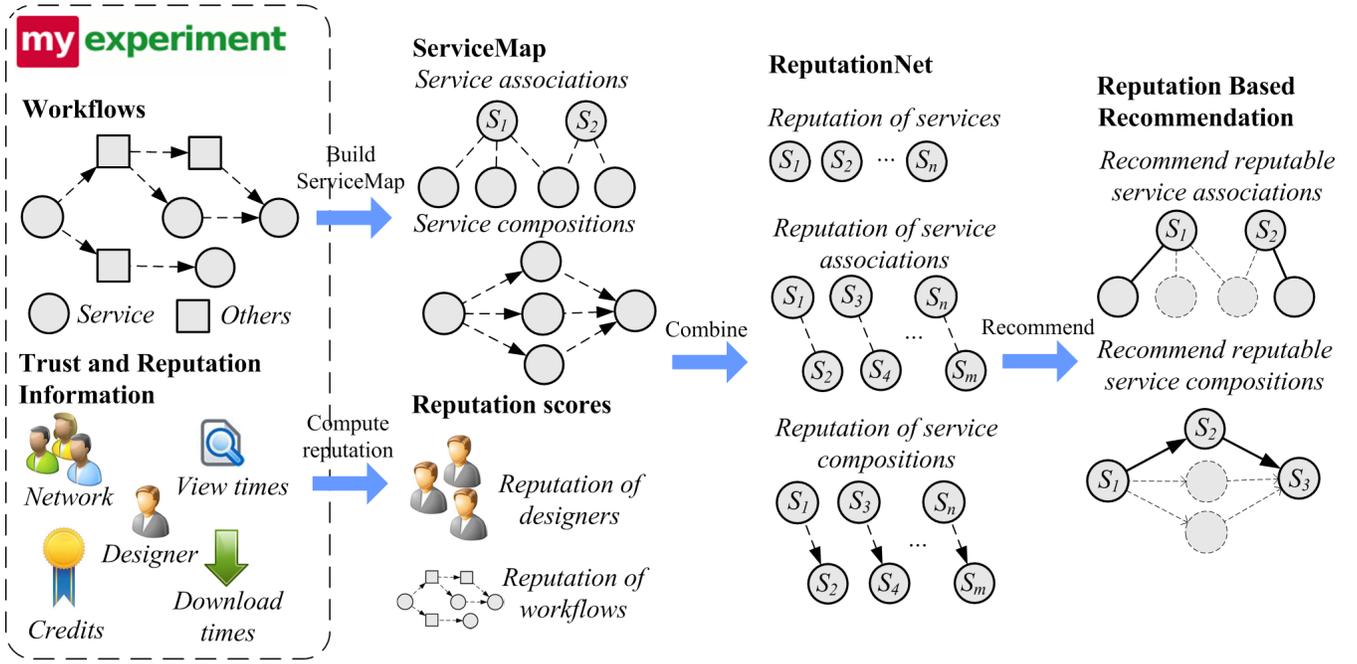


Figure 2. Reputation based approach for service recommendation

services, we can identify a path linking them that has the highest reputation (an extension to  $Q2$ ).

#### IV. MODEL OF SERVICE REPUTATION

In our model, we assume that all workflows are proposed by the members of a community. Each member can propose workflows and view the ones proposed by the peers. Moreover, new workflows can be developed based on existing workflows proposed by other members.

The reputations of the proposed workflows are, thus, evaluated according to two main factors, namely, the reputations of the designers ( $R_{D_i}$ ) and the popularities of the proposed workflows ( $p^{W_i}$ ). The reputation of the workflow proposed by designer is thus computed as

$$r_{D_i}^{W_i} = f(R_{D_i}, p^{W_i}) \quad (1)$$

where function  $f$  must be a monotonic function, i.e.  $R_{D1} > R_{D2}$  and  $p^{W1} > p^{W2}$  imply  $r_{D1}^{W1} > r_{D2}^{W2}$ . There is an obvious analogy between this modeling approach and the common empirical experience, whereby, the more reputable the designer is, and the more times the workflow has been used by others, the more likely it is that the workflow is of a higher quality.

As mentioned, the popularity of a workflow ( $p^{W_i}$ ) should reflect the degree of adoption by other users in the community. While it is difficult to establish exactly how many users have actually used this workflow, some indicating statistics are usually at our disposal, such as the number of times it is viewed ( $n_v^{W_i}$ ), or downloaded ( $n_d^{W_i}$ ), as well as the ratings or feedbacks given by other users ( $\{Rt^{W_i}\}$ ). Generally, the popularity can be computed as

$$p^{W_i} = f(n_v^{W_i}, n_d^{W_i}, \{Rt^{W_i}\}, z) \quad (2)$$

where  $z$  is a catchall variable that represents all other forms of statistics or feedbacks that may suggest the workflow's popularity.

The computation of the author's reputation should take into account his/her contributions to the community as well as his/her popularity among the users. The social network platform may have certain mechanisms to reflect one's contributions, as for myExperiment, each user has a credit score and an average rating. Moreover, the popularity of the workflows that the user has contributed in the past directly suggests his/her reputation as a credible designer. Similarly, a particular user's popularity in the community can be assessed according to the his/her connections. Hence, one's reputation in the community ( $R_{D_i}$ ) can be evaluated as

$$R_{D_i} = f(C_{D_i}, \{p_{D_i}\}, \{Cnt_{D_i}\}) \quad (3)$$

where  $C_{D_i}$  refers to the credit score (or other forms of evaluation) provided by the social network platform,  $\{p_{D_i}\}$  is the collection of the popularities of the workflows user  $D_i$  has contributed; and  $\{Cnt_{D_i}\}$  is the collection of the connections of  $D_i$ .

Each Workflow may contain one service or a collection of services, which, in turn, may form a service composition. Here, the term 'service composition' has a broader meaning, as in the scientific workflows services may comprise of other local non-service activities. Hence, in this context, service composition refers to a group of services linked by one or more operation paths. The reputation of the services or the service compositions can be derived from the reputation of the workflow in different ways. For intuitiveness, we adopt an *equal-share* model where the involved services ( $S$ ) and the service compositions ( $S^{\{ \}$ ) have the reputation of the

workflow.

$$\exists S \in W_i \quad r_S^{Wi} = r_{Di}^{Wi} \quad (4)$$

$$\exists S\{\} \in W_i \quad r_{S\{\}}^{Wi} = r_{Di}^{Wi} \quad (5)$$

The assumption underpinning the *equal-share* model is that if the groups of users are aware of the workflow, they are implicitly aware of the services contained within. In other words, the services, service compositions and the workflow are equally reputable. A more sophisticated approach is to adopt a *fair-share* model, whereby the reputation of the workflow is fairly distributed to the services and the service compositions according to their contributions to the workflow. More details about fair-share can be found in [22].

A particular service or a service composition may be used in multiple workflows designed. The instances of their use all contribute to their reputations in the community, i.e. the more times a service is used in different workflows, the more reputable this service will become. An aggregate reputation of the service or the service composition can be calculated according to the individual reputations derived from the workflows that incorporate them. This aggregate reputation suggests the reputation of the service or the service composition within this community. The aggregate reputation of a given service  $S_i(R_{Si})$  and the aggregate reputation of a given service composition  $S_i\{\}(R_{S_i\{\}})$  are computed as

$$R_{Si} = f(r_{Si}^{W1}, r_{Si}^{W2}, \dots, r_{Si}^{Wn}) \quad (6)$$

$$R_{S_i\{\}} = f(r_{S_i\{\}}^{W1}, r_{S_i\{\}}^{W2}, \dots, r_{S_i\{\}}^{Wn}) \quad (7)$$

Individual reputation scores are computed based on the reputation of the workflow designer and the popularity of the workflow. Thus, a service or service composition with a high aggregate reputation entails that i) it has been involved in the workflows that are used extensively by the users; or ii) it has been involved in the workflows that are designed by very reputable designers; or iii) both i) and ii).

#### A. Reputation bootstrapping

It is possible that for a workflow, one or both of the elements in the function in (1) are missing. For instance, newly posted workflows will have no popularity. In these cases we need to bootstrap the workflow reputation.

For newly posted workflows that have no popularity  $p^{Wi} = \phi$ , the solution is straightforward, as we can simply calculate the reputation of the workflow solely based on the reputation of the author, that is

$$r_{Di}^{Wi} = f(R_{Di}, \phi) = f(R_{Di}) \quad (8)$$

In this case, this workflow simply ‘inherits’ the reputation of the designer, which is an estimation based on his/her past contributions.

However, it is also recognized that new myExperiment members would be seriously disadvantaged if assigned zero reputation on joining. In order to encourage new users to contribute to the community, this disadvantage should be minimized if not eliminated. Bootstrapping one’s reputation refers to assigning or adding an initial value to a newcomer’s

reputation [16], so that his/her newly designed workflows are reasonably competitive among the workflows designed by others.

This can be achieved in multiple ways. For example, we can initialize one’s reputation based on the reputations of his connections. Similarly, if the community provider can incorporate endorsement techniques [17], newcomers can present the credentials of any existing users willing to endorse them. However these approaches will impose further assumptions on the community. An alternative, and a more generic approach can be taken by initializing one’s reputation with the average reputations of all users. It is demonstrated in [11] that such an averaging technique provides the best results in terms of fairness and accuracy. Taking this work as a starting position, we also adopt an averaging model in this case. The bootstrapping reputation of a newly joined user  $R_{Di}^{Bp}$  is

$$R_{Di}^{Bp} = \frac{1}{N} \sum_{j=1}^N R_{Dj} \quad \text{for all } j \neq i \quad (9)$$

where  $N$  is the total number of the users in the community. The newcomer’s reputation will be boosted by adding this average reputation to the real reputation computed in (3)

$$R_{Di}^{\#} = R_{Di} + R_{Di}^{Bp} \quad (10)$$

where  $R_{Di}^{\#}$  is the adjusted reputation for the newcomer  $D_i$ . In this way, the newly joined users are neither advantaged nor disadvantaged. According to (8) the new workflows designed by new users will have the same bootstrap reputation which is close to the average reputation value of all the newly designed workflows.

The bootstrapping reputation offers an advantage to the newcomers to offset their disadvantage inherent in not having any prior contributions. Clearly, this advantage should not be offered forever. This issue will be discussed in the next section when we introduce the *temporal sensitivity*.

#### B. Reputation fading

An important aspect in reputation assessment is the temporal sensitivity, as reputation information of a service decays with time [26]. For instance, given two workflows with the same popularity (i.e., with the same results in (2)), the workflow that is designed earlier should have a lower reputation, as it suggests that the newer workflow has gained equal popularity more rapidly and is likely more up to date. Similar idea can be applied to the reputation of the authors.

Along those lines, the advantage offered to the newcomers should diminish in time so that eventually only their true reputations are considered (equation (3)). Thus, a bootstrapping fading factor  $\alpha_{fd}^{Bp} \in [0, 1]$  must be incorporated into equation (10)

$$R_{Di}^{\#} = R_{Di} + \alpha_{fd}^{Bp} R_{Di}^{Bp} \quad (11)$$

$$\alpha_{fd}^{Bp} = f(\Delta t) \quad (12)$$

where  $\Delta t$  is the time difference between the present time and the time of the event. i.e. the time that the newcomer joined the

community. A linear or exponential function can be applied to reduce  $\alpha_{fd}^{Bp}$  from 1 eventually to 0.

The significance of popularity of a workflow fades in time; i.e. a workflow can gain a high popularity if it is downloaded or used extensively, however this popularity should decline if such trend discontinues. Given that the statistics of its usage usually cannot decrease (e.g. number of downloads) a popularity fading factor  $\alpha_{fd}^{Pop}$  should be incorporated into the model. The adjusted popularity  $p^{\#Wi}$  is then

$$p^{\#Wi} = p^{Wi} - \alpha_{fd}^{Pop} \quad (13)$$

$$\alpha_{fd}^{Pop} = f(\Delta t, z) \quad (14)$$

where  $\Delta t$  is the time interval difference between the present time and the time when this workflow is designed, and  $z$  is a catchall variable referring to other factors that may be considered, such as the pattern of the popularity growth.

Note that the popularity fading mechanism affects both the reputation of the workflow (in (1)) and the reputation of the authors (in (3)). This allows the convergence of reputation to a very small value as time passes if the workflow is no longer used or the author stops contributing.

## V. SERVICE RECOMMENDATION BASED ON REPUTATION

As noted above, two key recommendation needs have been recognized in ServiceMap, namely 1) recommend the services that have been used together - undirected association rules recommendation; and 2) recommend the path that links one service to another - directed service composition recommendation. Starting from ServiceMap, we now elaborate our approach to enhance the capability of such recommendations by utilizing the reputation mechanisms presented previously.

The association rules obtained can be used to suggest other relevant services usually used by peers, when scientists have already put some services into their partial-completed workflow. Feedback from caBIG [10] users shows that these rules are quite helpful in terms of introducing relevant services from a large set into their experiments. However, due to the limited number of frequently utilized sets, the number of association rules that can be discovered according to the number of occurrences is low. The reputation of a service, on the other hand, can be high even when it has only been used in one workflow, as it inherits the reputation of the workflow and hence of the author. A similar approach can be applied for associated service groups ( $S^{<>}$ ), that is

$$S^{<>} \in W_i \quad r_{S^{<>}}^{Wi} = r^{Wi} \quad (15)$$

Then, the aggregate reputation of the associated service group ( $R_{S^{<>}}$ ) can be derived as

$$R_{S^{<>}} = f(r_{S^{<>}}^{W1}, r_{S^{<>}}^{W2}, \dots, r_{S^{<>}}^{Wn}) \quad (16)$$

Based on (15) and (16), we have developed a simple algorithm, as shown in Algorithm 1, to find, for a particular service  $S_i$ , the set of the associated service groups  $\{S_{Sa}^{<>}\}$  that have reputations exceeding the threshold  $R^{Thres}$ .

Service composition recommendation provides a cross-workflow search technique. Our experience working with

---

### Algorithm 1: Choosing reputable service associations

---

**Input:** a set of workflows  $\{wf\}$ , a service  $S_a$   
**Output:** a set of associated service groups  $\{S_{Sa}^{<>}\}$

```

 $\{S_{Sa}^{<>}\} \leftarrow \phi$ 
foreach  $wf_i \in \{wf\}$  do
  foreach  $S^{<>} \in wf_i$  do
    if  $S_a \in S^{<>}$  and  $S^{<>} \notin \{S_{Sa}^{<>}\}$  then
      Compute  $R_{S^{<>}}$  using (16)
      if  $R_{S^{<>}} > R^{Thres}$  then
         $\{S_{Sa}^{<>}\} \leftarrow S^{<>}$ 
      end
    end
  end
end

```

---



---

### Algorithm 2: Choosing reputable service compositions

---

**Input:** the set of all service compositions  $\{S^{\{\}}\}, S_a, S_b$   
**Output:** the most reputable  $S^{\{\}}$  links  $S_a, S_b$

```

 $R_{S1^{\{\}}} \leftarrow 0, R_{S2^{\{\}}} \leftarrow 0, S_{rep}^{\{\}} \leftarrow \phi$ 
foreach  $S_i^{\{\}} \in \{S^{\{\}}\}$  do
  if  $S_a \in S_i^{\{\}}$  and  $S_b \in S_i^{\{\}}$  then
     $R_{S2^{\{\}}} \leftarrow$  compute  $R_{S_i^{\{\}}}$  using (19)
    if  $R_{S2^{\{\}}} > R_{S1^{\{\}}}$  then
       $S_{rep}^{\{\}} \leftarrow S_i^{\{\}}, R_{S1^{\{\}}} \leftarrow R_{S2^{\{\}}}$ 
    end
  end
end

```

---

caBIG community shows that this feature can be quite useful for scientists to explore best practices. Thus, in our view, given that the confidence of each recommendation is known, it would yield more insights for the users. The reputation of a service composition apparently provides indication of such measure.

As we have illustrated, the service compositions share the reputations of the workflows. For the service compositions that are used in different workflows that can be merged, we can synthesize the reputation of the merged larger service composition by merging their individual reputations. For instance, workflow  $W_a$  contains service composition  $S_a^{\{\}} : S_1 \rightarrow S_2$ , whilst workflow  $W_b$  contains service composition  $S_b^{\{\}} : S_2 \rightarrow S_3$ . Merging these two compositions, we obtain a larger composition  $S_c^{\{\}} : S_1 \rightarrow S_2 \rightarrow S_3$ . Although this composition may have not been used by any workflows designed, its reputation can be synthesized by utilizing the reputation of the two individual service compositions involved. That is

$$\exists S_c^{\{\}} = S_a^{\{\}} \cup S_b^{\{\}} \quad R_{S_c^{\{\}}} = f(R_{S_a^{\{\}}}, R_{S_b^{\{\}}}) \quad (17)$$

The above expression can be generalized to accommodate multiple service compositions, that is

$$\exists S^{\{\}} = S_1^{\{\}} \cup S_2^{\{\}} \dots S_n^{\{\}} \quad (18)$$

$$R_{S^{\{\}}} = f(R_{S1}, R_{S2}, \dots, R_{Sn})$$

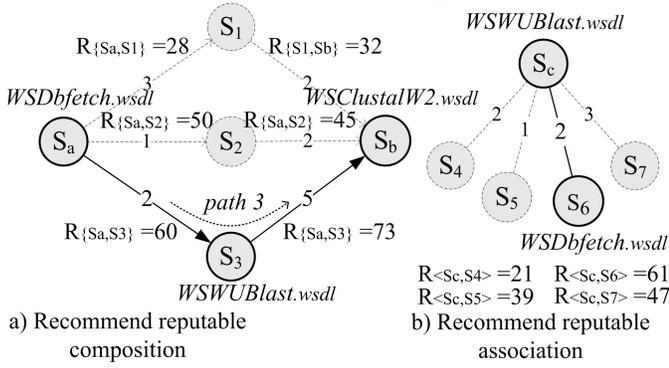


Figure 3. Examples of service recommendation

A simple algorithm shown in Algorithm 2 can be followed to find the service composition that links two specific services with the highest reputation. Note that it relies on the ServiceMap to develop the directed operation network to identify all possible service compositions ( $\{S^{\cup}\}$ ) contained in the workflow set.

To demonstrate the use of the recommendation algorithms, we have applied them to the example presented in Section 2 (Figure 1). In Figure 3, the service compositions and the associations in Figure 1 are displayed together with the scaled reputations (linearly scaled to a value between 0 and 100, the latter being the most reputable) of the individual directed/undirected links. We can see that, although the numbers of occurrences of the links are quite similar, their reputation scores are diverse. By using algorithm (1) and (2) we have selected the most reputable composition and the association, which are emphasized in the figure with thick solid lines, while other less reputable alternatives have been de-emphasized with thin dashed lines. Clearly, the most reputable composition links services  $S_a$  and  $S_b$  is through service  $S_3$  (*WSWUblast.wsdl*); whilst the most reputable association for service  $S_c$  is service  $S_6$  (*WSDbfetch.wsdl*) among the alternatives.

## VI. EXPERIMENTS

We conducted experiments with workflows on myExperiment to evaluate the effectiveness of our reputation based approach. We first downloaded all workflows on myExperiment to build the ServiceMap, which, as noted earlier consists of an undirected network of workflows-services and a directed network of service compositions. Next, we downloaded the metadata pertaining to all workflows (e.g., the rating, view times and download times, etc, of each workflow) as well as the information related to the authors (e.g., credit score, number of friends and ratings, etc.) to construct the ReputationNet, which is superimposed on the ServiceMap. For simplicity, we use addition in the equations in Section 4 and 5 to calculate the reputation scores, that is, all the elements in the functions are added.

We assume that the download times of a workflow to a certain extent, reflect its usage. We further assume that, in a social platform such as myExperiment, even without any recommendation mechanism, the community members will

eventually find the high quality services or service compositions to use in their workflows, either through a long try-and-error process or through peers' suggestions. Based on these two reasonable assumptions, it follows that, high download times suggest high workflow usages that, in turn, suggest the high quality of the services, associated service groups and service compositions contained within.

Figure 4a shows the plot of the scaled (linearly scaled to a value between 0 and 100, the latter being the most reputable or used) reputations of 258 WSDL-style services calculated using (6) versus their scaled usages (i.e., download times). Obviously, the distributions of their reputations and usages are not uniform, but the positive correlation between the usage and reputation is apparent. We calculated the least squares regression lines of the points. The slope of the regression line is about 0.93, suggesting a strong positive relation between the two. To further verify this relation, we exclude the popularity of the workflow in (1) and compute the reputations of the services solely using the reputations of the authors - bootstrapping the reputations for all workflows. In this approach, the effect of the download times on the reputations computed is greatly reduced. The results are plotted in Figure 4b. Although the plots in Figure 4b are more scattered, the positive correlation is still high, with the regression line slope of about 0.81. It is thus reasonable to conclude that the reputation is a good indicator of the service usage, and therefore its quality.

A cross-validation experiment is conducted to evaluate the effectiveness of the reputations computed for the associated service groups and the service compositions. The full set of workflows is partitioned into two subsets, i.e., the training set and the testing set. For workflows posted in each month, half of them are randomly selected into the training set while the remaining half goes to the testing set. This partition is to make sure the workflows in the training set and the testing set have similar timeline distributions. Next, we compute the reputations of the associated service groups and the service compositions using the workflows in the training set, and validate them with the usage calculated using the workflows in the testing set. By taking this approach, we can mostly (if not entirely) eliminate the effects of the subject's usage on the computing of its reputation. The results are plotted in Figure 4c and 4d. In the figures we have only plotted the subjects that appear in both training set and testing set. Due to the limited service reusability, a large portion of the associated service groups and service compositions only appear in one of the sets. However, apart from the sparseness, there appear to be a clear positive correlation between the usage and reputation in both figures. The slopes of the regression lines in Figure 4c and 4d are 0.73 and 0.77, respectively. These findings imply that the reputations computed can be used to provide effective and meaningful association rule and service composition recommendations as discussed in Section 4.

## VII. RELATED WORK

The concept of trust is not new. Trust has been studied in many disciplines including sociology [19], psychology [7], and computer science [12]. Each of these disciplines has

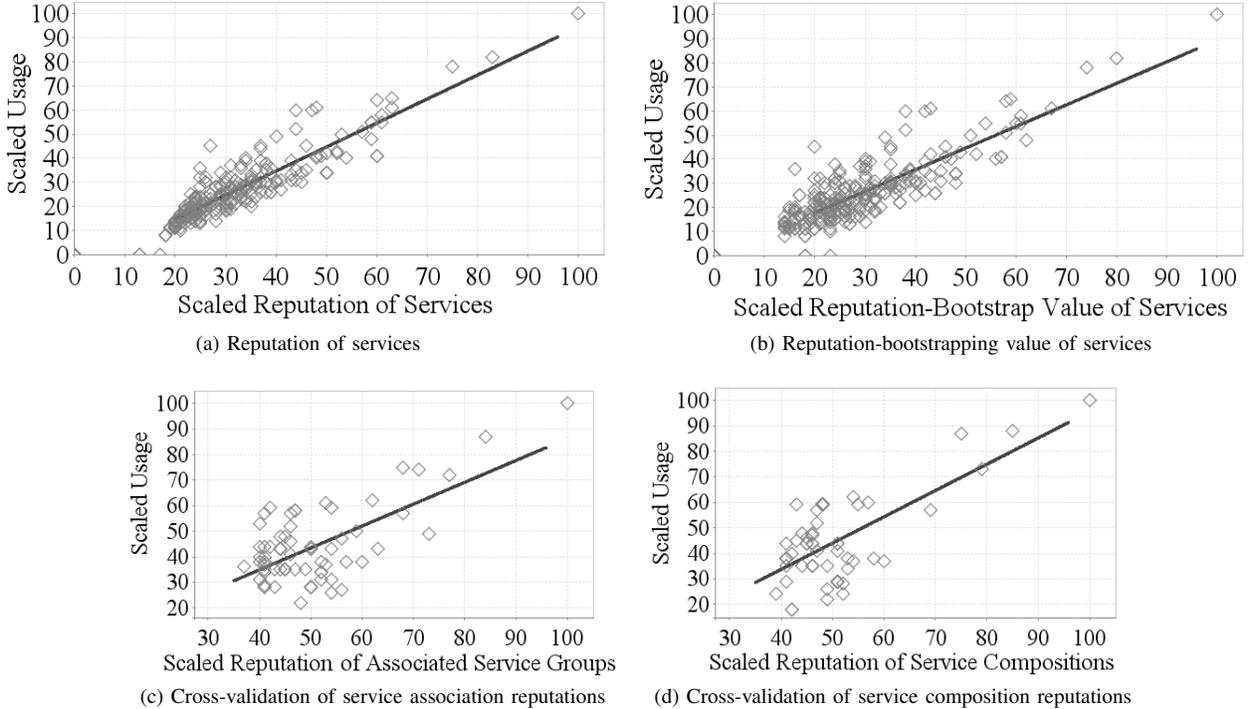


Figure 4. Performance testing

considered trust from different perspectives. There is no single consensus definition of trust in the literature. In general, trust is a measure of confidence that an entity or entities will behave in an expected manner. In this paper, we consider the reputation based trust [21].

Reputation systems have benefited electronic commerce in recent years. Amazon, eBay and Yahoo! Auction are examples of businesses that have deployed reputation systems successfully. These reputation systems use feedbacks from the consumers as the reputation measure, and have received considerable attention in the literature [5]. SPORAS [31] is one such centralized reputation model that extends the above mentioned models with more sophisticated characteristics to model trust dynamics. We have adopted and extended this concept and used along with reputation networks to develop a reputation based recommendation framework.

Trust based recommendation system has been studied in social networks and human computer interactions, where trust plays a major role in selecting the best nodes. [18] proposes a trust-based recommendation system where it is possible to search for trustworthy users by exploiting trust propagation over the trust network. [2] explore an axiomatic approach for trust-based recommendation and propose several recommendation models, some of which are incentive compatible. Hess [14] extend trust-based recommendations for single items such as movies to linked resources.

Automatic service composition is another intensively investigated topic in services computing. Various techniques have been developed to discover relevant services and compose them in a proper sequence [9][3]. Aalst proposed a framework named TomTom4BPM [1] that adopts process mining technique for various purposes, such as comparing the actual

process execution with pre-modeled ones and dynamically navigating during process exceptions. There are some work related to process mining, such as deriving patterns from past usage data to predict the most likely next-step in building visualization pipelines [15], and case base reasoning in finding a similar workflow and using it to suggest the next component to be included in a workflow [6].

The proposed approaches we have mentioned can yield good results when services do have complete metadata (input/output, pre/post conditions, QoS, etc), such that the composition problem can be translated into a well formalized one such as optimization and AI-based planning. In reality, however, many services are widely used without much metadata. Meanwhile, online workflow repositories (such as myExperiment) allow scientists to share successful experimental routines that contain best practices to compose services. Based on this observation, we have adopted a framework to derive the credibility of the authors and popularity of the services to develop a service reputation network, to provide recommendations based on empirical workflows. This work is developed based on our earlier works on ServiceMap [28][29] and service reputation [17][16][22].

## VIII. CONCLUSIONS

This paper presents the ReputationNet framework, as an extension to ServiceMap to employ trust and reputation mechanisms for service recommendations in building scientific workflows. We have developed a novel model to capture the trust and reputation aspects of scientific Web services and workflows. Based on the model, we have designed ranking algorithms to recommend reputable service associations and

compositions. We conducted experiments through the workflows on myExperiment. Our experiment results demonstrated a strong positive correlation between the reputation of the subject (services, service associations and service compositions) and its actual usage by the members of the myExperiment community. This result in turn confirms the validity of the reputation computed and the capability of our framework in terms of service recommendation. In future work, we plan to evaluate the reputation of services/workflows with more information, for example, by integrating with service registries like BioCatalogue [4], or consider the social network aspects of the workflows [8].

## IX. REFERENCES

- [1] W. M. P. Aalst, "Using process mining to generate accurate and interactive business process maps," in *International Business Information Systems Workshops*, 2009, pp. 1–14.
- [2] R. Andersen, C. Borgs, J. Chayes, U. Feige, A. Flaxman, A. Kalai, V. Mirrokni, and M. Tennenholtz, "Trust-based recommendation systems: an axiomatic approach," in *International Conference on World Wide Web*, 2008, pp. 199–208.
- [3] K. Belhajjame, C. Goble, S. Reyes, and D. De Roure, "Fostering scientific workflow preservation through discovery of substitute services," in *IEEE International Conference on eScience*, 2011.
- [4] J. Bhagat, F. Tanoh, E. Nzuobontane, T. Laurent, J. Orłowski, and M. Roos, "et al. biocatalogue: a universal catalogue of web services for the life sciences," *Nucleic Acids Research*, vol. 38, pp. W689–W694, 2010.
- [5] E. Chang, F. Hussain, and T. Dillon, *Trust and Reputation for Service-Oriented Environments: Technologies For Building Business Intelligence And Consumer Confidence*. John Wiley & Sons, 2005.
- [6] E. Chinthaka, "Cbr based workflow composition assistant," in *World Congress on Services*, 2009.
- [7] K. S. Cook, T. Yamagishi, C. Cheshire, R. Cooper, M. Matsuda, and R. Mashima, "Trust building via risk taking: A cross-societal experiment," *Social Psychology Quarterly*, vol. 68, pp. 121–142, 2005.
- [8] D. De Roure, S. Bechhofer, C. Goble, and D. Newman, "Scientific social objects: The social objects and multidimensional network of the myexperiment website," in *International Workshop on Social Object Networks*, 2011.
- [9] S. Dustdar and W. Schreiner, "A survey on web services composition," *International Journal of Web and Grid Services*, vol. 1, no. 1, pp. 1–30, 2005.
- [10] A. C. Eschenbach and K. Buetow, "Cancer informatics vision: cabig," *Cancer Informatics*, vol. 2, p. 22, 2006.
- [11] M. Feldman and J. Chuang, "The evolution of cooperation under cheap pseudonyms," in *Conference of E-Commerce Technology*, 2005, pp. 284–291.
- [12] M. Gamble and C. Goble, "Quality, trust, and utility of scientific data on the web: Towards a joint model," in *ACM International Conference on Web Science*, 2011.
- [13] C. Goble, J. Bhagat, S. Aleksejevs, D. Cruickshank, D. Michaelides, and D. Newman, "et al. myexperiment: a repository and social network for the sharing of bioinformatics workflows," *Nucleic Acids Research*, vol. 38, pp. W677–W682, May 2010.
- [14] C. Hess, "Trust-based recommendations for publications a multi-layer network approach," *IEEE Technical Committee on Digital Libraries Bulletin*, vol. 2, no. 2, 2006.
- [15] D. Koop, "Automating suggestions for visualization pipelines," *IEEE Trans. on Visualization and Computer Graphics*, vol. 14, pp. 1691–1698, 2008.
- [16] Z. Malik and A. Bouguettaya, "Reputation bootstrapping for trust establishment among web services," *IEEE Internet Computing*, vol. 13, no. 1, pp. 40–47, 2009.
- [17] Z. Malik and A. Bouguettaya, "Rater credibility assessment in web services interactions," in *Journal of World Wide Web*, 2008, pp. 1–23.
- [18] P. Massa and P. Avesani, "Trust-aware recommender systems," in *ACM conference on Recommender systems*, 2007.
- [19] G. Mollering, "The nature of trust: From geogimmel to a theory of expectation, interpretation and suspension," *Sociology*, vol. 35, pp. 403–420, 2002.
- [20] Montage. [Online]. Available: <http://montage.ipac.caltech.edu>
- [21] S. Nepal, Z. Malik, and A. Bouguettaya, "Reputation management for composite services in service-oriented systems," *International Journal of Web Service Research*, vol. 8, no. 2, pp. 29–52, 2011.
- [22] S. Nepal, Z. Malik, and A. Bouguettaya, "Reputation propagation in composite services," in *International Conference on Web Services*, 2009, pp. 259–302.
- [23] T. Oinn, M. Greenwood, M. Addis, M. N. Alpdemir, J. Ferris, and K. Glover, "et al. taverna: Lessons in creating a workflow environment for the life sciences," in *Concurrency and Computation: Practice and Experience*, 2006, pp. 1067–1100.
- [24] T. Oinn, P. Li, D. Kell, C. Goble, A. Goderis, and M. Greenwood, "et al. taverna/mygrid: Aligning a workflow system with the life sciences community," *Workflows in e-Science*, Springer, 2006.
- [25] M. P. Papazoglou, P. Traverso, S. Dustdar, and F. Leymann, "Service-oriented computing: State of the art and research challenges," *IEEE Computer*, pp. 38–45, 2007.
- [26] W. Sherchan, S. Nepal, and A. Bouguettaya, "A trust prediction model for service web," in *International Conference of TrustCom*, 2011, pp. 258–265.
- [27] W. Tan, K. Madduri, A. Nenadic, S. Soiland-Reyes, D. Sulakhe, I. Foster, and C. Goble, "cagrid workflow toolkit: A taverna based workflow tool for cancer grid," *BMC Bioinformatics*, vol. 11, p. 542, 2010.
- [28] 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, 2010.
- [29] W. Tan, J. Zhang, R. Madduri, I. Foster, D. De Roure, and C. Goble, "Servicemap: Providing map and gps assistance to service composition in bioinformatics," in *International Conference on Services Computing*, 2011, pp. 632–639.
- [30] I. Taylor, M. Shields, I. Wang, and R. Philp, "Distributed p2p computing within triana: A galaxy visualization test case," in *IEEE International Parallel and Distributed Processings Symposium*, 2003.
- [31] G. Zacharia and P. Maes, "Trust management through reputation mechanisms," *Applied Artificial Intelligence*, vol. 14, no. 9, pp. 881–908, 2000.