

# Opportunistic Information Flows Through Strategic Social Link Establishment

Florian Skopik, Daniel Schall, Schahram Dustdar  
Distributed Systems Group  
Vienna University of Technology  
Argentinierstraße 8/184-I, A-1040 Vienna, Austria  
{lastname}@infosys.tuwien.ac.at

**Abstract**—Social networks have emerged from niche existence to a mass phenomenon. Nowadays, their fundamental concepts, such as managing personal contacts and sharing profile information, are increasingly harnessed for businesses in professional environments. Similar to service-oriented networks, they allow flexible discovery on demand and loose coupling of participants. Establishing social links facilitates cooperation and enables selective sharing of information. Intuitively, one shares more information with his connected neighbors and less or even none with unrelated individuals. Today, information is one of the most important and valuable goods in business networks. Being informed about ongoing collaborations and upcoming trends is a key success factor. Thus, in professional networks, participants aim at strategically establishing connections to enable reliable information flows. In this paper, we especially highlight an opportunistic model that let mediators connect actually unrelated actors in order to benefit from information mediation. We further discuss a framework that implements this model for service-oriented professional virtual communities.

**Keywords**-social networks, information mediation, strategic link establishment, reciprocity, structural holes

## I. INTRODUCTION

The Web has evolved from a distributed document repository to an interactive medium in which people actively share and disseminate information. Parts of this evolution is often referred to as Web 2.0 and characterized by the emergence of knowledge sharing communities. The way people interact on the Web, especially in professional environments, is changing once more. Service-oriented computing takes off on its triumphal course to permit even human-centric business platforms [1]. Web services enable loosely-coupled cross-organizational collaborations, and are the ideal means to realize flexible discovery and binding of interaction partners. In such service-oriented collaboration environments, participants shape the availability of information and services.

Social networks typically emerge freely and independently without restricted paths and boundaries. Research has shown that the resulting social network structures allow for relatively short paths of information propagation. For example, ‘six-degrees of separation’ [2] refers to the idea that each node in a freely emerged people network can be reached by propagating items of information via six hops. While this is true for autonomously forming social networks,

the boundaries of collaborative networks are typically restricted due to organizational units and fragmented areas of expertise. In order to take advantage of social preferences, we propose social network principles to overcome limited information flows in collaborative environments. Particularly, in *service-oriented professional communities*, actors perform activities by interacting with other community members, e.g. to inquire information, exchange ideas, and delegate requests. Over time, actors establish social connections to their collaboration partners [3], [4]. Information are propagated along these links. Typically people share more information with well-known partners who proved their reliability earlier; and less or even none with unknown third parties. Hence, to motivate two unconnected users to exchange information, and thus, enable reliable information flows in networks, they need a mutually known intermediate actor.

According to the structural holes theory [5], people show the tendency to position themselves in networks as such *interaction mediators*. Their main incentives are to get valuable insights in ongoing collaboration and others’ work. Mediating interactions between network members further allows to influence and control partners and to build up distinct reputation and high visibility levels. However, acting as mediator also requires free capacities, e.g., in terms of time and effort, and since resources are usually limited it is a top priority to carefully decide about mediation targets.

In this paper we deal with the following contributions:

- *Information Mediation Model*. We design an analytical model that explains the fundamental concepts of opportunistic information flows. Our model introduces the notion of *utility* in service-oriented professional communities.
- *Social Link Establishment*. We formalize the incentives and motivation of human behavior regarding social link establishment in virtual communities established upon interaction analysis.
- *Evaluation and Discussion*. We evaluate the proposed model and its application for virtual communities, and derive general findings for designing applications for socially-enhanced service-oriented environments. Also, we discuss the application of and integration with available social network standards for open environments.

The remainder of this paper is organized as follows. Section II describes fundamental concepts of our approach and motivates the need for information mediation. Section III presents an algorithmic view to describe the motivation and incentives behind establishing social relations and deals with an information mediation model. Then, we evaluate and discuss our work in Section IV. Section V outlines a technical perspective on an SOA-based implementation and application. Section VI deals with related work and Section VII concludes the paper.

## II. SOCIAL INFORMATION MEDIATION

We discuss a layered *social information mediation model* and outline utilized major concepts on each layer.

### A. Overview

Basically, an actor has several *passive links*, such as FOAF<sup>1</sup> relations, that just express business/personal contacts (typically emerged from previous collaborations), but no interactions are performed along these links. An actor can activate these links (*active links*) by initiating a new collaboration, e.g., setting up a joint activity. However, due to resource constraints, members can only participate in a limited amount of concurrent activities, and thus, the number of simultaneously active links is limited. Hence, collaboration partners are selected carefully, considering required effort and received utility.

Figure 1 shows an overview of our layered approach. On the bottom layer, interactions are observed and collected to determine social relations. We designed the system to manage relations by evaluating occurring interactions and therefore, unburden network participants – at least partly – from managing their relations manually [3]. On the medium layer, direct relations are established to create a typical social network. Since single members usually build up strong relations to only a small amount of partners, reliable information flows through collaborative networks are limited. Therefore, on the top layer, social mediation is applied.

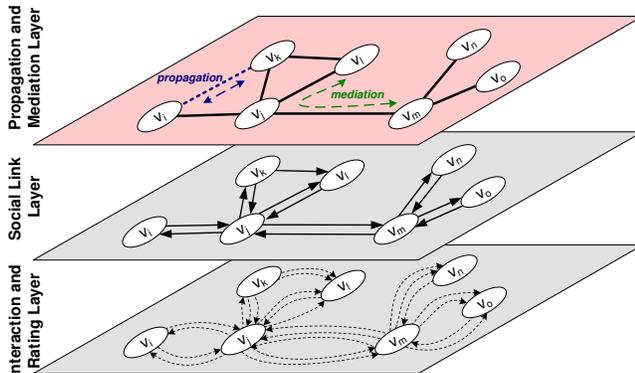


Figure 1. Model for social information mediation.

<sup>1</sup><http://xmlns.com/foaf/spec/>

We have studied the bottom and middle layer, as depicted by Figure 1, already in our previous work. We briefly outline the main principles of these layers to highlight the novelty of the proposed utility model.

- **Interaction Layer.** A professional virtual community (PVC) is a special kind of social network, where the single actors participate to perform activities. A community is modeled as a directed graph, where vertices  $V$  represent the actors that are connected through edges  $E$ . A directed edge from actor  $v_i$  to  $v_j$  is denoted as  $e_{ij}$ . Activities  $A$  are a fundamental part of our model; thus, we describe the graph model of a PVC as  $G = (V, E, A)$ . The concept of an activity  $a \in A$  is used to include a set of participants. Thus, in short, activities describe the collaboration boundaries and goals. Network members interact in scope of particular activities (i.e., to reach certain goals). Interactions are collected to determine (i) the center of interest of single network members by evaluating the frequency of used keywords [3], [6], and (ii) the strength of a social relation by determining the similarity of the center of interests [3].
- **Social Link Layer.** By issuing keyword-based queries, each actor's connectivity to other community members in a particular query context  $Q$  is determined [7]. The query context is described by a pool of keywords (e.g., describing certain expertise areas) picked from global taxonomies. Using logged interaction data and manual ratings the link strength from one actor  $v_i$  to another  $v_j$  in context of  $Q$  is calculated. For that purpose, various metrics, such as availability of actors, average ratings, responsiveness and interest similarities can be calculated dynamically.

### B. Information Mediation

Typically, network members share information along emerging direct relations. Sharing with known partners is beneficial for both, the information provider who can spread information but still knows the boundaries of potential receivers, and the receiver who directly knows the provider and can decide on the trustworthiness of delivered information. However, each network member does usually interact with only a small amount of distinct partners compared to the overall size of a large-scale system. Thus, information, such as announcements and invitations, cannot be reliably and also widely spread at the same time.

In order to compensate that issue, several propagation models have been proposed to establish artificial connections in networks that are inferred from existing ones [8]. The most common concept is *direct propagation*. Here, a node  $v_i$  has a strong relation to  $v_j$  which is tightly linked to  $v_k$  (see Figure 1). Propagation means that, although  $v_i$  and  $v_k$  have never personally interacted with each other, a synthetic social relation can be introduced using the existing network

structure. For instance, if these social links are considered as trust relations [3] one could say, because  $v_i$  trusts  $v_j$  and  $v_j$  trusts  $v_k$ , there is a high probability that  $v_i$  can also trust  $v_k$ ; e.g., due to their similar habits, attitudes, and values.

However, sharing information along synthetic links, has several implications and potential disadvantages for the involved actors. Typically, the information owner  $v_i$  shares information with the actually unknown sink  $v_k$ ,  $v_k$  receives information from an unknown source  $v_i$ , and  $v_j$  initially supports establishing a connection between  $v_i$  and  $v_k$  but is then bypassed and cannot control the actual information propagation over that supported link. Furthermore, there may be situations, where  $v_i$  does not want to reveal its identity to unknown network members (here:  $v_k$ ), however, still wants to distribute information. Hence, additionally to these propagation models, we further apply the concept of active *information mediation*.

### III. UTILITY-BASED MEDIATION MODEL

We begin with the basic mechanisms describing motivation and incentives for building relations on social Web platforms.

#### A. Link Establishment

Information mediation assumes that actors actively bridge isolated components in a social network. They mediate information to facilitate collaborations between actually unconnected network members. The most fundamental questions, however, are (i) what is a mediator's motivation for doing that and (ii) how do they select their interaction partners?

**Reciprocity.** Ultimately, each actor participates in the network for a certain reason, e.g., for improving its reputation, alleviating the access to information, or becoming a key player. In other words, actors aim at improving visibility and impact in the community. Intuitively, actors favor well-known partners for collaboration and information exchange over any unknown ones (social networking concept); i.e., partners who already proved their reliable and dependable behavior. Mutual social relations are required to maximize the probability of successful future interactions. Thus, our model utilizes the concept of *reciprocity* (Figure 2(a)).

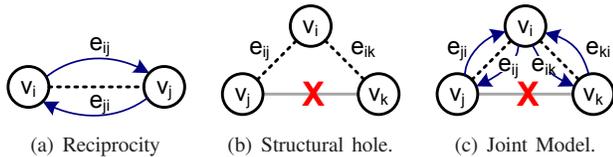


Figure 2. Social theory models.

The reciprocity theory [9] explains the motivation for reliable interaction behavior of actors with a mutual give and take relationship. The more reciprocal relations an actor has, the higher is its probability to find a reliable collaboration partner and thus, the higher is its benefit obtained from the

participation in the network. The utility  $U_R$  for an actor  $v_i$  from using existing reciprocal relations can be formulated as given in Eq. 1. Basically, this value is composed of the sum of weighted ( $w$ ) mutually strong ties; i.e., there exist equally weighted directed edges  $e_{ij}$  and  $e_{ji}$  between  $v_i$  and  $v_j$  as shown in Figure 2(a). Hence, the utility value of node  $v_i$  increases for each reciprocal relationship that it maintains. Utilized relations are determined by a query context  $Q$  to determine the utility in a *certain* expertise area.

$$U_R^Q(v_i) = \sum_{j=0}^{|V|-1} w_{ij}^Q w_{ji}^Q \quad (1)$$

Our model uses two metrics to determine link weights  $w$  (Eq. 2): (i) interest similarity *isim* [3], calculated from the similarity of interaction contexts (determined through tags), and (ii) the average mean of assigned ratings *avgr*. While *isim* is a globally valid metric, *avgr* is bound to distinct expertise areas. In particular, all activities whose description match at least one of the query keywords are considered. The function  $match(Q, a)$  determines the matching (within the range  $[0, 1]$ ) of an activity  $a$  to a query  $Q$ . Whether *isim* or *avgr* has more impact on the final link weight  $w_{ij}^Q$  can be configured through a globally valid weight  $\alpha \in [0, 1]$ . Currently we employ flat keyword-based matching only, however for more advanced ontology matching techniques see [10].

$$w_{ij}^Q = \alpha \cdot isim_{ij} + (1 - \alpha) \cdot \frac{\sum_{a \in A} avgr_{ij} \cdot match(Q, a)}{\sum_{a \in A} match(Q, a)} \quad (2)$$

**Structural Holes.** Another social theory based on self-interest is the concept of structural holes [5]. This theory states that there are actors who actively position themselves in beneficial positions within a community network. Such an actor filling a structural hole connects two previously unconnected (or at least loosely coupled) actors, and gains direct advantage by doing that. As depicted in Figure 2(b),  $v_i$  establishes an indirect link between the actually unconnected nodes  $v_j$  and  $v_k$  via edges  $e_{ij}$  and  $e_{ik}$ . In general, actors such as  $v_i$  collect larger amounts of contacts and represent the main hubs of information exchange [11]. Therefore, they are able to collect information on ongoing collaborations and can also exercise greater control on connected members. In the presented information mediation use case, the utility  $U_{SH}$  for  $v_i$  results from being able to control the communication of linked actors and collect exchanged information.

$$U_{SH}^Q(v_i) = \sum_{j=0}^{|V|-1} w_{ij}^Q \sum_{k=0}^{|V|-1} w_{ik}^Q (1 - w_{jk}^Q) \quad (3)$$

In an opportunistic network, the motivation of each actor is to maximize its utility. According to aforementioned concepts, well-known reliable actors are predominantly picked

and a mutual social relation is required to maximize the probability of successful future interactions. Each additional mediation role poses additional effort to the affected actor. However, each actor's mediation capacity is limited, thus the strategic positioning in the network is still of paramount importance.

**Joint Model.** Combining the concept of reciprocity and the structural holes theory, the overall utility value  $U$  in context of query  $Q$  is calculated as given in Eq. 4. Basically, the joint model has the same form as the structural hole model (Figure 2(c)), however, considers directed (and ideally balanced) relations between single actors. In case of social information mediation needs the utility value of each actor is evaluated. Thus, by periodically setting up new mediation activities and releasing unprofitable mediation partners (i.e, mediation targets) the utility value obtained from the network can be optimized, i.e., increased.

$$U^Q(v_i) = \sum_{j=0}^{|V|-1} w_{ij}^Q w_{ji}^Q \sum_{k=0}^{|V|-1} w_{ik}^Q w_{ki}^Q (1 - w_{jk}^Q)(1 - w_{kj}^Q) \quad (4)$$

### B. Mediation Cases

We study various cases of mediation needs as shown in Figure 3 to demonstrate the application of the proposed mediation algorithm. For those illustrative examples we neglect contextual constraints  $Q$  and thus assume that connections are established with a weight  $w = 1$ . Furthermore, assuming  $w = 1$  for the following examples better visualize the impact of certain links in the network.

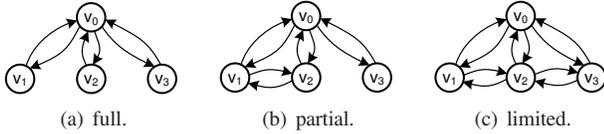


Figure 3. Various mediation needs for  $v_0$ .

In particular, we investigate the obtained utility of a certain  $v_0$  in simple networks consisting of only four nodes (Figure 3). Here,  $U(v_0)$  can be calculated as given in Eq. 5. Note, self connections are implicitly weighted with  $w = 1$ .

$$U(v_0) = \sum_{j=0}^3 w_{0j} w_{j0} \sum_{k=0}^3 w_{0k} w_{k0} (1 - w_{jk})(1 - w_{kj}) \quad (5)$$

Actually, four different cases of mediation needs exist from an algorithmic perspective (Figure 3 depicts three of them): (i) *Full Mediation* means that  $v_0$  is encouraged to mediate information between all of its neighbors because none of them are interconnected; (ii) *Partial Mediation* means that there are some mediation needs that can be fulfilled exclusively, e.g., between  $v_2$  and  $v_3$  in Figure 3(b); (iii) *Limited Mediation* is the case when no mediation is

required exclusively, because of available alternatives, e.g., mediation between  $v_1$  and  $v_3$  can be performed by  $v_0$  and  $v_2$  in Figure 3(c); (iv) *No Mediation* is the case, when there are neither exclusive nor alternative mediation needs, i.e., all neighbors of  $v_0$  are directly connected with each other.

Table I  
 $U(v_0)$  FOR DIFFERENT MEDIATION CASES.

case	$U_{j=0}$	$U_{j=1}$	$U_{j=2}$	$U_{j=3}$	$U(v_0)$
full	0	2	2	2	6
partial	0	1	1	2	4
limited	0	1	0	1	2
none	0	0	0	0	0

Table I shows the results of utility computation for node  $v_0$  in all of the mentioned cases. The first sum in Eq. 5 (with index  $j$ ) is split in single subresults shown for  $j = 0 \dots 3$ . The last column shows the utility result  $U(v_0)$ . Typically, every node that is exclusively connected to  $v_0$ , i.e., not connected to another one of  $v_0$ 's neighbors, adds a utility that is equal to  $|N_{v_0}| - 1$ , where  $|N_{v_0}|$  is the number of  $v_0$ 's neighbors in the network. In other words,  $v_0$  has the opportunity to mediate between this exclusively connected node and all other neighbors. If that exclusive nodes becomes connected to one of  $v_0$ 's neighbors, the obtained utility is reduced by 1 (see partial mediation case). Basically, supporting the emergence of connections between neighbors of  $v_0$  reduces its utility in the first place. However, allowing members to establish links may be reciprocated by affected actors, and thus, extend the number of known network members.

### C. Utility Optimization Algorithm

Each actor in the network periodically attempts to improve its current utility by replacing active collaboration partners that provide only low utility with known but currently

---

#### Algorithm 1 Utility optimization algorithm.

---

**Input:** context  $Q$ , node  $v$   
**Global:** activities  $A$ , network  $G = (V, E)$   
 $V^Q \leftarrow \text{passiveNeighbors}(v, Q)$   $\triangleright$  passive neighbors with free capacities  
 $V^A \leftarrow \text{activeNeighbors}(v, A, Q)$   $\triangleright$  actively collaborating neighbors  
 $U^Q \leftarrow \text{calcUtility}(v, V^A, Q)$   $\triangleright$  current utility in context  $Q$   
 $U_{max}^Q \leftarrow U^Q$   
 $\triangleright$  try to replace single active members with passive ones  
 $V_d^Q \leftarrow \text{determineDeLinkCandidates}(v, V^A)$   $\triangleright$  active low-utility partners  
**for each**  $v_d^Q \in V_d^Q$  **do**  $\triangleright$  for each partner that can be replaced  
  **for each**  $v_i^Q \in V^Q$  **do**  $\triangleright$  try all passive partners  
     $V_a^A \leftarrow V^A \setminus \{v_d^Q\} \cup \{v_i^Q\}$   $\triangleright$  modified temporary active set  
     $U^Q \leftarrow \text{calcUtility}(v, V_a^A, Q)$   $\triangleright$  alternative utility  
    **if**  $U^Q > U_{max}^Q$  **then**  $\triangleright$  apply changes if higher utility  
       $U_{max}^Q \leftarrow U^Q$   
       $V^Q \leftarrow V^Q \setminus \{v_d^Q\}$   $\triangleright$  remove candidate from passive set  
       $V^A \leftarrow V^A \cup \{v_i^Q\}$   $\triangleright$  make temporary set permanently active  
    **end if**  
  **end for**  
**end for**  
**end for**

---

passive network members. That procedure is described by Algorithm 1. First of all, when a node  $v$  executes this algorithm for a given context  $Q$  (e.g., expertise area), passive neighbors  $V^Q$  and active neighbors  $V^A$  are determined. Using the set  $V^A$ ,  $v$ 's current utility  $U^Q$  can be calculated. This value becomes the initial maximum value. Then a list of de-link candidates  $V_d^Q$ , e.g., actually unreliable collaboration partners or members of low utility, is determined. Now, the algorithm iterates through this set and tries to replace each single de-link candidate  $v_d^Q$  with one of the passive neighbors  $v_i^Q \in V^Q$ . If the utility value increases with this modification, the sets of active and passive neighbors are adapted accordingly. Notice, the assumption is that node  $v$  is in fact able to set up a collaboration (i.e., activity) with  $v_i^Q$  until the next update cycle. Otherwise, this replacement does not increase  $v$ 's utility and the current  $v_i^Q$  would become a de-link candidate in the next update period.

#### IV. EVALUATION AND DISCUSSION

In this section we outline the application of our approach and derive findings by studying various simulations.

##### A. Experiment Background

Resilience (robustness) of complex networks can be studied by analyzing the effects of node and edge removal. Social networks, as studied here, are one class of a complex networks. Other examples include the WWW, power grids, or biological networks. *Percolation analysis* [12] can be used in complex networks to understand the effects of node/edge removal; for example, to study the attack resilience of networks [13]. Here we take this approach to analyze the *utility evaluation* of mediators in social network under different conditions. The question we attempt to answer is how a mediator's utility is influenced by decreasing connectivity of the network. Based on gradually removed edges, we attempt to understand the effects of our utility metric. We expect that the utility of a mediator is particularly high if a network is not well-connected (see metrics such as average path length, density). Our approach is as follows:

- Take the initial network and calculate  $v$ 's utility.
- Remove edges between nodes that are *highly similar*, thereby introducing potential 'gaps' and mediation opportunities. Repeat the calculation of a node's (mediator) utility.
- Remove edges between nodes that are *less similar* to study the impact of profile similarity (mediation between similar nodes versus mediation between nodes with different interests).

##### B. Discussion and Findings

**Bootstrapping Collaboration Networks.** First, we outline a bootstrapping mechanism for social and collaborative networks. Our approach heavily relies on monitored interactions and collected data. For instance, the intensity, i.e.,

weight, of social relations is determined through various metrics – see Eq. 2. In particular we utilize *interest similarity* ( $isim$ ) to connect people with similar expertise (derived from manually defined profiles and tagging data), and *average rating* ( $avgr$ ) for evidence-based relations based on previous collaborations. The parameter  $\alpha$  controls which one of the parameters receive higher attention and influences the final weight more (Eq. 2). We distinguish between three cases: (i) *Bootstrapping Phase* ( $\alpha = 1$ ): here, only  $isim$  is used which describes a certain interest overlap between two people's profiles. Thus, potentially all nodes are connected to all others, independent from whether they collaborated in the past. Relations described by  $isim$  only, can be seen as a prediction of successful future interactions. (ii) *Formation Phase* ( $\alpha \in ]0, 1[$ ): here, we use a mix of  $isim$  and collected  $avgr$  to determine the weight. In other words, as members begin to interact, predictions based on  $isim$  lose importance while assigned ratings due to reliable interaction behavior mainly influence a link's weight. (iii) *Saturation Phase* ( $\alpha = 0$ ): after a while, rating-based structures emerge, which means that only  $avgr$  is used to determine a link's weight. However, these structures are in a constant flux and change due to periodic re-evaluation of personal utilities.

We use above mechanisms to create a synthetic collaboration network  $G = (V, E)$  with  $|V| = 100$  for our further studies. In the beginning, we assume that all nodes are connected to all others described by similarity relations ( $isim$  uniformly distributed in  $[0, 1]$ ) resulting in  $|E| = \frac{1}{2} \cdot |V| \cdot (|V| - 1) = 499500$ . Then, we utilize the *preferential attachment model* of Barabasi and Albert [14] to let rating-based graphs emerge with power-law distributed degrees. These structures are the basis for a realistic scenario and edge weight distribution among members. As typical for scale-free collaboration networks, we assume that 80% of interactions take place between 20% of the most active users.

**Simple Percolation Analysis.** Figure 4(a) depicts the aggregated utility (the percentaged amount of the maximum possible utility) of all nodes in the network when gradually decreasing the number of edges in the network. We start with a fully saturated graph (amount of edges is 100%). Two experiments show the difference between (i) binary relations only, i.e. weight =  $\{0|1\}$ , and (ii) weighted relations with

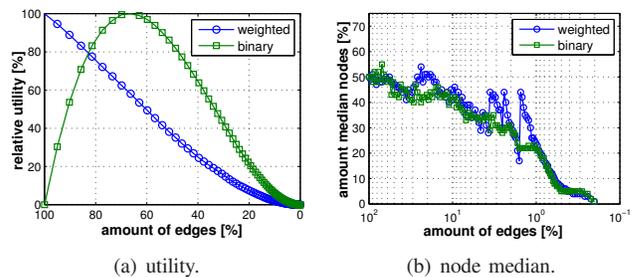


Figure 4. Utility and number of median nodes for varying edge saturations and weighting models (binary v.s. weighted edges).

weight =  $[0, 1]$ . In the binary case, the utility rises when edges are removed, because at the same time mediation opportunities emerge. At an amount of  $\approx 65\%$  the received utility is at a maximum. In the weighted case, the utility decreases strictly monotonic from the beginning, since there are always mediation opportunities between neighboring nodes, even if they are directly connected (except the case where mediation targets are directly connected with an edge weight = 1). Figure 4(b) demonstrates the amount of nodes that receive a utility above the median. Both cases produce similar results, i.e., removing 90% of edges still results in a fair distribution with 45% to 50% of nodes having an utility above the median.

We expect that the weights of relations have major influence on the received utility of a node. Thus, we apply percolation strategies where (i) lowly weighted edges are removed first (*LowToHigh*), and (ii) highly weighted edges are removed first (*HighToLow*). Figure 5 shows again the aggregated utility of nodes and the amount of nodes receiving utility above the median value. Obviously lowly weighted edges have less influence on the aggregated utility (utility drops slower) compared to highly weighted edges (utility drops faster). This is of particular interest, because one might assume that removing highly weighted edges first discovers new and better mediation opportunities. In fact, this is the case but also highly weighted connections to mediation targets are removed at the same time which results in lower average utility. The importance of weak links (even compared to strong ones) have been extensively studied by [15].

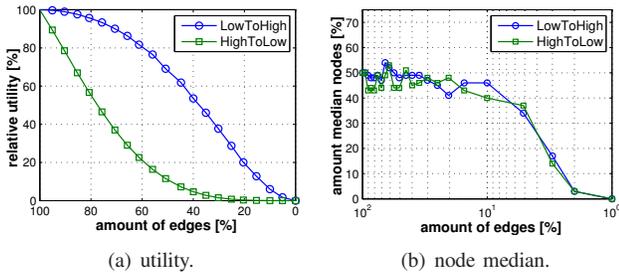


Figure 5. Utility and number of median nodes for varying edge saturations and different percolation strategies (LowToHigh v.s. HighToLow).

**Bounded Percolation Analysis.** Until now, we neglected the costs for maintaining relations. But in reality, keeping relations active results in high effort for serving requests and mediating information. Thus, usually only a small amount of all known partners are involved in active collaborations at the same time. We define that a node  $v$  has only a limited amount of resources to cope with *costs* for serving partners. We further define the *benefit-cost factor*  $\zeta(v) = \frac{U(v)}{costs(v)}$  that is the basis to decide whether node  $v$  should maintain or release a partner. In other words  $\zeta$  is the basis to determine how much effort  $v$  should invest and thus its re-linking behavior.

Typically, actors can cope with a predefined cost level, i.e., they have a limited amount of credits or resources (e.g., time) which they can invest. Thus, we argue that in most cases they attempt to keep  $\zeta \rightarrow max$ . Figure 6 visualizes the relative utility (amount from the highest possible utility when neglecting occurring costs) in a gradually disconnecting graph (decreasing amount of edges). For  $\zeta \rightarrow max$ , this experiment applies a cost function that allows a single member to collaborate with a maximum of 5 neighbors at the same time. Notice, in the beginning utility typically rises with decreasing amounts of edges when partners become served exclusively. Furthermore, only a limited number of neighbors can be served when accounting for costs. So, this curve has a shape similar to the binary weighted model in Figure 4(a) (binary edge weighting).

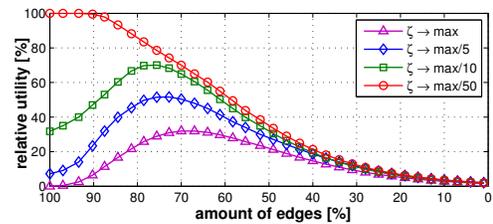


Figure 6. Utility for varying benefit-cost factors  $\zeta$ .

However, in certain cases actors may try to obtain a utility that is higher than in the case  $\zeta \rightarrow max$ . That means, they decide to invest comparatively more effort in order to obtain (even only slightly) higher utilities. Figure 6 shows the results for  $\zeta \rightarrow \{max, \frac{max}{5}, \frac{max}{10}, \frac{max}{50}\}$ . Note, we simplify the model and assume  $min \approx \frac{max}{50}$ ; i.e., the worst (lowest)  $\zeta$  factor means that costs are (nearly) neglected and each user is able to invest 50 times more resources than in the optimal case. As expected, the more effort is invested to maintain relations, the higher utility can be obtained. However, the amount of additional utility is comparatively low. For instance, if the network is still connected with 65% of the initial edges, then  $U_{\zeta \rightarrow max} \approx 0.5 \cdot U_{\zeta \rightarrow max/50}$ , i.e., the utility value can be doubled by coping with approximately 50 times higher costs.

**Utility Distribution.** We take the data from the previous experiments and study the case for an amount of edges of 65% (maximum utility for  $\zeta \rightarrow max$ ) in greater detail. Until now, we just studied the aggregated utility in the whole network (which possesses similar characteristics as the *average* utility of a single user depending on the amount of edges in the network). Here, we group members with respect to their obtained utility and highlight the distribution of actors in the single classes. Figure 7(b) depicts the relative amount of members from the whole population in distinct utility classes (reflecting users obtaining more than 10,20,30... percent of the maximum possible utility value).

Figure 7 shows the results. In particular, when (nearly) neglecting costs ( $\zeta \rightarrow \frac{max}{50}$ ) the distribution spans most

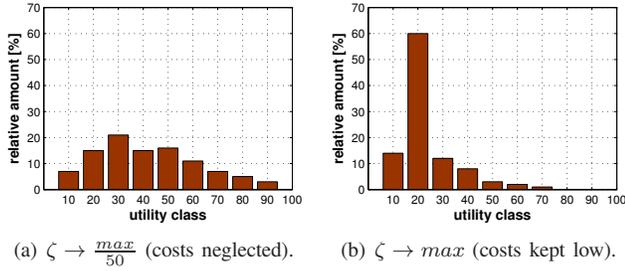


Figure 7. Node distribution in different utility classes.

utility classes. In other words, there are approximately as many actors who receive 20% of the maximum value as members with 50%. That distribution seems to be unfair. However, when accounting for costs, actors are not able to claim as many neighbors as possible (based on their structural connectivity) but need to carefully select their collaboration partners. So many partners are served exclusively and most actors are able to find a small amount of valuable partners. Here, approximately 60% of members are located in the 20% utility class. This demonstrates that applying the benefit-cost factor  $\zeta$  seems to be a feasible concept to ‘balance’ networks, i.e., to distribute gained utility in a fair manner among community members.

## V. SOA IMPLEMENTATION AND APPLICATION

We implemented the proposed model in a SOA-based collaboration environment (see [3], [16] for the big picture) to study the integration with existing standards, mainly FOAF, and its application in end-user environments.

### A. SOA-based Collaboration Environment

Figure 8 shows the basic building blocks located on three layers of different abstraction.

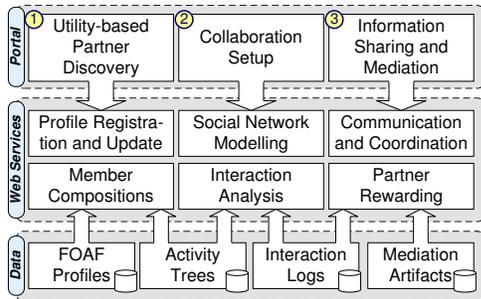


Figure 8. SOA-based Collaboration architecture.

**Portal Layer.** Here, we developed tools for (1) *Partner Discovery* using the approach of this paper; (2) *Collaboration Setup*, including the creation, update and deletion of activity structures [17] that describe the context of interactions; and (3) for *Information Sharing and Mediation*, i.e., to forward information selectively between mediation targets.

**Web Services Layer.** Services implement various functionalities in the back-end to enable the discovery and compositions of partners based on FOAF profiles, the creation

of social networks using monitored and fetched data (as explained in Section II), and communication and coordination facilities to exchange messages and files.

**Data Layer.** The bottom layer manages distributed data, including FOAF profiles, activities, SOAP logs, and artifacts (i.e., documents – see [16]) that are mediated between community members.

### B. Social Network Management with FOAF

Various concepts and protocols have been proposed to manage open social and collaborative networks in a distributed manner. The Friend-Of-A-Friend (FOAF) concept is one of the most popular ones on the Web. It allows to model user properties, interests and relations with a well-known ontology. We apply FOAF to facilitate the discovery process used to find potential collaboration partners and let users mediate information (modeled as documents) by linking them to their profiles. Listing 1 shows a simplified example of a public FOAF profile containing name, interests and knows-relations to other community members.

```

<?xml version="1.0"?>
<rdf:RDF xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:wot="http://xmlns.com/wot/0.1/"
  >
  <foaf:Person rdf:ID="me">
    <foaf:name>Florian Skopik</foaf:name>
    <foaf:nick>florian</foaf:nick>
    <foaf:mbox_sha1sum>a4b378...</foaf:mbox_sha1sum>
    <wot:haskey rdf:nodeID="KeyFS" />
    <foaf:interest rdf:resource="http://..." />
    <foaf:account>
      <foaf:OnlineAccount>
        <foaf:accountServiceHomepage rdf:resource="http://..." />
        <foaf:accountName>florian_skopik</foaf:accountName>
      </foaf:OnlineAccount>
    </foaf:account>
    <foaf:knows>
      <foaf:Person>
        <foaf:mbox_sha1sum>1a4578...</foaf:mbox_sha1sum>
        <foaf:name>Daniel Schall</foaf:name>
      </foaf:Person>
    </foaf:knows>
  </foaf:Person>
</rdf:RDF>

```

Listing 1. Example of public FOAF file.

Due to space limitations we highlight the usage of extended FOAF tags and certain add-ons that are used besides the basic profiles in our system:

- *External Document Links* (Document tags) are used to attach mediated information to personal profiles. For instance, one can attach a link to a document that he received from collaboration partners to his profile after reviewing and removing sensitive content. Friends can then retrieve the FOAF profile containing links to this information.
- *Public Key Infrastructures* are embedded in FOAF with the *Web Of Trust* ontology<sup>2</sup> to allow particular community members access to (parts of) one’s FOAF profile

<sup>2</sup><http://xmlns.com/wot/0.1/>

[18]. This way, linked documents can be encrypted for specific mediation targets only.

- *Enrichment with Collected Interaction Data*, such as manual ratings (tipjar tag) of users, similarity of defined interest tags, and weighting of know-relations based on interaction data provide more semantic meaning to control the mediation process (i.e., a kind of *information sharing* [16]).

## VI. RELATED WORK

Nowadays, virtual organizations [4] are more and more realized with SOA concepts, regarding service discovery, service descriptions (WSDL), dynamic binding, and SOAP-based interactions. A particular issue in such environments is the mediation and selective dissemination of information (SDI) [16]. The availability of rich and plentiful data on human interactions in social networks has closed an important loop [19], allowing one to model social phenomena and to use these models in the design of new computing applications such as crowdsourcing techniques [20].

Game-theoretic models [21] allow to explain the behavior of single actors. In coalitional games, people attempt to find collaboration partners to increase their benefits gained from the network. Depending on the environment, social relations are established based on the outcome of previous interactions [3] and interest similarity [22]. In our approach, various metrics express social behavior influenced by the context in which collaborations take place [3]. For instance, *reciprocity* [9] is a concept describing that humans tend to establish a balance between provided support and obtained benefit from collaboration partners. A further important concept in group formation is the *structural holes theory* [5]. This theory states that actors actively attempt to position themselves in beneficial positions within a community network. Many works have considered implications of this theory, for instance [23] in economics, and also mapped to Web environments [11].

## VII. CONCLUSION AND FUTURE WORK

In this paper we introduced an opportunistic approach that explains the formation process in open collaborative networks on the Web. We discussed two well-known theories, reciprocity and structural holes, and introduced a utility-based model. Several simulations showed the application of this model and its basic properties. Furthermore, we highlighted a technical point of view and the implementation and embedding of this model in a SOA-based collaboration environment.

Our future work includes the deployment of our implemented prototype in an end-user business environment to study its performance under realistic conditions. For that purpose we collaborate with large-scale enterprises in course of the EU project COIN.

## ACKNOWLEDGEMENTS

This work was supported by the European Union FP7 projects COIN (216256) and SM4ALL (224332).

## REFERENCES

- [1] Amazon.com, "Amazon mechanical turk," last access: 2010, available online: <http://www.mturk.com>.
- [2] J. Guare, *Six Degrees of Separation: A Play*. Vintage Books, 1990.
- [3] F. Skopik, D. Schall, and S. Dustdar, "Modeling and mining of dynamic trust in complex service-oriented systems," *Inf. Syst.*, vol. 35, pp. 735–757, 2010.
- [4] L. M. Camarinha-Matos and H. Afsarmanesh, "Collaborative networks," in *PROLAMAT*, 2006, pp. 26–40.
- [5] R. S. Burt, *Structural holes: The social structure of competition*. Harvard University Press, 1992.
- [6] D. Schall and S. Dustdar, "Dynamic context-sensitive pagerank for expertise mining," in *SocInfo*, 2010.
- [7] D. Schall and F. Skopik, "Mining and composition of emergent collectives in mixed service-oriented systems," in *CEC*. IEEE, 2010.
- [8] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust," in *WWW*, 2004, pp. 403–412.
- [9] A. Falk and U. Fischbacher, "A theory of reciprocity," *Games and Economic Behavior*, vol. 54, no. 2, pp. 293–315, 2006.
- [10] S. Castano, A. Ferrara, and S. Montanelli, "Matching ontologies in open networked systems: Techniques and applications," *J. Data Semantics V*, pp. 25–63, 2006.
- [11] J. M. Kleinberg, S. Suri, va Tardos, and T. Wexler, "Strategic network formation with structural holes," *SIGecom Exchanges*, vol. 7, no. 3, 2008.
- [12] D. Stauffer and A. Aharony, *Introduction to percolation theory*. CRC Press, 1994.
- [13] R. Albert, H. Jeong, and A.-L. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, July 2000.
- [14] A. Reka and Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, pp. 47–97, June 2002.
- [15] M. S. Granovetter, "The strength of weak ties," *The American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, May 1973.
- [16] F. Skopik, D. Schall, and S. Dustdar, "Trust-based adaptation in complex service-oriented systems," in *ICECCS*, 2010, pp. 31–40.
- [17] D. Schall, H.-L. Truong, and S. Dustdar, "Unifying human and software services in web-scale collaborations," *IEEE Internet Computing*, vol. 12, no. 3, pp. 62–68, 2008.
- [18] S. R. Kruk, S. Grzonkowski, A. Gzella, T. Woroniecki, and H.-C. Choi, "D-foaf: Distributed identity management with access rights delegation," in *ASWC*, 2006, pp. 140–154.
- [19] J. Kleinberg, "The convergence of social and technological networks," *Commun. ACM*, vol. 51, no. 11, pp. 66–72, 2008.
- [20] D. Brabham, "Crowdsourcing as a model for problem solving: An introduction and cases," *Convergence*, vol. 14, no. 1, p. 75, 2008.
- [21] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. MIT Press, 1999.
- [22] J. Golbeck, "Trust and nuanced profile similarity in online social networks," *TWEB*, vol. 3, no. 4, 2009.
- [23] S. Goyala and F. Vega-Redondo, "Structural holes in social networks," *Journal of Economic Theory*, vol. 137, no. 1, pp. 460–492, November 2007.