Journal of Universal Computer Science, vol. 14, no. 7 (2008), 1031-1047
submitted: 1/10/07, accepted: 21/1/08, appeared: 1/4/08 © *J.UCS*

Query Transformation Based on Semantic Centrality in Semantic Social Network[1]

Jason J. Jung
(Yeungnam University, Korea
j2jung@intelligent.pe.kr)

Abstract: Query transformation is a serious hurdle on semantic peer-to-peer environment. For interoperability between peers, queries sent from a source peer have to be efficiently transformed to be understandable to potential peers processing the queries. However, the problem is that the transformed queries might lose some information from the original one, as continuously traveling along peer-to-peer networks. We mainly consider two factors; i) number of transformations and ii) quality of ontology alignment. In this paper, we propose a new measurement of semantic centrality, i.e., the power of semantic bridging on semantic peer-to-peer environment. Thereby, we want to build semantically cohesive user subgroups, so that semantic affinities between peers can be computed. Then, given a query, we find out a path of peers for optimal interoperability between a source peer and a target one, i.e., minimizing information loss by the transformation. We have shown an example for retrieving image resources annotated on peer-to-peer environment by using query transformation based on semantic centrality.
Key Words: Semantic social network, Ontology alignment, Query propagation
Category: H.1.1, H.3.5, I.2.11

1 Introduction

In order to efficiently provide semantic collaboration and interoperability between people on peer-to-peer (p2p) network, each peer usually expresses a certain query with the semantic information from his own ontologies, and send it to other peers. However, semantic heterogeneity is a problem causing to misunderstand the queries and retrieve wrong answer set. Additionally, as the number of users and ontologies are dramatically increasing, the structure of semantic p2p networks are getting more complex.

More importantly, information retrieval process on the p2p networks has been performed by propagating a certain message containing a certain queries to neighbor peers and their neighbors. We assume that the queries for interactions between peers (from source peer to destination peer) are simply represented as a set of concepts derived from the ontology of source peer. For high accessibility, the queries have to be transformed into as many concepts of destination peer ontology as possible. The concepts in the original query can be replaced to the correspondent concepts resulting from ontology alignment between peer ontologies. More importantly, we propose a novel measurement of semantic centrality, which expresses the power of controlling *semantic* information on semantic p2p network, and show that it is applied to search for the most proper peers for concept-based query transformation [Jung 2007;Jung 2008].

[1] This research was supported by the Yeungnam University research grants in 2007.

Thereby, in this study, we introduce a three-layered structure [Jung and Euzenat 2007] made of superposed networks that are assumed to be strongly linked:

Social layer relating peers (or people) on the basis of common interest;

Ontology layer relating ontologies on the basis of explicit import relationships or implicit similarity;

Concept layer relating concepts on the basis of explicit ontological relationships or implicit similarity.

We may call this stack of interlinked networks a *semantic social network* (SSN). On semantic p2p network, users (or actors) and their own peer (or personal) ontologies are located in social and ontology layer, respectively. Implicit relationships between concepts defined in the ontologies (e.g., aligned correspondences) are contained in concept layer.

More importantly, in the three-layered model, we design to propagate the relational information (e.g., the distance or similarity) not only within a layer but also between layers. We have provided the principles for extracting similarity between concepts in different ontologies and propagating this similarity to a distance and an alignment relation between ontologies. We compute semantic affinities between peers, so that the semantic subgroups can be discovered. By using topological features of the discovered subgroups, two centrality measurements (e.g., local and global centralities) can be obtained. These centralities are applied to determine the best path on which the queries can travel in p2p network.

Especially, we need to discover the consensual ontology $\mathcal{CO}$ containing the most common substructures among peer ontologies. In fact, social network analysis (SNA) has regarded a consensus implying the central principles underlying the network as an important challenge [Wasserman and Faust 1994]. With respect to semantic interoperability between heterogeneous information sources, consensual ontology is playing a role of a "semantic pivot" between heterogeneous information sources [Stephens et al. 2004].

1.1 Network analysis

Now, we want to formalize the network structure on SSNs, and explain basic social measurement from SSNs. Generally, the networks will be characterized here as a set of objects (or nodes) and a set of relations.

Definition 1 (Network). A network $\langle N, E^1, \ldots E^n \rangle$ is made of a set N of nodes and n sets of object pairs $E^i \subseteq N \times N$ the set of relations between these nodes.

These networks can express the relationships between people or many other sort of items. As any graphs, the relations can also be represented by an adjacency matrix M

of which size is $|N| \times |N|$. Because the relation is directed, this matrix is asymmetric. Each element is given by

$$M^i_{e,e'} = \begin{cases} 1 \text{ if } e \text{ links to } e'; \\ 0 \text{ otherwise.} \end{cases}$$

As usual, a path p between node e and e' is a sequence of edges $\langle e_0, e_1 \rangle$, $\langle e_1, e_2 \rangle$, $\ldots$, $\langle e_{k-1}, e_k \rangle$ in which $e_0 = e$ and $e_k = e'$. The length of a path is its number of edges (here k) and the shortest path distance $spd(e, e')$ between two nodes e and e' is the length of the shortest path between them. By convention, $spd(e, e) = 0$.

Definition 2 (Distance network). A distance network $\langle N, E^1, \ldots, E^n \rangle$ is made of a set N of nodes and n sets of distance functions $E^i : N \times N \longrightarrow [0\ 1]$ defining the distance between nodes (so satisfying symmetry, positiveness, minimality, and triangular inequality).

It is clear that any network is a weighted network which attributes either 0 or 1 as a weight. The definition of social network analysis can be adapted to distance networks if each time the cardinality of a set of edges if used, it is replaced by the sum of its distances. The distance of a path is obtained by summing the distances of its edges.

1.2 Outline of this paper

The remainder of the paper is organized as follows. Sect. 2 introduces a novel platform for semantic social space. Then, Sect. 3 addresses inference of relationships between three layers. In Sect. 4, we will explain concept-based query transformation mechanism on our proposed system. Most importantly, Sect. 4.1 addresses the semantic centrality measurement from a given semantic social network with a simple example. Finally, Sect. 6 will discuss some issues and draw a conclusion.

2 Three-layered Semantic Social Space

We have proposed the three-layered architecture for constructing the socialized semantic infosphere, for uncovering potential links between unknown people from those that can be found from their knowledge [Jung and Euzenat 2007]. It consists of i) a social layer ($\mathcal{S}$), ii) an ontology layer ($\mathcal{O}$), and iii) a concept layer ($\mathcal{C}$), as shown in Fig. 1.

The characteristics of each layer and the relationships between layers are described below.

2.1 Social layer

In the social layer ($\mathcal{S}$), nodes are representing people, and relations are the connections between peoples. A social network $\mathcal{S}$ is a directed graph $\langle N_S, E_S^{knows} \rangle$, where N_S is

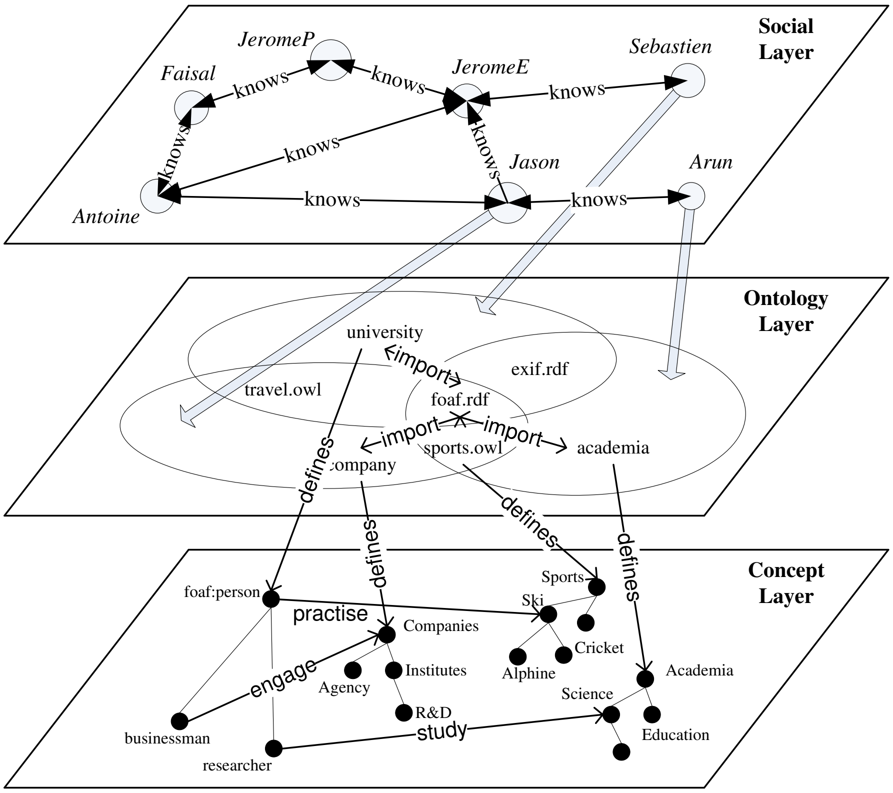

Figure 1: Three-layered semantic social network

a set of person and $E_S^{knows} \subseteq N_S \times N_S$ the set of relations between these persons. In most current applications, the relation used by social network analysis is the knows relation that can be found in FOAF.

Social network analysis [Wasserman and Faust 1994] has considered various measures on the networks between people (note that these measures apply only if the network is connected)[2]:

Closeness The inverse of average length of the shortest path between a node e and any other node in the network:

$$Closeness^i(e) = \frac{|N-1|}{\sum_{e' \in N} spd^i(e,e')} \quad (1)$$

Betweenness [Freeman 1979] The proportion of shortest paths between two nodes

[2] These measures are often normalized (between 0 and 1) but we present their simplest form.

Table 1: Closeness, authoritative and hub weights in Fig. 1

Weights	Closeness	Authoritative	Hub
$Arun$(AS)	0.5	0.21	0.01
$Antoine$(AZ)	**0.67**	0.45	**0.52**
$Faisal$(FAK)	0.5	0.37	0.27
$JeromeE$(JE)	0.6	**0.69**	0.32
$Jason$(JJ)	**0.67**	0.243	**0.54**
$JeromeP$(JP)	0.46	0.236	0.42
$Sebastien$(SL)	0.4	0.13	0.275

which contains a particular node (this measures the power of this node):

$$Betweenness^i(e) = \sum_{e',e''\in N} \frac{|\{p \in sp^i(e',e), p' \in sp^i(e,e'')|p \cdot p' \in sp^i(e',e'')\}|}{|sp^i(e',e'')|} \quad (2)$$

Hub and authority There are different but interrelated patterns of power: Authorities that are referred to by many and hubs that refers to many. The highest authorities are those which are referred to by the highest hubs and the highest hubs that those which refers to the highest authorities. Kleinberg [Kleinberg 1999] proposes an iterative algorithm to measure authority and hub degree of each node in interlinked environment. Given initial authority and hub degrees of 1, the degrees are iteratively computed by

$$Hub^i_{t+1}(e) = \sum_{e':\langle e,e'\rangle\in E^i} Auth^i_t(e') \text{ and } Auth^i_{t+1}(e) = \sum_{e':\langle e',e\rangle\in E^i} Hub^i_t(e') \quad (3)$$

Similarly to betweenness, the hub weight indicates the structural position of the corresponding user. It is a measure of the influence that people have over the spread of information through the network.

From the social network in Fig. 1, three social weight measurements of users, e.g., closeness, authoritative and hub, are shown in Table 1. Obviously, the highest hub weight is assigned to $Jason$, because he has an important and unavoidable role of bridging between the other users. As the hub weight [Kleinberg 1999] of a certain user is higher, he has more important and unavoidable role of bridging between the other users.

2.2 Ontology layer

The ontology network $\mathcal{O}$ is a network $\langle N_O, E^i_O\rangle$, in which N_O is a set of ontologies and $E^i_O \subseteq N_O \times N_O$ the relationships between these ontologies. There can be two main kind of relations at this stage:

- **import** when an ontology explicitly imports another ontology;
- **refer** when an ontology uses some concept defined in another ontology.

The objective relationship from $\mathcal{S}$ to $\mathcal{O}$ is through the explicit usage of ontology by a user which can be expressed by a relation: $Use \subseteq N_S \times N_O$. We can easily interpret the hubs as being the ontologies that combine a large number of other ontologies. These would be an interesting starting point for any newcomer willing to annotate a similar set of objects as his friend. Similarly, authority will be ontologies understood as *de facto* standards that are extended and imported by many different actors.

There is a difference between ontology networks and social networks though: while in social networks it is normal to be connected to several authorities, an ontology will only import one ontology on some topic. It would thus be useful to recognize those hubs that connects authorities on the same topics, these "ontologies" are likely to be the expression of an alignment between the two authorities.

2.3 Concept layer

In the concept layer ($\mathcal{C}$), nodes are concepts, and links are the numerous kinds of links that can be found in ontologies. The concept network $\mathcal{C}$ is a network $\langle N_C, E_C^i \rangle$, in which N_C is a set of entity of an ontology (classes, properties, individuals) and $E_C^i \subseteq N_C \times N_C$ the relationships between these entities. The concept relationships are far more numerous and depends on the kind of entity considered. If we restrict our attention to classes, they are the following:

- **subClass** linking a class to its subclasses;
- **superClass** (=subClass^{-1}) linking a class to its super classes;
- **property** (=domain^{-1}) linking a class to its properties;
- **instance** linking a class to its instances.

The objective relationship from $\mathcal{O}$ to $\mathcal{C}$ is through the definition of concept in an ontology which can be expressed by a relation: $Defines \subseteq N_O \times N_C$. However, this notion of definition is not easy to catch: it could be based on either the assertion of a constraint on some ontology entity or the namespace in which entity belongs. We will consider the namespace in the following.

We are here further away from social networking. As noted in [Alani and Brewster 2005], the notions of hub and authority cannot be understood in the same way for all the relations expressed in $\mathcal{C}$ (for instance, the root of concept hierarchies will certainly be hubs).

3 Inferring relationships

This three-level semantic social network does not bring in itself new improvement for our peer-to-peer sharing application. In order to provide new insight in the possible collaborations it is necessary to analyze these networks and to propagate information from one layer to another. We explain how, starting from the lower concept layer, it is possible to enrich the upper ontology and social layers with new relations from which social network analysis helps finding relevant peers.

3.1 Similarity on the concept layer

Beside the numerous relationships that can be found by construction of the concept layer, new relationships can be inferred between the entities. One particularly interesting relationship is similarity: in order to find relationship between concepts from different ontologies, identifying the entities denoting the same concept is a very important feature. As a matter of fact, most of the matching algorithms use some similarity measure or distance in order to match entities.

Some distances can be established from the local features of entities. For instance, the name of entities can be the basis for matching them. Many techniques have been developed for comparing strings [Shvaiko and Euzenat 2005], based on their structures (like edit distance), their morphology (through lemmatization), their entry in lexicons (using WordNet). Another kind of similarity can be established based on set of shared instances like in [Mika 2005].

Some other distances, more in the spirit of network analysis, can be defined from the structure of the network. For instance, [Euzenat and Valtchev 2004], defines similarities (e.g., Sim_C, Sim_R, Sim_A) between classes, relationships, attributes, and instances. It is based on the principle that the more features of two entities are similar, the more these entities are similar. Given a pair of classes from two different ontologies, the similarity measure Sim_C is assigned in $[0, 1]$. The similarity (Sim_C) between c and c' is defined as

$$Sim_C(c, c') = \sum_{E \in \mathcal{N}(C)} \pi_E^C MSim_Y(E(c), E(c')) \tag{4}$$

where $\mathcal{N}(C) \subseteq \{E^1 \ldots E^n\}$ is the set of all relationships in which classes participate (for instance, subclass, instances, or attributes). The weights π_E^C are normalized (i.e., $\sum_{E \in \mathcal{N}(C)} \pi_E^C = 1$).

If we restrict ourselves to class labels (L) and three relationships in $\mathcal{N}(C)$, which are the superclass (E^{sup}), the subclass (E^{sub}) and the sibling class (E^{sib}), Equ. 4 is rewritten as:

$$\begin{aligned} Sim_C(c,c') = \pi_L^C sim_L(L(A_i), LF(B_j)) \\ + \pi_{sup}^C MSim_C(E^{sup}(c), E^{sup}(c')) \\ + \pi_{sub}^C MSim_C(E^{sub}(c), E^{sub}(c')) \\ + \pi_{sib}^C MSim_C(E^{sib}(c), E^{sib}(c')). \end{aligned} \tag{5}$$

where the set functions $MSim_C$ compute the similarity of two entity collections. As a matter of fact, a distance between two sets of classes can be established by finding a maximal matching maximizing the summed similarity between the classes:

$$MSim_C(S,S') = \frac{\max(\sum_{\langle c,c'\rangle \in Pairing(S,S')} (Sim_C(c,c'))}{\max(|S|,|S'|)}, \tag{6}$$

in which $Pairing$ provides a matching of the two set of classes. Methods like the Hungarian method allow to find directly the pairing which maximizes similarity. The OLA algorithm is an iterative algorithm that compute this similarity [Euzenat and Valtchev 2004]. This measure is normalized because if Sim_C is normalized, the divisor is always greater or equal to the dividend.

A normalized similarity measure can be turned into a distance measure by taking its complement to 1 ($E_C^{dist}(x,y) = 1 - Sim_C(x,y)$). Such a distance introduces a new relation E_C^{dist} in the concept network $\mathcal{C}$. This relation in fact defines a distance network as introduced above.

3.2 From concept similarity to ontology similarity

Once such a distance has been introduced at the concept level, it can be used for computing a new distance at the ontology level. Again, a distance between two ontologies can be established by finding a maximal matching maximizing similarity between the elements of this ontology and computing a global measure which can be further normalized:

Definition 3 (Ontology distance). Given a set of ontologies N_O, a set of entities N_C provided with a distance function $E_C^{dist} : N_C \times N_C \longrightarrow [0\ 1]$ and a relation $Defines : N_O \times N_C$, the distance function $E_O^{dist} : N_O \times N_O \longrightarrow [0\ 1]$ is defined as:

$$E_O^{dist}(o,o') = \frac{max(\sum_{\langle c,c'\rangle \in Pairing(Defines(o), Defines(o'))} E_C^{dist}(c,c'))}{max(|Defines(o)|, |Defines(o')|}$$

The resulting measure is minimal ($\forall o \in N_O, E_O^{dist}(o,o) = 0$), but is not guarantee to be a distance unless we apply a closure with the triangular inequality. Figure 1, provides an example of the distance computation between ontology "university" and "academia". The concept layer shows the bests match between the tow ontologies, which is used for computing the distance of 0.2 between these ontologies.

This is the measure that is used in the OLA algorithm for deciding which alignment is available between two ontologies [Euzenat and Valtchev 2004]. However, other distances can be used such as the well known single, average and multiple linkage distances.

This ontology distance introduces a new relation on the ontology layer which provides a good idea of the distances between ontologies. It is, in turn, a clue of the difficulty to find an alignment between ontologies. It can be used for choosing to match the closest ontologies with regard to this distance. This can help a newcomer in a community to choose the best contact point: the one with whom ease of understanding will be maximized. This will be further developed in Section 3.4.

3.3 From concept similarity to alignment

It can however happen that people have similar but different ontologies. In order for them to exchange their annotations, they need to know the alignments existing within the ontology network. As the result of applying alignment algorithms, the similarity or distance on the network is the basis for many matching algorithms [Euzenat and Valtchev 2004]. Manually extracted alignments can also be added to this relation.

As a result, from concept similarity these algorithms will define a new relation E^{align} at the ontology level.

Definition 4 (Alignment relation). Given a set of ontologies N_O, a set of entities N_C provided with a relation $E_C^{dist} : N_C \times N_C$, and a matching algorithm $Match$ based on E_C^{dist}, the alignment relation $E^{align} \subseteq N_O \times N_O$ is defined as:

$$\langle o, o' \rangle \in E^{align} \text{ iff } Match(o, o') \neq \emptyset$$

If one has a measure of the difficulty to use an alignment or of its quality, this network can also be turned into a distance network on which all these measures can be performed. Of course, when an alignment exists between all the ontologies used by two peers, there is at least some chance that they can talk to each others. This can be further used in the social network.

This new relation in the ontology layer allows a new agents to choose the ontology that it will align with first. Indeed, the ontologies with maximal hub centrality and closeness for the alignment network are those for which the benefit to align to will be the highest because they are aligned with more ontologies. In the peer-to-peer sharing application, choosing such an ontology will bring the maximum answers to queries. For example, in the concept layer of Fig. 1, two alignments between $i)$ $university$ and $company$ and $ii)$ $company$ and $academia$ would make it possible for *Arun* and *Sebastien* to share information by using these existing alignments, even though they are not explicitly linked with each other.

This is the occasion to note the difference between the relations in the same network: in the ontology network, the hub ontologies for the import relation are rather complete ontologies that cover many aspects of the domains, while hub ontologies for the E^{align} relation are those which will offer access to more answers.

3.4 From ontology similarity to people affinity

Once these measures on ontologies are obtained, this distance can be further used on the social layer. As we proposed it is possible to think that people using the same ontologies should be close to each other. The affinity between people can be measured from the similarity between the ontology they use.

Definition 5 (Affinity). Given a set of people N_S, a set of ontologies N_O provided with a distance $E_O^{dist} : N_O \times N_O \longrightarrow [0\ 1]$ and a relation $Uses : N_S \times N_O$, the affinity is the similarity measure defined as

$$E^{aff}(p, p') = 1 - \frac{max\left(\sum_{\langle o,o'\rangle \in Pairing(Use(p),Use(p'))} 1 - E_O^{dist}(o, o')\right)}{max(|Use(p)|, |Use(p')|)} \tag{7}$$

Since this measure is normalized, it can be again converted to a distance measure through complementation to 1.

Introducing the distance corresponding to affinity in the social network allows to compute the affinity relationships between people with regard to their knowledge structure. Bottom-up inference from $\mathcal{C}$ allows to find out the semantic relationships between users based on this space. As shown in Fig. 1, based on the found similarity (or low distance) between "academia" and "industry" ontologies, *Sebastien* and *Arun* who do not know each others can know meet on the social network. Obviously, this is useful since matching these similar ontologies should be easy.

4 Transformation Path Selection

Affinity measurements between people (in Equ. 7) can play a role of the strength of social tie on a semantic social network. Then, we can apply various social network analysis methods to discover meaningful patterns from the social layer $\mathcal{S}$. In this study, by using cohesive subgroups (communities) identification [Newman 2004], the linkages on the p2p network should be re-organized to discriminate which peers are more proper to support interoperability among peers.

Basically, the interactions between peers are based on exchanging messages, including either a certain query or answer sets. To make queries understandable on heterogeneous peers, the queries have to be transformed with referring to the corresponding peer ontologies. The peer sending queries should select some other neighbor peers to ask query transformation with their own peer ontologies.

Definition 6 (Query). A query q can be embedded into a message $\langle p_{src}, p_{dest}, q\rangle$ sent from peer p_{src} to p_{dest}. The ontologies of two peers are denoted as $o_{src} = Use(p_{src})$ and o_{dest}. The query grammar is simply given by

$$q ::= c|\neg q|q \wedge q|q \vee q \tag{8}$$

where $c \in Define(o)$.

In this study, we are interested in queries consisting of a set of concepts from the peer ontologies, so that the queries can be transformed by *concept replacement* strategy based on correspondences discovered by ontology alignment.

Definition 7 (Correspondence). A set of correspondences discovered ontology alignment process between two ontologies o_i and o_j is given by

$$\{\langle c_i, c_j, rel\rangle | E^{align}(o_i, o_j), c_i \in Define(o_i), c_j \in Define(o_j)\} \tag{9}$$

where rel indicates a relation between two classes (e.g., equivalence, subclass, superclass, and so on).

For example, if there exist correspondences $\{\langle c^1_\alpha, c^3_\beta, =\rangle, \langle c^2_\alpha, c^4_\beta, =\rangle\}$ between peer ontologies o_α and o_β, a peer query "$q_\alpha = c_1 \vee c_2$" from α can be transformed to "$q_\beta = c_3 \vee c_4$" for β.

However, we have to deal with the problems;

- what if the correspondences are not enough to transform the queries sent?
- which peers can efficiently help this transformation process?

Thereby, main scheme of our approach is to find out the best transformation path, minimizing information loss from ontology alignment process. In order to reduce information loss caused by ontology mismatching during transforming queries, we can intuitively consider two heuristic criterion; i) minimizing the number of transformations (or length of transformation path), and ii) maximizing the semantic similarities (or correspondences) with neighbors. Instead of meeting these two objectives, we focus on searching for the most *powerful* peer, most likely to help them communicate with each other.

For example, in Fig. 2, imagine that peer x_1 (source) attempt to send a query to y_1 (destination). There are several possible candidates, e.g., $x_1 \rightarrow x_2 \rightarrow y_1$, $x_1 \rightarrow y_4 \rightarrow y_1$, $x_1 \rightarrow x_2 \rightarrow y_2 \rightarrow y_1$, and so on. By considering only the length of paths, the candidate peers would be x_2 and y_4 as shortest, but it is merely known how well the query will be semantically transformed.

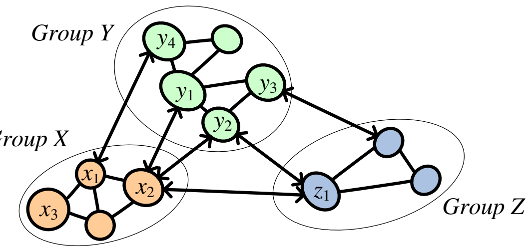

Figure 2: An example of query transformation on semantic p2p network; Links between peers should be larger than a threshold (i.e., minimum affinity τ_{aff}).

4.1 Measuring Semantic Centrality

When sending a query on semantic p2p network, we need to find out which peer (more exactly, peer ontology) is most useful to transform the query for interoperability between source and destination peer. Thereby, semantic centrality of each peer is measured by peer ontology alignment. By mapping peer ontologies, consensual ontology can be built and applied to semantic community identification.

Based on the strengths of social ties E^{aff} between pairs of peers, we can apply a non-parametric approach, e.g., nearest neighborhood method [Gowda and Krishna 1978]. As extending [Newman 2004], this task is to maximize "semantic" modularity function $Q^\diamond$ on social layer $\mathcal{S}$. With the number of communities k predefined, we find out that the given peer set in a social layer $\mathcal{S}$ can be partitioned into a set of communities (or subgroup) $\mathcal{G} = \{g_1, \ldots, g_k\}$. The users can be involved in more than one community. It means that a certain peer p in g_i can also be taken as one of members of g_j, because the semantics in his ontology is relatively close to both consensus semantics of g_i and g_j. Thus, the modularity function $Q^\diamond$ is formulated by

$$Q^\diamond(\mathcal{S}) = \sum_{i=1}^{k} \frac{\sum_{p_a \in g_i, p_b \in g_i} E^{aff}(p_a, p_b)}{|g_i|}. \tag{10}$$

The only pairs of peers where $E^{aff}(p_a, p_b) \geq \tau_{aff}$ should be considered. Thus, $\mathcal{G}(\mathcal{S})$ can be discovered when $Q^\diamond(\mathcal{S})$ is maximized. For computing this, in this paper, we applied an iterative k-nearest neighborhood methods. As changing k, consequently, the social layer is hierarchically re-organized. For example, Fig. 2 is illustrating when $k = 3$.

Generally, centrality measures of a user are computed by using several features on the social network, and applied to determine the structural power. So far, in order to extract the structural information from a given social network, various measurements such as centrality [Hagea and Harary 1995], pair closeness [Girvan and Newman 2002], and authoritative [Kleinberg 1999] have been studied to realize the social relationships among a set of users. Especially, the centrality can be a way of representing the geometrical power of controlling *information flow* among participants on p2p network.

We define two kinds of semantic centralities, with respect to the scope and the topologies of communities;

- *Local* semantic centrality $C^{\diamond}_{L}$, meaning the power of semantic bridging between the members within the same community, and
- *Global* semantic centrality $C^{\diamond}_{G}$, implying the power of bridging for a certain target community.

Local semantic centrality of peer $p \in g_i$ is easily measured by

$$C^{\diamond}_{L}(p, g_i) = \frac{\sum_{p,p' \in g_i, p \neq p'} E^{aff}(p, p')}{|g_i|}, \tag{11}$$

because we are concerning only $E^{aff}(p_a, p_b) \geq \tau_{aff}$ and regarding them as most potential transformation paths. This is similar to the closeness centrality.

On the other hand, global semantic centrality $C^{\diamond}_{G}$ of peer $p \in g_i$ toward a certain target community g_X is based on three factors; i) the number of available transformation paths (s.t. $E^{aff} \geq \tau_{aff}$), ii) the strength of each path E^{aff}, and iii) the local semantic centrality of the peer in target community. Thus, we formulate it as three different ways;

$$C^{\diamond}_{G}(p, g_X) = \frac{\sum_{p' \in g_X} E^{aff}(p, p') \times C^{\diamond}_{L}(p', g_X)}{|g_X|} \tag{12}$$

$$= \frac{\left[\max_{p' \in g_X} E^{aff}(p, p')\right] \times C^{\diamond}_{L}(p', g_X)}{|g_X|} \tag{13}$$

$$= \frac{\max_{p' \in g_X} \left[E^{aff}(p, p') \times C^{\diamond}_{L}(p', g_X)\right]}{|g_X|} \tag{14}$$

While Equ. 12 can take into account all possible paths to taget community by measuring the average centrality, Equ. 13 and Equ. 14 are focusing on only the maximum affinity path. We empirically evaluated these three different heuristic functions in Sect. 5.

4.2 Query Transformation Strategy

We establish query transformation strategy in accordance with the semantic position of peers in social layer $\mathcal{S}$. Query transformation between heterogeneous peers should be conducted by referring to the following strategies;

- If the peers p and p' are located in a same semantic community, a set of transformation paths $TP_L(p, p')$ between them can be evaluated (or ranked) by

$$\frac{\sum_{p'' \in TP_L} C^{\diamond}_{L}(p'')}{\exp(1 + |TP_L|)} \tag{15}$$

where p'' is on the transformation path TP_L. It means the best transformation path has to be chosen, as the length of the path is shorter and local semantic centralities of the peers on the path are higher.

- If the peers $p_i \in g_i$, $p_j \in g_j$ are in different semantic communities, a set of transformation paths $TP_G(p_i, p_j)$ between them can be evaluated (or ranked) by

$$TP_{L_i}(p_i, p'_i) + C^{\diamond}_{G}(p'_i, g_j) + TP_{L_j}(p'_j, p_j), \tag{16}$$

and this can be expanded as

$$\frac{\sum_{p''_i \in TP_{L_i}} C^{\diamond}_{L}(p''_i)}{\exp(1 + |TP_{L_i}|)} + C^{\diamond}_{G}(p'_i, g_j) + \frac{\sum_{p''_j \in TP_{L_j}} C^{\diamond}_{L}(p''_j)}{\exp(1 + |TP_{L_j}|)}. \tag{17}$$

A global transformation path is decomposed into two local transformation path and a transformation path with best global centrality. Exceptionally, when there is no path between communities, the social layer should be re-organized as decreasing the number of communities k.

Thereby, the best transformation path have to be selected by comparing all candidate ones. For example, in Fig. 2, let $C^{\diamond}_{L}(x_1)$, $C^{\diamond}_{L}(y_1)$, and $C^{\diamond}_{L}(z_1)$ be the maximum local semantic centrality in communities X, Y, and Z, respectively. Also, $C^{\diamond}_{G}(x_2, Y)$ and $C^{\diamond}_{G}(y_2, X)$ is the maximum global semantic centralities to the neighbor communities. Peer x_1 can be selected as the most powerful semantic bridge for transformation between x_2 and x_3. In contrast, from transformation from x_1 to y_2, we need to be supported by peers x_2 and y_1.

After the peers are selected, the concepts in the peer query can be replaces by referring to the correspondences. In case of the query $q_{x_2} = c_1 \vee \neg c_2$ from x_2, two sets of correspondences between peer ontologies $\mathcal{O}_{x_1}$ and $\mathcal{O}_{x_2}$, and between $\mathcal{O}_{x_1}$ and $\mathcal{O}_{x_3}$.

5 Experimental results

In order to evaluate the proposed approach, we invited seven students and asked them to annotate a given set of images by referring to any other standard ontologies (e.g., SUMO, WordNet and ODP). While annotating the images, we could collect peer ontologies for building semantic social space, as shown in Table 2.

5.1 From peer ontologies to social ties

Here, we want to show the experimental results of building our social semantic network (SSN) by ontology alignment. They are compared with simple co-occurrence patterns between the annotated images by Mika's social centrality C_M [Mika 2005], which is formulated by

$$C_M(U_i) = \frac{\sum \frac{\cap_{k=1,k\neq i}^{|U|}(\mathcal{R}_{U_k}, \mathcal{R}_{U_i})}{\mathcal{R}_{U_i}}}{|U| - 1} \tag{18}$$

Table 2: Specification of personal ontologies as testing bed

	AS	AZ	FAK	JE	JJ	JP	SL
Number of Resources ($\mathcal{R}_{User}$)	47	47	37	49	47	30	25
Number of Ontologies ($\mathcal{O}_{User}$)	3	5	2	6	1	1	2

Table 3: Experimental results of a) closeness centrality by co-occurrence patterns, and b) semantic affinity E^{aff} and centrality in semantic social network

(a/b)	AS	AZ	FAK	JE	JJ	JP	SL	C_M	$C_L^{\diamond}$
AS	-	0.98/0.65	0.62/0.33	0.94/0.73	1.00/0.26	0.60/0.32	0.23/0.62	0.73	0.49
AZ	0.98	-	0.62/0.49	**0.94/0.825**	0.98/0.31	0.62/0.3	0.26/0.52	0.73	0.52
FAK	0.78	0.78	-	0.70/0.57	0.78/0.28	0.54/0.22	0.30/0.32	0.65	0.37
JE	0.90	**0.90**	0.53	-	0.90/0.46	0.57/0.49	0.16/0.75	0.66	**0.64**
JJ	1.00	0.98	0.62	0.94	-	0.60/0.72	0.23/0.39	0.73	0.40
JP	0.93	0.97	0.67	0.93	0.93	-	0.13/0.51	**0.76**	0.43
SL	0.44	0.48	0.44	0.32	0.44	0.16	-	0.38	0.52

where $|U|$ is the total number of peers (or people) on social network. The results are shown in Table 3. We found out that the number of annotated resources are barely related to the social centrality. SL annotated the least number of resources, so that his centrality also lowest among people. But, even though JE's annotations were the largest one, JP has shown the most powerful centrality.

In semantic social network, we measured semantic affinity (Eq. 7). For doing this, the ontology distances E_O^{dist} between personal ontologies should be measured. We used edit distance measurement for comparing the strings. For instance, E_O^{dist} between JE and AZ is shown in Table 4.

Table 4: Ontology distance E_O^{dist} between JE and AZ; Mark '-' means no alignments between two ontologies.

	JE_foaf.owl	JE_Meteo.owl	JE_Picster.owl	JE_space.owl	JE_UrbanLand.owl	JE_World.owl
az_support-ontology.owl	0.03	-	-	0.17	-	-
az_hasSupplyLineOnt.owl	0.46	-	0.09	0.05	0.04	0.49
az_office.owl	0.47	-	0.04	0.05	0.06	0.04
az_people+petsB.owl	0.06	-	-	0.16	-	-
az_space-basic.owl	0.18	-	-	0.5	-	0.01

Table 5: Precision performance on query transformation strategies; stp means the simple shortest path on social layer.

	g_A				g_B				g_C			
	stp	Equ. 12	Equ. 13	Equ. 14	stp	Equ. 12	Equ. 13	Equ. 14	stp	Equ. 12	Equ. 13	Equ. 14
g_A	0.72	0.75 by Equ. 11			0.36	0.71	0.57	0.64	0.36	0.74	0.59	0.67
g_B	0.317	0.67	0.54	0.6	0.64	0.69 by Equ. 11			0.34	0.78	0.62	0.7
g_C	0.425	0.63	0.51	0.57	0.41	0.68	0.54	0.64	0.685	0.67 by Equ. 11		

5.2 Heterogeneous query processing

From the organized three groups $g_A = \{JE, AZ\}$, $g_B = \{JJ, JP\}$, and $g_C = \{AS, FAK, SL\}$ (the number of communities $k = 3$), we compared the image results retrieved by ten concept-based queries generated by every peers, according to the transformation strategies. In Table 5, we show "Precision" performance, because we are emphasizing the information loss effected from query transformation. We found out that Equ. 12 has outperform the others by about 19 % and 11%.

6 Discussion and Concluding remark

Semantic overlay network is a promising issue on semantic web. Various applications (Edutella, Bibster, and Oyster) for sharing resources on p2p network have been released. Most similarly, semantic overlay network [Crespo and Garcia-Molina 2004] concerns query processing for information sharing on p2p network, but it is based on simple keyword matching to estimate the relationships between nodes.

As another important issue, we want to carefully discuss information loss by semantic transformation [Jung 2007]. While equivalent correspondences (e.g., $\langle c, c', =\rangle$) are acceptable, subsumption correspondences make the transformed queries more specific, and the resources retrieved from peers may (possibly) show higher precision and lower recall results.

As a conclusion, in this paper, we claim a new centrality measurement for providing query-based interactions on p2p network. Especially, we found out very efficient transformation path selection mechanism (e.g., Equ. 12). Moreover, by peer ontology alignment, consensus ontology has been built and applied to identify some semantic communities. We believe that it will play a role of generating semantic geometry to quantify social roles on p2p network.

References

[Alani and Brewster 2005] Alani, H. and Brewster, C.: Ontology ranking based on the analysis of concept structures. In *Proceedings of the 3rd international conference on Knowledge capture (K-CAP '05)*, pages 51–58, New York, NY, USA, 2005. ACM Press.

[Crespo and Garcia-Molina 2005] Crespo, A. and Garcia-Molina, H.: Semantic overlay networks for p2p systems. In Gianluca Moro, Sonia Bergamaschi, and Karl Aberer, editors, *Proceedings of the 3rd International Workshop Agents and Peer-to-Peer Computing (AP2PC 2004), July 19, 2004*, volume 3601 of *Lecture Notes in Computer Science*, pages 1–13. Springer, 2005.

[Euzenat and Valtchev 2004] Euzenat, J. and Valtchev, P.: Similarity-based ontology alignment in OWL-Lite. In Ramon López de Mántaras and Lorenza Saitta, editors, *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI'2004), Valencia, Spain, August 22-27, 2004*, pages 333–337. IOS Press, 2004.

[Freeman 1979] Freeman, L.C.: Centrality in social networks: Conceptual clarification. *Social Networks*, 1:215–239, 1979.

[Girvan and Newman 2002] Girvan, M, and Newman, M.E.J.: Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99:7821–7826, 2002.

[Gowda and Krishna 1978] Gowda, K.C. and Krishna, G.: Agglomerative clustering using the concept of mutual nearest neighbourhood. *Pattern Recognition*, 10(2):105–112, 1978.

[Haga and Harary 1995] Haga, P. and Harary, F.: Eccentricity and centrality in networks. *Social Networks*, 17(1):57–63, 1995.

[Jung 2008] Jung, J.J.: Taxonomy alignment for interoperability between heterogeneous virtual organizations. *Expert Systems with Applications*, 36(4):2721–2731, 2008.

[Jung 2007] Jung, J.J.: Ontological framework based on contextual mediation for collaborative information retrieval. *Information Retrieval*, 10(1):85–109, 2007.

[Jung and Euzenat 2007] Jung, J.J. and Euzenat, J.: Towards semantic social networks. In Enrico Franconi, Michael Kifer, and Wolfgang May, editors, *Proceedings of the 4th European Semantic Web Conference (ESWC 2007), Innsbruck, Austria*, volume 4519 of *Lecture Notes in Computer Science*, pages 267–280. Springer, 2007.

[Kleinberg 1999] Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. *Journal of ACM*, 46(5):604–632, 1999. ISSN 0004-5411.

[Mika 2005] Mika, P.: Ontologies are us: A unified model of social networks and semantics. In Yolanda Gil, Enrico Motta, V. Richard Benjamins, and Mark A. Musen, editors, *Proceedings of the 4th International Semantic Web Conference (ISWC 2005), November 6-10, 2005*, volume 3729 of *Lecture Notes in Computer Science*, pages 522–536. Springer, 2005.

[Newman 2004] Newman, M. E. J.; Fast algorithm for detecting community structure in networks. *Physical Review E*, 69:066133, 2004.

[Shvaiko and Euzenat 2005] Shvaiko, P. and Euzenat, J.: A survey of schema-based matching approaches. *Journal on data semantics*, 4:146–171, 2005.

[Stephens et al. 2004] Stephens, L.M., Gangam, A.K., and Huhns, M.N.: Constructing consensus ontologies for the semantic web: A conceptual approach. *World Wide Web*, 7(4):421–442, 2004.

[Wasserman and Faust 1994] Wasserman, S. and Faust, K.: *Social Network Analysis.* Cambridge University Press, 1994.