Emergent Communities for Semantic Collaboration in Multi-Knowledge Environments: Methods and Techniques

Ph.D. Thesis Abstract

Stefano Montanelli

Università degli Studi di Milano
DICo - Via Comelico, 39, 20135 Milano - Italy
montanelli@dico.unimi.it

Abstract. The need of sharing data and resources to foster semantic collaboration is a key problem at the current stage of development of open distributed systems. In this context, *autonomous* and *self-organizing* communities of peers emerge by bringing together those peers that are interested in similar topics and plan to strengthen their cooperations. The Ph.D. thesis abstract illustrates a *semantic handshake process* based on ontologies and ontology matching techniques to handle consensus negotiation and peer community formation. Furthermore, we discuss the possible benefits deriving from semantic community adoption by describing the *community-aware query propagation strategy* for effective distribution of resource requests on a semantic basis within a committed community.

1 The research question of the thesis

The need of sharing data and resources to foster semantic collaboration is a key problem at the current stage of development of open distributed systems, like P2P networks, and semantic Grids [2]. In this context, the emergence of collaboration among peers requires dynamic capabilities of negotiating agreements on common interpretations within the context of a given task. This is typical for instance of peer-based systems, characterized by a set of independent peer parties without prior reciprocal knowledge and no degree of relationship, that dynamically need to cooperate by sharing their resources (e.g., data, documents, services). These collaboration scenarios are *multi-knowledge*, in that no centralized authorities are defined to manage a comprehensive view of the resources shared by all the nodes in the system, due to the high dynamism and variability of collaboration and sharing requirements. On the opposite, each peer is responsible of providing the knowledge description of the resources to be shared through its own ontology. In order to facilitate resource discovery in such highly dynamic and open contexts, the formation of *autonomous* and *self-organizing* communities of parties poses new issues to be investigated and some work in this direction has been appeared in the literature [1, 3, 8, 9]. Communities aim to handle the problem of high network traffic due to single-peer interactions and to

provide a coordination mechanism for processing and forwarding resource queries on a semantic basis, by exploiting available ontologies describing resources to be shared. In this respect, the intrinsically open nature of P2P systems, and thus of the communities, poses serious issues regarding the maintenance of the communities and requires policies and mechanisms to specify the rules regulating resource sharing and the conditions under which a peer is available to process incoming queries.

With respect to this scenario, the Ph.D. thesis will be devoted to investigate two main issues: i) the development of *consensus-driven* techniques which exploit ontological resource descriptions and ontology matching in order to form, maintain, and disband semantic communities in a P2P environment; and ii) the definition of a *community-aware* query propagation strategy for effective distribution of resource requests on a semantic basis to enforce coordinated sharing of distributed resources within a committed community. The Ph.D. thesis has focused till now on the formation of autonomous and self-organizing emergent semantic communities of peers. In particular, we have defined a *semantic handshake process* based on ontologies and ontology matching techniques to handle consensus negotiation and peer community formation. In this context, ontologies provide a semantically rich representation of the shared resources and enable peers to describe their involvement in one or more concepts of interest. The role of ontology matching techniques regards the semantic affinity evaluation between concepts provided by different peers in order to assess the level of match between nodes with similar interests. A key feature of our ontology matching techniques is related to their flexibility that makes them suitable for coping with the inherent dynamism of open systems, such as P2P systems [5].

The research methodology that is being applied for the Ph.D. thesis is based on the following main phases: i) *literature review* with the aim at providing a critical comparison of the state of the art solutions for managing peer communities and semantic routing in P2P systems, ii) *conceptual design* where requirements and foundational aspects related to the Ph.D. issues are formally addressed, iii) *experimentation* with the aim at validating the thesis results by means of simulation on a number of real test cases, and iv) *prototype implementation* where a P2P prototype tool is developed according to the results and final considerations of the Ph.D. thesis work.

2 Related work

Relevant research work with respect to the Ph.D. thesis regards community management and semantic query routing in peer-based systems.

Community management in peer-based systems. The idea of supporting peer communities by means of a semantics-based approach is at an initial stage of research, and few proposals have been appeared in the literature. In [4] a social collaboration model is proposed as the reference basis to develop the P2P Kex platform (Knowledge Exchange System) for supporting knowledge shar-

ing in federations of peers. In this system, knowledge is organized according to a XML-based syntax and a semantic matching algorithm is adopted to manage the different meanings provided by single peers and federations. In [8], peer communities are introduced as a generalization of peer groups to realize an efficient search query propagation strategy in a populated P2P space. Peer communities are formed on the basis of string-based interests that are used to determine the communities in which a peer would participate. Each peer adopts an escalation technique to advertise its interests and to take part to new communities. Trust and reputation information are used in [1] to define clusters of peers (i.e., communities) capable of providing relevant documents with respect to a given query. The network clustering emerges gradually by point-to-point interaction among the peers with the highest reputation. Each peer classifies the documents to share and constitutes its local knowledge represented by means of a concept hierarchy. Syntactic and structural matching techniques are adopted to compare an incoming query with peer local knowledge in order to identify possible common features. In [3], communities are defined as groups within or across organizations who share a common set of information, needs, or problems. In this approach, peer interactions (i.e., queries) are exploited to discover the communities and to populate the SWRC+COIN community ontology which describe the typical structure and the key entities of communities as well as their relationships. The community ontology can be queried in order to identify the most relevant peers with respect to a given request.

Original contribution of the thesis. We observe that most of the presented approaches recognize peer communities as a possible solution for improving system effectiveness (e.g., query propagation). String- or XML-based formalisms are used by peers to describe knowledge and interests. Syntax-based matching functionalities are generally provided even if the adoption of semantic matching techniques is becoming a key factor during the community discovery and formation phase. In this context, the main contribution of the thesis work is related to the development of a semantic handshake process for peer community formation capable of combining ontological descriptions of peer interests with dynamic ontology matching techniques. Such a solution provides semantic matchmaking capabilities in community formation that allow to overcome the limitations of exact matching techniques adopted in most approaches by contemporary addressing the dynamism and flexibility requirements of peer-based systems.

Semantic query routing in peer-based systems. Semantic query routing strategies are required to improve the performance and the effectiveness of discovery and search processes for resource sharing in P2P systems. In [10], the REMINDIN' multi-step query propagation protocol is described, to enforce selected propagation of queries by observing which queries are successfully answered by other peers, by storing these observations, and by subsequently using this information for peer selection. A similar approach is presented in [12] where the Intelligent Search Mechanism (ISM) is introduced to provide an efficient and scalable solution for improving the information retrieval problem in P2P systems.

Each ISM peer is composed of four basic elements: i) the *profiling structure* that is used to store the most recent replies of each known peer, ii) the *query similarity function* that is used to identify the similarity between different search queries, iii) the *RelevanceRank algorithm* which exploits the profiling structure to select the peers that can provide relevant answers with respect to a given query, and iv) the *search mechanism* that is used to send the query to the selected peers. In most recent work, some initial ideas to consider query routing as an application of peer communities have been appeared. In [11], the potential applications of communities are discussed and classified in *endogenous* and *exogenous* applications. Referrals networks based on a sociological metaphor are compared with bipartite communities based on link analysis in order to show the benefits of a collaborative approach for improving local performance in locating service providers. Agents (i.e., peers) adaptively select their neighbors and their query recipients by exploiting *sociability* and *expertise* information computed on previous interactions. The choices performed by the agents cause communities to emerge. Furthermore, the notion of P2P Semantic Link Network is introduced in [13] to emphasize the need of typed semantic links specifying semantic relationships between peers in order to maintain information about nodes with similar contents. Each peer defines its own XML Schema (source schema) describing the contents to share and adopts SOAP-based messages to communicate with the other members of the network. Semantic links are exploited with cycle analysis and functional dependency analysis in order to select the query recipients according to the types of the semantic links.

Original contribution of the thesis. Current P2P query propagation algorithms are essentially based on statistical observations and exploit, in some cases, a shared ontology, often just a taxonomy. The main contribution of the thesis work is related to the definition of the community-aware query propagation strategy to drive the selection of the best recipients for a given query by exploiting ontology knowledge of the sending peer. One important goal of the proposed approach is to address emergent semantics requirements, by extending current techniques to work in multi-ontology contexts and thus releasing the constraint of having an initial common shared ontology.

3 Preliminary results

In [8], the notion of peer community is introduced as a generalization of peer group involving peers that are actively engaged in sharing, communicating and promoting common interests. By extending such a notion, we define *a semantic community of peers as a set of nodes which show a common interest in a given topic and are organized in a structured way* (e.g., a tree).

Definition of semantic community. Formally, a semantic community SC is a 5-tuple of the form: $SC = \langle CID,\ ICard,\ Members,\ SPolicy,\ Status \rangle$, where:

- CID is the unique Community Identifier that characterizes the community SC.

- $ICard$ is the community Identity Card. The $ICard$ represents a subject category or topic area of interest and is defined as an ontology. The use of an ontology-based ICard provides a semantically rich description of a given topic area of interest and allows the characterization of the common interpretation (i.e., perspective) featuring the community.
- $Members$ is the set of participants that joins SC and spontaneously agrees with its $ICard$, since they have semantically relevant resources for the community.
- $SPolicy \in (strict \mid soft)$ defines the behavior that SC members have to observe in terms of resource availability. The *strict* policy requires that incoming requests are processed by all community members in cooperation. The $soft$ policy defines that each community member can autonomously choose the set of incoming queries to evaluate.
- $Status \in (potential \mid emerging \mid partially\ committed \mid committed \mid disbanded)$ represents the actual status of SC. During the consensus negotiation process, the community passes through the *potential*, *emerging*, and *partially committed* states. The *committed* and the *disbanded* states indicate that the community is effective and no more active in the network, respectively.

In our approach, we assume that each peer exposes to the system a *peer ontology* which provides a semantically rich representation of the resources that the peer exposes to the network, in terms of concepts, properties, and semantic relations [6]. Furthermore, each peer relies on the H-MATCH semantic matchmaker for matching ontologies in order to find which concepts match in different ontologies and at which level [5].

In the following, we discuss some preliminary results regarding i) semantic community formation based on the handshake techniques; and ii) the community-aware query propagation.

3.1 Semantic community formation

As described in Figure 1, a semantic community of peers emerges when a node, called *community founder*, invokes a *semantic handshake process* which is composed of the following transitions:

1. *ICard advertisement*. The founder P_f defines a CID and an ICard describing the topic area of interest of the emerging community, along with a set of *commitment constraints* specifying the conditions required for the community establishment (e.g., minimum number of member required, specific semantic affinity constraints). Then, the founder composes an *Invitation Message* containing the CID and the ICard created, as well as a TTL parameter defining the maximum number of hops allowed for the invitation propagation. Then, the invitation message is sent to all P_f neighbors in order to advertise the new community.

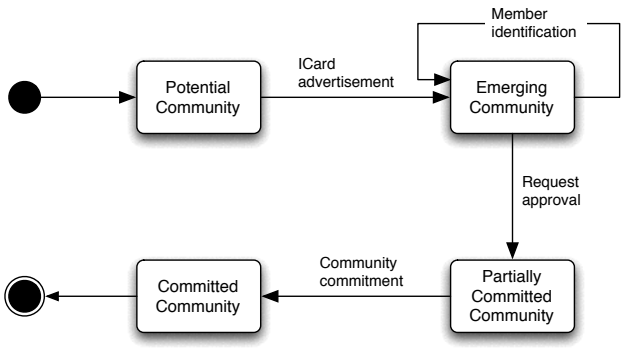

Fig. 1. The state transition diagram of the handshake algorithm

2. *Member identification*. Each invited peer P_i invokes the semantic matchmaker in order to compare the incoming ICard with its peer ontology. P_i is relevant for the community if the semantic matchmaker identifies concepts in the peer ontology with a high affinity with the ICard. In this case P_i replies to P_f with an *Interest Message* reporting the portion of its peer ontology related to the matching concepts found to be relevant for the community by the semantic matchmaker. Independently from the matchmaker results and if $TTL \geq 0$, P_i forwards the invitation message to all its neighbors, except for the peer from which the message has been received.
3. *Request approval*. Receiving the interest messages, the founder P_f has to evaluate which peers are admitted in the community. For this reason, P_f invokes its semantic matchmaker and compares each peer ontology portion received by the interested peers with its knowledge (i.e., its peer ontology). For each candidate peer, the goal of this comparison is to evaluate whether the provided knowledge matches the knowledge of the founder, and then to assess whether they share a common perspective of the community interests. If the matchmaker returns high matching results, P_f admits the peer in the community and sends an *Approval Message* to the admitted peer.
4. *Community commitment*. Once the Request approval phase is completed, the founder verifies that the commitment constraints are satisfied. In this case, a *Commitment Message* is sent to all the admitted peers and the semantic community is effectively established. If the committed constraints are not satisfied, the founder stops the community formation. In this case, the admitted peers wait for the commitment message until a predefined timeout expires and the community is considered as disbanded.

Appropriate techniques are also defined to address the main events that may occur during the semantic community life-cycle, such as insertion and deletion of participant, unexpected peer failure, and community disband. For further details

regarding the semantic community formation and management, the reader can refer to [7].

3.2 Community-aware query propagation

Committed communities are the reference for improving search and discovery capabilities in P2P networks. When a searching peer P_s needs to submit a query Q to the system, the communities of peers are exploited to select the query recipients that can provide resources matching the target. To this end, P_s exploits its joined communities in order to discover whether their ICards are related to the query target. P_s invokes the H-MATCH semantic matchmaker and evaluates the semantic affinity between the query Q and the ICard of each joined community. On the basis of H-MATCH results, we distinguish the following cases:

- P_s is member of one or more communities related to the query Q. For each community found to be relevant, P_s sends the query Q to its semantic neighbors in the community. Each receiving node P_r forwards the query Q to its community neighbors except for P_s, and invokes its semantic matchmaker to compare the query Q against its peer ontology in order to evaluate whether it can provide relevant knowledge to send back to P_s. The forwarding mechanism is iterated until the query Q reaches each community member.
- No semantic affinity exists between the query Q and the ICard of the communities joined by P_s. Q is sent to all the peers known by P_s according to the routing protocol of the underlying P2P infrastructure [1]. Each receiving peer invokes the semantic matchmaker and compares the contents of Q with the ICards it owns in order to renew the community-aware query propagation.

4 Ongoing and future work

In this abstract, we have presented the thesis work we are undergoing for semantic community formation and management in P2P systems. Future work will be devoted to finalize the development of the community formation process and to assess the effectiveness of the proposed community-aware query propagation techniques. In particular, two main issues will be considered:
Semantic handshake techniques. For what concern the semantic handshake algorithm, we plan to implement such a semantic community aggregation protocol and to develop appropriate *commitment policies* for allowing a community founder to specify the requirements to be satisfied by the potential member peers for the establishment of an emerging community. Moreover, we intend to refine the actual handshake process in order to share the responsibilities between the founder and the community members during the formation process. In particular, we are working on the definition of advanced consensus negotiation techniques in which the community ICard is the result of an active negotiation

[1] In this case, a peer-based semantic routing protocol can be used to define query recipients. Some initial results on this topic can be found in [6].

process where the founder and the interested peers interact and discuss changes to the community ICard until an agreement among them is established.

Community-aware query propagation. By using simulation techniques, we aim at comparing traditional P2P query propagation strategies with the community-aware propagation algorithm where queries are sent to the communities with the higher chance to provide relevant results according to the matching results. Further experiments will regard the comparison of the community-aware query propagation with a basic semantic routing protocol we have developed within the HELIOS framework for ontology knowledge sharing in P2P systems [6]. Finally, we are interested in developing *popularity-driven* community aggregation techniques, where a peer founder can advertise a new community on the basis of queries sniffed in the network. When a great number of queries in the network is due to similar requests, a peer can propose to found a semantic community regarding such a popular topic.

References

1. A. Agostini and G. Moro. Identification of Communities of Peers by Trust and Reputation. In *Proc. of the 11th Int. AIMSA Conference*, Varna, Bulgaria, 2004.
2. S. Androutsellis-Theotokis and D. Spinellis. A survey of peer-to-peer content distribution technologies. *ACM Computing Surveys*, 36(4), 2004.
3. S. Bloehdorn, P. Haase, M. Hefke, Y. Sure, and C. Tempich. Intelligent Community Lifecycle Support. In *Proc. of the Int. I-KNOW Conference*, Graz, Austria, 2005.
4. M. Bonifacio, P. Bouquet, G. Mameli, and M. Nori. Peer - Mediated Distributed Knowledge Management. In *Proc. of the Int. AMKM Symposium*, Stanford, CA, USA, 2003.
5. S. Castano, A. Ferrara, and S. Montanelli. Matching Ontologies in Open Networked Systems: Techniques and Applications. *Journal on Data Semantics*, V, 2006.
6. S. Castano, A. Ferrara, and S. Montanelli. *Web Semantics and Ontology*, chapter Dynamic Knowledge Discovery in Open, Distributed and Multi-Ontology Systems: Techniques and Applications. Idea Group, 2006.
7. S. Castano and S. Montanelli. Semantic Self-Formation of Communities of Peers. In *Proc. of the ESWC Workshop on Ontologies in Peer-to-Peer Communities*, Heraklion, Greece, 2005.
8. M. Khambatti, K. Dong Ryu, and P. Dasgupta. Peer-to-Peer Communities: Formation and Discovery. In *Proc. of the PDCS Conference*, Cambridge, USA, 2002.
9. P. Mika. Social Networks and the Semantic Web. In *Proc. of the IEEE/WIC/ACM Int. WI Conference*, Beijing, China, 2004.
10. S. Staab, C. Tempich, and A. Wranik. REMINDIN': Semantic Query Routing in Peer-to-Peer Networks based on Social Metaphors. In *Proc. of the 13th Int. WWW Conference*, New York, NY, USA, 2004.
11. P. Yolum and M.P. Singh. Dynamic Communities in Referral Networks. *Web Intelligence and Agent Systems*, 1(2), 2003.
12. D. Zeinalipour-Yazti, V. Kalogeraki, and D. Gunopulos. Exploiting Locality for Scalable Information Retrieval in Peer-to-Peer Networks. *Information Systems*, 30(4), 2005.
13. H. Zhuge, J. Liu, L. Feng, and C. He. Semantic-Based Query Routing and Heterogeneous Data Integration in Peer-to-Peer Semantic Link Networks. In *Proc. of the Int. Conference on Semantics of a Networked World*, Paris, France, 2004.