A Necessary Condition for Semantic Interoperability in the Large*

Philippe Cudré-Mauroux and Karl Aberer

School of Computer and Communication Sciences
Swiss Federal Institute of Technology (EPFL)
1010 Lausanne, Switzerland
{philippe.cudre-mauroux, karl.aberer}@epfl.ch

Abstract. With new standards like RDF or OWL paving the way for the much anticipated semantic web, a new breed of large scale semantic systems is about to appear. Even if research on semantic reconciliation methods is abundant, it is not clear how interoperable very large scale semantic systems can be. This paper represents a first effort towards analytically analyzing semantic interoperability in the large: By adapting a recent graph-theoretic framework, we examine the dynamics of large scale semantic systems and derive a necessary condition for fostering global semantic interoperability.

1 Introduction

Information systems are about to undergo profound changes through the wide adoption of a set of semantic standards comprising RDF, RDFS or OWL. These specifications aim at providing machine-processable information and should underpin the creation of systems where data are given well-defined semantics.

In [2], we introduced Semantic Gossiping as a new way of reconciling semantically heterogeneous domains in an evolutionary and completely decentralized manner. We have shown [3] that sets of pair-wise, local translations can be sufficient for creating a global self-healing semantic network where semantically correct translations get reinforced. A variety of related works, fostering global interoperability from local mappings (see for example [5,6,9]) have also proven to be successful, demonstrating the general validity of this approach recently termed as *Peer Data Management*. Even if much effort has recently been devoted to the creation of sophisticated schemes to relate pairs of schemas or ontologies (see [11] for a survey), it is still far from being clear how such large-scale semantic systems evolve or how they can be characterized. For example, even if a lack of ontology mappings clearly limits the quality of the overall semantic consensus in a given system, the exact relationships between the former and the latter are unknown. Is there a minimum number of mappings required to foster semantic

* The work presented in this paper was supported (in part) by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

R. Meersman, Z. Tari (Eds.): CoopIS/DOA/ODBASE 2004, LNCS 3291, pp. 859–872, 2004.

interoperability in a network of information sharing parties? Given a large set of ontologies and ontology mappings, can we somehow predict the impact of a query issued locally?

This paper represents a first attempt to look at the problem from a macroscopic point of view. Our contribution is two-fold: First, we develop a model capturing the problem of semantic interoperability with an adequate granularity. Second, we identify recent graph theoretic results and show how they are (with some slight adaptation) applicable to our problem. More particularly, we derive a necessary condition to foster semantic interoperability in the large and present a method for evaluating the propagation of a query issued locally. Also, we give some initial evaluation of our methods. The rest of this paper is organized as follows: We start by introducing a general layered representation of distributed semantic systems. Section 3 is devoted to the formal model with which we analyze semantic interoperability in the large. The main theoretical results related to semantic interoperability and semantic component sizes are detailed in Section 4 and Section 5. Finally, we discuss practical applications of our main results before concluding.

2 The Model

Large-scale networks are traditionally represented by a graph. In our case, however, a single graph is insufficient to accurately model the relationships between both the systems and their schemas. We present below a set of representational models for large-scale semantic systems which will then be used throughout the rest of this paper. We model information parties as peers related to each other physically (Peer-to-Peer model). Peers use various schemas or ontologies to annotate their resources (Peer-to-Schema model). Finally, schemas themselves can be related through mappings we term translation links (Schema-to-Schema model). Each of these models represents a distinct facet of the overall Peer Data Management System and can be quite independent of the other two (as, for example, in the *GridVine* system [4]).

2.1 The Peer-to-Peer Model

Peers represent autonomous parties producing and consuming information in a system. Each peer $p \in P$ has a basic communication mechanism that allows it to establish connections with other peers. We do not make any other assumption on this mechanism, except that any peer should be able to contact any other peer in the system – either by broadcasting (Gnutella) or by using a central (Napster), hierarchical (DNS) or decentralized (P-Grid [1]) registry. Furthermore, we assume that the information and meta-information (i.e., metadata, schemas and schema translations) available in the system are all indexed in a similar way, allowing a peer to retrieve any resource independently of its exact nature.

2.2 The Peer-to-Schema Model

We assume that peers produce annotations (metadata) related to resources available in the system. Each peer $p \in P$ organizes its local annotation database DB_p according to a set of schemas S_p. When a peer p organizes (part of) its annotation database following a schema s_i, we say that p is in the *semantic domain* of s_i: $p \leftrightarrow s_i$. Individual schemas are uniquely identified throughout the network and may be used by different peers (see for example Figure 1, representing such a bipartite Peer-to-Schema graph where p_3 annotates data according to schemas s_A and s_C).

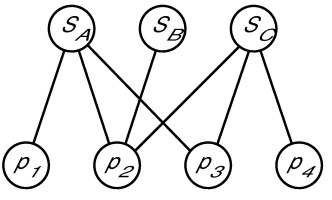

Fig. 1. The Peer-to-Schema model

We do not make any assumption on the languages used to express the metadata or schemas. Peers can for example use different mechanisms (e.g., XML Schema elements or RDFS/OWL classes) for categorizing resources. However, all peers should be able to locally issue queries $q_i \in Q$ against their databases using standard query operators in order to retrieve sets of specific resources.

2.3 The Schema-to-Schema Model

Finally, we allow peers to create translation links between schemas. We do not put any constraint on the origin of the translations: They might be automatically generated, written by domain experts, partially wrong, and may be provided by any peer, regardless of the schemas it uses for its own database. A translation link $T_{s_1 \to s_2}$ relates two schemas s_1 and s_2; Concretely, translation links may for example use mapping operations to relate two schemas: $s_2 = \mu_f(s_1)$ where f is a list of functions of the form $c_i := F(c_1, \ldots, c_k)$, with class names c_i from s_2 and $c_1, \ldots, c_k$ from s_1. The function F is specific to the mapping operations to be performed and can encompass syntactic reconciliation features. A special case is renaming of a class: $c_2 := c_1$.

Using a translation link $t_{s_1 \to s_2}$, a peer $p_1 \leftrightarrow s_1$ may transform a local query q on its database DB_{p_1} into a transformed query q' applicable to a second semantic domain s_2:

$$t_{s_1 \to s_2}(q(DB_{p_1})) \equiv q'(DB_{p_2}), p_1 \leftrightarrow s_1 \wedge p_2 \leftrightarrow s_2$$

Note that multiple transformations may be applied to a single query q. The composition of multiple transformations $t_1, \ldots, t_n$ is given by using the associative composition operator (specific to a given approach) $\circ$ as follows

$$(t_1 \circ \ldots \circ t_n)(q)(DB) \equiv q(q_{t_1} \ldots (q_{t_n}(DB))).$$

From a graph modelling perspective, translations may be viewed as edges interconnecting schema nodes. Figure 2 depicts a Schema-to-Schema graph. Note that the edges have to be directed in order to capture the peculiarities of the mapping operations, since mapping functions may not be invertible and since the properties of even the most basic translations can be dependent on the direction with which they are applied (e.g., relations between subclasses and super-classes). Also, note that a growing number of schemes use a metric to characterize the quality of the various mapping operations encapsulated by the translation links (see for example [8,13]). The resulting graph is therefore a weighted directed multigraph, i.e., a directed graph with (possibly) multiple, weighted edges (translation links) between two vertices (schemas).

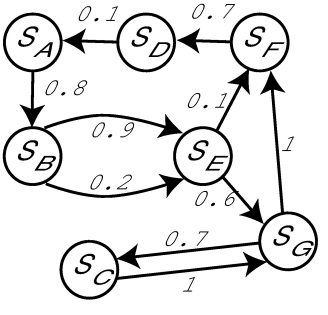

Fig. 2. The Schema-to-Schema model

3 Semantic Interoperability in the Large

The rest of this paper is devoted to the study of interoperability in our Peer-to-Peer setting, mainly through the analysis of a derived version of the Schema-to-Schema graph. A peer $p_i \leftrightarrow s_j$ may send a query to any peer in its own semantic domain, i.e., to any peer $p_k \in P \mid p_k \leftrightarrow s_j$ in the Peer-to-Schema model (supposing, again, that the Peer-to-Peer model allows to contact any peer in the network). The query may also be forwarded to peers in foreign semantic domains $s_l \neq s_j$ as long as there exist a translation $t_{s_j \rightarrow s_l}(q)$ or a series of translations $t_{s_j \rightarrow s_1} \circ \ldots \circ t_{s_n \rightarrow s_l}$ to transform the query adequately. Generalizing the above statement, we introduce the notion of semantic interoperability:

Definition (Semantic Interoperability). Two peers are said to be *semantically interoperable* if they can forward queries to each other, potentially through series of semantic translation links.

Note that the aforementioned definition does not characterize the quality of the semantic interoperability in any way; It simply acknowledges the existence of some semantic relationship between two peers on the basis of a translation link. If no semantic path exists to forward the query, we say that the two peers in question are *semantically unreconcilable*.

3.1 Semantic Connectivity

Analogously to physical network analysis, we define an intermediary layer accounting for the semantic connectivity of the system. Indeed, considering the definition given above, we can slightly relax our Schema-to-Schema model when analyzing semantic interoperability:

Unweighted model. Since our definition of semantic interoperability is based on the presence or absence of translation links, we ignore the weights in the Schema-to-Schema model.

No duplicate edges. From a vertex-strong connectivity point of view, duplicate edges between two vertices play no role. Thus, multigraphs may be replaced by their corresponding digraphs.

However, when analyzing semantic connectivity graphs, one has to account for two important specificities of large-scale semantic systems:

High clustering. Sets of schemas related to a given domain of expertise tend to organize themselves tightly and thus share many translation links, while being largely disconnected from schemas describing other domains. Therefore, we expect clustering coefficients in large-scale semantic graphs to be particularly high.

Bidirectional edges. Even if mappings used in translation links are essentially unidirectional, we can expect domain experts to create translations in both directions (to and from a given ontology) in order to foster semantic interoperability. Thus, a fraction of the links can be considered as bidirectional in our connectivity analysis.

Taking into account the points exposed above, we can finally propose our formal model for studying semantic interoperability:

Definition (Semantic Connectivity Graph). A *Semantic Connectivity Graph* is a pair (S, T) where

- S is the set of schemas in a large-scale semantic system
- T is a non-redundant, irreflexive set of ordered pairs $(s_i, s_j) \mid i \neq j \wedge s_i, s_j \in S$, each denoting a directed semantic translation link between two schemas.

Using this formalism, semantic systems can be represented by digraphs where S is a set of vertices and T a set of directed edges. A couple of statistical properties derived from these semantic connectivity graphs will be of particular interest for our upcoming analysis:

- The probabilities p_{jk} that a randomly chosen vertex has in-degree j and out-degree k
- The *clustering coefficient cc* defined as the average number of edges of a node's neighbor connecting to other neighbors of the same node
- The *bidirectional coefficient bc* defined as the average fraction of edges which can be considered as bidirectional, i.e., the fraction of translation $t = (s_i, s_j) \in T \mid \exists t' = (s_j, s_i) \in T$.

Remembering that a directed graph is strongly connected if it has a path from each vertex to every other vertex, one can easily determine wether or not a set of peers is semantically interoperable by inspecting the semantic connectivity graph:

Theorem 3.1. *Peers in a set $P_s \subseteq P$ are all* semantically interoperable *if $S_s \subseteq S$ is strongly connected, with $S_s \equiv \{s \mid \exists p \in P_s, p \leftrightarrow s\}$.*

Proof. If S_s is not strongly connected, there exists at least one vertex $s_l \in S_s$ which cannot be reached from another vertex $s_j \in S_s$. This means that a peer $p_i \in P_s$, $p_i \leftrightarrow s_j$ is semantically unreconcilable with a second peer $p_k \in P_s$, $p_k \leftrightarrow s_l$, and thus the set of peers is not semantically interoperable.

■

As a corollary, a network of peers is *globally* semantically interoperable if its semantic connectivity graph is strongly connected. This property may be satisfied in a wide variety of topologies. Introducing $|V_s|$ and $|E_s|$ as (respectively) the number of vertices and edges in a set of peers $P_s \subseteq P$, we can immediately derive two bounds on the number of translation links affecting the semantic interoperability:

Observation 1. A set of peers $P_s \subseteq P$ cannot be semantically interoperable if $|E_s| < |V_s|$.

Observation 2. A set of peers $P_s \subseteq P$ is semantically interoperable if $|E_s| > |V_s|(|V_s| - 1) - (|V_s| - 1)$.

The proofs of these two observations are immediate.

4 A Necessary Condition for Semantic Interoperability

4.1 Undirected Model

Real world graphs usually develop by following preferential attachment laws and exhibit properties (e.g., small-world, scale-free) specific to their statistical distribution. Thanks to recent advances in graph theory, it is now possible to study

arbitrary large graphs based on their degree distribution. However, there exists no model taking into account all the specificities of our semantic connectivity graph. In the following, we derive new results from the framework introduced in [10] to account for these specificities. Since we do not assume readers to be generally familiar with generatingfunctionologic graph theory, we start by introducing a simpler, undirected model before presenting the directed one.

Our approach is based on generating functions [12]; First, we introduce a generating function for the degree distribution of a semantic connectivity graph:

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k \tag{1}$$

where p_k is the probability that a randomly chosen vertex has degree k. This function encapsulates all the information related to the degree distribution of the graph, since

$$p_k = \frac{1}{k!} \frac{d^k G_0}{dx^k}\bigg|_{x=0}. \tag{2}$$

Theorem 4.1. *Peers in a set $P_s \subseteq P$ cannot be* semantically interoperable *if $\sum_k k(k-2-cc)p_k < 0$, with p_k the probability that a node has degree k in the undirected semantic connectivity graph of the set and cc the clustering coefficient.*

Proof. The average number of neighbors of a node is

$$z_1 =< k >= \sum_k k p_k = G_0'(1). \tag{3}$$

If we follow a randomly chosen edge, we arrive at a vertex with probability proportional to the degree of that vertex, i.e., proportional to kp_k. The correctly normalized degree distribution of the node we arrive at is

$$\frac{\sum_k k p_k x^k}{\sum_k k p_k} = x \frac{G_0'(x)}{G_0'(1)}. \tag{4}$$

If we start at a randomly chosen vertex and follow all the edges from that vertex to get to the set of direct neighbors, each of these first-order neighbors will have a degree distribution given by equation 4. Now, if we want to count the number of second-order neighbors from the original node we started at, we can consider the first-order neighbors as being one degree lower, since we do not want to take into account the edge connecting our original node to the first-order neighbor. Similarly, we can subtract on average *cc* degrees of the first-order neighbors to account for those links which connect first-order neighbors together. In the end, the distribution of the number of second-order neighbors we get from a first-order neighbor is

$$G_1(x) = \frac{1}{x^{cc}} \frac{G_0'(x)}{G_0'(1)} = \frac{1}{z_1} \frac{1}{x^{cc}} G_0'(x). \tag{5}$$

The probability distribution of the number of second-order neighbors is then obtained by multiplying 5 by the probability of the original node of having k first-order neighbors and by summing over these k neighbors. Remembering that the distribution of a distribution function summed over m realizations is generated by the m^{th} power of that generating function, we get

$$\sum_k p_k [G_1(x)]^k = G_0(G_1(x)). \tag{6}$$

The average number of second order neighbors is

$$z_2 = \left[\frac{d}{dx} G_0(G_1(x))\right]_{x=1} = G_0'(G_1(1))G_1'(1) = G_0'(1)G_1'(1) = \sum_k (k - 1 - cc)p_k \tag{7}$$

since $G_1(1) = 1$.

A necessary condition for a graph to be strongly connected is the emergence of a giant component connecting most of its vertices. It has been shown (see for example [10]) that such a component can only appear if the number of second-order neighbors of a graph is on average greater or equal than the number of first-order neighbors. Presently, if

$$z_2 \geq z_1 \Leftrightarrow \sum_k (k - 1 - cc)p_k \geq \sum_k k p_k \Leftrightarrow \sum_k (k - 2 - cc)p_k \geq 0. \tag{8}$$

If the condition in equation 8 is not satisfied, the undirected semantic connectivity graph cannot be strongly connected and thus the set of peers cannot be semantically interoperable.

■

We term $\sum_k (k-2-cc)p_k$ *connectivity indicator ci*. Figure 3 below compares this indicator with the size of the biggest connected component in a random undirected semantic connectivity graph of *10 000* vertices with a variable number of edges. Edges are generated randomly (each pair of distinct vertices has the same probability of being connected) such that the resulting graph approximates an exponentially distributed graph. We notice that *ci* is a very good indicator of the overall connectivity of a semantic graph, i.e., the graph is in a *sub-critical* phase when $ci < 0$ (no giant connected component) while it is in a super-critical phase when $ci > 0$ (after the percolation threshold).

4.2 Directed Model

We now turn to the full-fledge, directed model based on the semantic interoperability graph. Our methodology will be exactly the same as the one used above for the undirected case. Remember that p_{jk} is the probability that a randomly chosen vertex has in-degree j and out-degree k in our semantic connectivity

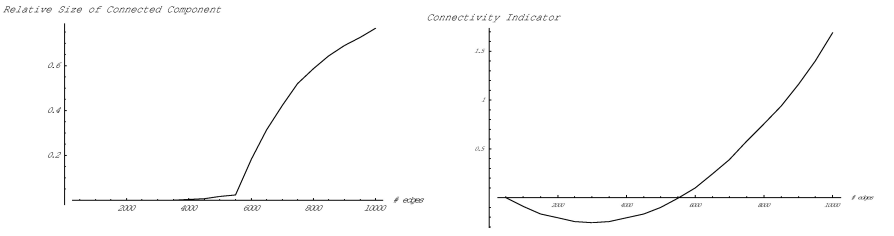

Fig. 3. Maximal connected cluster size and Connectivity Indicator for a random network of 10000 vertices

graph. We introduce $\mathcal{G}(x, y)$, a generating function for the joint probability distribution of in and out-degrees:

$$\mathcal{G}(x, y) = \sum_{j,k} p_{jk} x^j y^k \tag{9}$$

which has to satisfy

$$\sum_{jk} (j - k) p_{jk} = 0 \tag{10}$$

since every edge leaving some vertex has to enter another. This also implies that the average degree (both in and out) z_1 of vertices in the graph is

$$z_1 = \sum_{jk} j p_{jk} = \sum_{jk} k p_{jk} = \left. \frac{\delta \mathcal{G}}{\delta x} \right|_{x,y=1} = \left. \frac{\delta \mathcal{G}}{\delta y} \right|_{x,y=1}. \tag{11}$$

The joint probability p_{jk} is given by

$$p_{jk} = \left. \frac{1}{j!k!} \frac{\delta^{j+k} \mathcal{G}}{\delta^j x \delta^k y} \right|_{x=0,y=0}. \tag{12}$$

Again, the generating function encapsulates all the information contained in the discrete probability distribution p_{jk}.

Theorem 4.2. [Necessary condition for semantic interoperability] *Peers in a set $P_s \subseteq P$ cannot be* semantically interoperable *if $\sum_{j,k}(jk - j(bc + cc) - k)p_{jk} < 0$, with p_{jk} the probability that a node has in-degree j and out-degree k in the semantic connectivity graph of the set, bc the bidirectional coefficient and cc the clustering coefficient.*

Proof. The function generating the number of outgoing edges leaving a randomly chosen vertex is

$$G_0(y) = \mathcal{G}(1, y) \tag{13}$$

If we follow an edge chosen randomly, we arrive at a vertex with a probability proportional to the in-degree of that vertex. Normalizing on the degree distribution of that vertex, we obtain:

$$\frac{\sum_{jk} jp_{jk}y^k}{\sum_{jk} jp_{jk}} = x\frac{\delta\mathcal{G}}{\delta x}\bigg|_{x=1}\left(\frac{\delta\mathcal{G}}{\delta x}\bigg|_{x,y=1}\right)^{-1} \tag{14}$$

If we start at a randomly chosen vertex and follow each of the edges at that vertex to reach the k nearest, first-order neighbours, then the vertices we arrive at have a distribution of outgoing edges generated by 14, less one power of x to account for the edge that we followed. Thus, the distribution of outgoing edges after having followed a random edge is generated by the function

$$\mathcal{G}_1(y) = \frac{\delta\mathcal{G}}{\delta x}\bigg|_{x=1}\left(\frac{\delta\mathcal{G}}{\delta x}\bigg|_{x,y=1}\right)^{-1} = \frac{1}{z_1}\frac{\delta\mathcal{G}}{\delta x}\bigg|_{x=1}. \tag{15}$$

where z_1 is, as above, the average vertex degree. We can now determine the distribution of second-order neighbours by summing this expression over the probabilities of a node to have k outgoing edges, but we have to be careful of two facts:

1. Some of the edges leaving a first-order neighbor connect to other first-order neighbors (clustering effect). In our model, this occurs on average cc times for a given vertex. We should not to take these nodes into account when counting the number of second-order neighbors.
2. The edge going from our initial node to a first-order neighbor might be bidirectional. This happens with a probability bc in our model. We must subtract this edge from the number of outgoing edge of a first-order neighbor when it occurs.

Consequently, the distribution of outgoing edges from first to second-order neighbors is

$$G_1(y) = (1-bc)\frac{1}{y^{cc}}\mathcal{G}_1(y) + bc\frac{1}{y^{cc+1}}\mathcal{G}_1(y). \tag{16}$$

As for the undirected case, the average number of second-order neighbors is

$$z_2 = G_0'(1)G_1'(1). \tag{17}$$

Finally, the condition $z_2 > z_1$ yields to

$$\sum_{j,k}(jk - j(bc+cc) - k)p_{jk} > 0. \tag{18}$$

■

Equation 18 marks the phase transition at which a giant component appears in a semantic connectivity graph. By neglecting the bidirectional and the clustering coefficient ($bc, cc = 0$) and reorganizing the terms using Equation 11 we fall back on the equation for the appearance of a giant component in a directed graph derived in [10]. Neglecting these two terms has of course a negative influence on the precision of our method (e.g., in highly clustered settings, where links connecting first-order neighbors should not be taken into account for deriving the phase transition).

In a directed graph, the giant component can be represented using a "bow-tie" diagram [7] as in Figure 4: The *strongly connected component* represents the portion of the graph in which every vertex can be reached from each other, while the *links-in* and *links-out* respectively stand for those vertices which can reach the strongly connected component but cannot be reached from it and those which can be reached from the strongly connected component but cannot reach it. We call the union of the links-in and of the strongly connected component the *in-component* and the union of the links-out and of the strongly connected component the *out-component*.

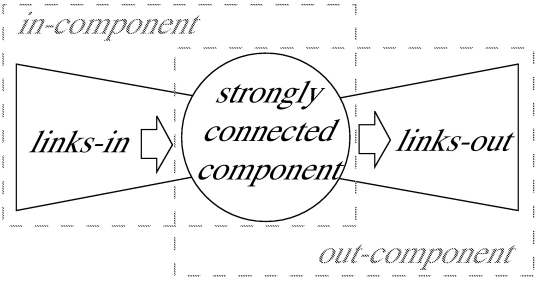

Fig. 4. The "bow-tie" diagram representing the giant component of a directed graph

Figure 5 below compares the evolution of the size of the biggest out-component in a random network of *10 000* vertices with the value of our new Connectivity Indicator $ci' = \sum_{j,k}(jk - j(bc - cc) - k)p_{jk}$ as the number of directed edges varies. The directed edges are added successively by choosing ordered pairs of vertices. At each step, we make sure that the graph remains non-redundant and irreflexive. As expected, the Connectivity Indicator becomes positive at the phase transition when a giant-component emerges and grows then with the size of that component.

5 Semantic Component Size

Even in a network where parties are not all semantically interoperable, a given peer can be tempted to send a query and observe how it gets propagated through

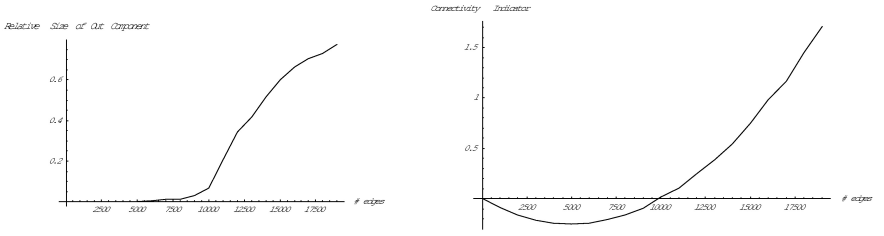

Fig. 5. Maximal out-component size and Connectivity Indicator for a random digraph of 10000 vertices

the different semantic domains. We can get a very good approximation of the degree of semantic diffusion of a query from our model.

Using a similar approach as described in [10] and taking advantage of our specific generating functions, we can calculate the relative size S of the subgraph which can be reached from the strongly connected component of the semantic connectivity graph (*out-component*):

$$S = 1 - G_0(u), \tag{19}$$

where u is the smallest non-negative real solution of

$$u = G_1(u). \tag{20}$$

Figure 6 shows the size of the out-component in a randomly generated digraph of *10 000* vertices with a varying number of edges. The two curves represent the relative size of the component (a) as evaluated using the degree distribution, the clustering coefficient and the bidirectional coefficient of the graph with the method described above and (b) as found in the graph. As the figure shows, the theory and practice are in good agreement (less than one percent of difference in the super-critical phase).

6 Use Case Scenarios

The methods described so far can readily be applied to study semantic interoperability of large-scale semantic systems in a global manner. Besides, we also believe in their high utility when used locally, e.g., by individual peers in the system. Peers can determine the statistical properties (degree distribution, clustering and bidirectional coefficients) of a semantic network in several ways:

- they can lookup the different values in the common registry of the system (see the Peer-to-Peer model in Section 2). This of course requires the different peers to insert their own local values in the repository beforehand.
- They can query a third-party tool (e.g., a semantic search engine) that regularly crawls the semantic graph to gather its statistical properties.

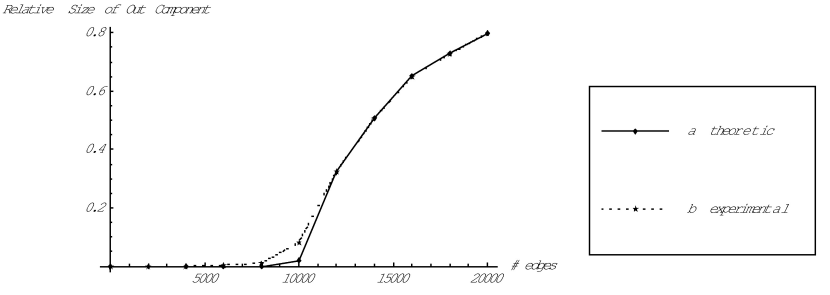

Fig. 6. Size Comparison of the out-component in a random network of *10 000* vertices

– They can approximate the statistical properties themselves, by gathering information from queries routed randomly through the semantic network (semantic random walkers).

Once gathered, the relevant data can be exploited in order to foster semantic interoperability in the large: When joining a semantic network, peers can determine wether the semantic network in question is semantically interoperable. If it is not, they can trigger the (automated or manual) creation of new translation links until the semantic connectivity subgraph moves to a super-critical phase ($ci > 0$). Such heuristics may have to be used periodically in environments where schemas and translations appear or disappear dynamically. Moreover, peers can evaluate the potential impact of a query based on a given schema: Once a network is semantically interoperable, peers can predict the degree to which a query will be forwarded through the Schema-to-Schema graph thanks to the component size analysis. Finally, note that our method could be applied at a finer granularity on classes also, to determine to which extent a given class c_i is known – in some form or another – throughout the network.

7 Concluding Remarks

So far, there exists little research on semantic interoperability in the large. Current approaches typically analyze a handful of schemas or ontologies at a time only. Research on large-scale systems (e.g., works on Web dynamics or social networks) cannot be directly applied to our problem because of its specificities (Section 2 and 3). We believe that new frameworks have to be developed in order to rightfully model the upcoming large-scale semantic systems. This paper pointed to one possible, and in our opinion promising, avenue by taking advantage of a recent graph-theoretic framework to analyze and iteratively realize semantic interoperability in a large network of information-sharing parties. This first work opens a whole range of extensions and improvements: Our next goal is to integrate weighted edges in the semantic connectivity model to analyze the

quality of translated queries. Also, we plan to integrate some of the heuristics presented above in our own semantic Peer-to-Peer system.

References

1. K. Aberer, P. Cudré-Mauroux, A. Datta, Z. Despotovic, M. Hauswirth, M. Punceva, and R. Schmidt. P-grid: A self-organizing structured p2p system. *ACM SIGMOD Record*, 32(3), 2003.
2. K. Aberer, P. Cudré-Mauroux, and M. Hauswirth. A Framework for Semantic Gossiping. *SIGOMD RECORD*, 31(4), December 2002.
3. K. Aberer, P. Cudré-Mauroux, and M. Hauswirth. Start making sense: The Chatty Web approach for global semantic agreements. *Journal of Web Semantics*, 1(1), December 2003.
4. K. Aberer, P. Cudre-Mauroux, M. Hauswirth, and T. van Pelt. GridVine: Building Internet-Scale Semantic Overlay Networks. In *International Semantic Web Conference (ISWC)*, 2004.
5. M. Arenas, V. Kantere, A. Kementsietsidis, I. Kiringa, R. J. Miller, and J. Mylopoulos. The Hyperion Project: From Data Integration to Data Coordination. *SIGMOD Record, Special Issue on Peer-to-Peer Data Management*, 32(3), 2003.
6. P. Bernstein, F. Giunchiglia, A. Kementsietsidis, J. Mylopoulos, L. Serafini, and I. Zaihrayeu. Data Management for Peer-to-Peer Computing : A Vision. In *International Workshop on the Web and Databases (WebDB)*, 2002.
7. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. http://www.almaden.ibm.com/cs/k53/www9.final.
8. S. Castano, A. Ferrara, S. Montanelli, and G. Racca. Semantic Information Interoperability in Open Networked Systems. In *International Conference on Semantics of a Networked World (ICSNW)*, 2004.
9. A. Y. Halevy, Z. G. Ives, P. Mork, and I. Tatarinov. Piazza: Data Management Infrastructure for Semantic Web Applications. In *International World Wide Web Conference (WWW)*, 2003.
10. M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev.*, E64(026118), 2001.
11. E. Rahm and P. A. Bernstein. A survey of approaches to automatic schema matching. *VLDB Journal*, 10(4), 2003.
12. H. S. Wilf. *Generatingfunctionology.* 2nd Edition, Academic Press, London, 1994.
13. H. Zhuge, J. Liu, L. Feng, and C. He. Semantic-Based Query Routing and Heterogeneous Data Integration in Peer-to-Peer Semantic Link Network. In *International Conference on Semantics of a Networked World (ICSNW)*, 2004.