ARTICLE IN PRESS

Computers in Human Behavior xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

How is the Semantic Web evolving? A dynamic social network perspective

Lina Zhou [a,*], Li Ding [b], Tim Finin [c]

[a] Department of Information Systems, Information Systems Department, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, United States
[b] Computer Science Department, Rensselaer Polytechnic Institute, United States
[c] Computer Science and Electrical Engineering, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore MD 21250, United States

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Semantic Web
Social network
FOAF
Dynamics
Evolution
Structural properties

ABSTRACT

Finding how the Semantic Web has evolved can help understand the status of Semantic Web community and predict the diffusion of the Semantic Web. One of the promising applications of the Semantic Web is the representation of personal profiles using Friend of a Friend (FOAF). A key characteristic of such social networks is their continual change. However, extant analyses of social networks on the Semantic Web are essentially static in that the information about the change of social networks is neglected. To address the limitations, we analyzed the dynamics of a large-scale real-world social network in this paper. Social network ties were extracted from both within and between FOAF documents. The former was based on knows relations between persons, and the latter was based on revision relations. We found that the social network evolves in a speckled fashion, which is highly distributed. The network went through rapid increase in size at an early stage and became stabilized later. By examining the changes of structural properties over time, we identified the evolution patterns of social networks on the Semantic Web and provided evidence for the growth and sustainability of the Semantic Web community.

1. Introduction

Semantic Web enables a new generation of decentralized knowledge management by enhancing information flow with machine-processable metadata (Cayzer, 2004). Since the vision for the Semantic Web was explicitly laid out in 2001 (Berners_Lee, Hendler, & Lassila, 2001), Semantic Web technologies have undergone significant advancement and the Semantic Web community has witnessed tremendous growth in scale and diversity. However, some argue that it is unrealistic to expect busy people and novice users to create and use enough metadata to make the Semantic Web work (cf. Borland, 2007). Additionally, it is not a trivial task to create and maintain ontologies that enable the representation of conceptual relationships. Therefore, it is important for both researchers and practitioners to understand how the Semantic Web community has actually been evolving.

Semantic Web is a platform that allows knowledge to be shared and reused across application, enterprise, and community boundaries. One way to generate new knowledge is by a process of socialization (Nonaka & Takeuchi, 1995). Socialization is one of the modes for transferring individual's knowledge and for expanding organizational knowledge (Nonaka, 1994). The emergence and wide adoption of social networking Web sites (e.g., del.icio.us for bookmarking and Flickr and YouTube for online media sharing) offers unprecedented opportunities for knowledge management. Semantic Web standards such as Friend of a Friend (FOAF) provide rich features that can be used to represent and infer social actors and ties. In this paper, we focus on studying social networks on the Semantic Web.

Despite a long standing history in studying traditional social networks in the physical world (e.g., Coleman, 1990; Freeman, 1979; Milgram, 1967), the availability of large online social networks has raised new questions and challenges. Much work to date has focused on the structure of a static snapshot of an evolving social network. To understand the evolving patterns of social networks requires longitudinal network data combined with information about individuals' attributes. However, longitudinal network data are rare, especially from the Semantic Web. Only a couple of recent studies have analyzed FOAF networks (Ding, Finin, & Joshi, 2005; Finin, Ding, Zhou, & Joshi, 2005). Their focuses were on cross-sectional or static analysis, or on social network metrics of individuals but not patterns of social structures. To address those limitations, we aim to investigate the evolution of Semantic Web technologies through the len of dynamic social network analysis.

Our contributions are: (1) We consider this work as an attempt to bridge the dynamics of a salient social network with social structures on the Semantic Web. (2) We provide the first empirical evidence for the evolution and growth of the Semantic Web community. (3) We discover the trend and patterns for social network evolution on the Semantic Web. (4) We lay the groundwork for

* Corresponding author. Tel.: +1 410 4558628; fax: +1 410 4551073.
E-mail addresses: zhoul@umbc.edu (L. Zhou), dingl@cs.rpi.edu (L. Ding), finin@umbc.edu (T. Finin).

0747-5632/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2010.07.024

building lifecycle models for knowledge management on the Semantic Web.

The rest of the paper is organized as follows. We first provide background on social networks on the Semantic Web and the evolution of social networks. Then, we analyze the evolution of social networks on the Semantic Web using a longitudinal real-world data set. Next, we discuss implications of our research findings and suggest future research directions. Finally, we conclude the paper by highlighting research contributions.

2. Background and research questions

We review related work and pose research questions in this section.

2.1. Social networks on the Semantic Web

A social network consists of a finite set or sets of actors and the relation or relations defined on them (Wasserman & Faust, 1994, p. 20). In case of human social networks, actors refer to individual persons while relations could be interpreted in a variety of ways. Depending on the source of data, relations in a social network could come from a verbal or written communication, scientific collaboration, kinship, physical or virtual proximity, and so on. Based on whether the links are explicitly described, we classify social networks into two types: salient and latent social networks. In latent social networks, social links are formed through shared resources or context such as co-membership and conversation relationship. As a result, two persons who are directly linked in a latent social network do not necessarily know each other. In contrast, in a salient social network, links are explicitly articulated in social networks such as FOAF, and such links generally reflect actual social relations.

On the Semantic Web, social network relations are represented with semantic information. FOAF defines a set of terms for letting users describe personal profiles including whom they know. Specifically, foaf:knows relations can form ties in social networks on the Semantic Web by directly linking two foaf:person. FOAF has been recognized as means of sharing social network data between social networking Web sites, and the ease of producing Semantic Web data is promoting this evolution (Golbeck & Hendler, 2006).

Social networks on the Semantic Web are considered as both online and salient. Online social networks are characterized by their openness and scale, where many actors come and go all the time. These features have an impact on the structure of social networks. Salient social networks in the traditional physical are featured with a small world phenomenon (Milgram, 1967) that there is a high tendency to form groups and communities. However, there is repeated evidence showing that many online social networks follow scale-free model rather than random or small world model (e.g., Faraj, Wasko, & Johnson, 2008; Finin, Ding, Zhou, & Joshi, 2005; Xu & Chau, 2006). In a scale-free network, a large percentage of nodes have just a few links, while a small percentage of the nodes have a large number of links. It remains unclear about the important properties of the social networks on the Semantic Web.

2.2. Dynamic social networks

Social networks are dynamic in essence and evolve over time (Doreian & Stokman, 1997). This is driven by the shared activities and affiliations of their members, by similarity of individuals' attributes, and by the closure of short network cycles (Watts, 1999). The dynamics of social networks may be considered in two ways – the dynamics of behavior within a network structure, or the evolution of the network itself over time (Metcalf & Paich, 2005). According to social network theory related to social dynamics, stability and dynamics operate in balance through many structural patterns (White, 2004).

Many of the extant studies were performed on static social networks whereas most real-world social networks are dynamic and evolving by nature. The dearth of papers that study the evolution of real-world social networks is partly due to the difficulty in obtaining time stamps for the arrival of every actor and tie arrival in an evolving real-world social network. Moreover, the study of network dynamics adds complexity to social network analysis. Social network analysis is an established field which proposes to analyze the relationships between social actors in a social network (Wasserman & Faust, 1994). The social network is often represented as a graph. In its simplest form, a social network graph contains nodes representing actors (generally people or organizations) and edges representing relationships or communications between the actors. Such information then enables reasoning about the individual actors and the network as a whole using graph-theoretic approaches. The analysis of dynamics social networks needs to go beyond traditional social network analysis by incorporating the temporal dimension of the network.

A typical way to address dynamics in a social network is to take snapshots of the network at various points in time and use these snapshots to make inferences about the evolutionary process (e.g., Kumar, Novak, Raghavan, & Tomkins, 2005). Additionally, each snapshot can be accumulative or constrained by a sliding window (Horn, Finholt, Birnholtz, Motwani, & Jayaraman, 2004). There are studies of structural properties of different snapshots of the world-wide web graph (Fetterly, Manasse, Najork, & Wiener, 2004; Ntoulas, Cho, & Olston, 2004). One of the studies (Leskovec, Kleinberg, & Faloutsos, 2005) gives insights into the evolution of graph properties over time. The dynamics of social ties in a social network can be shown by tracking the changes in large-scale data by periodically clustering data and examining the extracted temporal clusters (Kubica, Moore, Schneider, & Yang, 2002).

The properties of a dynamic social network could have great impact on the evolution of Semantic Web community and knowledge management practice on the semantic Web. However, extant social network research on the Semantic Web has mostly focused on static networks. It is assumed that all relations are essentially static in that all information about the time when the relations are forms is overlooked. The static nature of the findings can give incomplete and inaccurate information about structural patterns of the networks. Additionally, the static graph representation prevents us from even answering some fundamental questions about either the temporal patterns or the evolution of social structures. For example, how do the size and stability of social structures change over time? It remains unclear which model the evolution of semantic social networks follows. Therefore, we are mainly interested in the following research questions: Are the structural properties of social networks on the Semantic Web dynamic? How do they change over time? What's the lifecycle of Semantic Web documents?

To be able to answer these questions, we need to have information about when social relationships are formed, which allows us to examine social structure from the longitudinal perspective. This paper mainly attempts to address the discovery of temporal properties of social networks and the dynamics of social network patterns on the Semantic Web. As a secondary objective, the paper aims to understand the lifecycle of online social networks.

3. The evolution of social networks on the Semantic Web

In this paper we seek to explore how groups develop and evolve in large-scale social networks on the Semantic Web, specifically in the social networks formed between persons via foaf:knows relation within FOAF documents. Additionally, the ties between FOAF documents were created based on revision relations extracted from a longitudinal dataset.

There is a large body of work on identifying tightly-connected clusters within a given graph. The purpose of those work is to infer potential communities in a network based on the density of linkage. In contrast, social ties in FOAF social networks are explicitly identified and actors deliberately join together. In addition, we extract network data not just from selected communities but from the entire Web. Therefore, the questions addressed in the past literature are quite different from the focus of the current study.

3.1. A longitudinal Semantic Web dataset

We first detail our collection of a social network dataset. One needs a large, realistic social network containing a significant collection of explicitly identified groups, and with sufficient time-resolution that one can track their growth and evolution at the level of individual nodes (Backstrom, Huttenlocher, Kleinberg, & Lan, 2006). In this study, we take advantage of rich FOAF datasets and computational models for describing the process of group evolution.

The creation of the dataset took more than two years' effort spanning from January 2005 to March 2007. Compared with other types of Semantic Web documents (Ding et al., 2004), FOAF documents are unique in that FOAF offers a standard mechanism for directly expressing relationships between persons in triples. Therefore, we focus on studying social networks extracted from FOAF documents and their triples. In view that, the FOAF network is a salient social network, its dynamics is triggered by explicit revisions of FOAF documents. Thus, we captured the creation and revision history of FOAF documents on the Semantic Web, which has not been investigated in previous studies. The dataset consists of 688,298 FOAF documents, which totals about 1.3 M revisions comprising 270 M triples.The dynamic properties of the FOAF network were analyzed by taking snapshots at multiple time points. This approach involves generating a series of networks, each of which represents the FOAF network at a specific point in time. Each social network is represented as a directed time graph $G = (V, E)$, where V and E denote a set of nodes (e.g., persons) and edges (e.g., knows relation), respectively. For every node $v \in V$ and an associated directed edge $e = <u, v> \in E$ ($u \in V$), they have time stamps t on the time axis (t is a point in time within the given timeframe). v_t or $<u, v>_t$ indicates the time when node v or edge e is part of the graph. In particular, for any time t, there is a graph G_t that comprises the most recent versions of all the nodes and edges that remain alive by time t. We adopt the notion of time graph (Kumar, Novak, & Tomkins, 2006) in the analysis of properties that are specific to the dynamic networks and use graph dynamic networks the instantaneous network at any point in time such as the final network.

A window size of 9 months was used to divide the dataset. The choice was made base on the analysis of the distribution of revision cycles of FOAF documents that have ever been updated. Specifically, there are 454,678 (66%) FOAF documents only have one revision (i.e., they have remained unchanged ever since they were discovered), and the rest have 3.5 revisions on average. Additionally, based on the cumulative frequency distribution of revisions, as shown in Fig. 1, FOAF documents are revised at a relatively infrequent pace. For example, there are only about 10,000 documents that have over 10 versions and there are about 1000 documents that have over 20 revisions. In view of the observed periodicity of revision activities, sampling at nine-month interval would produce a reasonable approximation, of structural changes. Accordingly, we took four snapshots of the FOAF network, including $G_{2005\text{-}01}$, $G_{2005\text{-}09}$, $G_{2006\text{-}06}$, and $G_{2007\text{-}03}$, to analyze its evolution patterns.

3.2. Antecedents to network dynamics

In order to understand the magnitude of the antecedents to network dynamics (growth or loss), we analyzed 233,620 FOAF documents that have at least two revisions by tracking their changes from the first revision to the last revision. Fig. 2 summarizes the changes in the number of *person* instances and *knows* relations between the first revision and the last revision.

It is revealed from Fig. 2 that about half (50.82%) of the changes to FOAF documents (i.e., no change) was in semantic content other than FOAF *person* and *knows* relation or only in descriptions of FOAF persons. The above changes would keep social network intact. The rest of the changes to FOAF documents (49.185%) would contribute to the dynamics of social networks on the Semantic Web. These changes can be classified into four major categories: both growth, both loss, person growth only, and others. Among them, growth in both *person* instances and *knows* relations (35.94%) dominates the changes, and loss in both *person* instances and *knows* relations accounts for a relatively small percentage (10.52%). Compared with adding only persons to FOAF documents (2.13%), the chance of adding only knows relation is smaller (0.37%). This indicates that when a person is included in a FOAF document, he or she is likely to be connected via *knows* relation.

To help understand the overall changes of FOAF documents over time, Fig. 3 displays the trend of accumulative FOAF document revisions. The figure shows that the accumulation of FOAF document revisions increases steadily in a linear trend.

A further look into FOAF documents revisions reveals that the revisions can be classified into two major types: update of existing documents and creation of new documents. The distribution of the types of FOAF document revisions over time is reported in Fig. 4. It is noted from Fig. 4 that there is a spike of new creation activities in August 2005, which could have been caused by the upgrade of our crawler. There was a second spike around early 2006, suggesting an influx of new users to some FOAF online communities. The update frequency stays relatively constant but still reveals a general trend of high activities at the beginning and ending periods and low activities in between.

3.3. Structural properties of the FOAF network

Since we collected the FOAF dataset from a large number of resources on the Web, the network exhibits some unique structural properties and evolves in its own way. In this section, we first identify the structural properties of the entire FOAF network and then analyze their evolution in FOAF time graph by taking snapshots at different points in time.

Based on our analysis results, the FOAF network is mainly composed of two types of structural or (sub-)graph patterns: singletons and (connected) components. A *singleton* is a node that has no connection to other nodes, and a *component* is a set of nodes that are connected. As shown in Fig. 5, a singleton either has no edge or a self-loop edge. In fact, every node in the FOAF network has a self-loop because a person always knows herself. A *component* is a connected graph and disconnected with any other component. The components in the FOAF network can be further divided into four major sub-types namely star, non-star tree (ns-tree), twin, and complex components. A *tree* is a component having n nodes and $(n - 1)$ edges. A *star* is a special type of tree in which all but one node has zero out-degree. The remaining three types of components are collectively called non-star components. A *twin* is a component consisting of two nodes that are linked to each other, and a *complex* component is more complex than a *tree* or a *twin*.

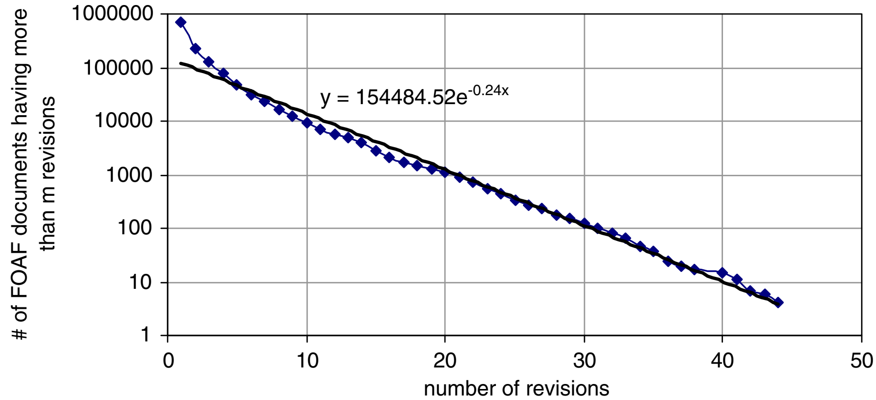


Fig. 1. The distribution of accumulative frequency of FOAF document revisions.

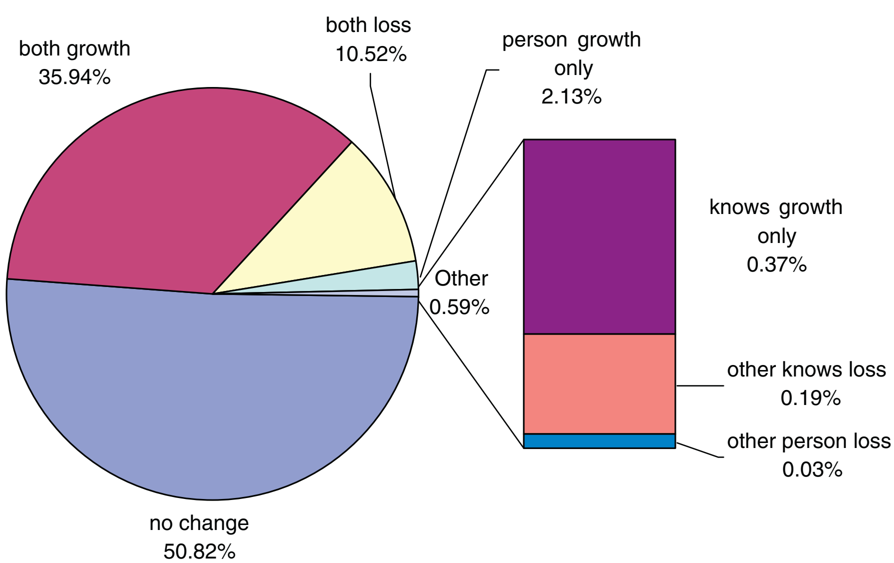


Fig. 2. Types of changes to FOAF documents.

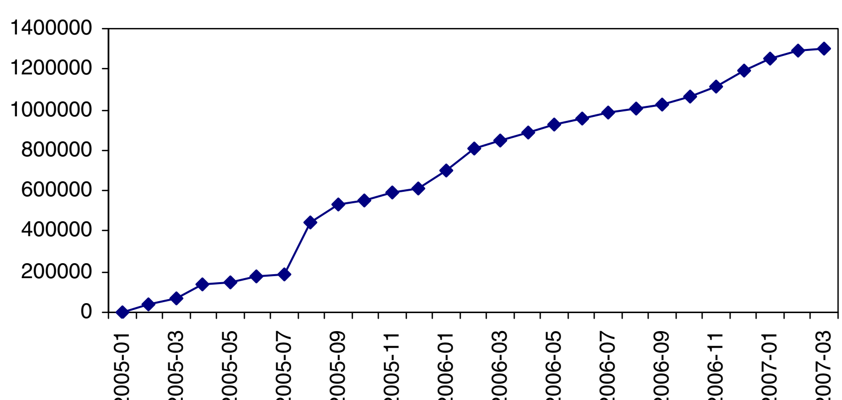


Fig. 3. The trend of accumulative FOAF document revisions.

The distribution of the above structural patterns is important information because it reflects the integration and complexity of the FOAF network. It will be more revealing to produce a time graph of the distribution. It is shown from Fig. 6 that the number of nodes in all types of structures keeps increasing over time, and the distribution of different types of structural patterns remains almost constant except for a slight change in the order between ns-tree and twin in $G_{2007\text{-}03}$.

Compared with other types of patterns, star is dominant across all the snapshots of the FOAF network. Star outnumbers the rest by at least one order of magnitude in the last three snapshots, and singleton is consistently ranked the second. In the last snapshot, star

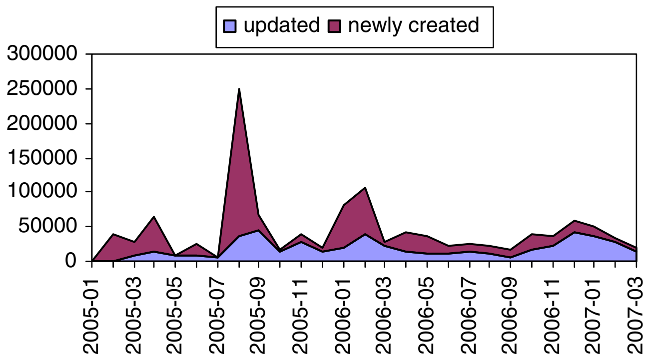


Fig. 4. The time graph of two types of FOAF document revisions.

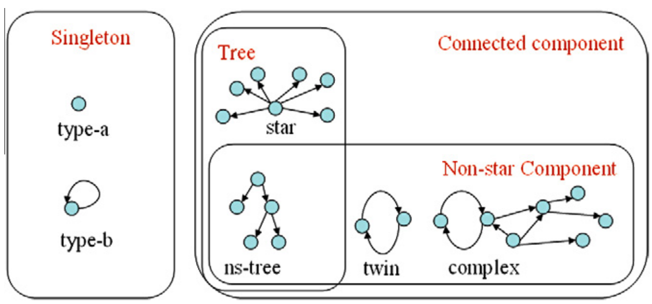


Fig. 5. Typical structural patterns of FOAF networks.

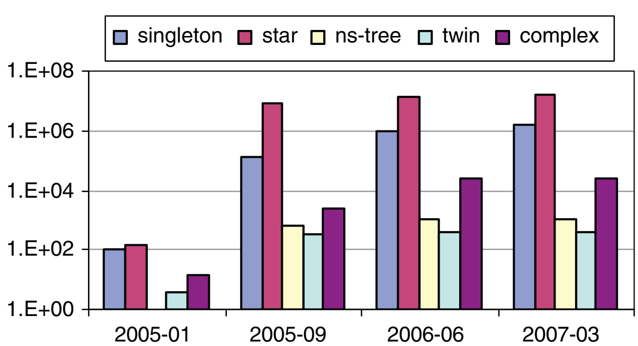


Fig. 6. The evolution of node distribution (grouped by structural patterns).

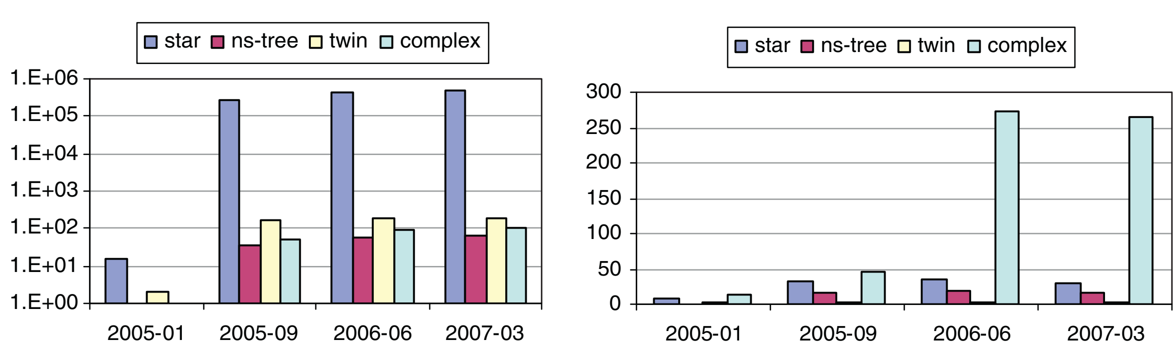
accounts for 90% of the nodes, and the other non-singleton components only totals less than 1% of the nodes. The number of non-star components is very small because most person instances (99.92%) in FOAF documents are defined as blank-nodes, which are local within their residing documents. Therefore, we hypothesize that the authors of FOAF documents do not want to publicize their friends due to privacy concerns, and if they do, they prefer to use other types of social identities rather than Uniform Resource Identifier (URIs) to link their friends. Such a decision could be explained from three perspectives: (1) it is harder to remember an URI than most other types of social identities; (2) the population who has adopted URI remains small; and (3) it is practically challenging for an individual to maintain a permanent URI on the Web despite that it is well intended.

Among the non-star components, complex components consistently contain the most nodes. In sum, Fig. 6 shows that the FOAF network is rather distributed and disconnected. If nodes do get connected beyond stars, they tend to have a complex structure. This motivates us to look deep into complex components in the next section.

The node distribution across different component patterns may be biased because some types of patterns contain significantly more nodes that others by nature. In order to provide a complete picture for the distribution of structural patterns, especially components, we generated two benchmark statistics: (1) the overall population (component frequency) and (2) the average size (number of nodes). Fig. 7a shows that *star* is by far the predominant type of component, outnumbering other types of components by about three orders of magnitude. Such an observation provides preliminary evidence in support of our hypotheses that users prefer to link to their friends indirectly. Fig. 7b shows that *complex* components on average involve much more nodes than any other types of components, especially after the space of FOAF documents has evolved over time. We also noted a non-trivial number of non-star trees and twins among the non-star components, and our decision on setting the two apart from complex components is mainly due to their simple structure and tiny size (see Fig. 7b).

Despite a general trend of growth in both the overall population and the average size of components, we noticed a slight drop in the average size of *complex* components in $G_{2007\text{-}03}$. This is mainly because, compared with $G_{2006\text{-}06}$, the relative increase in the number of components (12.2%) outweighs the relative increase in the number of nodes in $G_{2007\text{-}03}$.

Degree analysis helps us understand the level of connectivity in FOAF components. We first compute the average in-degree (or out-degree) for every component, and then derive the average of average in-degree for every type of structural patterns by aggregating all the components of that type. The results are plotted in Fig. 8. It is not surprising that the average of average in-degrees of *complex* components is significantly higher than others and continues to grow until $G_{2006\text{-}06}$. A slight drop in average degree of $G_{2007\text{-}03}$ is caused by the merging of some stars into existing complex



(a) overall population (b) average size

Fig. 7. The evolution of population and average size of components (grouped by structural patterns).

Please cite this article in press as: Zhou, L., et al. How is the Semantic Web evolving? A dynamic social network perspective. *Computers in Human Behavior* (2010), doi:10.1016/j.chb.2010.07.024

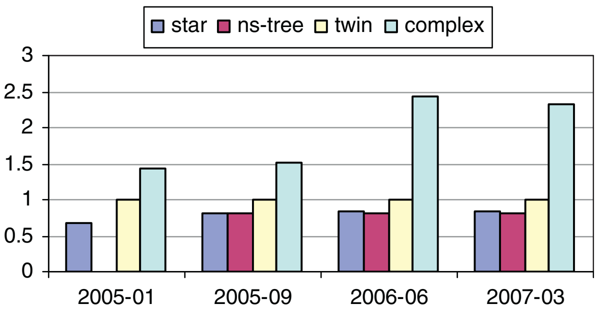


Fig. 8. The evolution of average in-degree distribution of components (grouped by structural patterns).

components. The average in-degrees of other types of components range between 0 and 1.

In view that the FOAF network is directed, we are interested in the extent to which ties are reciprocated. It is argued that a network with a predominance of asymmetric ties may be unstable and unequal (Borgatti, Everett, & Freeman, 2002). Thus, stars and trees are unstable by definition, whereas twins are in equilibrium state because ties in twins are always reciprocated. Thus, we focus on analyzing the reciprocity of complex components. Commonly used empirical measures of reciprocity are based on unweighted ties and differentiate between null and asymmetric dyads versus mutual dyads (Wasserman & Faust, 1994). We are concerned with the number of ties that are involved in reciprocal relations relative to the total number of actual ties. Results show that the average reciprocity of complex components increases from 30% for $G_{2005\text{-}01}$ to 35.3% for $G_{2005\text{-}09}$ and then to 43.1% for $G_{2006\text{-}06}$. There is a slight decrease in the average reciprocity to 41.3% for $G_{2007\text{-}03}$, as shown in Fig. 8.

Finally, we examine the density of FOAF components. The density of a network is the number of actual connections between nodes divided by the number of possible connections (Scott, 2000). Density values range from 0 to 1. The higher the density the more connected the group members are. The average density values of FOAF component time graphs are reported in Fig. 9. Generally speaking, the densities of FOAF components except twin are not high. The moving trends of average density are more diverse than other network properties. For example, the average density of star decreases, and that of twin stays the same, but those for non-star tree and complex components increase over time. The results indicate that, as FOAF components evolve, there is a tendency for non-star tree and complex components to move toward equilibrium but star tends to move in the opposite direction.

3.4. Structural properties of the largest component

By tracking network properties with respect of the largest FOAF component, it is possible to measure the rate and extent of integration of the community over time. Specifically, we analyzed the temporal patterns of component network properties, including average in-degree, reciprocity, diameter, density, clustering coefficients, and average distance.

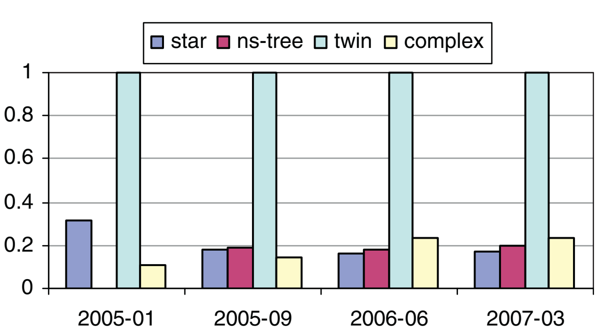


Fig. 9. The evolution of density distribution of components (grouped by structural patterns).

The evolution of the largest star and non-star components are plotted in Fig. 10. At each time point, components are listed from the left to the right in the descending order of their size. A component is represented as a rectangular block. To differentiate star components from non-star ones, the star components are marked with "∗" in the corresponding blocks. The number listed at the top of a blank denotes the identifier and the two numbers at the bottom denote the total number of persons and the total number of ties in the component, respectively. The evolution relationship between components for different points of time is depicted as arrows on the vertical dimension, which was derived through tracing the movement of persons in specific components over time.

We observed several interesting evolution patterns of components from Fig. 10. First, the largest non-star component remains as the largest over time. For example, components 38,382, 18,772, and 1532, which fall on the same evolution path, are also the largest in $G_{2005\text{-}09}$, $G_{2006\text{-}06}$, and $G_{2007\text{-}03}$ separately. The size of the largest component increases from 984 persons along with 1938 ties in $G_{2005\text{-}09}$ to 8140 persons along with 12,815 ties in $G_{2006\text{-}06}$, and then to 9035 persons along with 5381 ties in $G_{2007\text{-}03}$. Second, two or more non-star components could be combined into one largest component. For example, component 18,772 is combined with component 1532 in $G_{2006\text{-}06}$ and become one component in $G_{2007\text{-}03}$. Third, a star component can follow three evolving paths: merge into a non-star component such as component 178,110 in $G_{2005\text{-}09}$, grow on its own such as component 25,532,381 in $G_{2005\text{-}09}$, and completely disappear such as component 25,554,051 in $G_{2005\text{-}09}$. Since the power in a star component is highly centralized, the component could break down as along as the powerful person goes away.

Given the lack of diversity in the structures of star component, the analysis of the largest component focuses on non-star components. First, the overall reciprocity of the largest complex component in $G_{2005\text{-}09}$ was 2.43%, which was increased significantly to 13% in $G_{2006\text{-}06}$ and then to 14.29% in $G_{2007\text{-}03}$. The continual increase in reciprocity indicates that the ties in social networks on the Semantic Web move toward mutuality as in traditional social networks. Nonetheless, the level of reciprocity remains at a low level. We can find two alternative explanations for this phenomenon. First, there exists asymmetry in terms of the power or the social role of individuals. As a result, it is more likely for a person at a low power to link to other persons at high power positions than the other way around. Second, given the distributed nature of the Semantic Web, there is some natural delay in reciprocating social ties from different FOAF documents.

At the individual level, we also analyzed the diameter, mean node in-degree (or out-degree), and average shortest distance of the largest components. The results are reported in Fig. 11. The mean in-degree and mean out-degree are the same for a directed network. Nonetheless, the mean standard deviations of the out-degree are generally much higher than those of in-degree. For example, the ratio of standard deviation of in-degree to that of out-degree was 10.43:3.34. There is a slight decline in the mean in-degree in $G_{2006\text{-}06}$ and then the trend was reversed in $G_{2007\text{-}03}$. It suggests that the grow in the number of persons in a FOAF component precedes the grow in the number of ties. This also explains the trends of diameter and average distance, which show significant jumps at the beginning and then level out later.

Further, we also looked at the density of these graphs. The density of the largest component time graph is depicted in Fig. 12 where the value of density is increased by a factor of 100. Generally, the density of the largest remains quite low over time low.

Please cite this article in press as: Zhou, L., et al. How is the Semantic Web evolving? A dynamic social network perspective. *Computers in Human Behavior* (2010), doi:10.1016/j.chb.2010.07.024

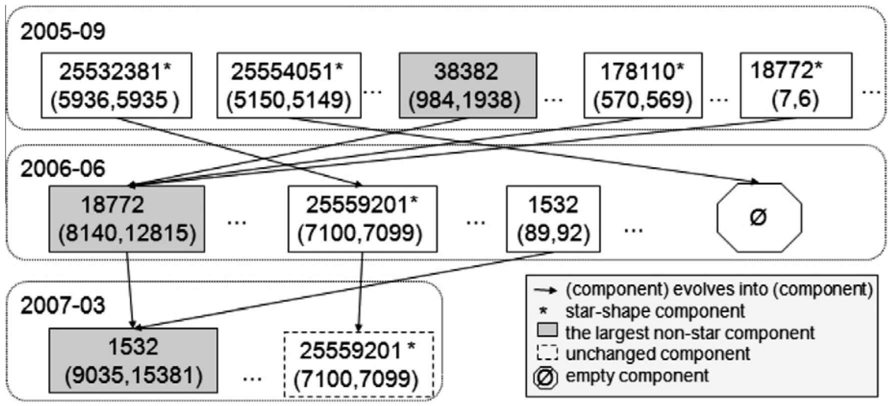


Fig. 10. The evolution of the largest non-star and star components.

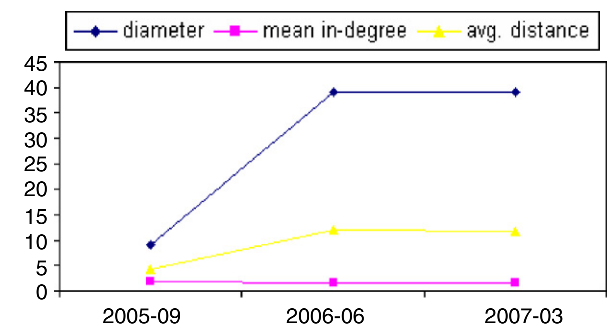


Fig. 11. The evolution of diameter, mean in-degree, and average distance of the largest component.

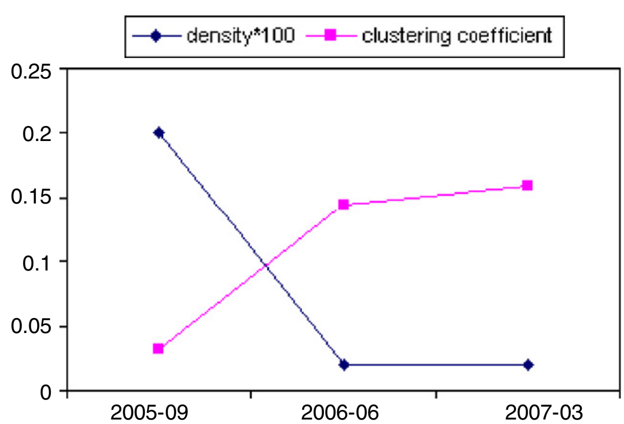


Fig. 12. The evolution of density and clustering coefficient of the largest component.

For example, the largest density is only 0.002, which occurs at the first time point. It implies that information diffusion in the current Semantic Web may not be efficient.

Unlike certain citation graphs that become denser over time (Leskovec et al., 2005), Online social networks exhibit three clearly marked stages (Kumar et al. 2006): an initial upward trend leading to a peak, followed by a dip, and the final gradual steady increase. Given that the density of the largest component in the FOAF social network undergoes a decline stage and then a stable stage, as shown in Fig. 12, we expect that the density is likely to increase and the network is likely to grow tighter in future. The FOAF community will finally set into a phase of organic growth in which both membership and ties increase.

Finally, the clustering coefficient is the likelihood that any two nodes that are connected to the same node are connected themselves (Watts & Strogatz, 1998). As shown in Fig. 12, the clustering coefficient follows an increasing trend, and the trend slows down over time. The result shows that there is a tendency for localized neighborhood or communities to emerge on the Semantic Web. The increase in the size of the community enhances the above trend.

4. Discussion

From the perspective of the Semantic Web community, an influx of new members could bring new ideas with them to stimulate Semantic Web research and adoption. On the other hand, high volatility of community members could endanger the formation of community identify. By examining the dynamics of structural properties of the FOAF network from the longitudinal perspective at both community and individual levels over time, we advance our knowledge about the evolution of FOAF social networks. We discuss the findings of this study and their implications to the Semantic Web and social network research in this section. We also discuss limitations and future research issues.

4.1. Findings and implications

There are several key takeaway points from this study. The first is that the Semantic Web community is highly distributed, which is similar to the traditional Web. It does not belong to the type of network with a highly connected core but rather contains a collection of disconnected components. Moreover the FOAF network grows in a "speckled" fashion. Further, as the Semantic Web continues to grow, there is a tendency for the community to become more integrated and stabilized.

Second, the FOAF social network community contains the majority of its mass outside the largest component, and the structure outside the largest component is largely characterized by singletons and stars. The large number of singletons and stars largely results from blank nodes where people do not directly link to their friends in FOAF documents.

Third, social networks in the Semantic Web appear to go through distinct stages of growth, characterized by specific behavioral patterns in terms of density, reciprocity, degree, and regularity of component structure. For example, the distribution of different types of structural patterns in components is fairly constant. Both the size and average degree of components undergo rapid increase, slow increase, and relatively stable or slow decline

Please cite this article in press as: Zhou, L., et al. How is the Semantic Web evolving? A dynamic social network perspective. *Computers in Human Behavior* (2010), doi:10.1016/j.chb.2010.07.024

phases. It would be instrumental to develop a more detailed model of the lifecycle of online social networks.

Fourth, our results also have research implications for time series and cross-sectional network analysis. This study shows that different network structural properties are in equilibrium (Doreian & Stokman, 1997) but exhibit varying levels of stability over time with respect to a 9-month elapse.

This study contributes to a growing body of research focused on large-scale networks. By studying FOAF social networks at both individual and community levels, we are able to provide a deeper and more complete understanding of how individuals manage implicit knowledge embedded in salient social networks. This study also contributes to knowledge management research by revealing the temporal patterns of FOAF document maintenance. Previous work has either used a single version of FOAF documents or fail to link different versions of the same documents together. Tracing the versioning relationship between documents would enable the discovery of lifecycle of social network management. Further, the findings of this study shed light on the evolution of online FOAF community in specific and the Semantic Web community in general. They reveal that the community has grown at a steady pace into a more integrated space. Nonetheless, there is much room for the community to grow into tight connection given the huge number of components and low average degrees. We should seek a balance between dynamics and stability to enhance the growth and the sustainability of the Semantic Web community.

4.2. Limitations and future research

As with any study, there are limitations that could be addressed in future studies. First, only the foaf:knows relationship was explored for social networks on the Semantic Web. FOAF provides additional personal information such as foaf:see also and foaf:interest that are potential candidates for building online social networks. Questions regarding how strong the relationships between FOAF members are require an analysis of a more complete set of relationships. Second, our analysis of revision history of online social networks is unique and appropriate to the present study of the evolution of dynamic social networks on the Semantic Web. It would require a deeper analysis of the evolving patterns of individual FOAF documents or expressed relationships to develop a formal lifecycle model for Semantic Web documents or dynamic online social networks. Third, although FOAF promotes the idea of associating each individual with a unique identifier, entity resolution (Aleman-Meza et al., 2006; Finin et al., 2005) remains to be an issue in processing real-world data. There is still a lack of perfect solutions for resolving identities in an online environment, which merits future improvement. Fourth, we assume that the social network relationships persist over time. In reality, some FOAF documents may temporarily go offline or may permanently disappear. It would be interesting to investigate the motivations for creating, revising, and deleting online social networks. Fifth, the number of time points selected for analysis restricts the types of dynamics that could be observed in this study. Nonetheless the duration of our data collection and the number of observations are unmatched in any of the previous studies of FOAF networks. We believe the duration of data collection is large enough to capture all the critical events. An analysis at a more granular level of time would help reveal more subtle dynamics of online social networks.

Although it is risky to generalize findings from FOAF social networks to the entire online social networks, this study highlights the need to examine the evolution of salient online social networks from the perspective of knowledge management. Future research would to investigate the impact of social network evolution on other types of knowledge management processes through social network analysis. For example, the analysis of structural changes and shared interests of dynamic social networks can provide insights into knowledge sharing and diffusion, which could be key drivers in the healthy growth of online social networks.

5. Conclusions

This study presents a first examination of the relationship between salient social networks and knowledge management to highlight the greatest social assets accrued to the Semantic Web community over the past few years. By exploring the ways in which communities in FOAF social networks dynamically grow over time – both at the level of individuals and at a global level of communities, this study discovers the evolution patterns of social structures and predicts future trend. It is hoped that the findings of this research will motivate a closer more investigations of the distinct characteristics of salient online social networks that can be used to support the management and sharing of implicit knowledge.

References

Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A. P., et al. (2006). *Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection.* Paper presented at the WWW, Edinburgh, Scotland (May 23–26).

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: Membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 44–54). Philadelphia, PA, USA: ACM Press.

Berners_Lee, T., Hendler, J., & Lassila, O. (2001). The semantic Web. *Scientific American, 284*(5), 34–43.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for windows: Software for social network analysis.* Harvard, MA: Analytic Technologies.

Borland, J. (2007). Part II: A smarter Web: New technologies will make online search more intelligent – and may even lead to a "Web 3.0". *Technology Review* (March 20).

Cayzer, S. (2004). Semantic blogging and decentralized knowledge management. *Communications of the ACM, 47*(12), 47–52.

Coleman, J. (1990). *Foundations of social theory*. Harvard.

Ding, L., Finin, T., & Joshi, A. (2005). Analyzing social networks on the semantic Web. *IEEE Intelligent Systems, 1*(9), 1–10.

Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Sachs, J., et al. (2004). *Swoogle: A search and metadata engine for the semantic Web.* Paper presented at the thirteenth ACM conference on information and knowledge management (CIKM), Washington DC (November).

Doreian, P., & Stokman, F. N. (Eds.). (1997). *Evolution of social networks*. New York: Gordon and Breach.

Faraj, S., Wasko, M., & Johnson, S. L. (2008). Electronic knowledge networks: Processes and structure. In I. Becerra-Fernandez & D. Leidner (Eds.), *Knowledge management: An evolutionary view of the field* (pp. 270–291). Armonk, NY: M.E. Sharpe Inc..

Fetterly, D., Manasse, M., Najork, M., & Wiener, J. L. (2004). A large-scale study of the evolution of web pages. *Software: Practice and Experience, 34*(2), 213–237.

Finin, T., Ding, L., Zhou, L., & Joshi, A. (2005). Social networking on the semantic Web. *The Learning Organization, 12*(5), 418–435.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215–239.

Golbeck, J., & Hendler, J. (2006). Inferring binary trust relationships in Web-based social networks. *ACM Transactions on Internet Technology, 6*(4), 497–529.

Horn, D. B., Finholt, T. A., Birnholtz, J. P., Motwani, D., & Jayaraman, S. (2004). *Six degrees of jonathan grudin: A social network analysis of the evolution and impact of CSCW research.* Paper presented at the proceedings of the 2004 ACM conference on computer supported cooperative work, Chicago, Il, USA.

Kubica, J., Moore, A., Schneider, J., & Yang, Y. (2002). *Stochastic link and group detection*. Paper presented at the Eighteenth national conference on Artificial intelligence, Edmonton, Alberta, Canada.

Kumar, R., Novak, J., Raghavan, P., & Tomkins, A. (2005). On the bursty evolution of blogspace. *World Wide Web Journal, 8*(2), 159–178.

Kumar, R., Novak, J., & Tomkins, A. (2006). Structure and evolution of online social networks. In *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 611–617). Philadelphia, PA, USA: ACM Press.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005). *Graphs over time: Densification laws, shrinking diameters and possible explanations*. Paper presented at the proceeding of the eleventh ACM SIGKDD international conference on knowledge discovery in data mining, Chicago, Illinois, USA.

Metcalf, S., & Paich, M. (2005). *Spatial dynamics of social network evolution*. Paper presented at the 23rd international conference of the system dynamics society, Boston, MA (July 17–21).

Milgram, S. (1967). The small world problem. *Psychology Today, 2*(May), 60–67.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science, 5*(1), 14–37.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company*. New York: Oxford University Press.

Ntoulas, A., Cho, J., & Olston, C. (2004). What's new on the web?: The evolution of the web from a search engine perspective. In *Proceedings of the 13th international conference on World Wide Web* (pp. 1–12). New York, NY, USA: ACM Press.

Scott, J. (2000). *Social network analysis: A handbook* (2nd ed.). Thousands Oaks, Calif: Sage Publications.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Watts, D. J. (1999). *Small worlds, the dynamics of networks between order and randomness*. Princeton: Princeton University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440–442.

White, D. R. (2004). Network analysis and social dynamics. *Cybernetics and Systems, 35*(2–3), 173–192.

Xu, J., & Chau, M. (2006). *Mining communities of bloggers: A case study on cyber-hate*. Paper presented at the international conference on information systems, Milwaukee, WI (December 10–13).

Please cite this article in press as: Zhou, L., et al. How is the Semantic Web evolving? A dynamic social network perspective. *Computers in Human Behavior* (2010), doi:10.1016/j.chb.2010.07.024