Network Analysis of Semantic Connections in Heterogeneous Social Spaces

Sheila Kinsella, Andreas Harth, John G. Breslin
Digital Enterprise Research Institute
National University of Ireland, Galway
Galway, Ireland

Abstract

The traditional idea of a social network as a set of individuals and relationships is deficient in that it does not consider the shared objects which bring people together. By taking into account the role of external objects in social relations, we can glean a more thorough understanding of the factors which influence the formation of interpersonal links. We investigate an object-centred social network extracted from the Semantic Web. We analyse the patterns of connectivity that exist among objects and people, and examine the role of objects in creating social ties.

1. Introduction

Many social network models are based on homogeneous graphs which are restricted to describing individuals and the links that exist between them. These representations do not include the external influences that explain why particular people share a relationship. While analysis of this network may reveal how people are connected, it can only produce limited results in terms of understanding why they are connected.

An alternative view of social networks has been described by Jyri Engestrom[2], based on the theory of sociologist Karin Knorr Cetina, which suggests that people do not just connect, they connect through shared objects. In real life, social ties are embedded in things like jobs and common activities. Online communication also revolves around objects, for example blogs and images. Often relationships are not explicitly stated, but exist via the medium of an intermediary object (see Figure 1). Therefore, in order to more accurately characterise the relations in a community, we must model social systems as heterogeneous networks, composed not only of people but also of objects.

The primary objective of this paper is to examine the topology of a social network composed of objects and people. We identify particular communities extracted from a datagraph gathered from the Semantic Web[1]. Our hypothesis is that the objects in the network play an important part in forming implicit links between people. We aim to validate this idea by creating and comparing two models of our dataset: a people-only social network, and an object-centred social network. We also identify and discuss the main types of objects and relationships present in our object-centred model.

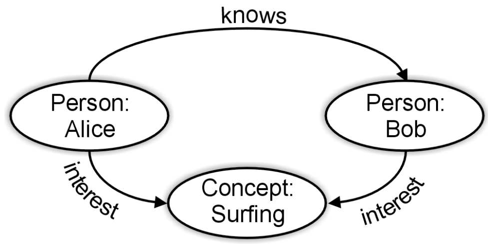

Figure 1. Direct and indirect links in a social network

2. Semantic Data Model

In order to construct an object-centred social network, we use data collected from the Semantic Web, an extension of the World Wide Web which enables the description of arbitrary objects and the relationships between them, using shared machine-readable formats. Each object has a unique Uniform Resource Identifier (URI), enabling it to be referenced across different sources. This common method of expressing information allows us to easily integrate diverse interlinked objects for analysis.

Our data is expressed using the Resource Description Framework (RDF)[1] language. An RDF document is composed of statements of the form $< subject, predicate, object >$, indicating a directed link from the subject node to the object node, where

1. http://www.w3.org/2001/sw/

the predicate link describes the relationship between them. The nodes represent instances of classes, and the links represent instances of properties. Classes and the possible properties which can exist between them are defined in schemas, which can be created or extended as required. Namespaces are used to indicate the schema to which classes and properties belong.

Schemas exist for describing concepts from many different domains. Central to our analysis is the Friend-of-a-Friend (FOAF) project[2], which defines a widely-used vocabulary for describing people and the relationships between them. Data expressed using this vocabulary provides the source of people in our social network. FOAF is integrated with a multitude of other vocabularies on the Semantic Web, from which we can extract the links that exist between people and objects. This will enable us to identify hidden links between people, and to analyse the type of objects around which relationships and communities form.

3. Data Extraction

We start with a massive RDF datagraph from many diverse sources. To obtain a specific community for analysis, we take the FOAF file of Tim Berners-Lee, inventor of the World Wide Web, and extract all documents within five links of this root node. We repeat the process for paper author Andreas Harth. To avoid results from any one source dominating the graph, we limit the number of documents from a single domain to 200 for Tim Berners-Lee, and 2000 for Andreas, therefore the dataset for Andreas is somewhat larger. We omit any instances whose class is unknown, as their inclusion would impede our investigation of relationships between people and objects. From each of the RDF graphs, we identify two subgraphs of interest - a people-only network, and an object-centred network.

- People-only network - Composed only of people who are either directly connected to the root node, or connected indirectly via other people.
- Object-centred network - Contains all of the nodes in our community who are directly or indirectly connected to the root node.

4. Network Analysis

We use the Java Universal Network/Graph Framework[3] to model and analyse both of the networks described above, for each of the RDF graphs. The object-centred network of Tim Berners-Lee comes from 55 different documents, and the object-centred network of Andreas comes from 1613 different documents.

2. http://www.foaf-project.org/
3. http://jung.sourceforge.net/

4.1. Effect of the inclusion of objects

Tables 1 (Tim Berners-Lee) and 2 (Andreas) show the sizes of the extracted networks. People still make up the majority of nodes in the object-centred networks. It is clear that the inclusion of indirect interpersonal links, via objects, has revealed hidden relations existing in our dataset. In fact the number of individuals in both object-centred networks are approximately double that of the respective people-only network.

Graph	Nodes (People)	Edges
People-Only	349 (349)	450
Object-Centred	1052 (691)	1446

Table 1. Statistics of the people-only and object-centred networks of Tim Berners-Lee.

Graph	Nodes (People)	Edges
People-Only	9011 (9011)	24875
Object-Centred	24513 (16671)	47286

Table 2. Statistics of the people-only and object-centred networks of Andreas Harth.

4.2. Analysis of object types

We analyse the frequency of the different classes and properties present in our dataset. The schemas most commonly found in our dataset are described in Table 3. XML qualified names are used as shorthand for full URI references. The prefix indicates the namespace, and the text after the colon is the name of the class or property. Table 4 lists the main classes present in Tim Berners-Lee's object-centred graph, while Table 5 lists the main properties. Table 6 and Table 7 show the same information for Andreas.

While the "knows" relationship forms the majority of connections between people in these subgraphs, the top classes and properties show that people are additionally related through their geographic proximity (based_near), their common interests and the topics associated with documents, through events they attend together, papers they co-author, etc.

5. Conclusion

This paper presents an empirical analysis of the networks formed by people through direct links and

Namespace	Namespace URI	Description
foaf	http://xmlns.com/foaf/0.1/	People and relationships schema
rss	http://purl.org/rss/1.0/	Web feed formats schema
wordnet	http://xmlns.com/wordnet/1.6/	English language lexical database
ical	http://www.w3.org/2002/12/cal/icaltzd#	Calendar data schema
rdfs	http://www.w3.org/2000/01/rdf-schema#	RDF vocabulary description language
rev	http://www.purl.org/stuff/rev#	Review vocabulary
trust	http://www.mindswap.org/~golbeck/web/trust.daml#	Trust ratings vocabulary
smw	http://smw.ontoware.org/2005/smw#	Semantic MediaWiki schema
geo	http://www.w3.org/2003/01/geo/wgs84_pos#	Geographical information vocabulary
ow	http://ontoworld.org/wiki/Special:URIResolver/	Ontoworld Wiki schema

Table 3. URIs and descriptions of the most common namespaces in the datasets

	Class	# of Instances
1	foaf:Person	691 (65.7%)
2	foaf:Document	76 (7.2%)
2	rss:channel	63 (6.0%)
4	rss:item	36 (3.4%)
5	rev:Review	33 (3.1%)
6	foaf:Agent	25 (2.4%)
7	foaf:PersonalProfileDocument	20 (1.9%)
8	foaf:Image	13 (1.2%)
9	wordnet:Airport	11 (1.0%)
10	ical:Vevent	8 (0.8%)
		976 (92.8%)

Table 4. Most frequently occurring classes in the Tim Berners-Lee object-centred graph.

	Class	# of Instances
1	foaf:Person	16671 (68.0%)
2	rss:item	1386 (5.7%)
2	foaf:PersonalProfileDocument	1263 (5.2%)
4	foaf:Agent	783 (3.2%)
5	foaf:Document	646 (2.6%)
6	smw:Thing	631 (2.6%)
7	rss:channel	532 (2.2%)
8	bio:Birth	389 (1.6%)
9	geo:Point	368 (1.5%)
10	foaf:Image	140 (0.6%)
		22809 (93.0%)

Table 6. Most frequently occurring classes in the Andreas Harth object-centred graph.

	Property	# of Instances
1	foaf:knows	583 (40.3%)
2	rdfs:seeAlso	136 (9.4%)
3	foaf:maker	82 (5.7%)
4	foaf:weblog	68 (4.7%)
5	foaf:interest	67 (4.6%)
6	foaf:topic	59 (4.1%)
7	foaf:made	41 (2.9%)
8	foaf:member	41 (2.9%)
9	rev:reviewer	33 (2.3%)
10	trust:trustsModerately	24 (1.7%)
		1134 (78.4%)

Table 5. Most frequently occurring properties in the Tim Berners-Lee object-centred graph.

	Property	# of Instances
1	foaf:knows	31791 (67.2%)
2	rdf-schema:seeAlso	2333 (4.9%)
3	foaf:maker	1495 (3.2%)
4	foaf:member	1312 (2.7%)
5	foaf:topic	1121 (2.4%)
6	foaf:weblog	910 (1.9%)
7	ow:Relation-3AParticipant_of	455 (1.0%)
8	bio:event	394 (0.8%)
9	foaf:based_near	370 (0.8%)
10	ow:Relation-3AMember_of	365 (0.8%)
		40546 (85.7%)

Table 7. Most frequently occurring properties in the Andreas Harth object-centred graph.

indirectly through the objects that they are connected to: e.g., that they create, share, are interested in. The dataset used is taken from a range of sources which have been collected through semantically-enabled sources of information. In summary, our work demonstrates a wider view of networks that is not normally apparent with disconnected social spaces.

6. Acknowledgements

This work has been partially funded by Science Foundation Ireland under grant number SFI/02/CE1/I131 and a Supplemental Equipment Grant. We acknowledge the work of the Semantic Web Search Engine group (http://www.swse.org/) which provided infrastructure services for data collection, storage, and retrieval.

References

[1] S. Decker, S. Melnik, F. van Harmelen, D. Fensel, M. C. A. Klein, J. Broekstra, M. Erdmann, and I. Horrocks. The semantic web: The roles of XML and RDF. *IEEE Internet Computing*, 4(5):63–74, 2000.

[2] J. Engestrom. Object-centered sociality. *Reboot 7*, 2005.