Proceedings of 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN 2018)

Complex Network Based Knowledge Graph Ontology Structure Analysis

Yuehang Ding
General Department
National Digital Switching System Engineering & Technological Research Center
Zhengzhou, China
739815262@qq.com

Hongtao Yu
General Department
National Digital Switching System Engineering & Technological Research Center
Zhengzhou, China

Ruiyang Huang
General Department
National Digital Switching System Engineering & Technological Research Center
Zhengzhou, China

Yunjie Gu
Third Department
National Digital Switching System Engineering & Technological Research Center
Zhengzhou, China

***Abstract*—Ontology is the core of knowledge graph. Traditional ontology description and ontology representation rely on ontology descriptional language. This kind of representation method makes it difficult for people to quickly grasp ontology's structure and then reuse it or segment it. To solve this problem, we proposed a method to transform ontologies into complex networks. This paper analyses ontologies' structural characteristics through ontology visualization and ontologies' degree distribution, clustering coefficient, average path length and eigenvector centrality. We observed that many ontologies have tree-like structures. Our analyses further revealed that a concept's importance is positively related to its degree and eigenvector centrality. Experiments in university ontology shows that our method has a good effect in intuitively understanding the ontology structure.**

Keywords*—knowledge graph, ontology, complex network, visualization, ontology reuse*

I. Introduction

Ontology originates from philosophy, which is a philosophical theory that explores the origins of the world. With the appearance of knowledge graph, ontology now has a new meaning. In information science, ontology means a set of concepts and relations between concepts [1]. With the rapid development of knowledge graph, ontology has received more and more attention due to its powerful abstract ability.

At present, research on ontology mainly focuses on knowledge sharing, description and retrieval [2]. Hoster et al. [3] took the lead in applying social network analysis techniques to ontology analysis. They suggested that ontologies are constructed with n kinds of nodes and k kinds of edges. Pujara et al. [4] regarded the concepts and relations in an ontology as nodes and edges respectively, constructing a directed acyclic graph. This method helps to further segment ontologies. Chen et al. [5] used the domain and range of the object as nodes and the relationships as edges, constructing ontology's schema graph. Z. Yang[2], D. L. Zhang[6], X. Liu [7], L. Xu [8], G. J. Xu [9] and other scholars regarded the concepts and entities in an ontology as nodes and the relations as edges, transferring ontologies into complex networks. They further analyzed ontology structure's small-world and scale-free attributes. However, their conclusions about ontology structure are inconsistent: Z. Yang and D. L. Zhang both analyzed the small-world and scale-free attributes of gene ontology, but obtained the opposite conclusion. The reason is that there is no specific standard for transforming an ontology into a complex network, and there are no hard indicators for evaluating the two attributes of a network. At present, there is no systematical theory and reasonable evaluating indicators for analysis of complex structures in ontologies. There are few researches on ontologies' conceptual structures[9].

To solve the above problems, this paper proposes a method to analyze ontology structure using complex network. We regarded concepts and entities as nodes, relations between concepts and entities as edges, transferring university ontology into a directed acyclic graph. The triples containing anonymous nodes are merged to ensure that there is no anonymous node in the transformed network. The transformed ontology network is visualized, which directly shows the ontology topology structure and information. We applied complex network theory to the ontology network for quantitative analysis, and analyzed the structural characteristics of the ontology network through various parameters.

II. Related theory

Ontology is the core layer of knowledge graph. This layer is the abstractive expression of concepts and relationships defined in the context of the resource description framework (RDF) [9]. Ontology is similar to complex network in structure. This section will describe ontology and complex network's basic concepts first, then analyzes the similarities and differences between the them.

A. Concept and Representation of Ontology

An ontology is an explicit, formalized summarization of a conceptual system, consisting of a limited list of terms and links between them [10].

There are currently two main ontology descriptional languages: RDF Schema and OWL. RDF Schema is extended from RDF, describing classes by defining a set of basic domain-independent structures. OWL2 is the second edition of OWL language, which is further divided into OWL2 Full and OWL2 DL. OWL2 Full extends directly from RDF. By

978-1-5386-4870-4/18/$31.00 ©2018 IEEE

increasing class membership, classification, equivalence relations, cardinality restrictions and consistency, OWL2 Full enhanced its expression capability and can be mapped to a known logic system, thus having reasoning ability. Owing to OWL2 Full is directly extended from RDF, it is fully compatible with RDF.

However, because OWL2 Full is so complete, it is undecidable. That is, logical reasoning may has no termination on an OWL2 Full file[10]. OWL2 DL achieves efficient reasoning by reducing the completeness of the logic system. OWL2 DL can be mapped to description logic so that it can directly reasoning on ontologies by combining existing reasoning engines.

B. The Concept and Representation of Complex Network

A complex network can be regarded as a collection of interconnected entities with independent features. Each entity can be regarded as a node in the graph, while interconnections between entities are considered as edges in the graph [11]. Complex network is a special graph. Specifically, complex networks are divided into four species: directed weighted, directed unweighted, undirected weighted, and undirected unweighted.

The main representations of complex networks include adjacency matrix representation, correlation matrix representation and arc representation.

The adjacency matrix representation uses a square matrix to represent the relationships between nodes. If there are n nodes in the network, a matrix with size $n \times n$ is constructed. If there is an edge between node i and node j, the matrix's value of i-th row and j-th column is 1, otherwise is 0.

The relation matrix representation uses a matrix to represent the relationships between nodes. If there are n nodes, m edges in the network, the size of the matrix is $n \times m$. If there is an edge k between node i and node j, the matrix's value of i-th row and k-th column is 1, otherwise is 0.

The arc representation represents nodes and relationship between nodes using a set of triples. This kind of representation fits the triplet representation in the knowledge graph. The three elements in the triple represent the starting node of the edge, the weight of the edge, and the ending node of the edge.

C. Similarities and Differences Between Ontologies and Complex Networks

1) Structure

a) Similarities

Complex networks are abstracted from real, large graphs with certain characteristics composed of nodes and edges. Ontology mainly reflects the relationship between concepts. Under the condition of removing constraints, the concept is abstracted into nodes, the relationship is abstracted into edges, and it can be mapped into the form of complex network.

b) Differences

In a complex network, the connections between nodes are various and there may not exists recurrence rules among them. However, ontology is derived from concepts, thus, in most cases, its structure is a directed acyclic graph. Otherwise there will be a conceptual cycle definition [2], which is not allowed.

2) Evolution

a) Similarities

Complex networks will evolve over time, and related concepts in the ontology will also evolve.

b) Differences

From the perspective of network dynamics, the connection between each pair of nodes in a complex network may change. Ontologies tend to be stable because it is high generalization of domain knowledge and tend to add nodes instead of changing edges.

III. Analysis of Ontology Structure from the Perspective of Complex Networks

If the concepts and entities in the ontology are regarded as nodes and the relations are regarded as edges, and the axiom constraints are discarded, the ontology can be transformed into a complex network and thus can be quantified through network analysis. The process of transforming an ontology into a complex network involves the extraction of tuples, the deletion of invalid tuples, and the processing of anonymous nodes. In this part, we will design a conversion algorithm based on the characteristics of OWL, calculate the main parameters based on network analysis, and explain its significance.

A. Tansform Ontology into Complex Network

1) Data Preprocessing

From the analysis of Section Ⅱ, it can be seen that although the ontology has similarities with the complex network in structure, there are still substantial differences between them. Complex network can be represented by nodes and edges, but ontology is usually characterized by an ontology language such as RDF Schema and OWL. The ontology language includes not only the triplet information required for conversion, but also the "structure-independent information" of the ontology, such as the license, comment and version of the ontology. Structure-independent information is deleted before conversion.

In addition, in order to clearly distinguish classes and entities, we ignored the links between them. A triple whose relationship is a data attribute, since it essentially represents the entity and its tag, carries semantic information rather than structural information. Thus we didn't integrate it into the ontology network.

2) Ontology Transformation Algorithm

Inspired from Xu Lei's paper, we designed an algorithm based on OWL2 DL to transform an ontology into a complex network. The algorithm's pseudocode is shown as follows:

Input: ontology file *owl_file*

Output: ontology network *out*

```
triples=file2triples(owl_file)
//transfer owl file to triples
fb=open('fb.owl')//to save triples containing anonymous nodes
out=open('out.csv')

for line in triples do
    if Valid(line) then//if the current triple is valid
        head=line[0]
        arc=line[1]
        tail=line[2]
        if head.isURI() and tail.isURI() then
            out.write(line)
        else:
            fb.write(line)
        end if
    end if
end for

for line in fb do
    if line[0].isURI() then
        head=line[0]
        arc=line[1]
        tail=line[2]
        getBlankNodeNet(head, arc, tail, fb, out)
    end if
end for
```

```
function  getBlankNodeNet(head, arc, tail, fb, out):
// Recursively merge anonymous nodes and enter the results to
the ontology network
    t1_o=getObject(tail, fb)
    // Finding objects whose subject is tail from file fb
    for x in t1_o do
    //for each [relation, object] tuple whose subject is o
        arc=arc+x[0]
        tail=x[10]
        if isURI(tail) then////if the object is an URI
            out.write([head,arc,tail])
        else:
            getBlankNodeNet(head, label, tail,fb,out)
        end if
    end for
```

The algorithm is divided into the following four steps:

(1) Convert OWL file to triples.

(2) For each valid triple, if its subject and object are represented as URI, add it to the current network.

(3) If a triple's subject or object is not represented by an URI, it means that the triple contains an anonymous node, thus need further processing.

(4) For a triple containing an anonymous node, start with the triple whose subject is an URI and object is an anonymous node. If it can be merged into a non-anonymous triple by chain rule (which means by chain, the object can be linked to an URI), add it to the current ontology network. For example, two triples (A, locatedIn, hidden_node1), (hidden_node1, partOf, B) containing anonymous node hidden_node1 can be merged into (A, locatedIn_partOf, B).

B. Ontology Structure Analysis

(1) degree distribution

Define $p(k)$ as the proportion of nodes whose degrees are k, $p(k)$ is the nodes' degree distribution [12]. Studies have shown that degree distribution of many networks in the real world obey power-law distribution [13], that is, in these networks, $p(k) \propto k^{-\gamma}$.

In ontology network, degree distribution can reflect concept importance. The greater the node's degree, the more connection the concept builds up with other concepts. Degree distribution means a distribution of connection.

(2) The average distance

The distance d_{ij} between the two nodes v_i and v_j is the length of the shortest path between the two nodes. The average distance of the graph $G(V,E)$ is the average of the distances between all pairs of nodes in the graph and can be expressed as:

$$< d >= \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij} \tag{3-1}$$

where N is the number of nodes in the graph.

The average distance of the ontology network can reflect the concepts' relevance in the ontology. The shorter the average distance, the greater the concept relevance in the ontology.

(3) Cluster coefficient

The cluster coefficient is defined as the ratio of the number of connected edges in the network to the maximum possible number of connected edges [12], which can be expressed as:

$$C_i = l_i \cdot \frac{2}{k_i(k_i-1)} \tag{3-2}$$

where l_i is the number of connected edges between node i's adjacent nodes, k_i is the degree of node i, and $\frac{k_i(k_i-1)}{2}$ the maximum possible number of edges between the adjacent nodes of node i.

The cluster coefficient of the entire network $G(V,E)$ is the average of the cluster coefficients of all nodes in the graph:

$$C = \frac{1}{N'} \sum_{k_i > 1} C_i \tag{3-3}$$

where N' is the number of nodes whose degree is greater than 1, k_i is the degree of node i.

In an ontology network, clustering coefficient can reflect the closeness of the connection between concepts in the ontology. The larger the cluster coefficient, the closer the relationship between related concepts.

(4) Centrality

The centrality of node is used to reflect the importance of nodes, which are divided into degree centrality, betweenness centrality, closeness centrality and eigenvector centrality, which represent the connection capabilities of nodes, the connected contribution of nodes in the diagram, the node's ability to communicate and node's ability to associate.

A node's degree centrality is the degree of the node, which is the simplest and most direct method to measure the importance of nodes [12].

A node's betweenness centrality is defined as the ratio of the number of shortest path passing through the node to all shortest path in the network [12]. Node v_i 's betweenness centrality can be expressed as:

$$B(v_i) = \sum_{s \neq i \neq t} \frac{n_{st}^i}{g_{st}} \tag{3-4}$$

where n_{st}^i is the number of the shortest paths from node s to node t passing through node i , g_{st} is the number of the shortest path from node s to node t . In the ontology network, betweenness centrality is suitable for mining the important concepts of the ontology.

A node's closeness centrality is the average of the shortest path between the node and all other nodes in the network [12].

$$Cc_i = \frac{\sum_j d_{ij}}{N} \tag{3-5}$$

In the ontology network, closeness centrality reflects the abstractness of a certain concept. In other words, the greater the closeness centrality of a concept, the better its generalization to other concepts in the ontology thus the easier it is to relate to other concepts.

Eigenvector centrality of a node depends not only on the number of its adjacent nodes, but also on the importance of its adjacent nodes [14].

$$x_i = \frac{1}{\lambda} \sum_{t \in M(v)} x_t \tag{3-6}$$

where $M(v)$ is the set of node i 's adjacent nodes, λ is a constant. A node's eigenvector centrality is equivalent to the weighted degree of the node, where the weight is the eigenvector centrality of its adjacent nodes.

IV. EXPERIMENTS AND ANALYSISSING

The experimental part will transform ontology file into complex network and perform visualization and parameter calculation on it. We use university ontology as our experimental data. Quantitive analysis, qualitative analysis and visualization are applied to the ontology.

A. Experiment on University Ontology

University Ontology is constructed by Hadjar et al. [17] according to the real situation of Ahlia University. After transformation of the file, we get a network of 1158 nodes and 4143 edges. The network's structure is shown in Figure 1.

Fig. 1. University ontology structure

As can be seen from Figure 1, the current ontology network consists of three disjoint subgraphs. It can be seen from the tag information that the three subgraphs have their own meanings. The three subgraphs are separately explained in the following part.

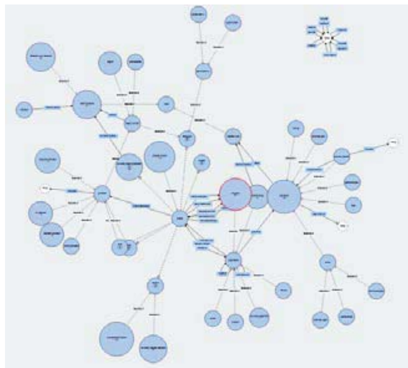

Fig. 2. Visualization through WebVOWL

Figure 2 shows the visualization through WebVOWL[18], an ontology visualization tool. Compared with Figure 1, Figure 2 shows the whole owl file without pre-treatment, which makes the ontology's structure indistinct. Besides, this method doesn't suitable for large-scale ontologies because neither can it visualize them clearly nor can it show the structure by calculating parameters under the complex network perspective.

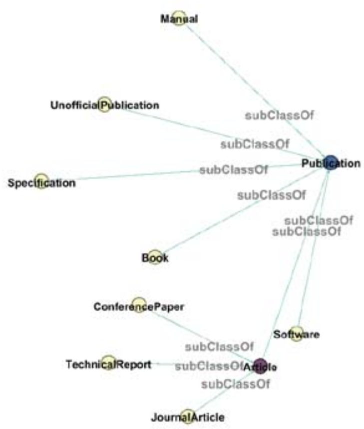

Fig. 3. Sub-ontology Publication's structure

Figure 3 shows one of the subgraphs of the ontology network. The subgraph's nodes and edges are tagged(for viewing clarity, the node position is slightly adjusted). It can be seen that the relationships in this subgraph are homogeneous: that is, each edge's tag is "subClassOf". Obviously, this subgraph represents ontology University's sub-ontology Publication. The concept of Publication is divided into several parts: Specification, Article, Manual, Book, Unofficial Publication, Software, and ConferencePaper. Furthermore, there are some sub-categories under Article. It can also be seen intuitively that this subgraph is a tree. Through visualization, the above information can be intuitively and quickly obtained, which greatly improves the efficiency of the ontology structure analysis

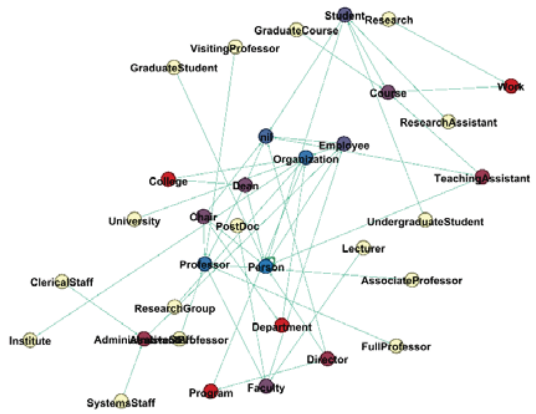

Fig. 4. Sub-ontology Relationship's structure

Figure 4 shows another subgraph of the ontology network. The subgraph's nodes are tagged(for viewing clarity, tags on edges are not shown). From the node's tag, it can be seen that this subgraph represents the interpersonal relationship in the university, which includes people, students, colleges, experts and other categories.

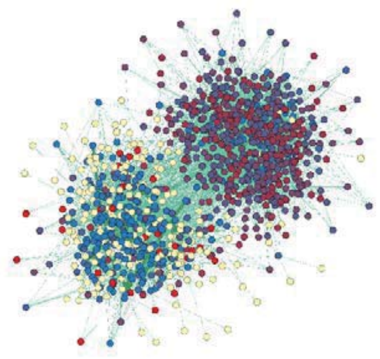

Fig. 5. Entity structure

Figure 5 shows the largest subgraph of the ontology network(for viewing clarity, tags on nodes and edges are not shown). This subgraph represents the relationship between individuals. Because the relationship between the class and the entity was deleted during the data preprocessing, this subgraph is not connected with the other two subgraphs.

Table 1 shows the top 20 rankings of all nodes in the university ontology according to their degree. As can be seen from table 1, "Department0" has the largest degree. However, the intuitively most important concept is "university". The reason for this result is that we have introduced too many instances in this ontology, thus weakening the real important concepts. Many triples in data layer are precisely related to "Department0", so this department has the greatest degree.

TABLE I. 20 NODES WITH LARGEST DEGREE

Num	Id	degree
1	www.Department0.University0.edu	730
2	Publication2	71
3	Publication0	64
4	Publication3	62
5	Publication7	60
6	Publication4	59
7	Publication1	57
8	Publication8	51
9	Publication6	50
10	Publication5	49
11	Publication10	48
12	Publication12	42
13	FullProfessor7	41
14	Publication9	41
15	Course57	38
16	Course53	37
17	Course4	36
18	Course34	35
19	Course20	35
20	Course21	35

Besides, top nodes are a series of publications. This shows that publications are very important to a university. In fact, students and teachers both publish academic articles and monographs. The courses they learn and teach also require the use of books. They really should have very close links with publications. Immediately thereafter, a large number of courses appeared in the ranking. In universities, courses are obviously important, and courses is also an important hub for connecting teachers and students.

Next, we calculate the network's eigenvector centrality. Table 2 shows the top 20 rankings of all nodes in the university ontology according to eigenvector centrality. Eigenvector centrality can be regarded as the weighted degree of the node, which reflects the association of current node and important nodes. Compared with Table Ⅰ, the courses in Table Ⅱ are generally ranked higher. The reason is that although the publications associate many nodes, their labels are "Publication Author". There may be many professors, teaching assistants and students who have published articles in a journal, but they may not all have a decisive position. However, each course in the ontology must include a full professor, an assistant professor and a number of students, possibly including associate professors. This data format ensures that each course is associated with at least one core node (professor). Therefore, the eigenvector centrality of each course is generally higher than the publications.

TABLE II. 20 NODES WITH LARGEST BETWEENNESS CENTRALITY

Num	Id	eigencentrality
1	www.Department0.University0.edu	1
2	www.University0.edu	0.119809
3	Course4	0.04536
4	Course21	0.044328
5	Course20	0.044211
6	Course57	0.044026
7	Course53	0.044022
8	Course34	0.043949
9	Course28	0.043088
10	Course42	0.042158
11	Course30	0.042109
12	Course2	0.040977
13	FullProfessor7	0.040974
14	Course18	0.040875
15	Course0	0.040266
16	Course15	0.039825
17	Course19	0.039288
18	Course12	0.039237
19	Course13	0.03858
20	Course10	0.037371

B. *Significance*

From the above analysis, we have enough reasons to believe that it is meaningful to transform knowledge graph into complex network. The reasons are as follows:

(1) Facilitate Ontology Visualization

The existing methods for ontology description are based on languages. Whether it is OWL language or RDF language, ontology users cannot grasp the ontology's structure in a short time. Even ontology experts are unable to quickly grasp the ontology's bulk of information. Visualization provides a user-friendly way for people to quickly grasp the ontology's structural information and meanings.

(2) Reuse Ontology

The original intention of the ontology is to share: everyone can publish their own ontology and data, and everyone can also reuse the ontology and data that others construct. It may even happen that A used the ontologies created by B to store the data created by C and build its own knowledge graph.

The ontology is a unified shared format. If there is a standard and unified ontology, the construction and integration of knowledge graphs will be more convenient. Obviously, ontology reuse is of great significance.

Converting ontology into complex network not only facilitates people to intuitively understand the structure of the ontology and select an appropriate ontology for reuse; it can also perform ontology segmentation and select the sub-ontologies that they need for reuse in order to achieve high efficiency. (If there are millions of vocabularies in Dbpedia ontology, it's impossible to reuse the whole ontology. If doing so, the storage space will be very large and the efficiency will be very low.)

(3) Facilitate the analysis of ontology structure

By transforming ontology into complex network, various parameters (clustering coefficients, average distances, etc.) on the network can be easily obtained. In this way, when the ontology's scale is large (such as gene ontology), the ontology structure and properties can be analyzed through these parameters.

V. CONCLUSION

This paper proposes a method that can transform ontologies into complex networks. Firstly, the structure-independent information in the ontology file is removed by preprocessing, then the ontology file is transformed into a complex network, at the same time, the triples containing anonymous nodes are integrated through concept merge. Through qualitative analysis of university ontology, we illustrated the effectiveness of studying ontology structure from the perspective of complex networks. Finally, the significance of transforming knowledge graphs into complex networks are summarized: visualization, ontology reuse and ontology structure analysis.

Although the proposed method can effectively show ontology structure, the visualization of large-scale ontology such as gene ontology is not clear. In the next step, we will enhance the visualization of large-scale ontology by showing segmented sub-ontologies.

ACKNOWLEDGMENT

This work has been supported by the National Digital Switching System Engineering & Technological Research Center. Research for this paper was financed by project “Abnormal Behavior Analysis and Warning”.

REFERENCES

[1] Asanee Kawtrakul, “Ontology Engineering and Knowledge Services for Agriculture Domain,” Journal of Integrative Agriculture, vol. 5, pp. 741-751, Nov 2012

[2] Z. Yang, “Complexity analysis on ontology structure based on complex network,” Computer Engineering and Applications, vol. 43, pp. 77-79, 2007

[3] B. Hoser, A. Hotho, R. Jäschke, C. Schmitz, & G. Stumme, “Semantic network analysis of ontologies,” Semantic Web Research & Applications, vol. 4011, pp. 297-305, 2006

[4] J. Pujara, H. Miao, L. Getoor, & W. Cohen, “Ontology-aware partitioning for knowledge graph identification,” The Workshop on Automated Knowledge Base Construction, vol.71, pp.19-24, 2013.

[5] Y. Chen, S. Goldberg, D. Z. Wang,& S. S. Johri, "Ontological Pathfinding," International Conference on Management of Data, ACM, pp. 835-846, 2016

[6] D. L. Zhang, Z. X. Wang, W. Liu& Z. Yang, "Complex network-based ontology structure analysis," Journal of Tongji University(Natural Science), vol. 37(2), pp. 258-261, 2009

[7] K. Liu, "Research on club division method of semantic networks for agricultural ontology," Anhui Agricultural University, 2013

[8] L. Xu, "A research on the structure of ontology networks and its evolution," Wuhan University, 2014

[9] G. J. Xu, S. W. Liu, X. Li, X. D. Zhang& W. G. Guan, "Study of the similarity on the properties of ontology structure and complex networks," Journal of Nantong University(Natural Science Edition), vol. 15(1), pp. 1-5, 2016

[10] G. Antoniou, F. Vanharmelen, "A Semantic Web Primer," MIT Press, vol. 3, pp. 47-54, 2008

[11] S. Z. Guo, "The basic theory of complex network," Science Press, vol. 2, 2012

[12] L. Y. Lv, T. Zhou, "Link prediction," Higher Education Press, vol. 1, 2013

[13] G. Caldarelli, "Scale-Free Networks: Complex Webs in Nature and Technology," Oup Catalogue, vol. 9, pp. 117-131, Feb. 2011

[14] X. F. Wang, X. Li, G. R. Chen, "An introduction of network science," Higher Education Press, vol. 1, 2012

[15] Author Unknown, 2018. http://www.geneontology.org/

[16] M. Jacomy, M. Bastian, S. Heymann, "Gephi: An Open Source Software for Exploring and Manipulating Networks," International Aaai Conference on Weblogs and Social Media, 2009.

[17] Hadjar K, "University Ontology: A Case Study at Ahlia University," Semantic Web. Springer International Publishing, 2016.

[18] Lohmann S, Link V, Marbach E, et al. "WebVOWL: Web-based Visualization of Ontologies," EKAW 2014: Knowledge Engineering and Knowledge Management, pp. 154-158, 2014