Ontology Design of Influential People Identification Using Centrality

Rolly Maulana Awangga, Muhammad Yusril, Helmi Setyawan

Applied Bachelor Program of Informatics Engineering, Politeknik Pos Indonesia, Indonesia

E-mail: awangga@poltekpos.ac.id, yusrilhelmi@poltekpos.ac.id

Abstract. Identifying influential people as a node in a graph theory commonly calculated by social network analysis. The social network data has the user as node and edge as relation forming a friend relation graph. This research is conducting different meaning of every nodes relation in the social network. Ontology was perfect match science to describe the social network data as conceptual and domain. Ontology gives essential relationship in a social network more than a current graph. Ontology proposed as a standard for knowledge representation for the semantic web by World Wide Web Consortium. The formal data representation use Resource Description Framework (RDF) and Web Ontology Language (OWL) which is strategic for Open Knowledge-Based website data. Ontology used in the semantic description for a relationship in the social network, it is open to developing semantic based relationship ontology by adding and modifying various and different relationship to have influential people as a conclusion. This research proposes a model using OWL and RDF for influential people identification in the social network. The study use degree centrality, betweenness centrality, and closeness centrality measurement for data validation. As a conclusion, influential people identification in Facebook can use proposed Ontology model in the Group, Photos, Photo Tag, Friends, Events and Works data.

1. Introduction

The semantic web standard proposed by World Wide Web Consortium (W3C) as the next-generation web. Ontology chose as a knowledge-based semantic description for base formalisation. Ontology used to communicate between website on the Internet. Ontologies give more power on conceptual with domain and relation to describe the problems. Many research on software engineering use ontology as a modelling language for formalisation design. Especially multiple sites for communicating between designs in remote team member [1] to avoid miss-configuration of software and mediating mechanism for accessing heterogeneous data and information sources [2]. If machine learning and software framework can solve healthy problem[3][4]. Data science can solve warehouse[5] and knowledge management problem[6]. Ontology is a solution for software engineering knowledge representation [1]. There was a classification of Ontology for Software Development Life Cycle phase [7]. Ontology is presenting more metadata schemas, offering a controlled vocabulary more than Unified Modelling Language (UML) [8].
Ontology can improve Social Network Analysis (SNA) measurement [9] and software engineering knowledge representation [10]. In the graph theory and network analysis, centrality used to determine the relative importance of a node within the graph [11]. Research topics in different aspects [12] using centrality analysis to get influential people. Beside for political issues [13] and terrorism information [14] on the social network [15], an online market business with higher values for three measurement degrees (degree, betweenness, closeness) obtain users who can target as productive customer for selling a product [16].

All the user-related graph on a social network mapped to semantic description model on ontology. Ontologies are a useful tool for modelling, sharing and reusing knowledge [10]. Ontologies are closely related to modern object-oriented software engineering with richer information model compared to UML and as a shared conceptualisation, naturally easy to develop in future work [8]. Centrality [17] use as graph theory with social network analysis [18] [19]. A person can be influential observed by how many friends list that he has, his ability to connect two networks of friends, or a person who knew a lot of others influential person. Degree, betweenness, and closeness centrality as popular centrality concept adopted in measuring graph. A graph consists a node and an edge. Node is defined person; edge defines a relation between nodes.

The objective of this research is modelling influential people from social network using OWL and RDF. Semantic description will use to modelling the concept of influential people. The relationships model used from a popular social network Facebook, which have ability providing user data in Application Programming Interface (API). The data gathered from Facebook using API [20]. Facebook allows third-party apps to take a user data from Facebook account who was interested in the application.

2. Methods
In the social network analysis, the data designed as structural characteristics of a social organisation. The Analysis need more information about the relation between nodes and edge [21]. Social network semantic make a computer understand the social network. In real social status, every person has a different relationship with different level and function. The Semantic describe more than attributes data and relation like characteristics of nodes and connections. Ontology is a philosophical concept to describe things. It is a modelling tool in artificial intelligence, which can describe conceptual models of information system in semantic and knowledge level [22] as a formal specification of a shared conceptualisation [23]. The definition relations in ontology illustrated as actor and relation attributes. The actor could be people, organisation or company and others [9]. Relation attributes describe various connection formed in social activities interact.

A Graph concept use for studying social phenomena. Actors and relationships are forming social networks data, the detailed explanation explains by the application of Wasserman and Faust (1994)[11] or Scott (1991)[24]. The most important thing in the analysis of social networks is vertices in the graph (Bavelas, 1984; Sabidusi, 1966; Freeman, 1979) [25]. The design was interpreting the interests into the graph. The formulation of centrality [10] obtained of the most influential people in a community. Centrality calculation divide into several calculation formulas.

Degree centrality is the most fundamental measure of centrality computation. Degree centrality has a formula essentially measures the amount of edge connection as a vertex of an object (equation 1).
Three Facebook classes create the domain on the social network (figure 2). The user is an account. Activities describe user activity in Facebook. Under activity class, there is a group where the user comes from and lives in recently. Activities describe user activity in Facebook. Under activity class, there is a group where the user joins in. Photo tags where the user was in the photo with someone else. Events which are user come to an invitation. The school who is user getting a school there.
Photos is a list of user photos. Work is the occupation of the user. We can look clearly at Onto-graph figure 3. Onto-graph is visualisation for domain and relation ontology, the connection between domain has many properties like in figure 4.

Super-properties in the social network ontology describing in five properties. HasActivity is a relation between users to the Activities superclasses, with HasUser as inverse properties. HasArea is a functional relation between user and Area Classes, with HasCitizen as inverse function. Also, HasFriend is a symmetric relation between users. There is a restriction in ontology, especially for the user. Property of come from has maximally related to one object in HasSignificantTo and LiveIn.
Restriction rules added in user class. The user has relation to hometown maximal is one relation, same as location and significantTo. The semantic description of ontology for influential people identification generated from this ontology. The research use HasFriend and HasActivity or HasUser as properties and user and activities as classes. The Ontology is explaining in figure 5.

Figure 5. Influential People Identification

Ontology.

Activities object class is superclasses with six classes under it. School, Works, Photos, PhotoTags, Groups, and Event is activities class related to the user with HasUser and HasActivity super properties. HasUser is an asymmetric relation between Activities class to User class. HasUser is super properties for HasEmployee, HasInviting, HasMember, HasOwner, HasPeople, HasStudents classes. Inverse object properties of HasUser are HasActivity. HasActivity are super properties for HasInvited, HasJoin, HasPhoto, HasSchool, HasTaggedIn, WorkFor. User class has symmetric irreflexive properties called HasFriend.

In figure 6 the blue box is a class, a solid arrow showing subclasses. Dashed Arrow is showing properties of each class well connected each other. This ontology used by Social Network Analysis calculation. Measuring influential people using Centrality on this ontology to get a value of importance user as nodes in social activities and users graph.

Centrality calculation was implementing into Facebook user data. The data gathered by Facebook Application which have 120283 friend connection. The highest frequency of appearance of account in Degree Centrality data measurement, sorted by frequency of occurrence. The Result is account A get 3, account IMB get 3, account JRJS get 2 and account AW get 1.

For Each measurement of betweenness centrality in Social Network Data. The top three nodes selected by frequency sorting. The result is: account IMB get 2, account AL get 2, account TIH get 2, account AN get 2, account AM get 1.

The result from every closeness centrality measurement. The data sorting by frequency. Top three selected from population data. Account AL gets 3; account MI get 2, account ER get 2, account ES get 2, account KYT get 1.
4. Conclusion

The highest Centrality calculation using Equation 1, 2, and 3 have results for influential people identification using ontology model in Figure 5 and 6. Ontology model in Figure 5 and 6 in the Group, Photos, Photo Tag, Friends, Events, and Works data shows the name of a person, this identified as influential people. The centrality calculation from Location, Hometown data shows the name of City so cannot identify influential people. Education Shows the name of an institution so cannot identify influential people. As a result, influential people identification in Facebook can use ontology on Figure 6 in the Group, Photos, Photo Tag, Friends, Events and Works data.

Influential people identification can be more varied with different social network data. Centrality measurement for influential people identification can expand with other calculation method using fuzzy logic or another intelligence system. The model can combine with another technique besides centrality for evaluation purpose on influential people identification. Another aspect of individual influential people measurement can use statistical methods.

References

[1] Wongthongtham P, Kasisopha N, Chang E and Dillon T 2008 Convergence and Hybrid Information Technology, 2008. ICCIT’08. Third International Conference on vol 2 (IEEE) pp 668–675

[2] Dillon T S, Chang E and Wongthongtham P 2008 Software Engineering, 2008. ASWEC 2008. 19th Australian Conference on (IEEE) pp 13–23

[3] Yulita I N, Purwani S, Rosadi R and Awangga R M 2017 Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), 2017 International Conference on (IEEE) pp 1–5

[4] Awangga R M, Fathonah N S and Hasanudin T I 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE) (IEEE) pp 255–260 ISBN 978-1-5386-0658-2 URL http://ieeexplore.ieee.org/document/8285506/

[5] Pane S F, Awangga R M and Azhari B R 2018 TELKOMNIKA (Telecommunication Computing Electronics and Control) 16

[6] Awangga R M, Pane S F, Tunnisa K and Suwardi I S 2018 TELKOMNIKA (Telecommunication Computing Electronics and Control) 16

[7] Zhao Y, Dong J and Peng T 2009 IEEE Transactions on Services Computing 2 303–317

[8] Siricharoen W V 2008 Computer Science and its Applications, 2008. CSA ’08. International Symposium on (IEEE) pp 320–325

[9] Li T, Yang H, He J and Ai Y 2010 Knowledge Acquisition and Modeling (KAM), 2010 3rd International Symposium on (IEEE) pp 258–261

[10] Hilera J R and Fernandez-Sanz L 2010 Software Engineering Advances (ICSEA), 2010 Fifth International Conference on (IEEE) pp 380–383

[11] Wasserman S and Faust K 1994 Social network analysis: Methods and applications vol 8 (Cambridge university press)

[12] Jamali M and Abolhassani H 2006 Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM International Conference on (IEEE) pp 66–72

[13] Ya-ting L and Jing-min C 2011 Consumer Electronics, Communications and Networks (CECNet), 2011 International Conference on (IEEE) pp 5441–5444

[14] Li B x, Li M j, Sun D y, Li J and Li W j 2012 Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on (IEEE) pp 193–193

[15] Wennerberg P O 2006

[16] Kumar P and Zhang K 2007 e-Business Engineering, 2007. ICEBE 2007. IEEE International Conference on (IEEE) pp 363–367
[17] Warmbrodt J, Sheng H and Hall R 2008 Hawaii International Conference on System Sciences, Proceedings of the 41st Annual (IEEE) pp 291–291
[18] Martin T, Zhang X and Newman M 2014 Physical review E 90 052808
[19] Scott J 2017 Social network analysis (Sage)
[20] Catanese S A, De Meo P, Ferrara E, Fiumara G and Provetti A 2011 Proceedings of the international conference on web intelligence, mining and semantics (ACM) p 52
[21] Chen L, Wei S and Qingpu Z 2010 E-Business and E-Government (ICEE), 2010 International Conference on (IEEE) pp 1936–1939
[22] Dorr C 2010
[23] Borst W N 1997 Construction of engineering ontologies for knowledge sharing and reuse (Universiteit Twente)
[24] Scott J 1991 Annual review of sociology 17 181–203
[25] Zalzala L 2013 ICMLG2013-Proceedings of the International Conference on Management, Leadership and Governance: ICMLG 2013 (Academic Conferences Limited) p 348

Acknowledgement
Thanks to Politeknik Pos Indonesia to support the publication.