

# Comparison of the social network weight measurements

Mahyuddin K M Nasution<sup>1,\*</sup>, Shahrul Azman Mohd Noah<sup>2</sup>

<sup>1</sup>Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Kampus USU, Padang Bulan 20155, Medan, Indonesia

<sup>2</sup>Knowledge Technology Group Research, Center for Artificial Intelligence Technology, FTSM, Universiti Kebangsaan Malaysia, Bangi, Selangor DE, Malaysia.

E-mail: \*mahyuddin@usu.ac.id

**Abstract.** To extract social networks from the information sources, it is necessary to measure the similarities. There are several measurements that can become application to form a social network based on information sources. On the other side, various the information sources with various conditions and are not homogeneous, but it is possible to cluster them into a homogeneous form. However, measurement about it require proof. To produce the proof, a comparison of results from various measurements gives one of the answers.

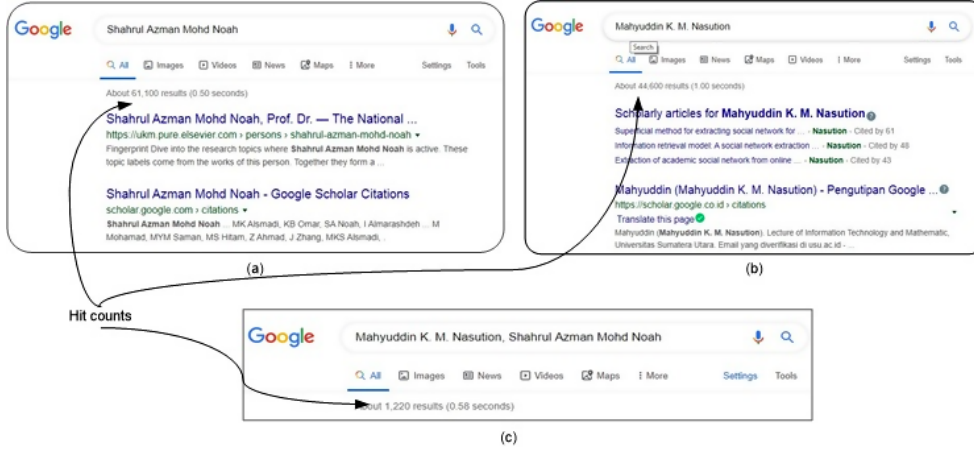
## 1. Introduction

Social network extraction from the Web by involving any search engine is necessary by extracting the information contained in the information sources [1, 2, 3]. Information sources such as the Web contain information that is not only dynamic, but also ambiguous and unstructured. The use of search engines to access such social network information: the information of social actors and their relationships, it requires a strategy or an approach that varies according to interests [4, 5]. However, in general the approach requires measurement which is generally also referred to as the similarities.

There are many similarity measures used in extracting social networks from the Web. Similarity targets various data objects. Any search engine like Geevv, Google, Yahoo, Bing, and others provide the hit count answers to questions in the query. The query contains question information from search engine users. Clustering involves queries, search engines, and the Web, this generates social networks. This paper aims to show the results of several measurements from the Web.

## 2. Some Similarities

After identifying the existence of social actors [6], the social networks exist when identifying the existence of relationships between social actors has a proof [7]. The existence of social actors through clustering involves a query, the query contains the name of the social actor and submits it to the search engine, and produces a hit count in occurrence which we then declare as  $|\Omega_a|$ , where  $a \in A$  is a set of social actors [8].



**Figure 1.** Occurrences and co-occurrence based on Google search engine

While the existence of social relationships through the use of queries, which contains a pair of social actor names, submit it to search engines, and produce a hit count in co-occurrence [9], which can then be stated as  $|\Omega_a \cap \Omega_b|$ ,  $a, b \in A$  [10].

For graph-based social networks representation [11]: When node  $v \in V$  has been declared, and edge  $e \in E$  can be only be expressed by calculating similarity weights so that each  $e \in E$  is a representation of the relations based on the weight calculation between two occurrences and one co-occurrence [12]. In the early stages of developing social network extraction from the web, the researchers use the Jaccard coefficient [13],

$$J_c = \frac{|\Omega_a \cap \Omega_b|}{|\Omega_a \cup \Omega_b|} \quad (1)$$

to determine the weight of the relation [14]. However, the existence of a relation between two social actors is based on the matching coefficient

$$M_c = |\Omega_a \cap \Omega_b|, \quad (2)$$

i.e. if  $M_c > 0$  [15]. While the mutual information [16],

$$M_i = \log \frac{m|\Omega_a \cap \Omega_b|}{|\Omega_a||\Omega_b|} \quad (3)$$

is to measure of the mutual dependence between two actors, or dependence of a social actor on other social actors. Use of dice coefficient [17],

$$D_c = \frac{2|\Omega_a \cap \Omega_b|}{|\Omega_a| + |\Omega_b|} \quad (4)$$

serves to maintain the heterogeneity of sensitive information between two social actors [18, 4].

Overlap coefficient

$$O_c = \frac{|\Omega_a \cap \Omega_b|}{\min(|\Omega_a|, |\Omega_b|)} \quad (5)$$

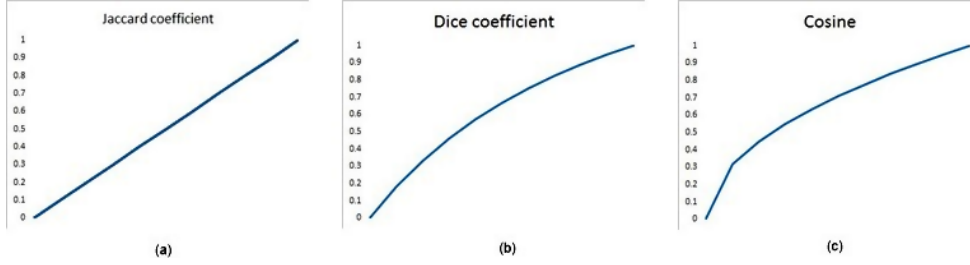
is to measure the overlap of information between two social actors [19], in terms of who influences and where the influence comes from, namely  $|\Omega_a| > |\Omega_b|$  assumes that the

**Table 1.** Calculation for two social actors based on some similarities.

Social Actors		Co-Occ	Similarity					
$ \Omega_a $	$ \Omega_b $	$ \Omega_a \cap \Omega_b $	$J_c$	$M_i$	$D_c$	$O_c$	$S_{cos}$	Appr
61,100	44,600	1,220	0.0117	0.6509	0.0231	0.0274	0.0234	0.0005

**Table 2.** Simulation in 10 numbers for  $J_c$ ,  $D_c$  and  $S_{cos}$ .

$ \Omega_x $	$ \Omega_y $	$ \Omega_x \cap \Omega_y $	$J_c$	$D_c$	$S_{cos}$
10	0	1	0.0	0.0000	0.0000
10	1	2	0.1	0.1818	0.3162
10	2	2	0.2	0.3333	0.4472
10	3	3	0.3	0.4615	0.5477
10	4	4	0.4	0.5714	0.6325
10	5	5	0.5	0.6667	0.7071
10	6	6	0.6	0.7500	0.7746
10	7	7	0.7	0.8235	0.8357
10	8	8	0.8	0.8889	0.8942
10	9	9	0.9	0.9474	0.9387
10	10	10	1.0	1.0000	1.0000

**Figure 2.** Three lines of similarities

influence comes from a social actor  $a \in A$  against other social actor  $b \in A$ . Besides, the cosine similarity [20],

$$S_{cos} = \frac{|\Omega_a \cap \Omega_b|}{\sqrt{|\Omega_a||\Omega_b|}} \quad (6)$$

is measurement to determine the opposition of a social actor to other social actors whereby the cosine angle serves to provide that opposition [18, 4].

Note that  $|\Omega_a| \geq |\Omega_a \cap \Omega_b|$  and  $|\Omega_b| \geq |\Omega_a \cap \Omega_b|$  such that  $J_c, D_c, O_c, S_{cos} \in [0, 1]$ , and  $m$  an arbitrary number such that  $M_i \in [0, 1]$ . For example, for two social actors Mahyuddin K. M. Nasution dan Shahrul Azman Mohd Noah, we obtain occurrences and co-occurrences (Co-Occ) as in Figure 1, and by using Eqs. (1), (3), (4), (5), (6) and an approach (Appr) to it we get the results such as Table 1 [21].

### 3. Simulation

To express the behaviour of each similarities in Eqs. (1), (3), (4), (5) and (6). For this reason, we generate a set of numbers randomly divided into three parts, namely

to represent two occurrences and one co-occurrence. Each of these parts is arranged in such a way that there is a maximum numbers, and 2 sets for twin numbers. The maximum collection of numbers is to represent the hit count for the social actor  $a$ , which is  $H_1 = \{|\Omega_a|_i | i = 1, \dots, n\}$ . The first set of twin numbers to represent the hit count for the social actor  $b$ , which is  $H_b = \{|\Omega_b|_i | i = 1, \dots, n\}$ . While the second set of twin numbers to represent the hit count for two social actors at once, namely  $H_c = \{|\Omega_a \cap \Omega_b|_i | i = 1, \dots, n\}$ . The sort of order of three sets of numbers from smallest to largest. Thus in general the members of all three arrangements apply  $|| \dots |_i \in H_c| \leq || \dots |_i \in H_a|$  and  $|| \dots |_i \in H_c| \leq || \dots |_i \in H_b|$ ,  $i = 1, \dots, n$  [22].

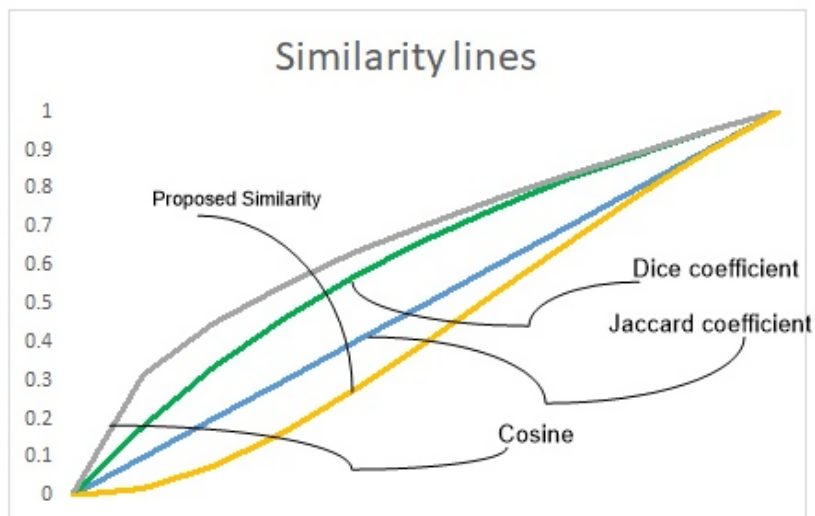
Numerically, for example, the simulation approach is as in Table 2. Diagonally, starting from 0 (left-down) to point 1 (right-up) the line from  $J_c$  cuts the area bounded by  $[0, 1]$  and the number  $n$  becomes two same part [23]. In other words, symmetrically the area by  $J_c$  is divided into two parts, see Figure 2(a). In contrast, the line formed by  $D_c$  from the point 0 to 1 diagonally divides the same area into two parts, but becomes one curve. The curve is reinforced at the initial numbers and descending to the other end, see Figure 2(b). Likewise, the line formed by computing  $S_{cos}$ , experiences more reinforcement at the starting point of the number and decreases again at the end, see Figure 2(c). An approach is possible to provide another impact on computing the numbers in a simulation, namely by reducing the strength at the beginning and experiencing reinforcement to the end.

#### 4. Discussion about similarity for social networks

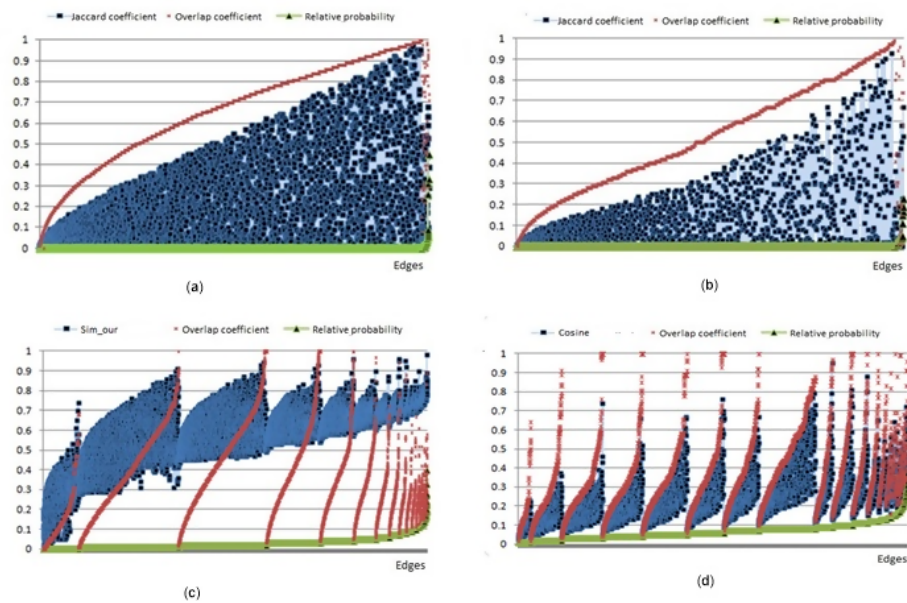
Similarity has an important role in generating social networks from information sources. The role is related to determining the relation between social actors in the form of weights [24, 25, 26]. As such, this role largely determines social networks presenting information that can be trusted or not. To the extent that role applies, similarities are a decisive measure other than input from information sources. Meanwhile, discussing the input applies on this occasion, in this case, it only involves the hit cont [2, 18]. The validity of a hit count depends on the capabilities and information structure of the search engine [4]. Thus, in a simulation the comparison between several similarity measurements is shown in Figure 3, and however there is a search for an approach that results in a comparison of similarity measurements whose results have so far been expressed as the proposed approach [27].

A experiment is to produce social networks from information space by involving several measures of similarity have produced some descriptions of the behaviour of measures of similarity [28]. A number of edges exist by the help of two occurrences and one co-occurrence for a pair of social actor, and then produce a set of edges for a number of social actors. A set of edges behaves as in Figure 4, and it is a result of that similarity [4].

Comparison between two measures of similarities by the help of relative probability is to show different behaviour of each measure. Whereas the comparison of two similar comparison structures shows different data behaviour [4, 21, 29]. Figure 4(a) presents the behaviour of  $J_c$  where all relations between social actors are distributed on one side of diagonal 0 (left-down) and 1 (right-up) in the order of relative probabilities, but the maximum limit of their weights are to follow the diagonal line of  $J_c$ . However, the edges density is distributed at the beginning of the measurement and  $o_c$  shows it based on a line that follows the  $J_c$  line curved. However, it is clear that the behaviour of the information source also affects the curve



**Figure 3.** Comparison of some similarity lines



**Figure 4.** Comparison of some similarity lines

$o_c$  according to the distribution density, see Figure 4(b) [3].

The involvement of measuring the proposal similarity  $Sim_{our}$ , which has been proposed on different occasions [2, 30],

$$sim_{our} = \frac{\log(2|\Omega_a||\Omega_b|)}{\log(|\Omega_a| + |\Omega_b|)}, \quad (7)$$

has shown different behavior by treating the same data to it, Figure 4(c). Edges come from results  $sim_{our}$  classified differently following the curve  $C_o$  according to the order of relative probabilities. Likewise, related to the measurement behavior of  $S_{cos}$  by the same treatment, also it produces the same pattern as  $Sim_{our}$ , but it generates the more detail classification to edges. This shows that all similarities have characteristics separate, and by different goals on the different results [31].

As a presumption there may be a measure of similarity based on the line pattern that borders the edges presented in Figure 4(b), it simulated in Figure 3 as a proposed similarity.

## 5. Conclusion

Measurement of the weight of relations in social network varies according to the emphasis on meaning and proof to be disclosed. Several measurement of similarity are used and comparisons between them indicate there is a specificity of each. Of course, based on the results of the comparison obtained there are three positions, namely the Jaccard coefficient in the midline position, while the Cosine and Dice coefficient are in the upper arch position, and finally the proposed measurement position is in the lower arch. Based on comparisons with the relative probability measurements of relations weights in social networks showed different behaviors.

## References

- [1] P Mika 2005 Flink: Semantic web technology for the extraction and analysis of social networks *Journal of Web Semantics* **3(2-3)**.
- [2] M K M Nasution and S A Noah 2010 Superficial method for extracting social network for academics using web snippets *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **6401** LNAI.
- [3] M K M Nasution 2013 Superficial method for extracting academic social network from the Web, *Ph. D. Thesis*, FTSM UKM Bangi, Malaysia.
- [4] M K M Nasution 2018 Social network extraction based on web: 2. Strategies in superficial methods *Journal of Physics: Conference Series* **1116(2)**.
- [5] E Suprptono, A Arfriandi, S I Rizqian 2019 The implementation of bellman-ford algorithm in Rembang tourism mobile application as a tourism information media *Data Science: J. of Computing and Appl. Informatics* **3(1)**.
- [6] M K M Nasution 2014 New method for extracting keyword for the social actor *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **8397** LNAI (Part 1).
- [7] M K M Nasution, O S Sitompul 2017 Enhancing extraction method for aggregating strength relation between social actors *Advances in Intelligent Systems and Computing* **573**.
- [8] M K M Nasution 2018 Singleton: A role of the search engine to reveal the existence of something in information space *IOP Conference Series: Materials Science and Engineering* **420(1)**.
- [9] H Kautz, B Selman, M Shah 1997 Referral Web: Combining social networks and collaborative filtering *Communications of the ACM* **40(3)**.
- [10] M K M Nasution 2018 Doubleton: A role of the search engine to reveal the existence of relation in information space *IOP Conference Series: Materials Science and Engineering* **420(1)**.
- [11] A Sala, L Cao, C Wilson, R Zablit, H Zheng, B Y Zhao 2010 Measurement-calibrated graph models for social network experiments *WWW '10 Proceedings of the 19th international conference on World wide web*.

- [12] P-C Lee, H-N Su 2010 Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis *Organization & Management* **12(1)**.
- [13] P Mika 2007 *Social Networks and the Semantic Web* Springer.
- [14] M Fire, L Tenenboim, O Lesser, R Puzis, L Rokach, Y Elovici 2011 Link Prediction in social networks using computationally efficient topological features *Privacy, Security, Risk and Trust (PASSAT), IEEE International Conference on and IEEE International Conference on Social Computing (SocialCom)*
- [15] A L Traud, P J Mucha, M A Porter 2012 Social structure of Facebook networks *Physica A: Statistical Mechanics and its Applications* **391(16)**.
- [16] Z Jia, X Sun 2012 Research on the evolution of development mode of trade network based on the economic globalization *Lecture Notes in Electrical Engineering LNEE* **163**.
- [17] P Velardi, R Navigli, A Cucchiarelli, F D'Antonio 2008 A new content-based model for social network analysis *IEEE International Conference on Semantic Computing*.
- [18] Y Matsuo, J Mori, M Hamasaki, T Nishimura, T Takeda, K Hasida, M Ishizuka 2007 POLYPHONET: An advanced social networks extraction system from the web *Journal of Web Semantics: Science, Services and Agents on the World Wide Web* **5**.
- [19] R Alguliev, R Aliguliyev, F Ganjaliyev 2011 Investigation of the role of similarity measure and ranking algorithm in mining social networks *Journal of Information Science* **37(3)**
- [20] E Spertus, M Sahami, O Buyukkokten 2005 Evaluating similarity measures: a large-scale study in the orkut social network *KDD '05 Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*.
- [21] M K M Nasution, O S Sitompul, S A Noah 2018 Social network extraction based on Web: 3. the integrated superficial method *Journal of Physics: Conference Series* **978(1)**.
- [22] M K M Nasution 2017 Modelling and simulation of search engine *Journal of Physics: Conference Series* **801(1)**.
- [23] M K M Nasution, S A Noah 2017 Social network extraction based on Web. A comparison of superficial methods *Procedia Computer Science* **124**.
- [24] Y Z Jin, Y Matsuo, M Ishizuka 2007 Extracting social networks among various entities on the Web *ESWC 2007 LNCS* **4519**.
- [25] H Young, R Dai, X Cui 2008 Extracting social network among various entities from Chinese news stories by content analysis *Annual IEEE International Computer Software and Applications Conference*.
- [26] M K M Nasution, S A Noah 2012 Information retrieval model: A social network extraction perspective *Proceedings - 2012 International Conference on Information Retrieval and Knowledge Management (CAMP'12)*.
- [27] M K M Nasution 2011 Kolmogorov complexity: Clustering objects and similarity *Bulletin of Mathematics* **3(1)**.
- [28] M K M Nasution 2011 Extraction of academic social network from online database *2011 International Conference on Semantic Technology and Information Retrieval (STAIR 2011)*.
- [29] M K M Nasution 2018 Social network extraction based on Web: 1. Related superficial methods *IOP Conference Series: Materials Science and Engineering* **300(1)**.
- [30] M K M Nasution, O S Sitompul, S Nasution, H Ambarita 2017 New similarity *IOP Conference Series: Materials Science and Engineering* **180(1)**.
- [31] M K M Nasution, R Syah, M Elveny 2017 Studies on behaviour of information to extract the meaning behind the behaviour *Journal of Physics: Conference Series* **801(1)**.