Journal of Theoretical and Applied Information Technology
15th July 2021. Vol.99. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MINING OF THE EXTRACTED SOCIAL NETWORK

[1]MAHYUDDIN K. M. NASUTION

[1]Data Science & Computational Intelligence Research Group, Universitas Sumatera Utara, Medan, Indonesia

E-mail: [1]mahyuddin@usu.ac.id

ABSTRACT

Social network mining (SNM) has become one of the main themes in the Semantic Web agenda. Social network can be extracted from different sources of information, and the resources – like documents/web pages - was growing dynamically not only require a flexible approach, but need behavior recognition. Each social network has the resources, but the relationship between resources and information sources requires explanation. It is SNM and it is not social network analysis, but it is possible to bridge social network and social network analysis. There is the behavior of resources of social network, and this article is to explore them by using the concept of clusters in theory and the statistical computation for conducting an experiment. That is using multiple-regression, where with applying graph as representation of relation between resources the growing the resources depend on each other. In a conclusion, there is a positive effect on the relations between resources for growing the social networks, where the behavior also indirectly indicates the extraction of engagement for the communities in the extracted social network.

Keywords: *Superficial method, stand-alone cluster, multiple-regression, association rule, positive effect, research group, social network analysis.*

1. INTRODUCTION

Extracting the social network (SN) from the Web-based on a superficial method is one of the approaches [1], where modality relations form a notion of the social networks that depend heavily on the co-occurrence for representing relations between individuals, groups, or organizations [2][3]. Modality, in this case, is like interests, requirements, exchange, hobby, etc., or something that causes them in textual appear on one Web page [4], for example [5]. In particular is the co-authoring relationship (co-authorship).

The extracted social networks are the results of the methods of social network extraction. Whereby actors, relationships, Web, vertices, until edges are the resources of SN that not only can use to manage any organization [6], nevertheless to administrate information also [7], that is the social network mining (SNM) is to provide a means of discovering the behavior of resources [8]. Namely, it is to trace early the formation of research groups following seed as the first authors in related scientific fields. Social network mining (SNM), therefore, is not the same social network analysis (SNA) [9]. Although SNA involves vertices and edges in a unit analysis such as size, density, degree, geodesic distance, and all properties associated with centrality [10], SNM plays a role as a bridge between the resources of SN and SNM by revealing different sides of the SNA. Of course, in order to uncover research groups based on seeds, the use of resources such as vertices and edges and the relationships between them is taken into consideration. Therefore, this article aims to address the behavior of resources and find out the relation between them. This article starts from concept development to theory in the section of the problem definition [11]. Then we propose an approach as a method to prove some assumptions and using the statistics computationally, multiple-regression for getting relations between resources and generating its behavior [6],[9],[12]. The section of experimental results is the disclosure of evidence through computation [13],[14].

2. PROBLEM DEFINITION

Through already used methods, extracting the social networks from information sources requires engineering, modeling, and casting [15]. Engineering requires a plan for implementing scientific superficial method [16]. Modeling is to determine the initial solution of a problem [17]. Meanwhile, casting as a siting or template is generally stated as data modeling [18]. It also needs to recognize the behavior of the information sources [19]. Therefore Web as a resource of SN is also an information source. The Web to be as big data keeps growing and dynamic has parts that can separate as a stand-alone cluster – in other literature it is introduced as a singleton [20], or the interconnected cluster – in the literature expressed as a doubleton [21]. The singleton and the doubleton go straight to hit counts, while the stand-alone cluster is to reveal what's behind the singleton and the doubleton [22,23].

Let the word "Web" or the phrase "World Wide Web" represents an object according to what we think [24]: The computer network is a social network [25],[26],[27],[28], then any social actor in the literal text is a term $t_k = (w_1w_2\ldots w_k)$ where w_k, $k = 1,2,\ldots,K$, are words. For example, t_x = Mahyuddin Khairuddin Matyuso Nasution. Let k is a number of words w and l is number of vocabularies (tokens) in t_k, then $l \leq k$ and $|t_k| = k$ is size of t_k [11]. In this case, $|t_x| = k = 4$ words and $l = 4$ tokens. In dynamic information space Ω, as Web, containing the ordered pair of the terms t_{ki} and web pages ω_{kj}: $(t_{ki}.\omega_{kj})$, where $i = 1,\ldots,I$ and $j = 1,\ldots,J$, or $\Omega_k = \{(t_k.\omega_k)_{ij}\}$ is the subsets of Ω, all of web pages indexed by a search engine be Ω and $|\Omega|$ is the cardinality of Ω [29], for example, $|\Omega_x| = 296$ for t_x in a query summited into Google search engine. Based on this concept, we will express some of the characteristics of big data formally as a behavior of resources [30],[31], as follows [32].

Definition 1: Suppose a search term t_a in q and q is a query. If an implication $\omega \Rightarrow t_a$ is TRUE, then a web page ω in Ω is relevant to q or $\Omega_a = 1$ if t_a in q is true at ω in Ω else $\Omega_a = 0$ otherwise, and Ω_a as a *cluster*.

Definition 1 describe that for each search term t_k, there a set of singleton search term of search engine Ξ such that t_k in Ξ, $(t_{ki}.\omega_{kj})$ is subset of dynamic information space Ω, or Ω_k is a subset of Ω, thus Ω_k is a singleton event of web pages that contain an occurrence of t_k in ω in Ω [11],[12]. Independently, we define it as follows [33],[34].

Definition 2: Suppose there be a set of attributes literally, $B = \{b_1, b_2, \ldots,b_{|B|}\}$. Let M_i is a set of transactions are subsets of attributes or M_i are subset or equal to B. The implication $\Omega_{b1} \Rightarrow \Omega_{b2}$ with two possible values TRUE and FALSE as an *association rule* if Ω_{b1}, Ω_{b2} are subsets of B and $\Omega_{b1} \cap \Omega_{b2} = \emptyset$.

Definition 3: Let $B = \{b_1, b_2, \ldots,b_{|B|}\}$ is a subset Ω. A cluster Ω_{b1} is as the stand-alone satisfies $\Omega_{b1} \Rightarrow \Omega_{b2}$ are subsets of B, $\Omega_{b1} \cap \Omega_{b2} = \emptyset$ (*association rule*).

Lemma 1: Suppose there be a set of actors, $A = \{a_1,a_2,\ldots,a_{|A|}\}$. If t_{ai} in q, $i=1,\ldots,|A|$, then Ω_{ai} is the stand-alone clusters for the social actors [11].

Proof. Based on Definition 1, we have ω in $\Omega \Rightarrow t_a$ in q such that $\Omega_a = \{(t_a,\omega_a)_{ij}\}$. Thus Ω_a is a cluster that contains web pages ω_a and t_a. Let there be t_{ai} in q_i and t_{aj} in q_j, we have $\Omega_{ai} \Rightarrow \Omega$ and $\Omega_{aj} \Rightarrow \Omega$, but based on Definition 2 and Definition 3: $\Omega_{ai} \cap \Omega_{aj} = \emptyset$, even though $\Omega_{ai} \Rightarrow \Omega_{aj}$ or $\Omega_{aj} \Rightarrow \Omega_{ai}$. Thus Ω_a has property as the stand-alone [11].

Definition 4: Suppose t_{ak} in q and t_{al} in q is a query. If an implication $\omega \Rightarrow t_{ak} \wedge t_{al}$ is TRUE, then ω in Ω is relevant to q or $\Omega_{akl} = 1$ if $t_{ak} \wedge t_{al}$ in q is true at ω in Ω, else $\Omega_{akl} = 0$ otherwise.

Lemma 2: Let A is a set of social actors. If t_{ak},t_{al} in q, then Ω_{akl} is a stand-alone cluster for a pair of social actors [11].

Proof. As applicable in Lemma 1, ω in $\Omega \Rightarrow \{t_{ak},t_{al}\}$ in q, ω in $\Omega \Rightarrow (t_{ak} \wedge t_{al})$ in q, then (ω in Ω t_{ak} in q) $\wedge$ (ω in Ω t_{al} in q), and we have Ω_{akl} $\{(t_{ak} \wedge t_{al},\omega_a)_{ij}\}$. If $i = k \wedge l$, then $\{(t_{ak} \wedge t_{al},\omega_a)_{ij}\}$, and $\Omega_{akl} = \{(t_a,\omega_a)_{kj}\} \wedge \{(t_a,\omega_a)_{lj}\}$. Thus Ω_a is a cluster that contains web pages ω_a, t_{ak} and t_{al}. However, based on Definition 2 and Definition 3, for $\Omega_{ak} \Rightarrow \Omega_{al} \Rightarrow \Omega$ and for $\Omega_{al} \Rightarrow \Omega_{ak} \Rightarrow \Omega$ such that $\Omega_{ak} \wedge \Omega_{al} \Rightarrow \Omega$ or

$$(\Omega_{akl} \Rightarrow \Omega) = ((\Omega_{ak} \cap \Omega_{al}) \Rightarrow \Omega). \quad (1)$$

Thus, there is a stand-alone cluster containing a pair of social actors.

Lemma 2 has stated that for two search terms t_x and t_y, there a set of doubleton search term of search engine Ξ such that t_x in Ξ and t_y in Ξ, $(\{t_{xi}.t_{yi}\}\omega_{xyj})$ is subset of dynamic information space Ω, or Ω_{xy} is a subset of Ω, thus Ω_{xy} is a doubleton event of web pages that contain an co-occurrence of t_x in ω in Ω and t_y in ω in Ω [11],[12]. Then, we have the direct consequence of Lemma 1 as follows.

Proposition 1: If Ω_{akl} is a stand-alone cluster for a social actor, then Ω_{akl} is representation of the actor [11].

Definition 5: Any cluster Ω_a has the behavior if there are a function γ that describes it one or more attributes b_1, b_2, … as descriptions of the cluster.

In a big-data context, (semi)-automatic extracting a social actor from Web is to produce information of the actor and to generate about another actor [35], i.e. the behavior of actor as resources of social network [36]. Therefore, the characteristics of big-data be the clues of the behavior of each social actor [37],[38],[39].

Proposition 2: If Ω_{akl} is a stand-alone cluster for a pair of social actors, then Ω_{akl} is a representation of relation between the actors [11].

Proof. Let Ω_{akl} is a stand-alone cluster of a pair of social actors. Eq. (1) states that $\Omega_{akl} = \Omega_{ak} \cap \Omega_{al}$, or $\gamma(\Omega_{akl}) = \gamma(\Omega_{ak} \cap \Omega_{al}) = \gamma(\Omega_{ak},\Omega_{al},\cap)$ where $\{\Omega_{ak},\Omega_{al},\cap\}$ is a set of attributes of Ω_{akl}. Whereas Ω_{ak} and Ω_{al}, respectively, are a stand-alone cluster, when $\cap$ as an attributes was connecting a cluster with another, i.e. the behavior of Ω_{akl} as a function, or there are a relation r such that $r(\Omega_{akl}) = \{\cap\}$ [11].

In a dynamic context, the automatic extraction method of social networks collects information about ties between all pairs (dyads) [40]. The ties connect a dyadic of actors by one or more relations in R (as a set of relationships), thus social network based on big data give a complete picture of relationships in the population [10],[41].

Theorem 1: If the behavior of a cluster describes the behavior of a social actor, then the behavior of other actors is expression of the relationship between the clusters [11].

A social network can be modeled very naturally by a graph $G(V,E)$ where SNM uses an approach to find out a behavior of the network that consists of vertices in V as a set of actors and edges in E as a set of relations for all e_j in E, $j = 1,\ldots,J$ [42]. Meanwhile, extracting social networks may be represented by $SN = \langle V,E,A,R,B,\gamma_1,\gamma_2 \rangle$ with the conditions as following [9]:

a. $\gamma_1(1:1) : A \rightarrow V$.

b. $\gamma_2 : R \rightarrow E$ in order to $e_j = \gamma_2(r_s(a_k,a_l)) = \gamma_2(B_{ak} \cap B_{al})$

where r_s in R, $R = A \times A$, a_k in A, a_l in A, B_{ak} is a subset of B, B_{al} is a subset of B, and $B_{ak} \cap B_{kl}$ is a subset of B. Especially, based on the superficial methods, r_s means the strength relations between two actors a_k and a_l in A by involving one or more of the similarity measurements [43]: mutual information, Dice coefficient, overlap coefficient, cosine, or for example Jaccard coefficient is

$$J_c = (|\Omega_{ak} \cap \Omega_{al}|)/(|\Omega_{ak}|+|\Omega_{al}| - (|\Omega_{ak} \cap \Omega_{al}|)) \quad (2)$$

in [0,1]. In this concept of similarity, $B_{ak} \cap B_{al} = (|\Omega_{ak} \cap \Omega_{al}|)/(|\Omega_{ak}|+|\Omega_{al}| - |\Omega_{ak} \cap \Omega_{al}|) = J_c$ such that e_j in E if $r_s > 0$. However, the behavior of r_s ($0 \leq r_s \leq 1$) depends on the behavior of: Ω_{ak} as a subset Ω, Ω_{ak} as a subset of Ω, and $\Omega_{ak} \cap \Omega_{al}$ as a subset of Ω, where $|\Omega_{ak}| \leq |\Omega_{al}|$ or $|\Omega_{ak}| \geq |\Omega_{al}|$, $|\Omega_{ak} \cap \Omega_{al}| \leq |\Omega_{ak}|$, and $|\Omega_{ak} \cap \Omega_{al}| \leq |\Omega_{al}|$ [11]. If another measurement concept is similar to J_c, then Theorem 1 is proved.

Corollary 1: If the behavior of social actors behaves in a cluster based on the information source then the behavior of the clusters is representation of extracting social network from the information source [11].

A social network, has been extracted *SN* have behavior that is based on the resources, whereby they can be recognized through statistical theory [44].

3. AN APPROACH

By population, in literal word or term in Web represent objects or things are discussed or events that occur in the real world [41],[45],[46]. Web as social media becomes an overview of the social behavior directly related to the personality of social actors or coherence of the social community [47]. One of the social communities that have an influence on changes in the world is the academic community [48],[49]. We will take therefore from the academic population some communities, namely the academic actors as the seed to form a growth picture of the social network [3]. The community of academic actors, as a sample, has continuously generated the web pages and then directly clustering of web pages. To explore the behavior of clusters we propose an approach based on statistical theory and graph theory [6],[9].

Suppose t_a = Zainal A, Hasibuan and t_a in q when q submitted to any search engine. It will generate a cluster Ω_a. It represents the hit count $|\Omega_a|$, or a number of an actor name already present on the Web. Similarly, t_a = "Zainal A, Hasibuan" and we have a cluster $\Omega_{\text{"}a\text{"}}$. It represents the hit count $|\Omega_{\text{"}a\text{"}}|$, is as number of an actor well-defined name already present on the Web [12].

That is a number of web pages that contain the name of an actor. Therefore, the hit count represents the behavior of social actors in the form of academic activities: dissemination of knowledge through scientific works and it forms a behavior of the clusters produced by a search engine for a query [50].

For engendering the behavior from the data and all of things represented by the data, the data should show measurable characteristics. That is representative sample: Quantitatively, for the sample size $n > 20$ sample distribution close to the average μ and variance σ of the normal distribution Z [51]. Furthermore, for a quantities sample which consists of n clusters may be measured by

$$Y = |\Omega_1| + |\Omega_2| + \ldots + |\Omega_n|$$

in order to produce a constant of internal consistency of the sample behavior in general, that is α Cronbach [52] or

$$\alpha = (n/(n-1))(1-\Sigma_{i=1\ldots n}(\sigma_{|\Omega i|})^2/(\sigma_Y)^2) \quad (3)$$

by which $(\sigma_Y)^2$ is the variance of the score total was observed while $(\sigma_{|\Omega i|})^2$ is the variance of i-component for sample Ω_i. Variance is calculated using $(\sigma_{|\Omega i|})^2 = (1/n)\ \Sigma|\Omega_i|-|\bar{\Omega}_i|$ in which $|\bar{\Omega}_i|$ is the average. In general, the rules of use α are to use the marker as follows [9,13,53]:

1. α≥0.9: The internal consistency of behavior is very good.
2. 0.7≤α<0.9: The internal consistency of behavior is good.
3. 0.6≤α<0.7: The internal consistency of behavior is acceptable
4. 0.5≤α<0.6: The internal consistency of behavior is poor.
5. α<0.5: The internal consistency of behavior is not accepted.

A social network that extracting directly from the dynamic resources, such as the Web, creating a social network dynamically. To avoid ambiguity and bias, extraction of social networks based on the name of a well-defined social actor as a seed [13]. Thus, expressing the behavior of growing social network initially is from a seed. It is based on using the multiple-regression to generate the behavior of relationship between resources [54]: actors/vertices, relationships/edges, and documents/web pages [55]. In the multiple-regression, the independent variables x_i $i = 1,\ldots,n$ and dependent variable Y, the average of $Y|x_i$ given by linear regression models [56]

$$\mu_{Y|xi} = b_0 + \Sigma_{i=1\ldots n} b_i x_i$$

and the estimation of responses obtained from the regression equation of a sample is

$$\hat{y} = \beta_0 + \Sigma_{i=1\ldots n} \beta_i x_i, \quad (4)$$

Suppose there are 3 (three) independent variables, x_1, x_2, and x_3, in equation (4), the relationship of each variable to the other builds a directed graph, the arrangements of vertices in graph for determining effects namely $s_1 = (x_1x_2)$, $s_2 = (x_1x_3)$, $s_3 = (x_1y)$, $s_4 = (x_2x_3)$, $s_5 = (x_2y)$, $s_6 = (x_3y)$, and $s_7 = (x_1x_2x_3)$. Direct effect of the seeds toward edges is s_3, while the indirect effect is $s_1 + s_2 + s_7$.

The first case yields $\hat{y}$ as a representation of papers growth to be regressed against one or more factors [57]:

1. x_1 as a representation of web pages or number of documents (NoP).
2. x_2 as a representation of vertices growth or number of social actors (NoA).
3. x_3 as a representation of edges growth (NoE).

The second case, in a similar way that is by involving the accretion of seeds in each year, building a social network for an academic community is realization. But it did not involve the hit count.

Revealing the behavior of relations between resources can be modeled naturally by a direct graph $G(V,E)$ as an approach to visualize the behavior that consist of: the vertices in V as a set of factors and the direct edges in E as a set of relations between resources with weights, i.e. β_i, $i = 1,\ldots,n$ of each linear regression models [58],[59]. Based on the direct graph we can compute a total relation tr as follows

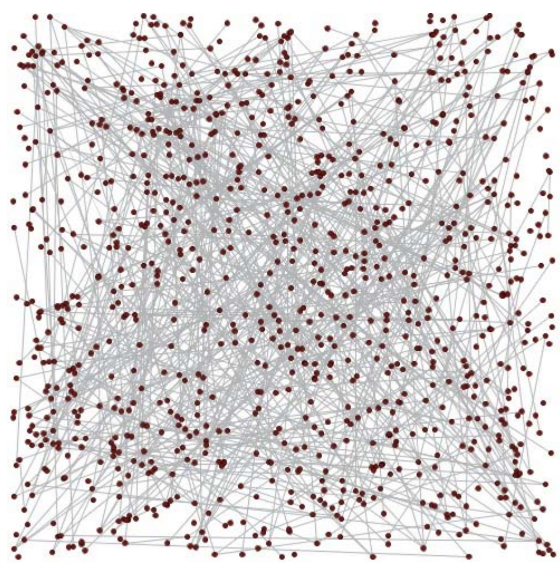

Figure 1: A social network based on seeds.

$$tr = \Sigma_{i=1\ldots n}\Pi_{j=1\ldots I}\ \beta_j \quad (5)$$

where $\beta_1 = \Pi_{i=1}\ \beta_i$ means a direct effect and $\Pi_{i=2\ldots n}\ \beta_i$ means the indirect effect from a resource to a destination.

Generally, for mining the extracted social network, the steps in this approach are as follows:

1. Determine the seeds and set as collection of social actors, A.
2. Collection of documents based on seeds from DBLP where there are authors and co-authors, titles of documents, and others.

Journal of Theoretical and Applied Information Technology
15th July 2021. Vol.99. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Collection of authors and co-authors into *A* follow the time of appearance.

3. Collection another actors from set of co-authors in step 2.
4. Generate $|\Omega_a|$ or $|\Omega_{"a"}|$ for each actor social *a* in *A*.
5. Compute α in equation (3) for number of documents from DBLP (*P*), $|\Omega_a|$, and $|\Omega_{"a"}|$, make the marker and determine initial behavior.
6. Generate social networks following time of year for each seed, and accumulative social network by reducing the overlap vertices and edges.
7. Calculate the prediction of the model by using multiple-regression for each seed and for all seeds.
8. Compute relations between resources of social networks for generating β factors.
9. Based on relations between resources, determine behavior of social actors, community, and social networks.

4. RESULTS AND DISCUSSION

For an experiment, we have defined several 36 names of actors as the seeds and then generate other actors and build social networks like Figure 1, which consists of 17 Professors (Pr) and 19 Associate Professors (AP) with a transition *tr*. For each seed, we collect from DBLP: the number of papers p_i in *P*, $i = 1,\ldots,36$, the cardinalities of clusters (hit count) with query *q* with and without "…", i.e. $|\Omega_{"a"}|$ and $|\Omega_a|$, respectively, see Table 1.

We obtain the average of p_i, $|\Omega_{"a"}|$ and $|\Omega_a|$ are $\bar{p} = 21.194$, $|\bar{\Omega}_{"a"}| = 15226.64$, and $|\bar{\Omega}_a| = 53422.22$, respectively. We provide label *lb* for item of data: If $p_i \leq \bar{p}$, then $lb = -1$, otherwise $lb = 1$; and similarly for $|\Omega_{"a"}|$ and $|\Omega_a|$. We have taken a community (of a university): 31 actors of the first sequence are a sample. The rest is coming from different academic communities as the additional sample. In this case based on *T* in Table 1., we have grouped each of the data set by the run *r* ⌂ {18,8,14}. Generally, for significance level α = 0.05, all values *Z* in $Z_{-½\alpha} < Z < Z_{½\alpha}$. Therefore, the behavior about resources: Papers or web pages randomly represent either the seeds or the social networks. Each cluster represents the behavior of actor as a seed. It is also supported by the internal consistency of behavior, that is very good, where each of the data set (Table 1) have $\alpha_C = 1.0286 > 0.9$. In other words, each cluster Ω_α of the second and third samples can generally be represented by the sample in the collection of papers. Figure 1 shows that adjacent vertices are social actors that do not have a stronger relationships with distant vertices topologically or in a graph generally represented as leaves. It is based on different treatment by using unsupervised method, distant relation extraction has been an effective way only for texts in corpus and using supervised method [60].

Table 1: Seeds, number of documents and hit count for each seed.

Set	Data
T	Pr, Pr, Pr, Pr, AP, AP, AP, Pr, AP, Pr, AP, AP, Pr, AP, AP, AP, AP, AP, AP, AP, AP, AP, AP, AP, Pr, Pr, AP, Pr, AP, Pr, AP, Pr, Pr, Pr, Pr, Pr
P	28, 53, 66, 11, 8, 13, 4, 48, 7, 29, 6, 8, 28, 8, 16, 13, 8, 18, 14, 18, 12, 9, 11, 56, 48, 37, 40, 12, 10, 33, 12, 33, 7, 9, 12, 10
$\|\Omega_{"a"}\|$	6830, 5410, 4690, 28200, 380000, 1590, 336, 14300, 1610, 6720, 2120, 1920, 3900, 5640, 1890, 2030, 1840, 1870, 1970, 2930, 2690, 1630, 2140, 1280, 5750, 2850, 267, 3100, 2150, 2740, 3610, 10700, 4150, 2030, 16400, 876
$\|\Omega_a\|$	383000, 131000, 5290, 93300, 487000, 2130, 14700, 15900, 1920, 238000, 6400, 4240, 19200, 1890, 2340, 13300, 2120, 2300, 2670, 10800, 25100, 2220, 7700, 4700, 40100, 19000, 17700, 1920, 4770, 20000, 12000, 102000, 4540, 3950, 17000, 243000

Concentrated vertices randomly in the center are vertices of on different sides. Topologically, vertex behavior will underlie the formation of structured communities, where maybe an actor occurs connected or not with other actors [61]. Other vertices growth with following years, but stem from seeds in particularly and differently. The vertices are clustered around the seed, and it is possible to form a research group.

In the context of exploration, the behavior of growth of the social network bases on the prediction models of resources takes heavy a position as a conduit of information about the dynamism of social network. Multiple regression is one of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

methods is to determine the causal relationships between factors as resources of the social network.

Table 2: Number of documents in DBLP for Prof. Shahrul Azman Noah (SAM) [9] *and Tengku M. T. Sembok (TMTS).*

++	SAM			TMTS		
Year	NoP	NoA	NoE	NoP	NoA	NoE
1990				1	2	1
1995	1	2	1			
1996				2	4	4
1998	2	2	1			
1999	3	2	1			
2000	4	3	2	4	7	11
2002	5	5	5	5	10	17
2003	6	12	33	11	26	69
2004	8	13	37	13	27	74
2005	14	19	49	16	30	83
2006	15	20	51	20	35	92
2007	17	23	55	21	36	94
2008	18	24	56	22	37	96
2009	23	40	195	25	49	227
2010	32	55	235	26	51	230
2011	41	71	272	33	56	243
2012	47	74	278	34	62	277
2014	48	78	301	40	71	321

From Table 2., we conduct the modeling: In accumulation into a year, the dependent variable y successively is the number of papers (NoP), number of actors (NoA), and number of edges (NoE), while the first independent variable is the number of seeds (NoS) and then on the following equations respectively. The independent variables involved are NoP and NoA such that the models are

$$\acute{y} = -57.4284 + 12.0279x_1 \quad (6)$$

$$\acute{y} = 183.3478 - 0.3471x_1 + 1.2190x_2 \quad (7)$$

$$\acute{y} = -1454.4764 - 768333745.9175x_1 + 11.0897x_2 + 10.4489x_3 \quad (8)$$

Based on equations (6), (7) and (8), we have the relations between resources: β_1 ⌂ {(NoS(Year)-NoP), (NoS-NoA), (NoS-NoE)}, β_2 ⌂ {(NoP-NoA), (NoP-NoE))}, β_3 ⌂ {(NoA-NoE)}.

Data collection from DBLP is based on the following principles: Every year there are additional published documents, which are accumulatively stated the number of documents up to that year. If different authors appear, there will be additional social actors accumulatively, but if the documents come from the same author, there will be no additional social actors. Thus, automatically the number of edges follows the addition of vertices, and the addition of edges is not proportional to the addition of social actors. If there are no additional documents in that year, it means that they are cumulatively the same as the previous year and do not need to be stated in a table, such as Table 2. For each seed, the data collected is related to the NoP, NoA, and NoE, see also Table 3. To reveal the overall growth of the seeds, the data collected were NoS, NoP, NoA, and NoE, see Table 4. In the same way, equations of multiple-regression: (6), (7) and (8) generate relations between resources for the seeds SAM, TMTS, ZAH, and another, respectively. Similarly by using equations of multiple-regression for generating relation between resources based on all of seeds.

Table 3: Number of documents in DBLP for Prof. Zainal A. Hasibuan (ZAH) and …

++	ZAH			Another seed		
Year	NoP	NoA	NoE	NoP	NoA	NoE
2003	2	3	2	…	…	…
2004	3	6	8	…	…	…
2005	6	10	18	…	…	…
2006	8	13	22	…	…	…
2012	9	14	23	…	…	…
2013	10	15	24	…	…	…
2014	12	18	31	…	…	…

Table 4: Number of documents in DBLP for all of seeds

++	All of Seeds			
Year	NoS	NoP	NoA	NoE
1990	1	1	2	1
1992	1	2	4	2
1994	1	3	6	3
1995	2	5	9	4
1996	2	6	13	10
1997	3	8	16	13
1998	3	10	19	22
1999	4	13	22	25
2000	4	18	29	37
2001	7	21	37	46
2002	12	35	57	79
2203	16	50	71	144
2004	17	60	91	165
2005	24	84	129	304
2006	25	101	148	336
2007	28	123	175	413
2008	30	154	200	468
2009	31	224	273	729
2010	34	284	334	891
2011	36	460	559	1473
2012	36	517	610	1629
2013	36	567	673	1818
2014	36	633	730	1998

By using equation (5), we get the behavior of the social networks in Table 4, i.e., the behaviors of growth of social network based on the seeds: For all

Journal of Theoretical and Applied Information Technology
15th July 2021. Vol.99. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of the seeds as generator of the social networks, there is a positive effect of the seeds on the growth of papers, so there is a positive effect of papers on the growth of vertices and edges, and then a positive effect of actors for growing the edges in social network, although generally there is no positive relationship between the resources of social networks in growth based on a single seed, i.e.,

[Seed-1]: SAM, there are direct effect

$$\beta_1 = -736280.5940, \quad (9)$$

indirect effects

$$\beta_1\beta_2 = -75.6156, \quad (10)$$

$$\beta_1\beta_3 = 8.7688, \quad (11)$$

$$\beta_1\beta_2\beta_3 = -736496.2024, \quad (12)$$

and

$$tr = -736496.2024. \quad (13)$$

[Seed-2]: TMTS, there are direct effect

$$\beta_1 = -471371.6258, \quad (14)$$

indirect effects

$$\beta_1\beta_2 = -56.8788, \quad (15)$$

$$\beta_1\beta_3 = 100.6473, \quad (16)$$

$$\beta_1\beta_2\beta_3 = -297.7641, \quad (17)$$

and

$$tr = -471826.9160. \quad (18)$$

[Seed-3]: ZAH, there are direct effect

$$\beta_1 = -2253.4364, \quad (19)$$

indirect effects

$$\beta_1\beta_2 = -7.2707, \quad (20)$$

$$\beta_1\beta_3 = -10.6096, \quad (21)$$

$$\beta_1\beta_2\beta_3 = -76.6508, \quad (22)$$

and

$$tr = -2194.6659. \quad (23)$$

[Seed-…]: another.

In this case, there is an effect of a collection of papers on the growth of edges, where β in multiple-regression. The behavior of each cluster Ω_a (like in Table 1) also contains the relationship between the actors who produce the edges in social networks. In other words, If every web page representing the activity of a social actor, then the class of web pages represent the behavior of a social actor. The behavior is hidden behind the source of the information [12]. Thus, data with a system that is incorporated in the web as a source of information becomes a document that has an influence on the formation of social actor behaviors [62]. Equations: (9) until (23) or etc., case by case reveal addition evidence experimentally against Theorem 1. That is, it is an implication: SN => V,E; V => A and E => $A\times A$, a in A => ω in Ω. It states that any added webpage or document effect the behavior of the social network. Meanwhile, social network behavior determines the behavior of a social actor.

Furthermore, every web page represents the communication (modality) of social actors, where the web pages represent direct the behavior of the relation between social actors. Thus, the individual behavior of social actors also accumulates into social network behavior as expressed through social actors and emphasized by the relationship between social actors [12]. The behavior of different resources, on the growth of social networks dynamically, has proven Corollary 1 by experimental data, where there are similarities in behavior [63], i.e., a similarity between events and similarity between structures of web such as a consequence of computer networks.

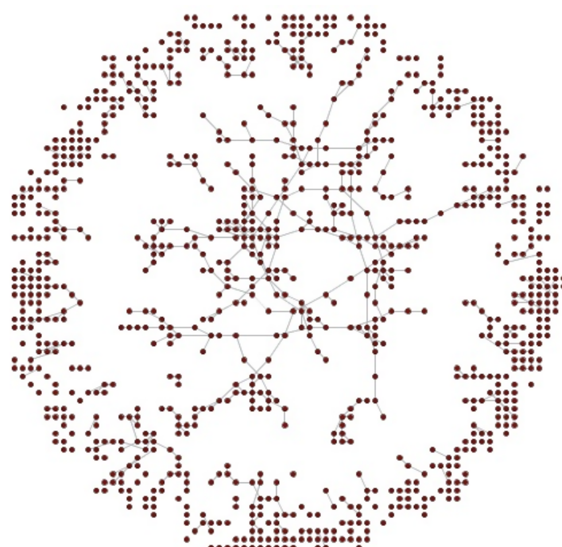

Figure 2: The communities in a social network.

Based on statistic computation of Table 4, direct effect

$$\beta_1 = -768333745.9000, \quad (24)$$

indirect effects:

$$\beta_1\beta_2 = 133.3858, \quad (25)$$

$$\beta_1\beta_3 = -3.6268, \quad (26)$$

and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\beta_1\beta_2\beta_3 = 153.2019, \quad (27)$$

where

$$tr = -768333463.0000. \quad (28)$$

In implication: Behavior of equations (9)-(23) means that behavior of equations (25)-(28).

In other words, in context of that behavior, it is shown computationally statistics on the equations (9) until (28) that the social network based on seeds have behavior like in Figure 2. Figure 2 reveals the formation of communities through social networks in Figure 1. All seeds form the core of social networks, although among the seeds also form communities and it reveals that academically there are research groups. Meanwhile, the members of the same community are outside the core boundaries of the community. The members form a community of newcomers who are particularly close to the core community. Member behavior is the same as core behavior. The behavior of the same members in particular expresses the closeness of their fellowmen as the formation of self-concentrated vertices in Figure 2.

Members of community individually or collectively engagement with certain levels with seeds individually or the core community [64], where there is collaboration between individuals as the core of the community. But also, when the core communities collaborated, and this was followed by other members of the community, where each a tree has engagement with another and then become a forest of communities in a social network [65,66].

Indirectly, social network extraction through mining carries out engagement extraction to reveals the attractiveness of each community [67],[68],[69]. It is what Figure 2 shows structurally, representing the social network in Figure 1. Thus, research groups that interact with each other and have the same scientific mission form a center of excellence. For example, Center for Artificial Intelligence Technology (CAIT), *Fakulti Teknologi & Sains Maklumat* (FTSM), *University Kebangsaan Malaysia* (UKM), *Bangi, Selangor,Malaysia* (http://www.ftsm.ukm.my/cait/) [70], wherein some of the seeds in Table 1 are Pr and AP which developed the research groups of CAIT. Meanwhile, research group outside of CAIT based on the concept of attraction are in a collaborative position and look out for the outside of the core social network. Likewise, members of the research group as students initially and later become alumni and free from the research group are also on the outside of the core social network. Figure 3 shows one of flyers of CAIT-FTSM-UKM.

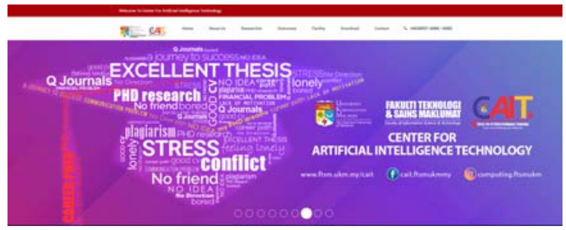

Figure 3: One of flyers on Website of CAIT, FTSM, UKM, Malaysia.

Therefore, the extraction of social networks from information sources based on seeds has demonstrated the growth behavior of the resources and the relations between them, and that the behavior of each seed supports the behavior of a community as a whole. Data processing of the extracted social network based on multiple regressions is one way of mining social networks, and it differs from applying the degree of vertices in social network analysis [25]. Thus, it indicates the extraction of engagement as a consequence of social network mining based on scientific publication documents.

5. CONCLUSION

In this social network study, we have presented a different way for finding behavior of resources of social network where it is not social network analysis, so it is a social network mining. A statistical measurement of the experiment is to complete the proof of the proposed theory. Each actor has the influence for concept of clustering web pages, where its representation is the behavior of the actor or their socials. Based on statistic computation, the resources have the similar behaviors towards growth of social network, where there are the positive effects on relation between the resources based on prediction models of multiple-regression. It has shown that social network behavior based on community behavior and the community behavior based on seeds behavior. Then social network extraction also reveals engagement extraction. The future work will involve the extraction of a social network to describe the research collaboration.

REFERENCES:

[1] A. Agarwal, "Social network extraction from texts: A thesis proposal", *ACL HLT 2011 – 49th Annual Meeting of the Association for Computational Linguistics: Human Language*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Technologies, Proceedings of Student Session, 2011, pp. 111-116.

[2] A. Agarwal, O. Rambow, and R. J. Passonneau, "Annotation scheme for social network extraction from text", *ACL 2010 – LAW 2010L 4th Linguistic Annotation Workshop, Proceedings*, 2010, pp. 20-28.

[3] M. K. M. Nasution, and S. A. M. Noah, "Superficial method for extracting social network for academic using Web snippets", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 6401 LNAI, 2010, pp. 483-490.

[4] R. Potapova, and L. Komalova, "Lexico-semantical indices of "deprivation – aggression" modality correlation in social network discourse", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10458 LNAI, 2017, pp. 493-502.

[5] A. K. Bhowmick, and B. Mitra, "Listen to me, my neighbors or my friend? Role of complementary modalities for predicting business popularity in location based social networks", *Computer Communications*, Vol. 135, 2019, pp. 53-70.

[6] M. K. M. Nasution, M. Elveny, R. Syah, and S. A. Noah, "Behavior of the resources in the growth of social network", *Proceedings of the 5th International Conference on Electrical Engineering and Informatics*, IEEE, 2015, pp. 551-554.

[7] M. Brambilla, and A. Bozzon, "Web data management through crowdsourcing upon social networks", *IEEE/ACM International Conference on Advances in Social Network Analysis and Mining*, 2012, pp. 1123-1127.

[8] P. Moradianzadeh, M. Mohi, and M. Sadighi Moshkenani, "Digital social network mining for topic discovery", *Communications in Computer and Information Science*, 6 CCIS, 2008, pp. 1000-1003.

[9] M. K. M. Nasution, "Social network mining (SNM): A definition of relation between the resources and SNA", *International Journal on Advanced Science, Engineering and Information Technology*, Vol. 6, No. 6, 2016, pp. 975-981.

[10] S. Siersdorfer, P. Kemkes, H. Ackermann, and S. Zerr, "Who with whom and how? Extracting large social networks using search engine", *International Conference on Information and Knowledge Management, Proceeding*, CIKM'15, 2015, pp. 1491-1500.

[11] M. K. M. Nasution, "Extracted social network mining", *Proceeding of International Conference on Information Technology and Engineering Application*, 5th ICIBA, 2016, pp. 19-20. arXiv: 1604.06976 [cs.SI].

[12] M. K. M. Nasution, M Hardi, and R. Syah, "Mining of the social network extraction", *Journal of Physics: Conference Series*, Vol. 801, No. 1, 2017.

[13] M. K. M. Nasution, O. S. Sitompul, E. P. Sinulingga, and S. A Noah, "An extracted social network mining", *Proceedings of 2016 SAI Computing Conference*, SAI 2016, 2016, pp. 1168-1172.

[14] M. K. M. Nasution, "Social network mining: A discussion", *Journal of Physics: Conference Series*, Vol. 1235, No. 1, 2019.

[15] T. Beck, C. Biermann, D. Fritz, and R. Dillmann, "Robust model-based centerline extraction of vessels in CTA data", *Progress in Biomedical Optics and Imaging – Proceedings of SPIE*, Vol. 7259, 2009.

[16] P. Joshi, S. Chaudhary, and V. Kumar, "Information extraction from social network for agro-produce marketing", *Proceedings – International Conference on Communication Systems and Network Technologies*, CSNT 2012, 2012, pp. 941-944.

[17] R. Santana, R. Armañanza, C. Bielza, and P. Larrañag, "Network measures for information extraction in evolutionary algorithms", *International Journal of Computational Intelligence Systems*, Vol. 6, No. 6, 2013, pp. 1163-1188.

[18] S. K. Endarnoto, S. Pradipta, A. S. Nugroho, and J. Purnama, "Traffic condition information extraction & visualization from social media twitter for android mobile application", *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*, ICEEI, 2011.

[19] H Roy, S. E. Kase, and J. Knight, "Transforming a research-oriented dataset for evaluation of tactical information extraction technology", *Proceedings of SPIE – The International Society for Optical Engineering*, 2016.

[20] G. Skobeltsyn, F. P. Junqueira, V. Plachouras, and R. Baeza-Yates, "ResIn: A combination of results caching and index pruning for high-performance web search engines", *ACM SIGIR 2008 – 31st Annual International ACM SIGIR*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Conference on Research and Development in Information Retrieval, Proceedings, 2008, pp. 131-138.

[21] C. S. Meredith, A. S. Trebitz, and J. C. Hoffman, "Resolving taxonomic ambiguities: Effects on rarity, projected richness, and indices in macroinvertebrate datasets", *Ecological Indicators*, Vol. 98, 2019, pp. 137-148.

[22] M. K. M. Nasution, "Singleton: A role of the search engine to reveal the existence of something in information space", *IOP Conference Series: Materials Science and Engineering*, Vol. 420, No. 1, 2018.

[23] M. K. M. Nasution, "Doubleton: A role of the search engine to reveal the existence of relation in information space", *IOP Conference Series: Materials Science and Engineering*, Vol. 420, No. 1, 2018.

[24] V. Vijayarajan, M. Dinakaran, and M. Lohani, "Ontology based object-attribute-value information extraction from web pages in search engine result retrieval", *Smart Innovation*, *Systems and Technology*, 27(Vol 1), 2014, pp. 611-620.

[25] Barry Wellman, "For a social network analysis of computer networks: a sociological perspective on collaborative work and virtual community", *Proceedings of the ACM SIGCPR Conference*, 1996, pp. 1-11.

[26] B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, C. Haythorntkwaite, "Computer networks as social networks: Collaborative work, telework, and virtual community", *Annual Review of Sociology*, Vol. 22, 1996, pp. 213-238.

[27] B. Wellman, "Computer networks as social networks", *Science*, Vol. 293, No 5537, 2001, pp. 2031-2034.

[28] B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, C. Haythornthwaite, "Computer networks as social networks: Collaborative work, telework, and virtual community", *Knowledge and Communities*, 2009, pp. 179-208.

[29] G. Cordasco, L. Gargano, A. A. Rescigno, and U. Vaccaro, "Evangelism in social networks: Algorithms and complexity", *Networks*, Vol. 71, No. 4, 2018, pp. 346-357.

[30] F. Arifin, H. Hariadi, and M. Anshari, "Extracting value and data analytic from social networks: Big data approach", *Advanced Science Letters*, Vol. 23, No. 5, 2017, pp. 5286-5288.

[31] V. Kakulapati, and S. Mahender Reddy, "Mining social networks: Tollywood reviews for analyzing UPC by using big data framework", *Advances in Intelligent Systems and Computing*, Vol. 851, 2019, pp. 323-334.

[32] M. K. M. Nasution, and S. A. Noah, "Information retrieval model: A social network extraction perspective", *Proceedings – 2012 International Conference on Information Retrieval and Knowledge Management*, CAMP'12, 2012, pp. 322-326.

[33] M. K. M. Nasution, and S. A. Noah, "Extraction of academic social network from online database", *2011 International Conference on Semantic Technology and Information Retrieval*, STAIR 2011, 2011, pp. 64-69.

[34] P. Ampornphan, and S. Tongngam, "Exploring technology influencers from patent data using association rule mining and social network analysis", *Information* (Switzerland), Vol. 11, No. 6, 2020.

[35] C. Makris, and G. Pispirigos, "Stacked community prediction: A distributed stacking-based community extraction methodology for large scale social networks", *Big Data and Cognitive Computing*, Vol. 5, No. 1, 2021.

[36] L. Jintao, R. Licheng, Y. Yinpu, and X. Guangdong, "Research on the evolutionary mechanism of social network based on competitive selection", *IEEE International Conference on Computational Aspects of Social Networks*, 2010, pp. 524-628.

[37] J. Golbeck, and J. Hendler, "Accuracy of metrics for inferring trust and reputation in semantic web-based social networks", *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, Vol. 3257, 2004, pp. 116-131.

[38] S. Downes, "Semantic networks and social networks", *Learning Organization*, Vol. 12, No. 5, 2005, pp. 411-417.

[39] P. Abrosimova, I. Shalyaeva, and L. Lyadova, "The ontology-based event mining tools for monitoring global processes", *IEEE 12th International Conference on Application of Information and Communication Technology, AICT 2018 – Proceedings*, 2018.

[40] W. S. Wan, O. Dastane, N. S. M. Satar, M. Y. Ma'arif, "What we chat can learn from whatsapp? Customer value proposition development for mobile social networking (MSN) apps: A case study approach", *Journal of Theoretical and Applied Information*

Technology, Vol. 97, No. 4, 2019, pp. 1091-1117.

[41] M. Bendechache, S. Svorobej, P. T. Endo, A. Mihai, and T. Lynn, "Simulating and evaluating a real-world elasticsearch system using the recap des simulator", *Future Internet*, Vol. 13, No. 4, 2021.

[42] R. Harrison, G. Hernandez, and R. Muñoz, "A discrete model of market interaction in the presence of social networks and price discrimination", *Mathematical Social Sciences*, Vol. 102, 2019, pp. 48-58.

[43] M. Lubani, and S. A. M. Noah, "Text relation extraction using sentence-relation semantic similarity", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in BioInformatics)*, Vol. 11909 LNAI, 2019, pp. 3-14.

[44] G. Robins, P.Pattison, and S. Wasserman, "Logit models and logistic regression for social networks", *Psychometrika*, Vol. 64, 1999, pp. 371-394.

[45] M. Biasini, V. Carmignani, N. Ferro, P. Filianos, M. Maistro, and G. M. di Nunzio, "FullBrain: A social e-learning platform", *CEUR Workshop Proceedings*, Vol. 2816, 2021, pp. 25-41.

[46] P. Abrosimova, I. Shalyaeva, and L. Lyadova, "The ontology-based event mining tools for monitoring global processes", *IEEE 12th International Conference on Application of Information and Communication Technology, AICT 2018 – Proceedings*, 2018.

[47] A. N. Sergeev, "Learning in online communities: Behavioral strategies of the users of educational social networks", *International Journal of Criminology and Sociology*, Vol. 10, 2021, pp. 84-92.

[48] R. Palmieri, and C. Giglio, "Informal learning in online social network environments: An evidence from an academic community on facebook", *Turkish Online Journal of Educational Technology*, 2015, pp. 279-287.

[49] E. L. Atieh, K. L. Chun, R. Shah, F. Guerra, and D. M. York, "Creation of academic social network (ASNs) for effective online elearning community", *ACS Symposium Series*, Vol. 1217, 2016, pp. 106-126.

[50] M. K. M. Nasution, and S. A. Noah, "Social network extraction based on Web. A comparison of superficial methods", *Procedia Computer Science*, Vol. 124, 2017, pp. 86-92.

[51] M. Coscia, "Noise corrected sampling of online social networks", *ACM Transactions on Knowledge Discovery from Data*, Vol 15, No. 2, 2021.

[52] S. Torkian, A. Shahesmaeili, N. Malekmohammadi, and V. Khosravi, "Content validity and test-retest reliability of a questionnaire to measure virtual social network addiction among students", *International Journal of High Risk Behaviors and Addiction*, Vol. 9, No. 1, 2020, pp. 1-5.

[53] S. H. Hoseini, and N. Tavrah, "The study of effect of social networks to attract banking customers by taking the competitive advantage index (case study: Bank Refah, South of West Azerbaijan Province)", *International Business Management*, Vol. 10, No. 7, 2016, pp. 1120-1224.

[54] M. Talukder, A. Quazi, and D. Djatikusumo, "Impact of social influence on individuals' adoption of social networks in smes", *Journal of Computer Science*, Vol. 9, No. 12, 2013, pp. 1684-1694.

[55] J. H. Koskinen, and T. A. B. Snijders, "Bayesian inference for dynamic social network data", *Journal of Statistical Planning and Inference*, Vol. 137, No. 12 SPEC. ISS, 2007, pp. 3930-3938.

[56] D. Hassan, "Predicting the degree of collaboration of researchers on co-authorship social networks", *Advances in Intelligent Systems and Computing*, Vol. 923, 2020, pp. 75-84.

[57] T. Y. Chin, and M.M. Singh, "Multi-facet trust factors ranking with correlation analysis technique: A case study on online social network (OSN)", *Proceedings of the 26th International Business Information Management Association Conference – Innovation Management and Sustainable Economic Competitive Advantage: From Regional Development to Global Growth*, IBIMA 2015, 2015, pp. 1812-1822.

[58] L. De-Marcos, E. Garcia-Lopez, A. Garcia-Cabot, J.-A. Medina-Merodio, A. Dominguez, J.-J. Martinez-Herraiz, and T. Diez-Folledo, "Social network analysis of a gamified e-learning course: Small-world phenomenon and network metrics as predictors of academic performance", *Computers in Human Behavior*, Vol. 60, 2016, pp. 312-321.

[59] K. Hedge, M. Magdon-Ismail, B. Szymanski, and K. Kuzmin, "Clustering, prominence and social network analysis on incomplete

networks", *Studies in Computational Intelligence*, Vol. 693, 2017, pp. 287-298.

[60] Z, Zhou, Y. Cai, J. Xu, J. Xie, Q. Li, H. Xie, "A distant supervised relation extraction model with two denoising strategies", *Proceedings of the International Joint Conference on Neural Networks*, Vol. 2019-July, 2019.

[61] K. Taha, "Detecting disjoint communities in a social network based on the degrees of association between edges and influential nodes", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 33, No. 3, 2021, pp. 935-950.

[62] R. R. Saedudin, S. Kasim, H. Mahdin, I. T. R. Yanto, "A comparative analysis of rough sets for incomplete information system in student dataset", *International Journal on Advanced Science, Engineering and Information Technology*, Vol. 7, No. 6, 2017, pp. 2078-2084.

[63] D. S. Sisodia, "Augmented session similarity based framework for measuring web user concern from web server logs", *International Journal on Advanced Science, Engineering and Information Technology*, Vol. 7, No. 3, 2017, pp. 1007-1013.

[64] F. Banhawi, N. Mohamad Ali, and H. M. Judi, "Measuring user engagement levels in social networking application", *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*, ICEEI, 2011.

[65] F. Banhawi, and N. M. Ali, "Measuring user engagement attributes in social network application", *2011 International Conference on Semantic Technology and Information Retrieval*, STAIR, 2011, pp. 297-301.

[66] G. S. Saragih, Z. Rustam, D. Aldila, R. Hidayat, R. E Yunus, and J. Pandelaki, "Ischemic stroke classification using random forests based on feature extraction of convolutional neural networks", *International Journal on Advanced Science, Engineering and Information Technology*, Vol. 10, No. 5, 2020, pp. 2177-2182.

[67] G. Wang, Y. Zhao, X. Shi, and P. S. Yu, "Magnet community identification on social networks", *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012, pp. 588-596.

[68] K. Vasssakis, E. Petrakis, I. Kopanakis, J. Makridis, and G. Mastorakis, "Location-based social network data for tourism destinations", *Big Data and Innovation in Tourism, Travel, and Hospitality: management Approaches, Techniques, and Applications*, 2019, pp. 105-114.

[69] M. Kodad, "Social networks: Generator of performance test and measurement of big data publications' influence", *Journal of Theoretical and Applied Information Technology*, Vol. 98, No. 20, 2020, pp. 3272-3279.

[70] M. K. M. Nasution, S. Azman Mohd Noah, O. Salim Sitompul, "The excellent center: Center for Artificial Intelligence Technology", *Journal of Physics: Conference Series*, Vol. 1235, No. 1, 2019.