Conference Information

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2017 J. Phys.: Conf. Ser. 801 011002
(http://iopscience.iop.org/1742-6596/801/1/011002)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 36.66.254.225
This content was downloaded on 24/03/2017 at 15:38

Please note that terms and conditions apply.
Conference Information

Dates: December 14 - December 15, 2016
Organizer: Faculty of Computer Science and Information Technology
University of Sumatera Utara
Venue: Hotel GranDhika Setiabudi Medan
Jalan Dr. Mansyur no. 169
Medan, Sumatera Utara
Indonesia
Phone: +62-61-42081234
Fax: +62-61-42081234
Email: reservation@grandhika-hotel.com
Website: www.grandhika-hotel.com
Official Language: English
Secretariat: Faculty of Computer Science and Information Technology
University of Sumatera Utara
Jalan Universitas No. 9A, Kampus USU
Medan, 20155
Sumatera Utara, Indonesia
Email: iccai@usu.ac.id
Conference Website: http://iccai.usu.ac.id
Committee

Honorary Chair
Prof. Dr. Runtuung Sitepu, S.H., M.Hum (Rector – University of Sumatera Utara)

Steering Committee
Prof. Hiromi Homma (Toyohashi University of Technology, Japan)
Prof. Dr. Masashi Daimaruya (Muroran Institute of Technology, Japan)
Prof. Dr. Rahmat Budiarto (Al Baha University, Saudi Arabia)
Muhammad Ferdi Pasha, B.Comp.Sc., M.Sc., Ph.D (Monash University, Malaysia Campus)
Prof. Dr. Hideki Kawai (Muroran Institute of Technology, Japan)
Prof. Zainal A.Hasibuan, Ir., MLS, PhD (University of Indonesia, Indonesia)
Prof. Dr. Opim Salim Sitompul (University of Sumatera Utara, Indonesia)
Drs. Mahyuddin K. M. Nasution, M. IT, PhD (University of Sumatera Utara, Indonesia)
Prof. Dr. Herman Mawengkang (University of Sumatera Utara, Indonesia)
Prof. Dr. Saib Suwilo (University of Sumatera Utara, Indonesia)
Dr. Ir. Rinaldi Munir, M.T (Bandung Institute of Technology, Indonesia)
Prof. Dr. rer.nat. Dedi Rosadi, M.Sc. (University of Gadjah Mada, Indonesia)
Benny B. Nasution, Ph.D (Politeknik Negeri Medan, Indonesia)
Dr. Fitri Arnia, M.Eng.Sc (Syiah Kuala University, Indonesia)

Editorial Board
Dr. Eng. Himsar Ambarita (University of Sumatera Utara, Indonesia)
Drs. Mahyuddin K. M. Nasution, M. IT, PhD (University of Sumatera Utara, Indonesia)
Romi Fadilah Rahmat, B. Comp. Sc, M.Sc (University of Sumatera Utara, Indonesia)
Prof. Dr. Masashi Daimaruya (Muroran Institute of Technology, Japan)
Prof. Dr. Rahmat Budiarto (Al Baha University, Saudi Arabia)
Muhammad Ferdi Pasha, B.Comp.Sc., M.Sc., Ph.D (Monash University, Malaysia Campus)
Dani Gunawan, S.T, M.T (University of Sumatera Utara, Indonesia)
Cheng-Yuan Chang (Chung Yuan Christian University, Taiwan)
Chair
Dr. Eng. Himsar Ambarita (University of Sumatera Utara, Indonesia)

Co. Chair
Romi Fadilah Rahmat, B.Comp.Sc., M.Sc (University of Sumatera Utara, Indonesia)
Suherman, PhD (University of Sumatera Utara, Indonesia)

Executive Members
Dr. Erna Budhiarti Nababan, M. IT (University of Sumatera Utara, Indonesia)
Emerson Sinulingga, PhD (University of Sumatera Utara, Indonesia)
Dr. Fahmi F (University of Sumatera Utara, Indonesia)
Mohammad Fadly Syah Putra, M.Sc (University of Sumatera Utara, Indonesia)
Muhammad Anggia Muchtar, ST, MMIT (University of Sumatera Utara, Indonesia)
Dani Gunawan, S.T, M.T (University of Sumatera Utara, Indonesia)
Dr. Eng Taufiq Bin Nur, ST, M.Eng.Sc (University of Sumatera Utara, Indonesia)

Local Members
Prof. Dr. Muhammad Zarlis (University of Sumatera Utara, Indonesia)
Dr. Poltak Sihombing, M. Kom (University of Sumatera Utara, Indonesia)
Dr. Syahril Effendi, M. IT (University of Sumatera Utara, Indonesia)
Dr. Sawaluddin, M. IT (University of Sumatera Utara, Indonesia)
Dr. Elviawaty Muisa Zamzami, S.T, M.T, M.M (University of Sumatera Utara, Indonesia)
Dr. Maya Silvi Lydia, M. Sc (University of Sumatera Utara, Indonesia)
Amalia, S.T, M.T (University of Sumatera Utara, Indonesia)
Ainul Hizriadi, S.Kom, M.Sc (University of Sumatera Utara, Indonesia)
Indra Aulia, S.TI, MKom (University of Sumatera Utara, Indonesia)
Dedy Arisandi, S.T, M.Kom (University of Sumatera Utara, Indonesia)
Dian Rachmawati, S.Si, M.Kom (University of Sumatera Utara, Indonesia)
Baihaqi Siregar, S.Si, M.T (University of Sumatera Utara, Indonesia)
Ivan Jaya, S.Si, M.Kom (University of Sumatera Utara, Indonesia)
Jos Timanta Tarigan, S.Kom, M.Sc (University of Sumatera Utara, Indonesia)
Sarah Purnawati, S.T, M.Sc (University of Sumatera Utara, Indonesia)
Mohammad Andri Budiman, S.T, M.Comp.Sc, M.E.M (University of Sumatera Utara, Indonesia)
Ulfı Andayani, S.Kom, M.Kom (University of Sumatera Utara, Indonesia)
Sri Melvani Hardi, S.Kom, M.Kom (University of Sumatera Utara, Indonesia)
Amer Sharif, S.Si, M.Kom (University of Sumatera Utara, Indonesia)
Seniman, S.Kom, M.Kom (University of Sumatera Utara, Indonesia)
Venue Map

Hotel GranDhika Setiabudi Medan
Jalan Dr. Mansyur no. 169
Medan, Sumatera Utara
Indonesia
Phone : +62-61-42081234
Fax : +62-61-42081234
Floor Plan

![Floor Plan Diagram]
Mining of the social network extraction

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2017 J. Phys.: Conf. Ser. 801 012020
(http://iopscience.iop.org/1742-6596/801/1/012020)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 125.162.210.150
This content was downloaded on 24/03/2017 at 15:25

Please note that terms and conditions apply.

You may also be interested in:

Design of slow extraction system for therapy synchrotron
Zhang Jin-Quan, Song Ming-Tao and Wei Bao-Wen

Review of Road Extraction Methods from SAR Image
N Sun, J X Zhang, G M Huang et al.

Analytic signal extraction approach based on 2D Grating Interferometer and systematic comparison between 2D GI and 1D case
Z. Ju, Y. Wang, P. Li et al.

Electron Beam Assisted Extraction of Highly Charged Ions from an ECRIS
S Runkel, O Hohn, L Schmidt et al.

Positive ion extraction across the superfluid-vapor helium interface
S Purushothaman, K Peräjärvi, M Ranjan et al.

Fabrication of Atomic-Scale Structures on Si(111)-7×7 Using a Scanning Tunneling Microscope (STM)
Dehuan Huang, Hironaga Uchida and Masakazu Aono

Development of an algorithm to analyze cartographic features extraction methods in digital images
Guilherme Pina Cardim and Erivaldo Antônio da Silva

The Appearance of Radial Electric Field in Source Plasma Due to Extraction
Masao Naito and Hideo Akimune

Enhancement of Light Extraction Efficiency in Light-Emitting Diodes with Prism-Type Textured Top Surface
Myung-Whun Chang, Hyong Sik Won, Hyunjun Kim et al.
Mining of the social network extraction

M K M Nasution1*, M Hardi2, R Syah3
1,2,3Informatics Technology, Fasilkom-TI, Universitas Sumatera Utara, Padang Bulan USU Medan 20155, Indonesia.
*Email: mahyuddin@usu.ac.id

Abstract. The use of Web as social media is steadily gaining ground in the study of social actor behaviour. However, information in Web can be interpreted in accordance with the ability of the method such as superficial methods for extracting social networks. Each method however has features and drawbacks: it cannot reveal the behaviour of social actors, but it has the hidden information about them. Therefore, this paper aims to reveal such information in the social networks mining. Social behaviour could be expressed through a set of words extracted from the list of snippets.

1. Introduction
Based on the superficial method for extracting social network from information sources we have the social network as a resultant [1]. This method is pre-processing for transforming the raw data into the social network whereby Web as one of information sources [2]. Web be a picture of social change dynamically, and Web therefore represent the activities of social actors [3]. In other words, exploring Web as the social media is to allow us acquire the behaviour of either social actor personally or the community of social actors [4]. However, there is no extraction of social networks connected with the exploration of the behaviour of social actors. Thus, this paper aims to disclose formally the mining of the social network extraction: That is to develop the concept formally for extracting social network and propose some problems as basic concept and motivation; to propose an explanation of the problems and prove its importance in an approach systematically; and to conduct the extraction of social network and the related behaviour in an example.

2. Basic Concept and Motivation
Each network naturally we can model it as a graph \( G< V, E > \) whereby \( V \) is a set of vertices \( \{ v | i = 1, \ldots, I \} \) and a set of the name labels \( \{ a | i = 1, \ldots, I \} \), and \( E \) is a set of edges \( \{ e | j = 0, \ldots, J \} \), a set of weights \( \{ b | j = 0, \ldots, J \} \) and a set of labels \( \{ l | k = 0, \ldots, K \} \), or we can define the template of network as follows [5].

\[ \text{Definition 1. } T_N = G< V,E,L_1,L_2,B > \text{ is a template of network with conditions as follows} \]

\[ T_1 \ V = v | i = 1, \ldots, I \text{ as vertices in } T_N. \]
\[ T_2 \ E = e | j = 0, \ldots, J \text{ as edges in } T_N. \]
\[ T_3 \ L_1 = a | i = 1, \ldots, I \text{ as labels of vertices.} \]
\[ T_4 \ L_2 = l | k = 0, \ldots, K \text{ as labels of edges.} \]
\[ T_5 \ B = b | j = 0, \ldots \text{ as weights of edges.} \]

Each network not only has a set of vertices and a set of edges, but also it has a set of labels for vertices/edges and a set of weights for edges. Therefore, to develop a social network from information
sources, we choose the tools that are used to access information. Suppose for a social actor 'Mahyuddin K. M. Nasution' we use search engines, if query $q_1 = 'Mahyuddin K. M. Nasution'$ or $q_2 = "Mahyuddin K. M. Nasution"$ submitted to any search engine, then we get the hit counts like Table 1. So there is a difference between $q_1$ and $q_2$. The results of search engine have been contaminated by ambiguity in documents / web pages [6] or the semantic problems: meronymy, holonomy, hyponymy, synonymy, or polysemy [7]. The result of $q_1$ explained that each social actor socially has many names, aliases or attributes, while the result of $q_2$ still contains the possibility of same name for different people, although the exact pre-defined name for somebody. Therefore, to gauge out the appropriate information from the stack of documents, we always add an eligible keyword [8].

<table>
<thead>
<tr>
<th>Search engine</th>
<th>$q_1$</th>
<th>$q_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>16,100</td>
<td>1,110</td>
</tr>
<tr>
<td>Yahoo</td>
<td>1,980</td>
<td>244</td>
</tr>
<tr>
<td>Bing</td>
<td>213,000</td>
<td>2,090</td>
</tr>
<tr>
<td>Yandex</td>
<td>57,000</td>
<td>634</td>
</tr>
</tbody>
</table>

In the extraction of social networks, first we declare the social actors with names that are properly defined, to ensure every vertex $v$ in $V$ in template $TN$ represents a social actor: A set of actors $A = \{a_i|i=1,\ldots,l\}$. So there is a function $\gamma_i$ that maps one by one actor to vertex, or $\gamma_i(1:1): A \rightarrow V$ [9]. In this case, if a query $q$ contain a term for $a_i$ in $A$, then the query classify web pages into one class that is recognizable as a singleton event $\Omega_{w}$ whereby number of occurrences (hit counts) is $|\Omega_w|$ [2].

**Lemma 1.** If every web page representing the activity of a social actor, then the class of web pages represent the behavior of a social actor.

In addition, for $|\Omega_w| > 0$ the search engine also bring back (return) a list of snippets, every snippet contains $\pm$ 50 words, and the label of social actors can be generated from this list [5].

In the classical social networks, the relations between social actors was transformed directly from the data, but in the extraction of a social network, a collection of relationship that may be contained in the document and translated through a search engine in the form of co-occurrence [5]: If the query $q$ contains two terms $a_i,a_j$ in $A$, then the query classifying web pages into one class that can be recognized as a doubleton event $\Omega_{w} \cap \Omega_{w}$ whereby number of co-occurrences (hit counts) is $|\Omega_{w} \cap \Omega_{w}|$, with conditions $|\Omega_{w} \cap \Omega_{w}| \leq |\Omega_{w}|$ and $|\Omega_{w} \cap \Omega_{w}| \leq |\Omega_{w}|$ [2].

**Lemma 2.** If every web page representing the communication of social actor, then the class of web pages represents the behavior of the relation between social actors.

Let $|\Omega_{w} \cap \Omega_{w}| > 0$, the search engine returns a list of snippets that also contains information about 50 words in each snippet, and label of a relation can be generated from it through a procedure. All relations that may exist between social actors are accumulated into one and recognizable as weight of the strength relation by using the concept of similarity, for example Jaccard coefficient

$$sim_{ij} = \frac{|\Omega_{w} \cap \Omega_{w}|}{|\Omega_{w}| + |\Omega_{w}| - |\Omega_{w} \cap \Omega_{w}|}$$

**Definition 2.** $SN = G(V,E,A,R,\Gamma)$ is an extracted social network with an operation $\Gamma$ which satisfies the conditions

C1 $\gamma_i(1:1): A \rightarrow V$, $\gamma_i$ in $\Gamma$.
C2 $\gamma_i: R \rightarrow E$, $\gamma_i$ in $\Gamma$. 
Gamma function contains a set of processes to gain information from the sources: $\gamma_1$ for example is to declare the social actors based on MM method [8], and $\gamma_2$ is the steps that must be done for exploring the relations between the social actors like the superficial method [10].

**Definition 3.** $SN_c = G\langle V, E, A, R, I \rangle$ is an labeled social network based on extraction method $I$ which satisfies the conditions C1, C2 and/or

C3 $\gamma_2(M; 1): L_1 \rightarrow V, \gamma_5$ in $\Gamma$ and $L_1 = \{t_j | j = 0, \ldots, J\}$ as labels of vertices, or

C4 $\gamma_2: B \rightarrow E, \gamma_4$ in $\Gamma$ and $B = \{b_j | j = 1, \ldots, K\}$ as weights of edges, or

C5 $\gamma_2(M; 1): L_2 \rightarrow E, \gamma_5$ in $\Gamma$, and $L_2 = \{t_i | i = 0, \ldots, T\}$ as labels of edges.

**Theorem 1.** If the behavior of a cluster describes the behavior of a social actor, then the behavior of other actors expressed by the relationship between the clusters.

3. An Approach

In Web as social media, each web page can represent an actor in a way that is the name of actor is a part of text in web page and other words in the same web page as actions, ethics and attitudes of the actor [15]. In literal, the name of actor is a term $t_a$ consists of at least one or a set of words or $t_a = (w_0, \ldots, w_l)$, $l \leq k$, $k$ as a number of words $w$, $l$ as a number of vocabularies in $t_a$, $|t_a| = k$ is the size of $t_a$, and a cluster based on a term we can generate it as follows [12].

**Definition 4.** Let $S$\textsuperscript{-} Mega\$ as a space of the search engine for indexing the web pages. For a search term $t_a$, where $t_a$ in $\Sigma$, is a set of singleton search term of search engine. There are a dynamic space $\Omega_a$ containing the ordered pair of the term $t_a = 1, \ldots, I$ with web pages $\omega_{ij} = 1, \ldots, J$, i.e. $\omega = (t_a, \omega_{ij})$ such that $\Omega_a = \{(t_a, \omega_{ij})\}$ is a subset of $\Omega$ is a singleton search engine event of web pages (briefly we call it as singleton) that contains the occurrences of $t_a$ in $\omega$.

A set of web pages as the cluster is the representation of the actor formally (Definition 4), whereby $|\Omega|_a$ is the cardinality of $\Omega_a$ (hit count), while the expression of behavior we can mine from a list of snippets.

**Proposition 1.** If $|\Omega|_a > 0$ for a search term $t_a$, then there is a set of the weighted words $W = \{(w_i, b_i) | i = 1, \ldots, J\}$, $w_i$ is a word and $b_i$ is a weight of word.

Proposition 1 as the direct consequence of $\gamma_3$ in Definition 3 whereby $b_i$ of words are generated by using term frequency (tf) or probability of word $p(w)$ in collection of words from the list of snippets. Therefore, we can rewrite Lemma 1 formally as $|\Omega|_a > 0 \rightarrow \Omega_a \wedge \{(w_i, b_i)\}_a$ and Proposition 1 as the proof of Lemma 1.

**Definition 5.** [13] Let $\Omega$ as a space of the search engine for indexing the web pages. For two search terms $t_{ai}$ and $t_{aj}$, that is a set of doubleton search term of search engine. There are a dynamic space $\Omega_a \cap \Omega_{aj}$ containing the ordered pair of the terms $t_{ai} = 1, \ldots, I$ and $t_{aj} = 1, \ldots, J$ with web pages $\omega_{aij}$, i.e. $\{(t_{ai}, \omega_{aij}) \wedge \Omega_{aj} \cap \{(t_{ai}, \omega_{aij})\} \wedge \Omega_{aj}$ is a subset of $\Omega$ is a doubleton search engine event of web pages (briefly we call it as doubleton) that contains the co-occurrences of $t_{ai}, t_{aj}$ in $\omega$.

**Proposition 2.** If $|\Omega_a \cap \Omega_{aj}| > 0$ for two search terms $t_{ai}$ and $t_{aj}$, then there is a set of the weighted words $W = \{(w_i, b_i) | k = 1, \ldots, K\}$, $w_k$ is a word and $b_i$ is a weight of word.

Similar to $\gamma_3$, Proposition 2 as a consequence of $\gamma_3$ in Definition 3 and it prove Lemma 2 if $|\Omega_a \cap \Omega_{aj}| \leq |\Omega_a|$ and $|\Omega_{aj} \cap \Omega_{aj}| \leq |\Omega_{aj}|$ whereby $|\Omega_a \cap \Omega_{aj}|$ is the cardinality of $\Omega_a \cap \Omega_{aj}$, while $\{(w_i, b_i) | k = 1, \ldots, K\}$ as the expression of the relation behavior between actors [14].

$|\Omega_a \cap \Omega_{aj}| > 0$ proves the existence of relationship between two social actors, whilst using Eq. (1) gives the strength relation as a weight. A weight of the relation $\gamma_4$ is a behavior of same actors based on the relationship between the clusters, whereby $\gamma_4(\text{sim}) \times b_i$ be expression of behavior. It also can proved by the similarity of $W_{ai} = \{(w_i, b_i) | k = 1, \ldots, K\}$ and $W_{aj} = \{(w_i, b_i) | k = 1, \ldots, K\}$, $W_{ai} \cap W_{aj} \neq \phi$. 


Theorem 1 is proved. Therefore, for exploring the behavior of the social actors and the relation between them as the social network mining [15, 16] we have algorithm as follows.

Algorithm SN_i:

Declare a query for each social actor, \( q_a \) for \( a \) in \( \mathcal{A} \),
\( \mathcal{A} \) is a set that consist of \( n \) social actors.
Submit a query to the search engine, \( |\Omega_a| \) \( = q_a \) in \( \mathcal{A} \).
Collect the snippets for \( t_{ai} \), \( \mathcal{L}_{ai} \), \( \mathcal{L}_{ai} \) is list of snippets for \( a_i \) in \( \mathcal{A} \).
Give a weight to every word in \( \mathcal{L}_{ai} \), \( \mathcal{W}_{ai} \), \( \mathcal{W}_{ai} \) is a set of words for \( a_i \) in \( \mathcal{A} \).
Submit a query to the search engine, \( |\Omega_j| \) \( = q_j \) in \( \mathcal{A} \).
Collect the snippets for \( t_{aj} \), \( \mathcal{L}_{aj} \), \( \mathcal{L}_{aj} \) is list of snippets for \( a_j \) in \( \mathcal{A} \).
Give a weight to every word in \( \mathcal{L}_{aj} \), \( \mathcal{W}_{aj} \), \( \mathcal{W}_{aj} \) is a set of words for \( a_j \) in \( \mathcal{A} \).
Collect the snippets for \( t_{ai} \) and \( t_{aj} \), \( \mathcal{L}_{ai,aj} \), \( \mathcal{L}_{ai,aj} \) is list of snippets for \( a_i \) and \( a_j \) in \( \mathcal{A} \).
Give a weight to every word in \( \mathcal{L}_{ai,aj} \), \( \mathcal{W}_{ai,aj} \), \( \mathcal{W}_{ai,aj} \) is a set of words for \( a_i \) and \( a_j \) in \( \mathcal{A} \).
Compute the strength relations by using \( |\mathcal{L}_{ai}|, |\mathcal{L}_{aj}|, \) and \( |\mathcal{L}_{ai,aj}| \).
Compute the similarity between \( \mathcal{W}_{ai} \) and \( \mathcal{W}_{aj} \).
Compute the average for \( \mathcal{W}_{ai,aj} \).

4. Extracting social network: A discussion

For the Google search engine, whenever a query containing the search term \( t_a \) submitted, we will get hit count \( |\mathcal{L}_a| \) and a list of snippets. Each snippet consists of the URL address of the web page, the title of the web page, and the abstract of web page. For example, for two social actors obtained a number of snippets, number of words from snippets, and the hit counts like Table 2.

<table>
<thead>
<tr>
<th>Search term</th>
<th>“Abdullah Mohd Zin”</th>
<th>“T Mohd T Sembok”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of snippets</td>
<td>204</td>
<td>155</td>
</tr>
<tr>
<td>Number of words</td>
<td>5,982</td>
<td>4,680</td>
</tr>
<tr>
<td>Hit count</td>
<td>5,850</td>
<td>2,740</td>
</tr>
<tr>
<td>List of words</td>
<td>( w_{ai} )</td>
<td>( p(w_{ai}) )</td>
</tr>
<tr>
<td>malaysia</td>
<td>0.502</td>
<td>malaysia</td>
</tr>
<tr>
<td>journal</td>
<td>0.4099</td>
<td>prof</td>
</tr>
<tr>
<td>ismail</td>
<td>0.2649</td>
<td>university</td>
</tr>
<tr>
<td>nazri</td>
<td>0.2177</td>
<td>ahmad</td>
</tr>
<tr>
<td>university</td>
<td>0.1738</td>
<td>abdullah</td>
</tr>
<tr>
<td>computer</td>
<td>0.1043</td>
<td>halimah</td>
</tr>
<tr>
<td>science</td>
<td>0.0921</td>
<td>journal</td>
</tr>
<tr>
<td>prof</td>
<td>0.0733</td>
<td>badioze</td>
</tr>
<tr>
<td>network</td>
<td>0.0652</td>
<td>zaman</td>
</tr>
<tr>
<td>university</td>
<td>0.0405</td>
<td>ukm</td>
</tr>
<tr>
<td>information</td>
<td>0.0276</td>
<td>abu</td>
</tr>
</tbody>
</table>

Naturally a social network is built with the social actors. The number of social actors who may be extracted from the Web cannot be predicted, but for resolving this constraint always the extraction of social network based on the community of social actors [4]. In the superficial method, the social network for two social actors formed from two singleton and one doubleton, or formed from three clusters of Web. The clusters have differences and similarities. Each difference causes their behavior
differs, for example, if the word "computer" does not exist in the list of words for "T Mohd T Sembok", then it became an expression of behavior differentiator for clusters based on term search "Abdullah Mohd Zin" [5], and this cluster has the weight as follows

\[ |Q_{\text{Abdullah Mohd Zin}}| \cdot p(w) = 5.850 \cdot 0.1043 = 610 \]

or vice versa

\[ |Q_{\text{T Mohd T Sembok}}| \cdot p(w) = 2.740 \cdot 0.1005 = 275 \]

Thus, the behavior of any social actor be formed by the expression. It defined by a number of words with its weight that affects the cluster \(|Q_c|\) [9].

![Figure 1. The strength relation between AMZ ("Abdullah Mohd Zin") and TMTS ("T Mohd T Sembok") in the extracted social network](image)

The computing toward three clusters using Eq. (1) yield a strength relation between "Abdullah Mohd Zin" and "T Mohd T Sembok", that is 0.1561 where \(|Q_c \cap Q_d| = 1160\), whereas the relationship behavior affects the behavior of social actor personally or the social actors have the relationship with others. This relation behavior are defined by an expression through a number of words in the same cluster. It is also indicated by the similarity of expression between social actors although from the different clusters, see Figure 1. By using Eq. (1) for two set of words with their weights we obtain the strength relation between "Abdullah Mohd Zin" and "T Mohd T Sembok" based on similarity of words, i.e. 0.0785, while the average of word weights based on list of snippets (of \(Q_c \cap Q_d\)) is 0.1432. Therefore, the social network that is extracted through two singletons and a doubleton, can be mined by involving the snippets to give meaning to the strength relation that have been generated.
5. Conclusion
In particular, the behavior of social actors and the relationships between them based on the Web depends on the expression and the behavior of clusters. The superficial method can be developed to express these behaviors and to mine it by following the extraction of social networks. Further study is to reveal the properties associated with the data, information, and analysis of social networks.

References
[16] S Vaithyanathan 1999 Introduction: Data mining on the Internet Artificial Intelligence Review