

COVER SHEET

2010 Fall Semester

Instructor: CLASTER William

Class: Mathematics for Information Technology EA

Report: NodeXL Assignment

APM 1st year student

12410179

HOANG Nguyen Phong

Due Date: January 25, 2011 16:00

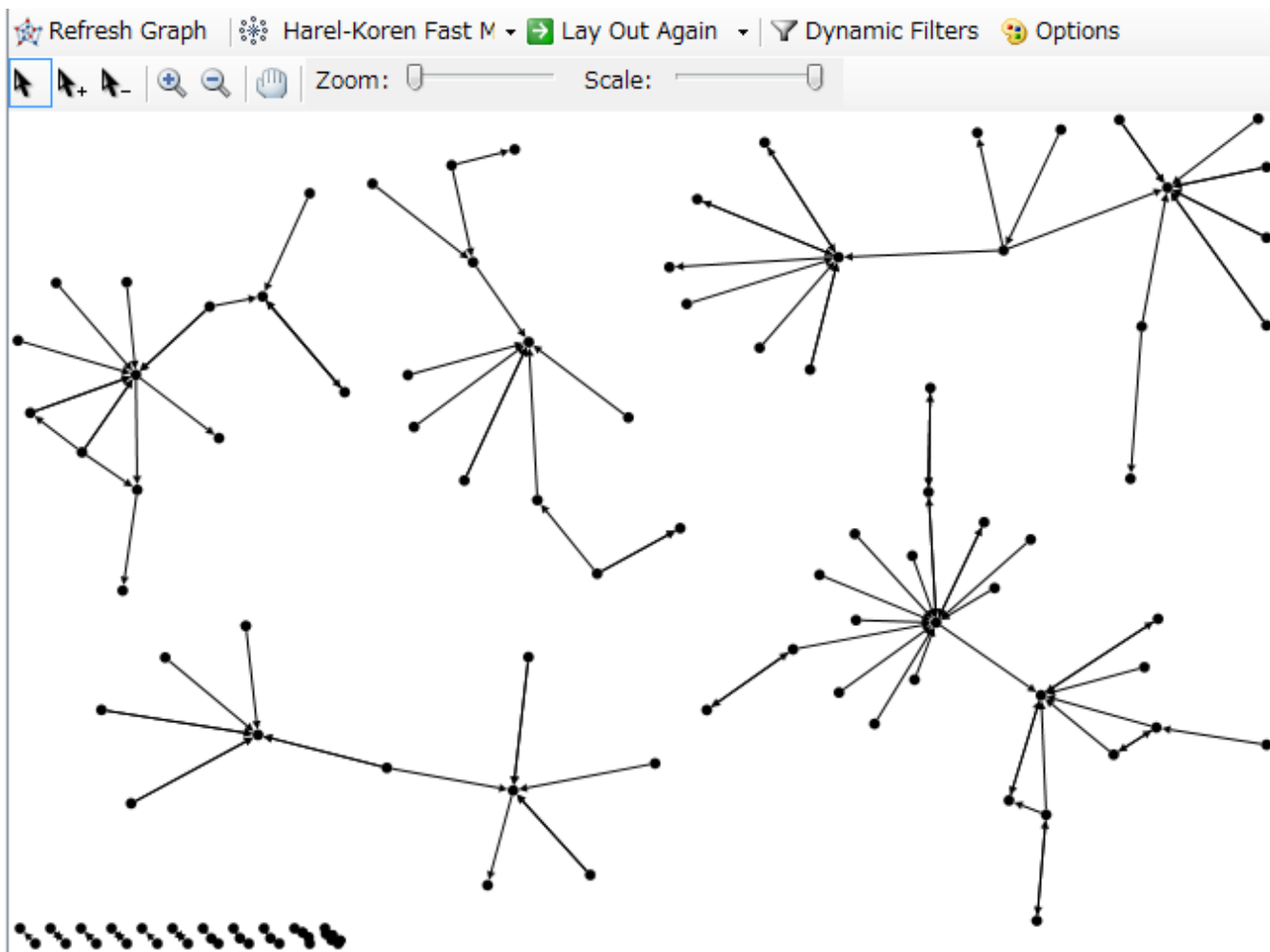
I. Getting and Processing data from Twitter:

Firstly, by “*import tool*” in NodeXL plug-in for Microsoft Excel, I searched for people whose tweets contains the key word “*Julian Assange*”, and ticked to all of three boxes “*Follows*”, “*Replies-to*”, and “*Mentions relation*” with the 100 people limitation so that I could get a stronger-connected graph. However, in some first searches, I could not get a reasonable graph as I expected due to 2 reasons. The first one is the limitation number was small, and I chose the box “*I don't have a twitter account*” was another reason. And the graph was really weak-connected because there were only 4 or 5 edges and about over 90 single points. So I was not able to apply the “*graph metrics*” contents to consider these graphs.

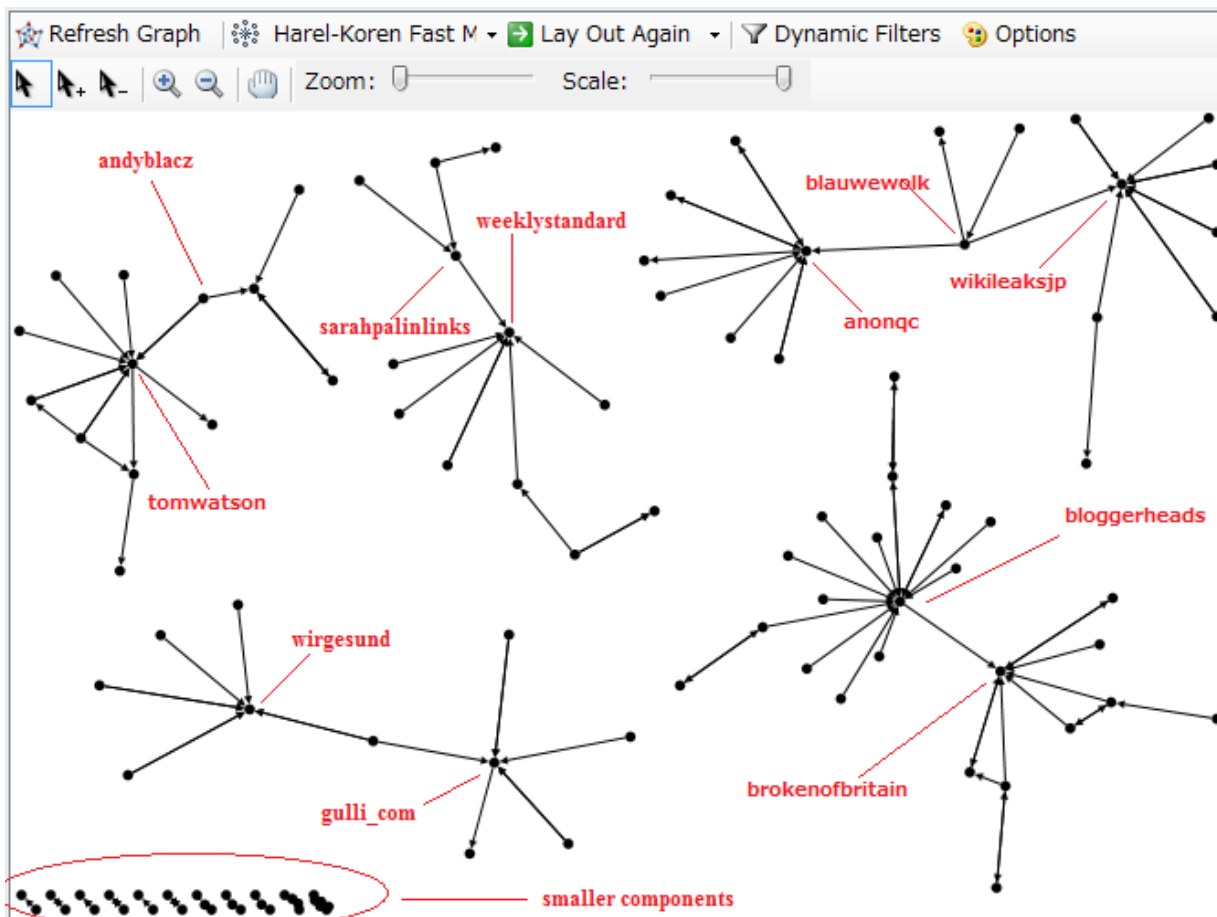
Then I registered to use twitter account to import more than 100 people data. This time I also used the key word “*Julian Assange*” but with the limitation up to 300 people. After that, I got 300 people whose tweets contain “*Julian Assange*”, and the graph was stronger-connected. Then I calculated all numbers by “*Graph Metrics*” tool, and removed some points which have zero degree in both of “*in-degree*” and “*out-degree*” in order that we can examine and deduce information more easily from the graph.

II. Examining the graph:

As for the Graph, under “Harel-Koren Fast Multiscale” algorithm and in “Layout Options” task, I chose the box “Move the graph’s smaller components into boxes at the bottom of the graph” with “Maximum size of the components to move” is 6, I got the below graph which has all the significant components appear clearly, and all the smaller components, which have less than 6 vertices, are moved to the bottom of the graph.



Then, according to the “*Betweenness Centrality*” numbers which is obtained from “*Graph Metrics*” tool, I labeled the “important” vertices in the Graph as below:



And basing on the above labeled graph and the results of the examination by “*Graph metrics*” tool in NodeXL, we can deduce many things:

<p><i>The important person or well-known organization</i></p>	<p><i>The relevant things to the key word “Julian Assange” and some interesting information on their pages.</i></p>
<p>1. The twitter user with nick name “<i>bloggerheads</i>” is the most important person in this set of people, because his “<i>Betweenness Centrality</i>” point is</p>	<p>Going to his page, I found that his field of expertise is online communities and their impact on search engine results, but he is also interested on the media and politics.</p>

<p>430, which is the largest in the graph. And he also has the largest “<i>In-Degree</i>” is 12.</p>	<p>Therefore, he has his own technic ways to obtain much information about communications, media and also politics which lead to the relation with the key word “<i>Julian Assange</i>” who revealed on media field many diplomatic secrets of the United State of American last year. This is the reason why many twitter users follow him in oder to up-date the lastest information.</p>
<p>2. The second important vertex is not a person; it is an account of an organization which has the “<i>Betweenness Centrality</i>” point is 284 and “<i>In-Degree</i>” is 7.</p>	<p>Surfing their page, it is found that they are disable people in UK, who want to reveal some hidden things which is hidden by government or an invisible force. Their page is for disable people and by disable people, so it tweets a lot about “<i>Julian Assange</i>” and his wikileaks’ hidden secrets.</p>
<p>In graph metrics aspect, both of the above vertices are really important. Because they are not only the centers of their groups, but also is a bridge (<i>each of them is a cut point</i>) between two big groups. Hence, if we replace the edge (<i>the connection</i>) between those, two groups will be disconnected, or if we remove both of them, all the points in these two group maybe become the single points.</p>	
<p>3. The next vertex which has a third largest “<i>Betweenness Centrality</i>” point is “<i>wikileaksjp</i>” with the point is 180 and the “<i>In-Degree</i>” is 7.</p>	<p>Basing on the name of this vertex, it is clear that it must have a close relation to the key word “<i>Julian Assange</i>” (the founder of wikileaks website) and it is one of many mirror versions of Wikileaks, and on its twitter page I found that they tweet</p>

	<p>many things about wikileaks in both Japanese and English.</p>
<p>4. The fourth one is “<i>blauwewolk</i>” vertex which has 174 point of “<i>Betweenness Centrality</i>” 180, but only 1 “<i>In-Degree</i>”.</p>	<p>This is an account of a female who seems to like freedom in information and support for Wikileaks. Therefore, there are many “<i>Julian Assange</i>” words mentioned on her page.</p>
<p>5. The fifth and sixth vertices which have the same “<i>In-degree</i>” 6 and 162 and 110 point of “<i>Betweenness Centrality</i>” respectively.</p> <p>They are “<i>tomwatson</i>”(a well-known American author of CauseWired) and another one is “<i>anonqc</i>” (an anonymous organization)</p>	<p>Because Tom Watson is an author, blogger, and also a journalist, so on his page, every he tweets so much about America news such as President Obama, and also many other information which is relevant to America, including wikileaks and “<i>Julian Assange</i>”.</p> <p>Besides, on the page of “<i>anonqc</i>” which is an anonymous organization, and against the “<i>internet censorship</i>” and ask for freedom in information, and so, they tweet a lot about “<i>Julian Assange</i>” and Wikileaks as well.</p>
<p>In Graph Metrics view of point, all vertices “<i>wikileaksjp</i>”, “<i>blauwewolk</i>”, “<i>anonqc</i>”, and “<i>tomwatson</i>” are quiet important components in the graph. However, “<i>blauwewolk</i>” is more important, while its “<i>In-Degree</i>” is only 1, and the others’ are 7 or 6. Why?</p> <p>We are now take it into account:</p> <p>At the first glance, “<i>blauwewolk</i>” seems to be not important because it is not a center vertex of a big group, and there is only one person follow it. However, if we examine its “<i>Betweenness Centrality</i>”, it really plays a vital role in the connection of two big</p>	

groups which “*wikileaksjp*” and “*anonqc*” are centers. Thus, in total, “*blauwewolk*” is such an important vertex in this graph.

III. Conclusion:

By applying the “*Graph Metrics*” method, I can found out easily the necessary sources of information which I need. And these results will be more accurate if I do the research with the limitation up to 1000 people and examine more values which are obtained by “*Graph Metrics*” calculating tool. In Graph Metrics aspect, I found that the importance of a vertex depends on not only one number or a rate but also on many other factors such as “*In-Degree*”, “*Betweenness Centrality*” and so on...Therefore, in order to obtain a believable result, we have to spend more time on examining all the other numbers, and rates from “*Graph Metrics*” tool.