

Exploring “0day” mentioning by twitter user using NodeXL

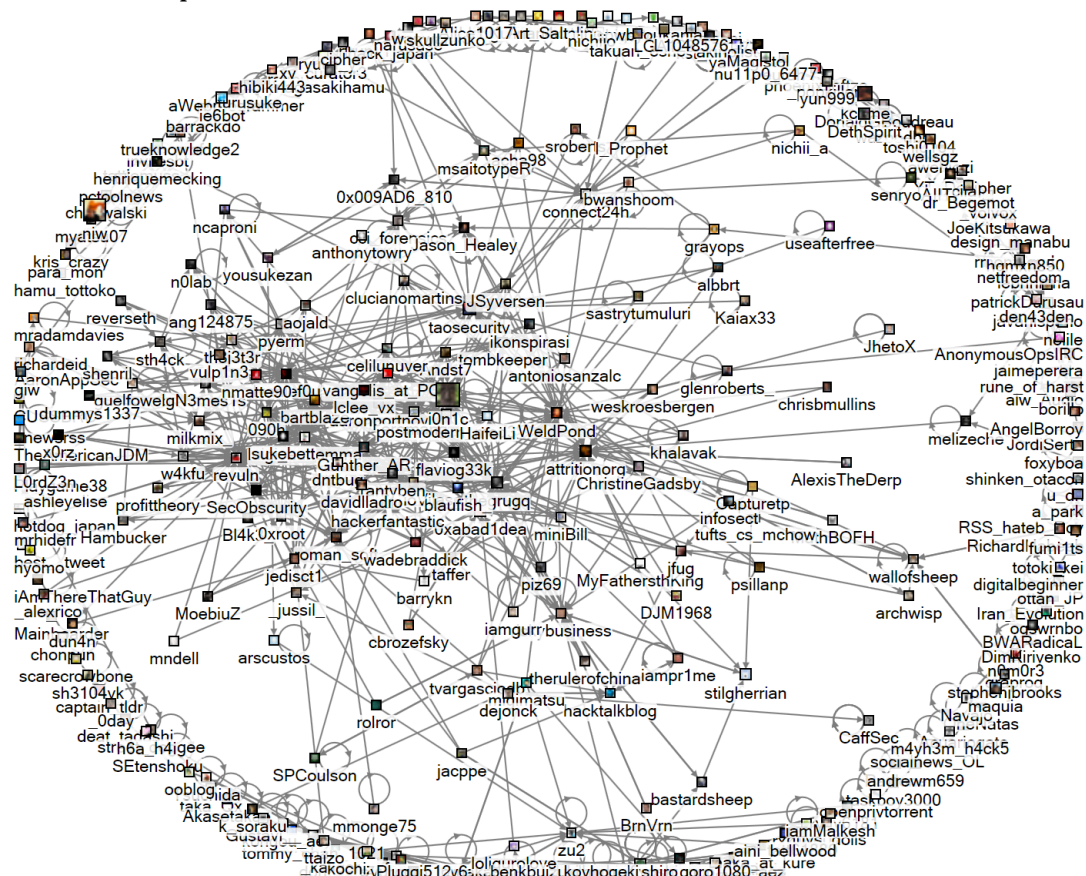
Introduction

This assignment asks us to use NodeXL to create a network graph and explore it. NodeXL is a free, open-source template for Microsoft Excel. After installing the NodeXL plugin and getting familiar with it, I decided to create a twitter search network. I tried some key words such as “Galaxy S4” but it turned out that the vertex didn’t have too much connection with each other. Then I came up with the idea that I can choose “0day” as the search key word. As far as I know, hackers and security specialists often mention “0day” on twitter. They are small groups of people and they also have strong relationships. So I decide to explore “0day” mentioning by twitter user.

Three interesting characteristics of the network

1. Relationships between twitter users who mentioned “0day”

In order to get a graph that can show the relationships between users who mentioned “0day” clearly, I choose the Fruchterman-Reingold algorithm to lay out the graph. The vertexes of the graph are twitter users who mentioned “0day” and the edges are connections (follower, tweet, mentioned) between them. As the graph shows, the users on the circle are those who do not have too many relationships with others. And the users in the circle have more relationships with others. From the density of the vertexes we can figure out the group of user who interact with others most frequently. The interesting thing is that this group of people contains many the most famous hackers in the world, such as @i0n1c and @tomkeeper.

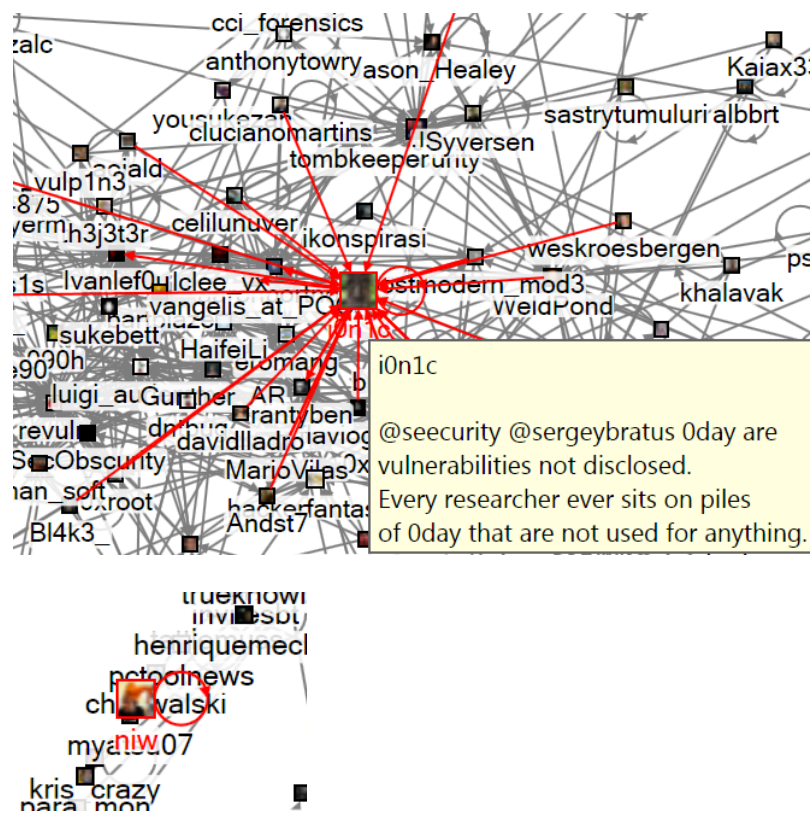


2. Popular twitter user who mentioned “0day”

In order to represent the popular users in the graph, I set the vertex size as the follower in the autofills columns settings. So the more follows the user has, the bigger vertex is. In the dataset figure we can find out the relative followers number, which is showed below.

	Vertex	Color	Shape	Size
2				
3	i0n1c			10.0
4	niw			8.5
5	kcome			3.6
6	best_tweet			3.1
7	taosecurity			2.8
8	WeldPond			2.4
9	stilgherrian			2.2
10	attritionorg			2.2
11	hackerfantastic			2.1
12	maquia			2.1

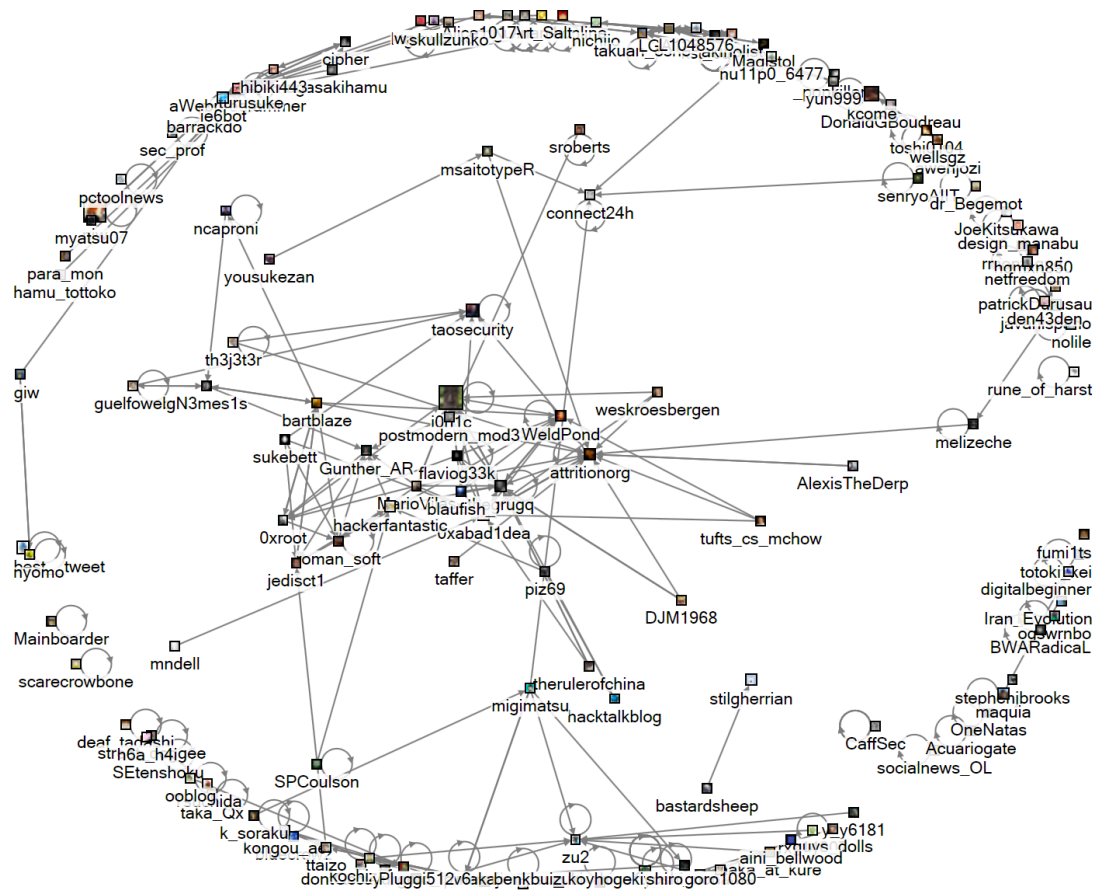
And from the graph we can see clearly that the number of follower of a user has nothing to do with the times he mentions about “0day”



3. Relationship between total tweet number and “0day” tweet number

As this paper is to explore the “0day” mentioning by twitter user, I want to explore relationship between total tweet number and “0day” tweet number. In

the dynamic filter settings I set the number of tweets ranges from 5000, so people who has less than 5000 tweets will be removed from the graph. The interesting thing is that, compared with the first graph, the number of people on the circle and in the circle are just changes equally. It indicates that some people who has few tweets on “0day” might be a active twitter user because the total number of their tweets is big. Also some people may not have too many tweets, but most of their tweets focus on “0day”.



Critique of the system

There are two shortcomings of this system. One is the time limit. As the NodeXL can only import data of recent few days, I can't analysis the long-term data. In my exploration of "0day" mentioning, there might be some error that some people just happened to mention "0day" a lot in the recent days.

Another shortcoming is the API limit of twitter. As twitter limit the API to prevent software such as NodeXL from getting to much data, I can only set the user no more than 300. Otherwise the twitter will force a pause and I have to wait 1 hour to continue getting data. Due to this limit I cannot do analysis on a large amount of people.