ISSN: 2393-8528

Contents lists available at www.ijicse.in

International Journal of Innovative Computer Science & Engineering

Volume 4 Issue 4; July-August-2017; Page No. 01-05

# Pokémon Go and Socialization: An Analysis of Twitter Network

Xin Song[1], Shicheng Zhao[2], Xilin Jiang[3], Yiheng Sun[4], Chengda Lyu[5]

[1]Hunan University (China), [2]The Windsor School (USA),

[3]Dalian No. 24 H.S. (China), [4]Dalian No. 8 H.S. (China), [5]JFK C.H.S. (USA)

## ABSTRACT

Given the popularity of Pokemon Go worldwide, a social network analysis method was used to evaluate connections among users. The Twitter historical database was used to search for suitable tweets, and the search term of "Pokémon Go," and "friends" were used. We used software tool NodeXL, which uses the Twitter API connection to search Twitter database. There were 975 tweets, 790 mentions and 185 replies. There were 1108 closed subgraphs. There were 785 single users which are not connected with anyone. Maximum size of a subgraph was 69 users, and the maximum number of connection in a subgraph was 98. The maximum number of connections to get from one user to another was 6, with the average number of 1.8 connections. It was hard to recognize a clear network structure, but few central users were identified. Most clusters were broadcast networks, and there were many famous individuals or organizations with large numbers of followers that were not closely connected. The network grew over time, and the growth momentum was closely dependent on influential users.

**Keywords:** Social Network Analysis, Pokemon Go, Twitter, NodeXL

## Introduction

Decline in the game console market has energized mobile games, a vast market (Penttinen, Rossi & Virpi, 2010). Then came Pokemon Go, a huge social phenomenon (Dorward, Mittermeier, Sandbrook & Spooner, 2016). This new way of playing has certain consequences, and notable features include data gathered from mobile devices. In addition, the method of collaboration is different: in person and outdoor. From the security point of view, the most vulnerable population is children (Sharma & Vassiliou, 2016). Players may not know intentions of other players, and the mere fact that games are played outdoors can make it a risky activity. On the other hand, Pokemon Go has an appealing profit model, and the collaboration can lead to new friends.

Given the rise of Pokemon Go, this study examines the developing social network through the social network analysis (Lin, Cao, Liu, Papadimitriou, Sun & Yan, 2009). For this, Twitter and Tweets are used.

The social network analysis method is an important tool to recognize and visualize different types of research subjects. In the last few years it has spread to different field of research, including sociology, biology, information technology, and economics. Some of the best examples of social network analysis include citation networks, egocentric networks, and twitter networks (Wang & Li, 2013).

This study limits the focus to a selected number of users so that the graph will be visually understandable and to identify underlying groups of users connected based on their tweets (mentions and retweets). The study identifies central users and dynamic element of network growth. The tools include NodeXL for social network analysis and visualization and Excel for data preparation (Hansen, Shneiderman & Smith, July 2009).

## Methods

Twitter historical database was used to search for suitable tweets. Search term of "Pokémon Go," and "friends" was consequently used as the most common simplification of a term connected to the meaning of socializing.

We used software tool NodeXL, which uses the Twitter API connection to search Twitter database. Twitter has set so called "rate limitation" which actually means the search and export has limitation in quantity of tweets and users and the time frame. Twitter API allows to search only thru last 7 days of tweets history.

*Corresponding author: Xin Song

NodeXL can go through search results and automatically enable network analysis and graph visualization. It does differentiate between three different type of relationship: tweet, mention and retweet. Logic behind twitter is sharing, so we can treat list of tweets as indicator of user activity but only the frequency of mentions and replays give us the measurement of influence of a user. It is important what kind of reach user tweets have. List of followers are important to understand how long is the reach of a user voice. This so called central or influential positions of users will be clearly identified in graphs in results section.

Network graph is formed based on lines connecting different users and representing the mentions or replays. There is also another type of network which can be drawn. Compared to graph where source of connection is actual tweets itself we can draw graphs based on users who are followers and following. In this way, we can identify user network with most important, central users and underlying clusters of similar users. Where measure of similarity is shared list of followers or following. Most commonly used in branding or marketing actions. In this way with little effort and selection of right user one can easily and quickly spread the news around. These central users are known as influencers and they have a role of being a hubs and bridges in the network.

As in any conversation and especially in social network analysis we can always identify different type of conversation. Pew Research center in association with Social Media Research Foundation (Smith, Rainie, Himelboim & Shneiderman, 2014) has identified 6 different type of twitter conversation: divided, unified, fragmented, clustered, in-hub & spoke, and out-hub & spoke.

**NodeXL Procedure**

NodeXL has the capability to create, analyze and visualize different type of networks.

We use built-in Twitter API service which connect us directly to a Twitter database (Soni & Mathai, 2015). In the search box we entered the search terms: Pokémon and friends.

Output does create two type of tables. Vertices and edges, first being the list of all users found in tweets and the edges being the connections between this user. Connections are actual content of the tweets.

These two tables are our data sources to create a network graph. Vertices sometimes called nodes are users (dots) and edges are lines (mentions and replays) connecting different users. After the graph drawing the actual social network analysis begin. With different types of graph layout algorithms, we can search the best visual representation of our network.

We apply different visualization techniques based on metrics to create graph. We can enlarge dots based on number of connections (in or out) each user has or number of followers. We color nodes to identify clusters or to add different user attributes. We enlarge edges (lines) to emphases the importance of connections for this we can use different type of weights which represents the power of connection between the users.

**Results**

The sample has list of tweets from 5.9.20016 until 12.9.2016. There is 1937 users (vertices) and 1984 connections (edges) and 1768 unique connections. We counted 975 tweets, 790 mentions and 185 replies. Graph has 1108 closed subgraphs, these are so called connected components and in our case these are users connected with each other with two type of connections (mentions and replies). It also has 785 single users which are not connected with anyone. Maximum size of a subgraph is 69 users and maximum number of connection in a subgraph is 98. The maximum number of connections to get from one user to another is 6 with the average number of 1.8 connection. Our sample graph is a very sparse graph. Maximum density of a graph is 1 and minimum is 0, our is 0.00025 which lead us to the conclusion that sample of users sharing same search term are not closely connected.
In this basic graph visualization, it is hard to recognized network structure but few central users could be identified. In another instance, we did adjust layout and there are clusters visible and large number of users as loops.

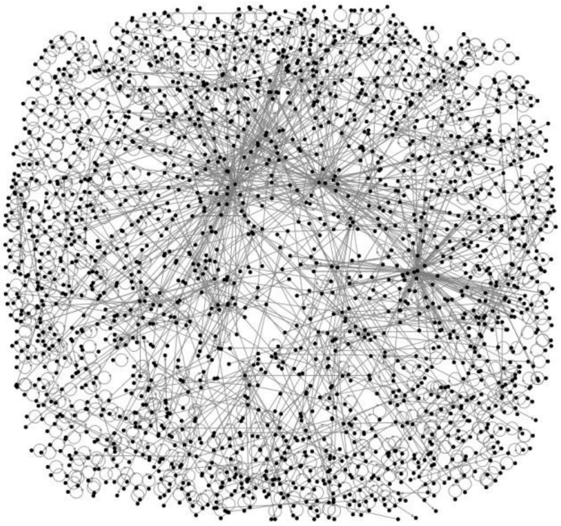
**Visualization**

For better understanding we created a clustered layout so the structure of our network is easier to read (Fig. 1). Below is the graph with applied clusters and different layout where we can identify large number of isolated subgraphs. Loops or as we saw single tweets have no visual value for us. We will filter them out and focus on groups with more connections.

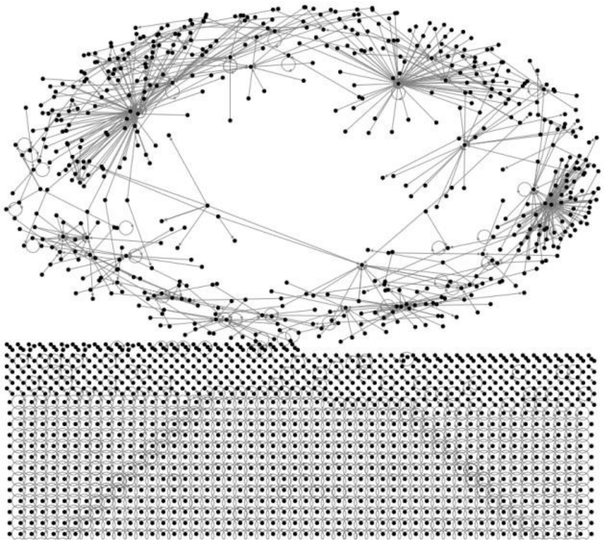
After applying filter on relationship type and use only mentions and replays we are getting more clear visual representation of our network based on selected search term.

Most of our cluster is broadcast networks which means they are star shaped and in central position we often have famous individuals or organizations with large number of followers which are not closely connected.

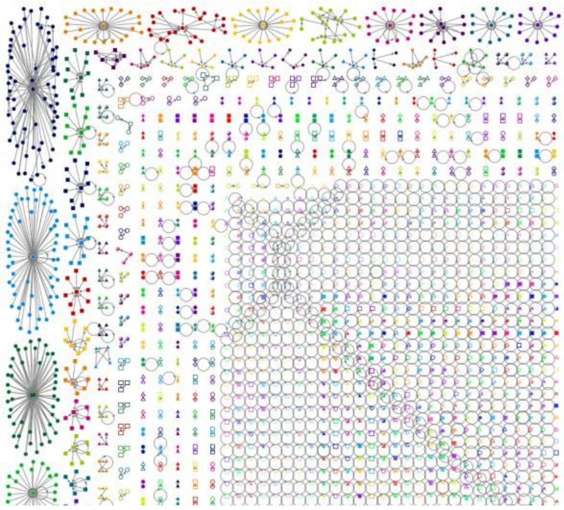
Draft layout



Three form layout



Clustered layout



Clustered layout without the loops

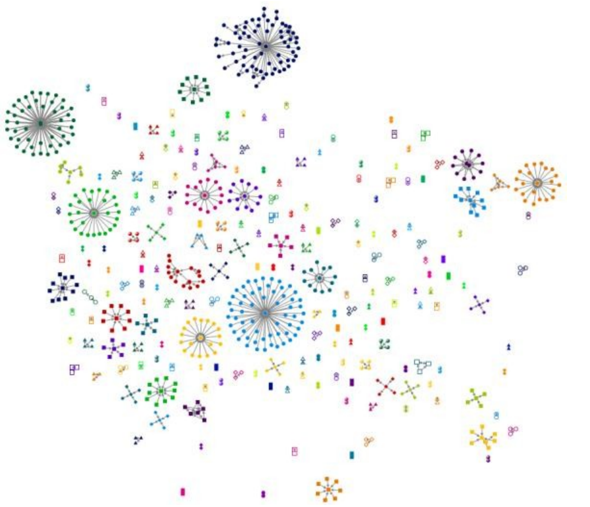


**Fig. 1: Network Visualization**

**Network Growth over Time**

As we said dynamic component of network growth is highly dependent from the time of the news/topic/event is entered in a tweeter space. Network is growing in time but the momentum of growth is highly dependent from the influential users (Fig. 2).

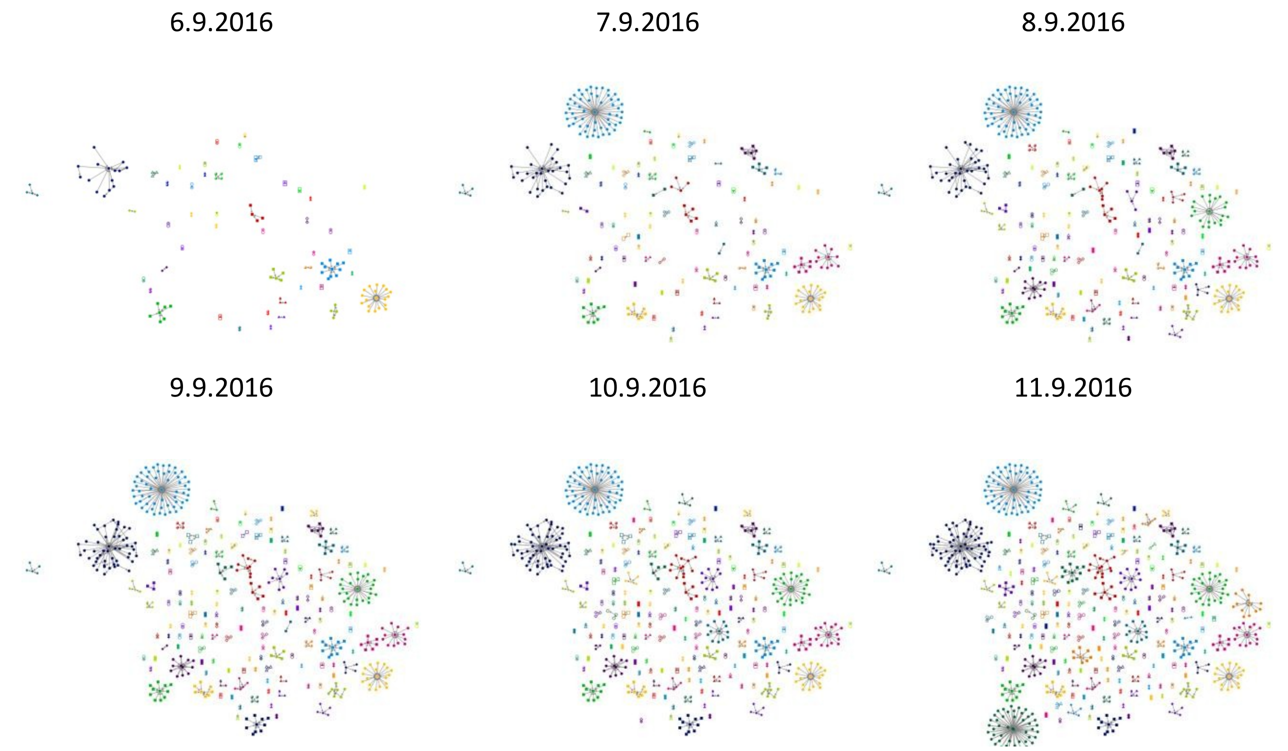



**Fig. 1: Network Growth over Time**

**Cluster Example: @YOUTUBE**

An example of single cluster would allow us to show different attributes of network. Youtube channel is used as the popular broadcasting source and in our case powerful medium enabling user to post and comment video content.

We used same procedure as with the main network in search box we just add @youtube with Pokemon and friends (Fig. 3).

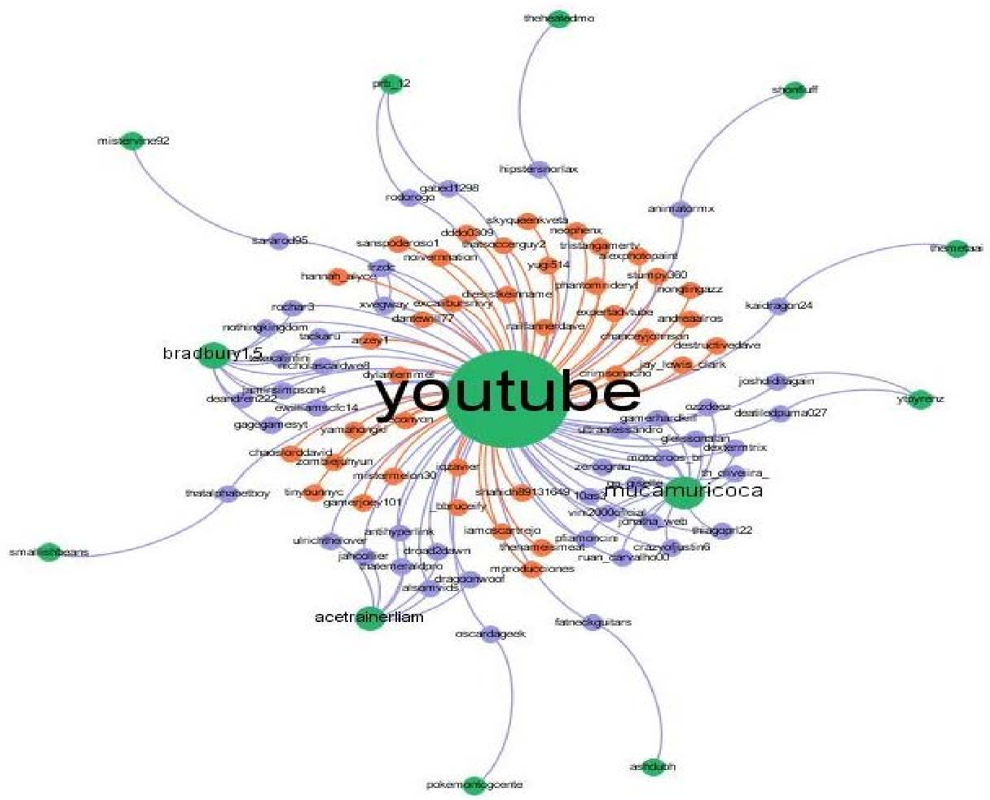



**Fig. 3: YouTube**

In graph above central positon of channel YouTube is evident. We used size of the nodes as number of mentions each user had. Green color users have a role of source or as we stated above broadcasters. Purple color users are mostly followers but compared with orange users they are connected with two different sources/broadcasters so they have more than one source of information. Orange users are connected just with one central user youtube. It is interesting to see that green users in outside rim has no direct connection with main youtube user. They represent the hubs or bridges where their followers have connection with youtube user. In this way network of youtube user can have very large reach.

**Discussion**

At this stage we were able to draw our graph and recognized it as highly clustered network. Pokémon is at this point a very trendy topic on the other hand it might be too soon to analyze sociological issues affected by this new phenomenon.

Never the less we can identify the fact that it does affect how players are functioning and there are some implications of the fact that new friendship are formed and also old friendships are broken. Causal relationship between these subjects needs more thorough research research which goes beyond this graph analysis.

We did identify different clusters and key influential users. This gives us the insight for future research how to identify suitable sample and with content analysis deeper insight of Tweets. Simple sentiment analysis showed us that content of tweets is positively oriented when consider friends and playing Pokemon.

**References**


1. Dorward, L., Mittermeier, J., Sandbrook, C., & Spooner, F., "Pokemon Go: Benefits, Costs, and Lessons for the Conservation Movement," Conservation Letters, November 2016
2. Hansen, D., Shneiderman, B., & Smith, M., "Analyzing Social Media Networks: Learning by Doing with NodeXL," NodeXL – Network Overview, Discovery and Exploration for Excel," July 2009
3. Lin, C., Cao, N., Liu, S.X., Papadimitriou, S., Sun, J., & Yan, X., "SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence," IEEE International Conference on Data Engineering, June 2009
4. Penttinen, E., Rossi, M., & Virpi, K. T., "Mobile Games: Analyzing the Needs and Values of the Consumers," Journal of Information Technology Theory and Application, Volume 11, Issue 1, March 2010
5. Sharma, P., & Vassiliou, V., "Pokémon Go: Cardiovascular benefit or injury risk?" Oxford Medical Case Reports, Editorial Comment, October 2016
6. Smith, M., Rainie, L., Himelboim, I., & Shneiderman, B., "Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters," Pew Research Center, February 2014
7. Soni, R., & Mathai, K., "Improved Twitter Sentiment Prediction through Cluster-then-Predict Model," International Journal of Computer Science and Network, Volume 4, Issue 4, August 2015
8. Wang, H., & Li, W., "Online Egocentric Models for Citation Networks," Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, July 2013