# Exploring Tag-based Like Networks

**Kyungsik Han**
College of Information Sciences and Technology
The Pennsylvania State University
kuh178@psu.edu

**Jin Yea Jang**
College of Information Sciences and Technology
The Pennsylvania State University
jzj157@psu.edu

**Dongwon Lee**
College of Information Sciences and Technology
The Pennsylvania State University
dongwon@psu.edu

## Author Keywords
Instagram; Like network; social network; tags; Likes

## ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.


CHI'15 Extended Abstracts, Apr 18-23, 2015, Seoul, Republic of Korea
ACM 978-1-4503-3146-3/15/04.
http://dx.doi.org/10.1145/2702613.2732908

## Abstract
The emergence of social media has had a significant impact on how people communicate, interact, and socialize. People engage in social media in different ways by not only adding content such as photos, texts, and videos, but also adding tags, *Likes*, comments, and following others. Through these activities, people form and develop social connections and networks. In this paper, we present a two-dimensional Like network formed and developed by people who have a same tag in their photos. Based on the dataset consisting of 51K photos posted by 36K users in Instagram, we present the *structural* and *relational* aspects of tag-based Like networks. Our study results highlight that Like networks have different sizes and degrees of network components depending on a tag type. We also found that a large portion of Likes came from random users for all networks.

## Introduction
The recent increase in the usage and prevalence of social media has led to the creation and sharing of a significant amount of information in various formats such as texts, photos, and videos [4][10]. It has become commonplace for people to actively access, appreciate and interact with shared content.

Especially, tagging and Liking are one of the popular activities that people do everyday across ages [9]. People add tags to their contents (e.g., photos) to describe what those items mean or what they want to express through the items [11]. People add Likes to

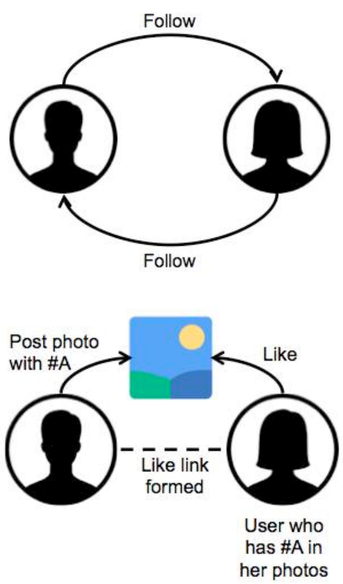


**Figure 1.** Two networks formed by a follow activity (top) and a Like activity (bottom)

| ID | Tag | Type |
|---|---|---|
| 1 | #blackfriday | Seasonal |
| 2 | #halloween | Seasonal |
| 3 | #thanksgiving | Seasonal |
| 4 | #tbt | Popular |
| 5 | #pennstate | Interest-based |
| 6 | #gameofthrones | Interest-based |

**Table 1.** Characteristics of six tags used in data collection.

indicate their shared interest in particular content [5] or acknowledge the user who uploaded the content by pressing a Like button, which can be done easily and quickly compared to adding comments.

Research has found that these activities of adding tags and Likes imply social and interpersonal connections [1][5]. For example, people who have the same tag posted on their content or those who added Likes to each other's content may have same or similar interests. Thus, a social network can be formed and developed from Like activities and interactions among a group of people who have the same tag. Tags can be used to find people with the same interest, and Likes can be used to find connections among those people.

This paper aims to understand a network formed based on two dimensions – *tags* and *Likes*. Previous research has mostly considered the social networks based on one's friendship or followships [7][8] (Figure 1, top), and relatively little has articulated the network formed by same interests in tags. We call this a "tag-based Like network" as illustrated in Figure 1 (bottom). For data collection and analysis, we chose Instagram, because Instagram is one of the most popular social network sites with a sufficient user base [6] and provides a relatively well-designed and easy-to access programming API[1] that facilitates the process of data collection. With the data collected, we explored the following research question.

*RQ: What are the structural and relational aspects of a tag-based Like network?*

[1] http://www.instagram.com/developer/

## Study Procedure

We chose a total of six tags that imply different characteristics as shown in Table 1. First three tags (i.e., #blackfriday, #halloween, and #thanksgiving) are the ones that people mostly add to their photos during a specific time period (e.g., #blackfriday is mostly added during a Thanksgiving holiday) every year. #tbt means "Throw Back Thursday." This is the name of a weekly social media-posting theme that many people participate in as part of a very general "throwback" activity for posting content. The throwback theme can be anything that happened in the past. This is also one of the most popular tags that people add in social media.

Lastly, #pennstate and #gameofthrones are the tags that imply a strong interest-based connection. #pennstate is used by people who are currently or used to be affiliated with Penn State or live in a Penn State local community, and #gameofthrones is one of the most popular TV shows that has hundreds of thousands of fans around the world.

| Tag ID | # users who posted photos | # Like links in networks | # users (%) in networks |
|---|---|---|---|
| 1 | 5,663 | 886 | 784 (13.8%) |
| 2 | 6,675 | 1,039 | 964 (14.4%) |
| 3 | 7,088 | 462 | 491 (6.9%) |
| 4 | 6,469 | 736 | 813 (12.6%) |
| 5 | 4,989 | 12,612 | 4,102 (82.2%) |
| 6 | 5,790 | 34,551 | 5,320 (91.9%) |

**Table 2.** # users who posted photos, # Like directed links (edges) and # users (nodes) who are in the Like networks.

With these six tags, we first collected a total number of 51,066 photos for all tags (8,511 photos per tag) via a PhotoSearch API. Then we extracted the information of the users who posted photos via a UserSearch API. Each

| Tag | Ave. Degree | Ave. Weighted Degree | Network Diameter | Modularity (# of communities) | Ave. Clusterting Coefficient | Reciprocity (Hybrid) | Transitivity (Triad) |
|---|---|---|---|---|---|---|---|
| #blackfriday | 0.96 | 1.13 | 7 | 0.93 (180) | 0.025 | 0.165 | 0.095 |
| #halloween | 0.97 | 1.08 | 7 | 0.83 (118) | 0.006 | 0.174 | 0.167 |
| #thanksgiving | 0.84 | 0.94 | 3 | 0.94 (108) | 0.009 | 0.191 | 0.100 |
| #tbt | 0.84 | 0.90 | 4 | 0.93 (210) | 0.011 | 0.189 | 0.250 |
| #pennstate | 2.33 | 3.07 | 20 | 0.43 (45) | 0.201 | 0.204 | 0.030 |
| #gameofthrones | 4.05 | 6.49 | 10 | 0.36 (10) | 0.276 | 0.220 | 0.084 |

**Table 3.** Summary of network components of the Like networks for six tags. Interest-based tags (#pennstate and #gameofthrones) showed a higher degree, but smaller modularity and lower transitivity than other tags.

tag has users between around 5,000 and 7,000 (2st column, Table 2). Next, we collected Like activities that were occurred and shared among these users (3nd column, Table 2) via a PhotoLikeSearch API. For the data analysis, we only considered users who had the record of Like activities (4rd column, Table 2). We also used a Relationship API to measure follow relationships among users.

## Results

### Like networks

Table 2 shows the number of users who posted photos, the number of directed Like links (edges) that users have, and the number of users (nodes) in the Like network. It was found that the number of users who posted photos for all tags ranged from 5,000 to 7,000, but when it comes to the number of users in Like networks, the four tags (i.e., #blackfriday, #halloween, #thanksgiving, and #tbt) showed a significant drop (<15%). Interestingly, however, the number of users who have Like edges in the other two tags (#pennstate and #gameofthrones) showed a small drop, holding more than 80.0% of users. This might be because of the fact that #pennstate and #gameofthrones are more interested-based tags, and the users in those tag groups seem to have many interactions among themselves and have a shared online community.

*Structure*

We first measured several network components for each tag. We used UCINET for the analysis [3]. Table 3 shows an average degree, average weighted degree, network diameter, modularity, average clustering coefficient, reciprocity, and transitivity. This result is consistent with the one presented in Table 2 in a way that #pennstate and #gameofthrones showed the networks with higher density and connections, because more users are presented in the Like network of those tags. For example, #pennstate and #gameofthrones showed both very high average and average weighted degrees. They also showed the smaller number of communities (modularity) with a higher clustering

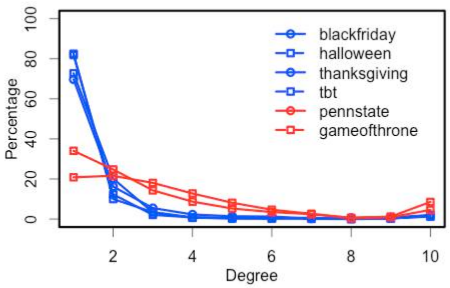


**Figure 2.** Ratio of the number of edges for each Like network.

| Tag ID | R0 | R1 | R2 |
|---|---|---|---|
| 1 | 399 (45.0%) | 171 (19.3%) | 316 (35.7%) |
| 2 | 910 (86.7%) | 56 (5.3%) | 83 (7.9%) |
| 3 | 322 (69.7%) | 32 (6.9%) | 108 (23.4%) |
| 4 | 504 (68.5%) | 63 (8.5%) | 169 (23.0%) |
| 5 | 8697 (69.0%) | 1445 (11.5%) | 2470 (19.5%) |
| 6 | 19062 (55.0%) | 5722 (16.7%) | 9767 (28.3%) |

**Table 5.** Number of Like edges found based on three-types of friendships. Grey area indicates the highest result.

coefficient, indicating that there are more people clustered together in fewer groups.

There is no significant difference among tags when it comes to reciprocity. Interestingly, however, transitivity from those two tags is lower than others. We expected that transitivity would be higher in #pennstate and #gameofthrones, because they have more nodes (users) and edges (Likes). This result raised an interesting question about a structure inside tag-based Like networks. Because #pennstate and #gameofthrones showed high degrees and clustering coefficients but low transitivity, perhaps their networks contain many "stars" as sub-networks. Figure 2 shows the percentage of the different number of edges that users had for each tag. As we expected, #pennstate and #gameofthrones (red lines) showed an overall higher ratio on the degree above 1 than other four tags (blue lines), which mostly had degree-1 networks.

*Relationship*

We investigated the relationship between users and the Like network. Variables from users can be many, including age, gender, level of engagement (e.g., how many photos each user has posted; how many Likes each user has added or received), popularity (e.g., how many followers each user has), friendship (e.g., are two users following each other). In this paper, we considered user's *friendship* as a user component. For each tag, we clustered edges that any two users have into three groups - *R0: No relationships; R1: One-way relationships; R2: Mutual relationships*. Table 5 shows the summary of the results.

One interesting result in Table 5 is that R0 is the highest relationship for all tags, showing that people tend to receive a large portion of Likes from people whom they do not know (i.e., random people). What makes this result more interesting is that people identified in this study are the ones who are already linked somehow through the tag; thus, they are not totally random people. Even for #pennstate and #gameofthrones which showed a large number of users and interconnected Like activities among users as described in the previous section, when it comes to their relationship, it was found that many of them do not follow each other.

In addition, we found that #blackfriday (Tag ID: 1) has the highest R2 percentage (35.7%) and the lowest R0 (45.0%). Perhaps users in this tag are likely to add Likes to the photos from close friends or families for holiday shopping. Conversely, #halloween (Tag ID: 2) has the highest R0 (86.7%) and the lowest R1 (5.3%) and R2 (7.9%). Although #halloween implies social activities among friends and family members, when it comes to interactions in online space, perhaps more users interact with random users, because they will be able to access many unique and interesting Halloween related photos such as costumes, festivals, events, and so on.

## Visualizing Like networks

Based on the findings in the previous sections, we also visualized the Like networks by using Gephi [1].

Figure 3-5 (best viewed in color) show directed Like networks in #halloween, #tbt and #pennstate. Nodes represent users, and edges represent Like links between two users. Blue edges represent the Likes from two users with no relationship (R0), whereas red edges represent two users who have either one-way (R1) or mutual relationships (R2).

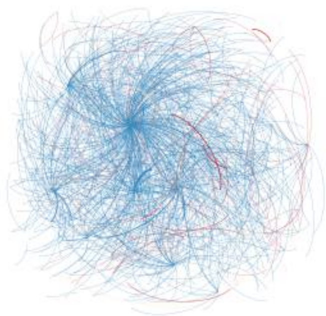

**Figure 3.** Like networks in #halloween.

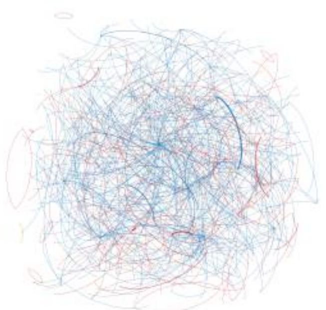

**Figure 4.** Like networks in #tbt.

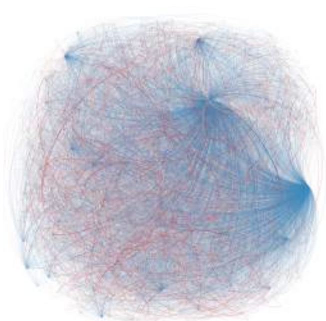

**Figure 5.** Like networks in #pennstate.

Overall, all networks contain interwoven links and connections. The network in #pennstate is bigger and more interconnected than other two networks, but there seem to be similar patterns observed in all networks. First, similar to what we have discussed before, there are a number of users who have many edges to others; in other words, having big star sub-networks. We can see more big stars especially in an interest-based network (Figure 5). In addition, it seems that there are relatively more stars formed exclusively in blue edges (i.e., no-relationship) than red ones. This again indicates that people in Instagram tend to add a lot of Likes to the photos posted by random users.

## Discussion and Conclusion

Our study highlights the structural and relational aspects of tag-based Like networks in social media, summarized as follows.

First, the network analysis on Like networks shows that the networks developed from the interest-based tag (i.e., #pennstate and #gameofthrones) had more users who added Likes one another, more Likes, smaller modularity, and higher clustering coefficient than other types of the tags such as seasonal (#blackfriday, #halloween, and #thanksgiving) and popular (#tbt) that do not have a specific boundary. Although there was no significant difference in reciprocity among all networks, interestingly, there was a very low transitivity in interest-based Like networks, and we found that this is because each network had a different ratio of stars as sub-networks.

Second, a large number of Likes came from random users whom one does not know. Even the networks in #pennstate and #gameofthrones, which initially presented a very dense network of people from the original dataset, showed a high percentage of Likes received from random users.

Then an interesting question is the reason why we had these results. One possible answer can be related to a “one-click” interface in Instagram. Instagram originally started from the mobile application, and its use through the mobile application is much higher than the webpage. In this regard, we investigated the Instagram mobile interface. We found that the application has been designed to allow people to easily navigate a set of photos by simply clicking the tag, meaning that people can see a collection of random photos that have the same tags of their own photos. The mobile application also provides a “search function” where people can easily look for and check any photos related to a keyword that they type in a search bar. Figure 6 shows the screenshots of the Instagram mobile application. By clicking a tag in the photos (#gameofthrones colored in blue; Figure 6, top), users can see a set of other photos with the same tag. We can easily imagine a scenario where people regularly check the photos that they posted and click the tags in their photos to explore other photos with the same tag posted by random users.

Furthermore, we found that there are a number of regular or ad-hoc online events or promotions hosted by Instagram[2] or many other third-party websites[3]. They encourage people to add a special tag to their photos (similar to adding hashtags in Twitter) as well as provide a summary page that shows a set of tags shared by many people based on time, location, topic and so on, allowing people to add Likes to many types of photos very easily. These insights all support the

[2] http://blog.instagram.com/
[3] http://iconosquare.com/


**Figure 6.** Screenshot of Instagram interfaces. Users can easily navigate a collection of photos that have the same tag as their own photos. User can also search for the specific tag easily through the search function.

idea that having friendship is not necessarily for Like interactions. We plan to conduct follow-up research studies in order to articulate reasons behind this trend.

This study has several limitations that also suggest many interesting future directions. First, the results presented in this paper may neither be generalized in other social media (e.g., Flickr, Pinterest, etc.) nor represent the whole Instagram community. We also acknowledged that all six tags are US culture-based. To address this issue, we plan to collect data from more diverse tags and re-run the same comparison analysis. We also plan to extend our study to other social media platforms and study if they present similar trends. Second, some of our analyses are somewhat speculative; therefore, we plan to investigate reasons more specifically by taking both qualitative and quantitative approaches. Third, we want to further explore the relationship between Like networks and other user attributes. In this paper, we only considered followships; thus, we plan to expand the analysis based on age [9], gender, popularity, and engagement in Instagram. Lastly, we want to more specifically investigate individual users who added a lot of links to others. After we have a list of special users like them, we can manually visit their Instagram page and see who they are and what kind of engagement they have presented and what their roles are in their Like network compared to other users.

In summary, this paper tackles unique and interesting perspectives from tag and Like activities in social media and networks. We plan to investigate them more in order to understand and articulate a variety of social phenomena created by people in online space.

## Acknowledgements
This research was in part supported by NSF CNS-1422215 award.

## References
[1] Ames, M., & Naaman, M. Why we tag: motivations for annotation in mobile and online media. In *Proc. CHI*, ACM (2007), 971-980.

[2] Bastian, M., Heymann, S., & Jacomy, M. Gephi: an open source software for exploring and manipulating networks. In *Proc. ICWSM* (2009), 361-362.

[3] Borgatti, S.P., Everett, M.G. and Freeman, L.C. *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies, 2002.

[4] boyd, d., & Ellison, N. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication 13*, 1 (2007), 210-230.

[5] boyd, d., Golder, S., & Lotan, G. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In *Proc. HICCS* (2010), 1-10.

[6] Duggan, M. Photo and Video Sharing Grow Online. *Pew Research*, 2013.

[7] Gruzd, A., Wellman, B., & Takhteyev, Y. Imagining Twitter as an Imaged Community. *Journal of American Behavioral Scientist 55*, 10 (2011), 91-98.

[8] Huberman, B.A., Romero, D.M., & Wu, F. Social networks that matter: Twitter under the microscope. *Peer-reviewed Journal on the Internet 14*, 5 (2009).

[9] Jang, J., Han, K., Shih, P.C., & Lee, D. Generation Like: Comparative Characteristics in Instagram. In *Proc. CHI*, ACM (2015).

[10] Kaplan, A.M. & Haenlein, M. Users of the world, unite! The challenges and opportunities of social media. *Journal of Business Horizons 53*, 1 (2009), 59-68.

[11] Wu, C. & Zhou, B. Analysis of tag within online social networks. In *Proc. Group*, ACM (2009), 21-30.