Do You Really Like Her Post?: Network-Based Analysis for Understanding Like Activities in SNS

Junho Song
Hanyang University
junonjuno@bigdas.hanyang.ac.kr

Hyekyoung Park
Hanyang University
hk719@bigdas.hanyang.ac.kr

Kyungsik Han
Ajou University
kyungsikhan@ajou.ac.kr

Sang-Wook Kim*
Hanyang University
wook@hanyang.ac.kr

ABSTRACT

As social network services (SNS) are expanding from friend-based to interest-based, users form a new type of relationships, namely *interest-based relationships*, with friends and others through *social activities* (e.g., likes, comments). Although such relationships are highlighted in the common-identity theory and have important values in theoretical and practical aspects, little evidence exists in the literature pertaining to the explanation of social activities as a central component for social network analysis and an association with friendship. In this paper, we build *like networks* in Instagram and analyze them through the lens of two salient aspects – friendship and interest – that constitute social networks. Our study results (1) show ambiguous interpretations of the like activities between users who are friends, based on the comparative analysis between friend- and non-friend-based like networks, and (2) demonstrate strong signals of the hashtag characterizing the interest-based relationships in users and content. Our research substantiates and gives insights on the common-identity theory applied in online social networks through data-driven, empirical analysis.

CCS CONCEPTS

• **Human-centered computing → Social network analysis**; *Social networks*; Social networking sites; Social tagging systems.

KEYWORDS

Common-identity theory; Like network; Hashtags; Instagram

ACM Reference Format:
Junho Song, Hyekyoung Park, Kyungsik Han, and Sang-Wook Kim. 2020. Do You Really Like Her Post?: Network-Based Analysis for Understanding Like Activities in SNS. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3340531.3412166

*Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CIKM '20, October 19–23, 2020, Virtual Event, Ireland
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-6859-9/20/10…$15.00
https://doi.org/10.1145/3340531.3412166

1 INTRODUCTION

As social network services (SNS) are expanding across the border of services from friend-based to interest-based, SNS users form a new type of *interest-based* relationships through social activities (i.e., referring to one's actions such as 'like' or 'comment' performed on the content created by someone else). According to the *common-identity theory* in sociology, a social group is formed and maintained by the minimum identity of people such as common-interest [12]. This minimum identity is also observed in SNS [2], represented by the interest-based relationships through social activities.

Such *interest-based relationships* have important values in both of theoretical and practical perspectives. Researchers can study various characteristics (e.g., relationship types, personal/societal impacts, online/offline behavioral influences) of a social group of users who have common identity [7]. Marketers can identify the customers who will be likely to purchase their products, which directly influences profits [9], since social networks have a significant impact on consumers' behavior and decision-making.

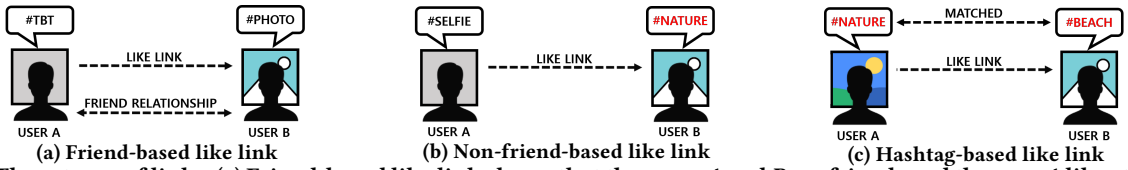
However, despite the big potential of such interest-based relationships, they have been mostly theoretically and conceptually understood. Even prior research in the network analysis has used social activities merely as secondary information to analyze friend-based networks [3, 8]. Little research has looked into the relationships based on the real SNS data and identified their salient network characteristics through a data-driven, empirical analysis and comparison with other networks (e.g., friend-based network).

In this paper, we study the *interest-based relationship* formed by like activities and analyze *like networks* [11] in Instagram. With like activity logs performed over a year from 3, 567 public Instagram users, we investigate the characteristics of the like networks, together with friendships and hashtags, in two aspects.

(1) The existence of friendship. We compare and analyze a Friend-based Like Network (FLN) and a Non-Friend-based Like Network (NFLN) to investigate how the existence of friendships among users affects their like activities. In addition, we categorize hashtags of posts, which are the keywords explaining the content of the posts, into *topics*, and analyze the topic distribution of FLN and NFLN.

(2) The existence of common interest (hashtags). We analyze a Hashtag-based Like Network (HLN) to see how common interest (represented by hashtags) between a pair of users affects their like activities. To this end, we build HLN composed of the users who used the hashtags of the topics categorized in (1) above.

The results from the (1) and (2) above provide us the following observations: (1-1) NFLN shows more nodes, higher modularity [4],



(a) Friend-based like link (b) Non-friend-based like link (c) Hashtag-based like link

Figure 1: Three types of links. (a) Friend-based like link shows that the users *A* and *B* are friends and the user *A* likes the user *B*'s post. (b) Non-friend based like link shows that the user *A* likes the user *B*'s post without friend relationship. (c) Hashtag-based like link shows that the user *A* likes the user *B*'s post where they have used hashtags in the same topic category.

and higher transitivity [4], while FLN does more edges, higher average degree, and higher reciprocity [4]; (1-2) the hashtags in NFLN tend to reflect users' real interests such as cosmetics, photography, and outdoors; (1-3) FLN users tend to use the hashtags for content-promotion such as #wcw, #notifier, and #tbt, which are less related to their true interest; (2-1) the modularity, transitivity, and reciprocity of HLN exhibit similar tendency to those of NFLN; (2-2) NFLN users and HLN users are greatly overlapped (98%).

In summary, these results indicate that (1) FLN users perform more like activities than NFLN users, but *the like activities in FLN are performed irrespective of interests*; the content of FLN posts mainly contains the hashtags that entail the like activities; that is, *like activities of FLN may not necessarily mean that the users really like the content*; conversely, NFLN users perform like activities on content, based on *their real interest*. In addition, (2) hashtags play an important role in *characterizing interest-based relationships and networks* regardless of the existence of friendship. Study results and interpretations are expected not only to provide researchers with an empirical evidence of the characteristics of the like network along with their association with friendships and hashtags, but also to give practitioners insights on how to strategically leverage social like networks to achieve their goals (increasing advertising effect).

2 RELATED WORK

Interest-based analysis. Social activities on SNS are recently under the spotlight because they consist of many forms (e.g., likes, comments, direct messages) and can be interpreted in various ways when analyzed alone or together. [3] reported that such activities are conducted regardless of the relations such as follows and friends, showing that social activities imply users' intentions. [6] showed that comments and users' interests are correlated through the analysis on the comment network.

Research directions and goals. In this paper, we study Instagram through 'like' activities. Unlike prior research that considers like activities *as secondary information for understanding friend-based networks* [1] or studies basic characteristics of like activities or like networks themselves [4, 5], we focus on *the characteristics and meanings of like activities in forming interest-based relationships* through the lens of two salient aspects in SNS, *friendship* and *common interest* (hashtags). Our study results highlight the ambiguity of friend-based activities in representing interests, complementary roles of likes, and roles of hashtags featuring the like network.

3 STUDY PROCEDURE

Survey. To confirm which factors influence like activities, we conducted a survey with 300 Instagram users through Amazon Mechanical Turk [1]. The survey consists of the following 4 questions with a 5-point scale (1: strongly disagree; 5: strongly agree).

Q1. Do you consider the *content* when you add a like to the post?
Q2. Do you consider the *poster* when you add a like to the post?
Q3. Do you consider the *friendship* more importantly than the content when you add a like to the post?
Q4. Do you have an experience in adding a like because the post is posted by your friend regardless of content?

Data collection. From Instagram, we crawled the direct and indirect friends from a random seed user until we reached 1,000 users. We then randomly chose 100 users among them, and again crawled the direct and indirect friends of the 100 users until we had 30,000 users. From the 30,000 users, we again randomly sampled 5,000 users. With these steps, we minimized the bias in sampling a homogeneous population. Finally, we filtered out advertisers, celebrities, and bots, whose 'like' activities are abnormal in aspects of general users. As a result, we obtained 3,567 public users and crawled their log of 'like' activities and their posts for the analysis.

Analysis of networks. For building networks, we defined three types of links as shown in Figure 1: friend-based like link, non-friend-based like link, and hashtag-based like link.

Based on these three types of links, we build three networks; Friend-based Like Network (FLN), Non-Friend-based Like Network (NFLN), Hashtag-based Like Network (HLN) where a node indicates a user, and an edge indicates one of the three types of links. Therefore, FLN represents the network of users who like others' posts and have friend relationships with them, NFLN represents the network of users who like others' posts but do not have friend relationships with them, and HLN represents the network of the users who like others' posts and have hashtags in the same category as those of their posts. We analyzed and visualized the networks using the Python library, Network X [2].

4 RESULTS AND DISCUSSIONS

4.1 Survey (factors influencing like activities)

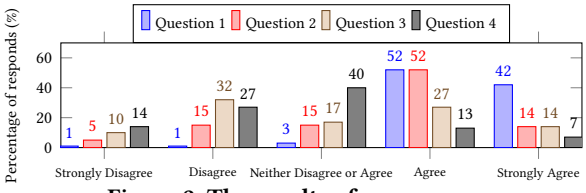


Figure 2: The results of our survey.

As shown in Figure 2, most respondents (94%) answered that they are affected by the content itself when adding likes (Q1). However, more than half of the respondents (66%) answered that they are affected by the poster who wrote the content (Q2), and 20% have experienced adding a like to content just because the poster is their friend (Q4). Related to Q4, about half of respondents answered that they add a like to content if the poster is their friend regardless of the interest in the content itself (Q3).

[1] https://www.mturk.com/

[2] https://networkx.github.io/

Table 1: Structural characteristics of friend-based (FLN) and non-friend-based like networks (NFLN).

	#Nodes	#Edges	Avg. degree	Avg. weighted degree	Avg. in-degree	Avg. out-degree	Modularity	Avg. clustering coefficient	Transitivity	Reciprocity
FLN	799	**4451**	**11.141**	**193.945**	7.144	6.199	0.571	0.328	0.086	**0.612**
NFLN	**846**	3839	9.076	37.326	5.730	6.084	**0.665**	0.185	**0.121**	0.152

These results indicate that, although one's interest in content is a significant factor that motivates like activities, *such activities are also influenced by the existing relationships* (i.e., friendship). The interpretation of like activities and the investigation of interest-based networks are required to clearly understand interest-based relationships in SNS. Towards this end, it is challenging to distinguish *real* like activities which express true interest of users from those affected by friendship. This highlights the importance of addressing the ambiguity of interpreting like activities.

4.2 The existence of friendship

To investigate how the existence of friendships affects like activities, we build FLN and NFLN, and compare them with respect to the structural characteristics and topic distributions.

Structural characteristics. Table 1 summarizes the results of comparing the structural characteristics of FLN and NFLN. First, FLN shows more edges (4,451), higher average weighted degree [4] (193.94), and higher reciprocity [4] ($F(1,198)=4.79$, $p < 0.05$), while NFLN does more nodes (846), higher modularity [4] ($F(1,198)=6.55$, $p < 0.05$), and higher transitivity [4] ($F(1,198)=1.52$, $p = 0.12$, marginally significant).

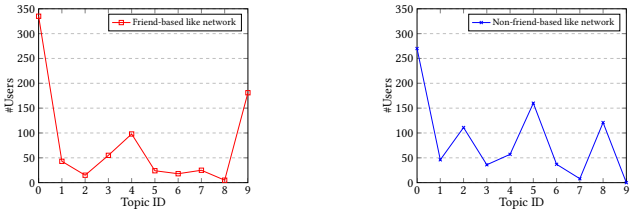
Through the results, we can infer that FLN users add likes more actively to the content than NFLN users. However, those like activities are highly *reciprocal* and *repetitive*, which means that those activities are used to *tighten their friendship*. On the other hand, NFLN users who add likes with real interest of content are separated into a few groups and have been tied stronger.

(a) FLN **(b) NFLN**

Figure 3: Visualization of FLN and NFLN.

To further confirm the characteristics of the two networks, we visualize FLN and NFLN using forceatlas2, which is the force-directed graph drawing algorithm. The result is illustrated in Figure 3. Here, the size of a node corresponds to the weighted degree of the node, and the distance between nodes corresponds to a weight on the edge, and the color of a node represents a community to which the node belongs. To detect the communities, we use a modularity based algorithm [10].

As shown in Figure 3a, we found a very large star user in FLN. Most of nodes are connected to this star user. In addition, although there are a few different communities, they are close to each other and centered on the star user, indicating that those communities may be interconnected by friendship with the star user. Whereas, as shown in Figure 3b, we found that NFLN has evenly-distributed communities and no star user. With the visualization of the two networks, we confirmed our inference again that FLN users add likes to tighten their friendship, while NFLN users do to have interest-based relations.



(a) FLN (Kurtosis: 1.663, Skewness: 1.696) **(b) NFLN (Kurtosis: 0.423, Skewness: 1.129)**

Figure 4: Topic distributions in FLN and NFLN.

Topic distributions. We investigate the topic distributions of FLN and NFLN. For this, we extract 10 topics by using the topic modeling method, LDA, over the hashtags used by all users in our data. Figure 4 shows the distributions of the extracted topics. It shows that hashtags in NFLN (Kurtosis: 0.423, Skewness: 1.129) tend to spread over more-various topics than FLN (Kurtosis: 1.663, Skewness: 1.696). Through this result, we found that the communities in NFLN are more evenly distributed than those in FLN since each topic might correspond to an interest of users in a community. On the other hand, in FLN having the star user, we confirmed that the distribution of topics is highly concentrated on a few topics in FLN having the star user.

Table 2: Top-3 hashtags in top-3 topics from FLN and NFLN.

	Top-3 Topics	Top-3 Hashtags		
FLN	**Content-promotion**	#wcw	#nofilter	#dogofinstagram
	Art	#flower	#artist	#drawing
	Makeup	#mua	#makeupartist	#beauty
NFLN	**Cosmetics**	#maccosmetics	#macsnowball	#vivaglam
	Photography	#photography	#art	#nature
	Outdoor	#summer	#travel	#beach

To further investigate the differences between FLN and NFLN, we select the top-3 topics in the order of frequency of topics in the two networks. Table 2 shows the top-3 frequent topics and the top-3 frequent hashtags in each topic. The top-3 topics of NFLN are cosmetics, photography, and outdoors, and the top-3 topics of FLN are content-promotion, art, and makeup. Interestingly, the most frequent topic of FLN is content-promotion, and its top-3 hashtags are #wcw, #notifier and #dogofinstagram, which are not related to interest. These hashtags are often used to promote their content or to share the memories between friends; this indicates that the users of FLN show like activities for other intentions rather than expressing their interest.

In summary, FLN users perform like activities more than NFLN users, but these activities are not related to their interest. In addition, the content in FLN mainly contains common hashtags that are not related to users' true interest. That is, *the like activities of FLN can be ambiguous regarding their association with true interest in the post*, given that a user adds likes because of the content and/or the

Table 3: The network structure comparison of hashtag-based like networks along the categories.

	Categories	#Nodes	#Edges	Avg. degree	Avg. weighted degree	Avg. in-degree	Avg. out-degree	Modularity	Avg. clustering coefficient	Transitivity	Reciprocity
HLN	Beauty	918	2,344	5.107	44.100	2.978	5.773	0.743	0.063	0.089	0.198
	Art	650	1,291	3.972	32.086	2.343	4.332	0.813	0.055	0.050	0.189
	Event	1,212	2,225	3.672	32.721	2.179	4.230	0.830	0.038	0.041	0.174

poster. On the other hand, the like activities in NFLN are likely to be related to a user's interest in the post.

4.3 The existence of common interest

To investigate how hashtags between a pair of users affect their like activities, we build HLN, analyze its structural characteristics, and compare it with FLN and NFLN. From the 10 topics of the previous analysis, we select 3 interest-based topics that have the networks of a size similar to those of FLN and NFLN (the cosmetics and makeup topics are merged into a beauty topic).

Table 4: Topic categories and their top-5 frequent hashtags.

Categories	Top-5 frequent hashtags (Frequency)				
Beauty	#beautiful (890)	#beauty (680)	#makeup (643)	#hair (558)	#fashion (533)
Art	#art (817)	#disney (363)	#artist (336)	#painting (333)	#starwars (327)
Event	#love (1279)	#family (1102)	#friends (905)	#happy (837)	#christmas (829)

The results of categorization are shown in Table 4. 'Beauty' and 'Art' contain the hashtags such as #beauty, #makeup, #fashion, #art, #artist, and #painting, which represent the interest. Besides, 'Event' contains the hashtags such as #love, #happy, and #christmas, which represent the emotion or anniversaries. We then build 3 HLNs and analyze their structural characteristics. Table 3 shows that HLNs have a structure closer to that of NFLN than FLN (Table 1), because of low avg. weighted degree (NFLN: 37.326), high modularity (NFLN: 0.665), high transitivity (NFLN: 0.121), and low reciprocity (NFLN: 0.152). In addition, HLNs' avg. degree is lower than that of NFLN (9.076) despite the bigger size of networks, and their avg. in-degree is lower than their avg. out-degree.

From these characteristics, we can infer that HLNs tend to express users' interests more than of NFLN. For example, in Instagram, when users search for the post through hashtags they are interested in, they may add a 'like' if the searched post matches their interest. Also, they may click on the hashtag in the post to see other posts with the same hashtag and add 'likes' to some of them. Such 'like' activities are not necessarily related to friendship (reciprocal relationship) but rather to their pure interest and curiosity. In this sense, the 'like' activities of users in HLNs can be interpreted as the *pure* interests.

To confirm this, we extract the users who have the same hashtags of the 3 topics from NFLN and examine the number of users overlapped between 3 HLNs and NFLN. Table 5 shows very high proportions (98%, 98%, 96%) of the users who belong to NFLN also belong to 3 HLNs, indicating that most users in HLNs are not friends. Thus, we can confirm that the hashtags play an important role as a *filter* that removes the like edges for the pairs of the users in friendship from the like network, thereby enabling to find a group of users who have true interests in common.

Table 5: The overlap of users between HLN and NHLN.

Categories	Users in HLN / Users in NFLN	Overlap ratio
Beauty	253 / 259	98%
Art	209 / 213	98%
Event	230 / 239	96%

5 CONCLUSIONS

Our study presents the importance of interpreting 'like' activities in SNS through the network-based analysis. It aims to better understand interest-based relationships among SNS users formed by social activities and also discussed in the *common-identity theory*. Our study results (1) point out ambiguous interpretations of the like activities between the users who are friends, based on the comparative analysis between friend- and non-friend-based like networks (Sections 4.1 and 4.2), and (2) demonstrate that hashtag-based like networks well reflect the characteristics of the interest-centric SNS (Section 4.3). We expect that our results give valuable insights to researchers and practitioners in industry as well as academia.

ACKNOWLEDGEMENTS

This work was supported by (1) the National Research Foundation of Korea (NRF) grant (No. NRF-2020R1A2B5B03001960), (2) the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant (No. 2020-0-01373, Artificial Intelligence Graduate School Program (Hanyang University)), and (3) the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) (No. NRF-2017M3C4A7083678).

REFERENCES

[1] Yong-Yeol Ahn, Seungyeop Han, Haewoon Kwak, Sue Moon, and Hawoong Jeong. 2007. Analysis of Topological Characteristics of Huge Online Social Networking Services. In *Proceedings of the International Conference on World Wide Web*. 835–844.

[2] Pasquale De Meo, Fabrizio Messina, Domenico Rosaci, and Giuseppe ML Sarné. 2017. Forming Time-stable Homogeneous Groups into Online Social Networks. *Information Sciences* 414 (2017), 117–132.

[3] Bluma Gelley and Ajita John. 2015. Do I Need to Follow You?: Examining the Utility of the Pinterest Follow Mechanism. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1751–1762.

[4] Kyungsik Han, Jin Yea Jang, and Dongwon Lee. 2015. Exploring Tag-based Like Networks. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. 1941–1946.

[5] Jin Yea Jang, Kyungsik Han, and Dongwon Lee. 2015. No Reciprocity in Liking Photos: Analyzing Like Activities in Instagram. In *Proceedings of the ACM Conference on Hypertext and Social Media*. 273–282.

[6] Bo Jiang, Yun Ling, and Jiale Wang. 2010. Tag Recommendation Based on Social Comment Network. *International Journal of Digital Content Technology and its Applications* 4, 8 (2010), 110–117.

[7] Jan H Kietzmann, Kristopher Hermkens, Ian P McCarthy, and Bruno S Silvestre. 2011. Social Media? Get Serious! Understanding the Functional Building Blocks of Social Media. *Business Horizons* 54, 3 (2011), 241–251.

[8] Yeon-Chang Lee, Nayoun Seo, Kyungsik Han, and Sang-Wook Kim. 2020. ASiNE: Adversarial Signed Network Embedding. In *Proceeding of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 609–618.

[9] Andrew Lipsman, Graham Mudd, Mike Rich, and Sean Bruich. 2012. The Power of "like": How Brands Reach (and Influence) Fans through Social-media Marketing. *Journal of Advertising Research* 52, 1 (2012), 40–52.

[10] Mark EJ Newman. 2004. Fast Algorithm for Detecting Community Structure in Networks. *Physical review E* 69, 6 (2004), 066133.

[11] Hyekyoung Park, Junho Song, Kyungsik Han, and Sang-Wook Kim. 2018. I Like Your Tagged Photos, But Do We Know Each Other?: Analyzing the Role of Tags in Like Networks. In *Proceeding of the IEEE International Conference on Web Services*. 335–338.

[12] Henri Tajfel. 1981. *Human Groups and Social Categories: Studies in Social Psychology*. Cup Archive.