

Introducing Hidden Nodes in a Relationship Graph of Instagram Users

Unmanee Unmuang and Sukree Sinthupinyo

Abstract—Online social networks (OSNs) like Facebook, Twitter, and Instagram assist their users in establishing new social relations by making recommendations. In order to make relevant recommendations to users, social network providers need to understand the relations between them.

Many studies have discovered unknown relations between users using similarity-based metrics. The most frequently used similarity-based metric is a common neighbor (CN) metric. However, such a metric assumes that the networks are homogeneous, while real-world social networks actually contain various types of entities and relations, making them heterogeneous. It also only considers structural information without leveraging the contextual information of the networks. Consequently, the discovered relations contain no hidden meaning.

In this paper, we analyze the relations between Instagram users considering both structural and contextual information. As a result, we leverage the heterogeneity of the network and discover hidden nodes that carry the semantics of the relations. We perform an analysis in two steps as follows: 1. we consider the structural information of the networks by using common neighbors between two users. We select the top 40 user pairs that share the most common neighbors to execute the next step; 2. We perform a contextual analysis between each user in the user pairs by using posts and comments. We then calculate the term frequency of each token in the comments. We observe that tokens with a high-term frequency value represent contextual information between two users. Finally, we represent these tokens as hidden nodes in the relation between two users using the heterogeneous relation graph.

Index Terms—Hidden node, users relationship, social media, heterogeneous graph, structural information, contextual information, directed networks, combination, sequence matcher, content-based.

I. INTRODUCTION

Online social networks (OSNs) play a vital role in modern life. They have become ubiquitous in our everyday activities, enabling new ways of following the news, acquiring and transferring knowledge, sharing experiences, and communicating with others. Social networks have introduced several new concepts into our society, such as online social relations (e.g., friends, following/follower), and interactions (e.g., liking, sharing, and commenting).

Understanding the relationships among users can help social network providers in assisting their users to establish new social relations by recommending unseen users. To understand the relationships between users, we can derive knowledge from the classical social network analysis (SNA)

problems, namely link prediction. Liben-Nowell and Kleinberg define the link prediction problem as the following. Given a snapshot of a social network at time t , we seek to accurately predict the edges that will be added to the network during the interval from time t to a given future time t' [1]. In short, can we predict unknown links based on the current information of the network? The meaning of a link varies depending on the social network characteristics.

We can characterize social networks from many viewpoints. From the relationship aspect, we can classify social networks as either *directed* or *undirected* networks. In *undirected* networks, such as Facebook¹, two users establish a relation only if both users accept the friendship. The relation is mutual, thus *undirected*. On the contrary, *directed* networks, such as Twitter², allow users to establish a relationship with another user without mutual agreement. On Twitter, a user can follow another user to receive status and activity updates. Moreover, we can also classify social networks into *homogeneous* and *heterogeneous* networks based on a variety of relations and entities in the network.

In this paper, we focus on studying the relationship between users on Instagram³. Instagram has a mixed combination of features from multiple social media and online social networks. It contains various types of nodes/entities (e.g., users, posts) and consists of multiple types of links/relations (e.g., liking, commenting). Instagram has a followed-follower relationship. Thus, Instagram can be a *directed* and *heterogeneous* network.

As real-world online social networks are complicated and often modeled as heterogeneous interactions, a majority of researches on link prediction focus on heterogeneous networks. Yet, there has been little focus on studying directed networks [2]-[4].

Schall proposed new link prediction techniques using graph patterns (triads) and introduced a similarity-based metric called Triadic Closeness [2].

Bütün *et al.* proposed a new topological metric called Candidate Triad Score (CTS), which considers the direction of links, as well as temporal and weighted information [3].

Chen *et al.* transformed link prediction into a binary classification problem. The authors used the topological features of an ordered pair of nodes as input. The binary classification problem was solved by AUC optimization [4].

While these researches proposed techniques that take into account the direction of the links, they only examined the structural features of the networks without considering contextual information, such as profile information and user-generated content. On the other hand, a number of researchers employed contextual information to link prediction without leveraging structural information simultaneously [5]-[7].

Manuscript received October 21, 2019; revised March 25, 2020.

The authors are with the Department of Computer Engineering, Chulalongkorn University, 254 Phayathai Rd., Phatumwan Bangkok, 10330 Thailand. (e-mail:6170975121@student.chula.ac.th, sukree.s@chula.ac.th).

Recently, researchers proposed hybrid approaches that analyze both structural and content information [8], [9].

Wang *et al.* integrated content and structural information using a user-to-user topic inclusion degree (TID) score. The authors fused the TID networks and the original network with the probabilistic matrix factorization techniques and performed link prediction with the unified matrix [8].

Muniz *et al.* proposed three weighting criteria that combine contextual, temporal and topological information. The authors combined topological, contextual and temporal aspects of social networks in weight calculation and used this combination to predict links [9].

Following the intuition from previous studies, we hypothesize that incorporating both the structural and content information can reveal hidden information regarding the social relations between users. Hence, we analyzed both the structural information and content information of 2,000 public Instagram users. The analysis was done in two steps: 1. We examined the structural information of the network using the common neighbors (CN) [10] of two users; 2. We selected the top 40 user pairs that shared the most common neighbors. The content information, such as comments on Instagram posts of each user, was collected and tokenized. For each user pairs' comments, we calculated the term frequency of each word and analyzed it.

Our analysis results show that words extracted from the comments with high-term frequency can represent a social relation between two Instagram users. The high term-frequency words, such as legend or Lisoo (denotes the relation between two celebrities, named Lisa and Jisoo), contain hidden information about the relation. These words are hidden nodes that can connect two users. We represent the hidden nodes (e.g., words in a comment) in social relations between two sample Instagram users using a heterogeneous relations graph.

The rest of the paper is organized as follows: Section II summarizes the related works to link prediction techniques that consider either or both structural information and contextual information of the networks. In Section III, we present the methodology used in this paper, starting from data collection, extraction, and analysis. Section IV shows the results of our analysis with example representations. Finally, we conclude the paper and discuss possible future works in Section V.

II. RELATED WORK

As previously stated, we follow the intuition from the link prediction problem to analyze the hidden information in social relations between users.

Link prediction methods define a similarity function for a pair of nodes. Similarity functions rely on either or both structural information and contextual information of the networks.

A large number of link predictions consider only the structural information of the networks. One of the most commonly used similarity functions is the common neighbors (CN), which was proposed by Newman [10]. A similarity function of common neighbors is calculated by the number of shared neighbors between two nodes. Yet, common neighbors do not incorporate link direction in the

case of a directed network. The following works considered the direction of links in the networks:

Schall proposed a new link prediction technique using graph patterns (triads) and introduced a similarity-based metric called Triadic Closeness. The author performed experiments on three different directed social networks, including GitHub, GooglePlus, and Twitter. The results showed that the proposed metrics, Triadic Closeness, and the proposed framework performed better predictions when compared with well-known undirected similarity metrics [2].

Chen *et al.* transformed link prediction into a binary classification problem. The authors used the topological features of an ordered pair of nodes as input. The authors treated link prediction as a binary classification, where the label of each node pair was determined by the existence of a directed link between the node pair. Then, the binary classification problem could be solved by the area under the receiver operating characteristic curve (AUC) optimization. The authors performed experiments on the largest connected component in the following benchmark data: the US airport network (USAir), political blogs (PB), a protein-protein interaction network (PPI), and the Internet (INT). According to the empirical results, this method achieved high-quality predictions [4].

Recently, researchers proposed hybrid approaches that analyze both structural and content information.

Wang *et al.* integrated context and structural information using a new similarity score. The authors defined a user-to-user topic inclusion degree (TID) score based on the dissemination of the published content in the following/followed network structure, and then constructed a new TID-based network. The authors then fused the information of the original following/followed network and the TID-based network in a unified probabilistic matrix factorization framework. They experimented with two types of social networks, Twitter and Weibo. The result showed that the proposed approach is effective for link prediction [8].

Muniz *et al.* proposed three weighting criteria that combine contextual, temporal, and topological information. The authors combined topological, contextual, and temporal aspects of social networks in weight calculation and used this combination to predict links. The authors evaluated three weighting criteria: 1) The Temporal-Topological (TT) criterion, 2) The Contextual-Topological criterion (CT), and 3) The all features criterion (CTT, The Contextual-Temporal-Topological) with two common weighted similarity functions (Adamic-Adar and Common Neighbors) in ten networks frequently used using link prediction. The results showed that only combining all features criterion (CTT) enhanced link prediction [9].

Overall, many researchers have conducted experiments on the link prediction problem on either or both structural information and contextual information of the networks. However, recent works have shown that analyzing both the structural and content information can improve link prediction performance.

III. METHODOLOGY

This paper aims to discover and introduce hidden information that connects users using contextual information,

such as post comments. We analyze the relationship of a large number of following users on Instagram by randomly scraping users' information and collect the information of users and their relations with other users.

The results we obtained comprise a list of celebrities that have relationships between them. Then, we selected a pair of celebrities that have a relationship to find the pictures that a user tag the other. We collected captions, comments, hashtags, and the media URL of pictures to identify the nodes that can represent both of them and the information about their relationship (see Fig. 1).

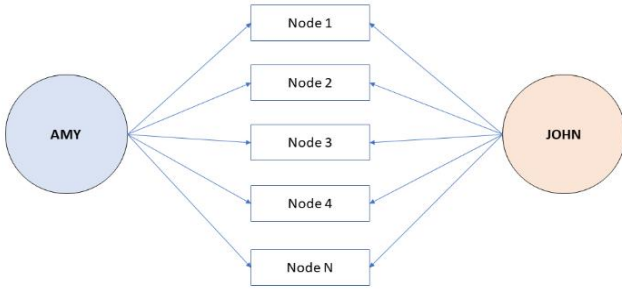


Fig. 1. Hidden nodes between user relationship.

Retrieving information from Instagram used to be done using Graph API, an official API from Instagram, but it no longer exists. Hence, we used Python Selenium Tool to scrape element data and collect information from Instagram.

We separate the methodology into five steps as follows:

First, we began randomly scraping 10 seed users' information, such as names from Instagram users who set their account profile as public and had less than 1,000 followers. Then, we scraped the following information from those users, such as name, user URL, number of followers, number of following, and the names of following users under the condition that they set their account profile as public (see Fig. 2). We collected a total of 2,353 users in this step.

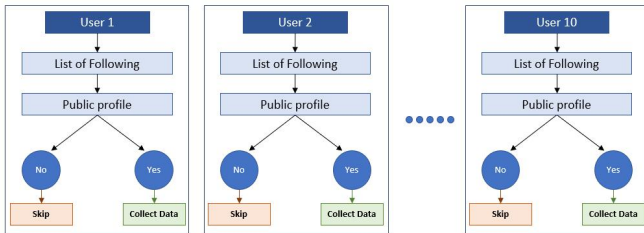


Fig. 2. Steps for the process of collecting following information.

Second, we grouped users' data from the previous step by the name of the users whom they followed (see Table I). We then ranked each user group by the number of members. We could identify the celebrities by their ranking (see Table II). U_i denotes the i^{th} user and C_i denotes the i^{th} celebrity.

TABLE I: LIST OF USERS FOLLOWED

Users	Users followed
U_1	$C_1, C_2, U_2, U_{19}, U_{10}, C_{15}, U_{13}, C_{18}, U_{17} \dots U_n$
U_2	$U_2, C_1, C_2, U_{20}, U_{21}, C_{28}, U_{22}, C_{10}, U_{27} \dots U_n$
U_3	$U_8, C_1, C_2, U_{39}, U_{30}, C_{25}, U_{30}, C_{20}, U_{37} \dots U_n$
U_4	$U_2, C_1, C_2, U_{39}, U_{40}, C_{15}, U_{30}, C_{30}, U_{40} \dots U_n$
U_5	$U_7, C_3, C_4, C_5, C_6, C_7, U_1, U_2, U_{99}, U_{65} \dots U_n$
U_6	$U_8, U_{78}, C_5, U_4, C_1, C_{20}, C_{30}, U_{50}, U_{84} \dots U_n$
U_{2535}	$U_n, U_n, C_n, C_n, U_n, C_n, U_n, C_n, C_n \dots U_n$

TABLE II: LIST OF CELEBRITIES FOLLOWERS

Celebrities	Celebrities followers
C_1	$U_1, U_2, U_3, C_{15}, U_4, C_{18}, U_6, \dots U_n$
C_2	$U_1, U_2, U_3, U_4, U_{21}, C_{28}, U_{22}, C_{10}, \dots U_n$
C_3	$U_5, U_{81}, U_{82}, U_{83}, U_{30}, C_5, U_{30}, C_{20}, \dots U_n$
C_4	$U_5, C_1, C_2, U_{39}, U_{40}, C_{15}, U_{30}, C_{30}, \dots U_n$
C_5	$U_5, U_6, C_4, C_5, C_6, C_7, U_1, U_2, U_{99}, \dots U_n$
C_6	$U_5, U_6, C_5, U_4, C_1, C_{20}, C_{30}, U_{50}, \dots U_n$
C_n	$U_n, U_n, U_n, U_n, U_n, U_n, U_n, C_n, C_n \dots U_n$

Third, the celebrities from the previous step were processed in pairs. We analyzed the combination of celebrity pairs to find the pairs that are related using the number of users who follow them (see Table II).

We observed the pair of users who shared a large number of common followers, such as (C_1, C_2) and (C_5, C_6) .

Fourth, we selected the top 40 celebrity pairs that shared the most following users. Then, we scraped the pictures that had been tagged with either one or both celebrities in each pair. We extracted users' comments and tags from these pictures. The network that we scraped can be represented in a heterogeneous graph, as shown in Fig. 3.

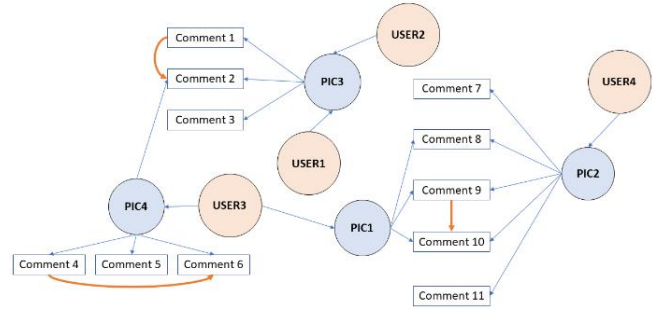


Fig. 3. Heterogeneous Graph that contains users, pictures, comments, and relations between them.

Last, we used a word tokenizer to remove stop words, and then we joined the separated tokens back into one comment. Subsequently, we performed the longest common substring search between the comments of each pair of celebrities, as the following examples.

Comment 1: My codes are mysteries.

Comment 2: My life is also a mystery.

Filtered comment 1: codes **mysteries**

Filtered comment 2: life **mystery**

Example 1: We discovered the longest common substring "mystery" from the two comments.

Comment 1: My codes are mysteries.

Comment 3: But my life is uncertain.

Filtered comment 1: codes mysteries

Filtered comment 3: life uncertain

Example 2: No common substring was found between the two comments.

We then counted the occurrences of each longest common substring in the comments of each celebrity pair. We kept only the substrings that occurred more than once in the search process. These substrings contain contextual information about the relationship and act as hidden nodes that connect the two celebrities. We represent the discovered hidden nodes using heterogeneous relation graphs that consist of users and comments entities as well as the relationship between them.

IV. RESULTS

We present the results of the study in two steps. First, we show how we utilized common followers to identify pairs of celebrities with our collected data. We represent the structural relationship between two celebrities using the relation graph. Second, we present the contextual information that connects two celebrities, namely "Hidden Node." We represent the hidden nodes between the two celebrities with a heterogeneous relation graph.

A. Identifying Celebrities and Their Relationship

We gathered the information from the followee of the random seed of 10 users under the condition that their profile was public. We gathered a list of 2,353 users and collected their names and follower-following information. We then grouped these users from the users whom they followed. To identify the relations between two celebrities, we selected the top 40 users who shared the most common followers. We demonstrated the result in the following heterogeneous graphs.

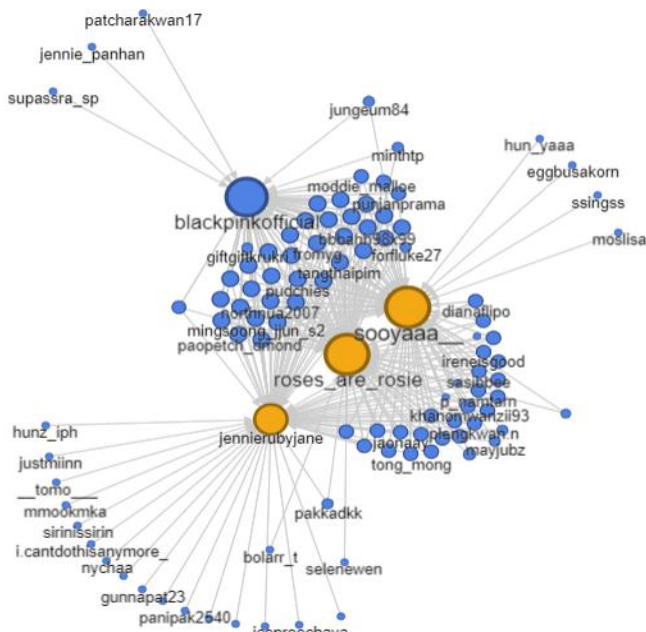


Fig. 4. Celebrities have been in a relationship.

In both graphs, the orange nodes represent celebrities, while blue nodes represent other entities, such as followers and official account.

As in Fig. 4, we observed that a majority of blue nodes had relations with multiple orange nodes. Several blue nodes connected to every orange node. We interpreted that celebrities who shared many common followers were related in some ways. Particularly, the example celebrities are members of the same band, called "Blackpink", as shown in Fig. 4. We also observed the relationship between the official band account and the celebrities. Likewise, the official account and the celebrities shared a considerable number of common followers.

In Fig. 5, there is no apparent pattern between blue and orange nodes. Some blue nodes connect to multiple orange nodes, while several blue nodes only connect to one orange node. Unlike Fig. 4, the percentage of common followers between each pair of celebrities is not apparent. We interpreted that there is no self-evident relationship between

any pair of celebrities.

The result from Fig. 4 and Fig. 5, showed that common followers could be a reliable indicator of the relationship between two users in a directed network, such as Instagram and Twitter. The practical feasibility of this "common followers" indicator can be studied in future works.

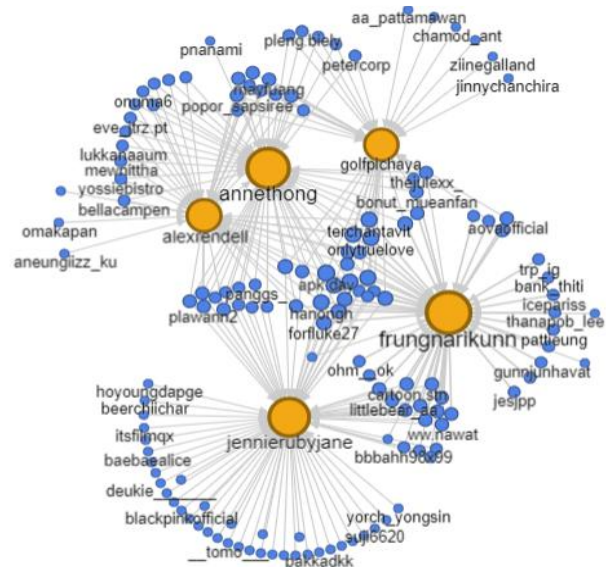


Fig. 5. Celebrities have not been in a relationship.

B. Hidden Node

We scraped additional information from Instagram. We collected users' information about the celebrities whom we interpreted as being in a relation, such as names, post links, captions, user tags, hashtags, and comments.

We used both the hashtags and comments of the celebrities. We then searched for the longest common substrings from the comments. We counted the occurrences of both hashtags and the substrings in the comments of the related celebrities. The most occurrences of hashtags or substrings represented a relationship between celebrities.

As in Table III, the data indicates that both celebrities are in a relationship as close friends. We interpreted from the hashtags and comments' substring, such as "bestfriend," "saranghae," and "#blackpink." We define this information as a "hidden node."

TABLE III: OCCURRENCES OF WORDS IN THE COLLECTED COMMENTS

Words	Count
#blackpink	6
roseiscut	4
saranghae	2
soon	3
sooya	3
sweloveyousoso	4
your	4
bestfriend	2

To represent the weighted relations and nodes with modern visualization tools, we explored the results for the records of user 1, user 2, and hidden nodes (see Table IV).

The relationship between users and hidden nodes is undirected. Hence, we transform the record into only a user-hidden node record, as in Table V.

We represent the hidden nodes and the celebrity, so the nodes are represented. From the collected data, we can

observe multiple networks of celebrities and hidden nodes.

We demonstrate three examples from different networks, including family, music band, and sports team. We demonstrate the results in a heterogeneous graph in Fig. 6-Fig. 8.

TABLE IV: EXPLODED USER-NODE RECORDS

User 1	User 2	Hidden nodes
blackpinkofficial	lalalalisa_m	lisoo
blackpinkofficial	lalalalisa_m	lisoo
blackpinkofficial	lalalalisa_m	lisoo
blackpinkofficial	lalalalisa_m	lisoo
blackpinkofficial	lalalalisa_m	lisoo
blackpinkofficial	lalalalisa_m	lisoo
...
blackpinkofficial	lalalalisa_m	roseiscut

TABLE V: TRANSFORM THE RECORD INTO ONLY A USER-HIDDEN NODE

User	Hidden nodes
blackpinkofficial	lisoo
blackpinkofficial	lisoo
blackpinkofficial	lisoo
lalalalisa_m	lisoo
lalalalisa_m	lisoo
lalalalisa_m	lisoo
...	...
lalalalisa_m	roseiscut

In Fig. 6, the graph represented the relationship between four celebrities, with three of them being brothers: “boy_pakorn”, “cmcamel”, and “hanong”. “Momomama1234” is their mother. The hidden nodes have the meaning related to family, such as #พาลูกหาข้าวกิน, ครอบครัว (family), กบ้าน (home), คุณแม่ถึงกับเวทนา, พี่น้อง (brothers), แม่จ่าย (mom pays), and #พี่บุฟุ้งนี้รับน้อง (brother will pick up our niece tomorrow). Since boy_pakorn is a well-known celebrity, we observe that a majority of the hidden nodes had connected with him. Addition to the family-related nodes, the hidden nodes between brothers have the complimentary meaning, such as หล่อ (handsome), น่ารัก (cute). This also represent some contextual relationship between them. However, we observed that there were still some wrongly separated tokens and duplicated nodes in the relation graph. These issues can be improved in future works by applying a different word segmentation and word occurrences counting method.

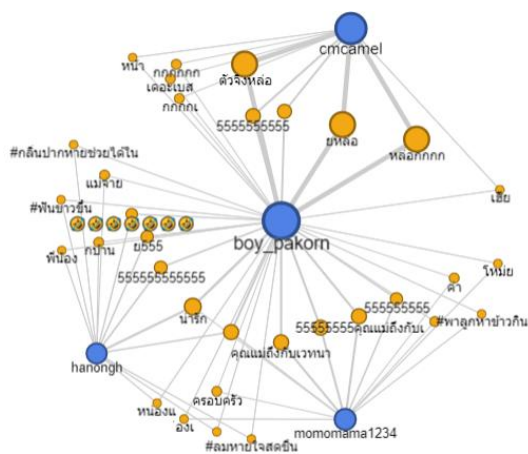


Fig. 6. Members in a family.

Fig. 7 represents the members of a music band, namely

“Blackpink”. The celebrities' usernames are “lalalalisa_m”, “sooyaaa”, “roses_are_rosie”, and “jennierubyjane”. Moreover, the official band username is “blackpinkofficial”. The hidden nodes represent a close friendship between band members, comprising chaelisa, queens, chaesooooo, you, lisoo, sooya, janlisis, love, roseiscut, saranghae, and soon.

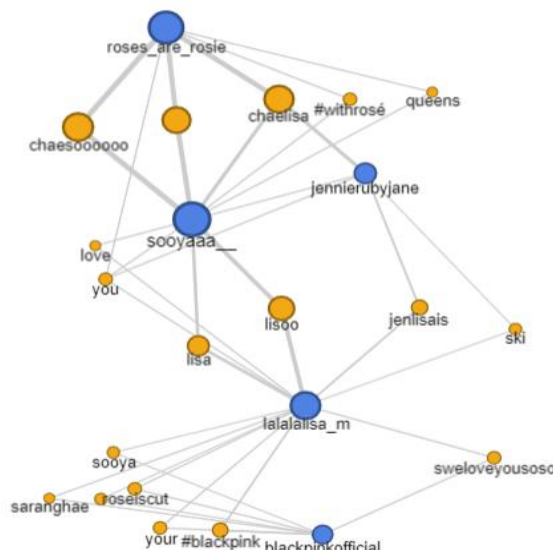


Fig. 7. Members of a band.

In Fig. 8, the graph represents the relationship between the two celebrities, “nootsara13” and “onuma6” who are on the same volleyball team. The hidden nodes have semantic information about sport, such as สู้ (keep on fighting), รัก, ตำนานวอลเล (volleyball legend), and คิดถึง.

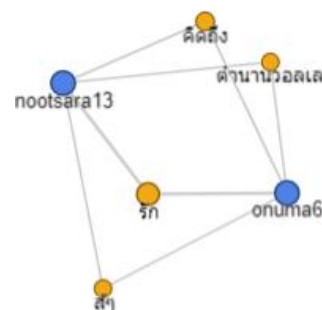


Fig. 8. Members of a volleyball team.

Using these hidden nodes gives us an advantage in many applications, such as providing semantically connected users to the search functionality, recommending unseen relations to new users based on their interests.

Our study obtained both celebrities, identified by common followers, and the hidden nodes that contain contextual information. We can leverage these results by combining them. For example, the two celebrities’ networks that share no common followers can be connected by the semantically equivalent hidden nodes.

We can leverage these results by combining them. For example, the two celebrities’ networks that share no common followers can be connected by the semantically equivalent hidden nodes. In an example case, the members of the Thai women's volleyball team in Fig. 8 are connected with the hidden node “ตำนานวอลเล”, which means volleyball legend. This hidden node can be a connection between the members

of the Thai women's volleyball team and the members of other Thai volleyball teams, even though they might not share the same followers.

V. CONCLUSION

In this paper, we collected data of 2,353 public Instagram users and whom they had followed. We then grouped the followed users by their followers and observed that a pair of users who share a significant number of common followers were related in some ways. On the other hand, a pair of users who had no self-evident relationship did not share a notable number of common followers. From this observation, we concluded that structural information, such as a common follower, can be used to identify the relationship between two users in a directed network. However, structural information alone cannot explain the context of the relationship.

In this paper, we utilized contextual information by extracting users' comments and tags from Instagram posts of the two users who share common followers. The comments and tags were split into words. We then counted the occurrences of each word or its substring within the comments and tags. We observed that words with the top occurrences could represent the context of the relationship between two users. These words were considered the "hidden nodes" in the relationship graph.

We represent the hidden nodes in a relationship graph of Instagram users. These hidden nodes can be used to improve the efficiency of recommendations on Instagram. They can help improving the suggestions of media and users that are related based on their interests. For example, users who interested in Thai volleyball begin to follow the team members on Instagram. In this paper, we can indicate the nodes that imply the relationship among Thai volleyball team members, for examples "volley" and "lovevolley" are the hidden nodes that we can find. Both keywords won't represent only Thai volleyball team members but may represent the relations between other teams from other countries that punished one or more keywords in the image captions or comments on their social media.

One direction for future work is to compare the proposed approach with other variations of methodology parameters, such as using term frequency-inverse document frequency (TF-IDF) instead of word occurrences to measure the importance of each word proportionally to its occurrences across multiple documents. To evaluate the performance of each approach, we can perform a survey on network experts who are knowledgeable in each network on how appropriate the hidden nodes are. To speed up the evaluation process, Amazon Mechanical Turks should be considered.

CONFLICT OF INTEREST

We declare no conflict of interest.

AUTHOR CONTRIBUTIONS

U.U. conceived and carried out the experiments, analyzed data and wrote the manuscript. S.S. supervised the research, provided critical feedback and helped shape the research, analysis and the final version of the manuscript.

REFERENCES

- [1] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," presented at the Proceedings of the twelfth international conference on Information and knowledge management, New Orleans, LA, USA, 2003.
- [2] D. Schall, "Link prediction in directed social networks," *Social Network Analysis and Mining*, vol. 4, pp. 1-14, 2014.
- [3] E. Butun, M. Kaya, and R. Alhaji, "A new topological metric for link prediction in directed, weighted and temporal networks," in *Proc. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016, pp. 954-959.
- [4] B. Chen, Y. Hua, Y. C. Yuan, and Y. Jin, "Link prediction on directed networks based on AUC optimization," *IEEE Access*, vol. 6, pp. 28122-28136, 2018.
- [5] A. Anderson, D. P. Huttenlocher, J. M. Kleinberg, and J. Leskovec, "Effects of user similarity in social media," *WSDM*, 2012.
- [6] P. Bhattacharyya, A. Garg, and S. F. Wu, "Analysis of user keyword similarity in online social networks," *Social Network Analysis and Mining*, vol. 1, pp. 143-158, 2010.
- [7] C. G. Akcora, B. Carminati, and E. Ferrari, "User similarities on social networks," *Soc. Netw. Anal. Min.*, vol. 3, no. 3, pp. 475-495, 2013.
- [8] Z. Wang, J. Liang, and R. Li, "Exploiting user-to-user topic inclusion degree for link prediction in social-information networks," *Expert Syst. Appl.*, vol. 108, pp. 143-158, 2018.
- [9] C. P. M. T. Muniz, R. R. Goldschmidt, and R. Choren, "Combining contextual, temporal and topological information for unsupervised link prediction in social networks," *Knowl.-Based Syst.*, vol. 156, pp. 129-137, 2018.
- [10] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Physical review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 64, no. 2, p. 025102, 2001.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Unmanee Unmuang was born in Bangkok, Thailand in 1993. She received a bachelor's degree in computer science from the Department of Science, Kasetsart University and is currently studying for a master's degree in computer science at the Department of Computer Engineering, Chulalongkorn University. Her research interests include social network analysis, social network mining, data and web mining, and information retrieval.



Sukree Sinthupinyo was born in Bangkok, Thailand in 1975. He received bachelor's, master's and doctoral degrees in computer engineering from the Department of Computer Engineering, Chulalongkorn University. His research interests include artificial intelligence, machine learning, innovation social network analysis, and social network mining.