Mobile Social Big Data: WeChat Moments Dataset, Network Applications, and Opportunities

Yuanxing Zhang, Zhuqi Li, Chengliang Gao, Kaigui Bian, Lingyang Song, Shaoling Dong†, Xiaoming Li
Peking University, Beijing, China
†Fibonacci Data Consulting Services Inc, Shenzhen, China
{longo, lizhuqi, gaochengliang, bkg, lingyang.song, lxm}@pku.edu.cn, †dong@szzbmy.com

Abstract

In parallel to the increase of various mobile technologies, the mobile social network (MSN) service has brought us into an era of mobile social big data, where people are creating new social data every second and everywhere. It is of vital importance for businesses, government, and institutes to understand how peoples’ behaviors in the online cyberspace can affect the underlying computer network, or their offline behaviors at large. To study this problem, we collect a dataset from WeChat Moments, called WeChatNet, which involves 25,133,330 WeChat users with 246,369,415 records of link reposting on their pages. We revisit three network applications based on the data analytics over WeChatNet, i.e., the information dissemination in mobile cellular networks, the network traffic prediction in backbone networks, and the mobile population distribution projection. Meanwhile, we discuss the potential research opportunities for developing new applications using the released dataset.

I. INTRODUCTION

Owing to the furious rate at which proprietary mobile technologies and networks evolve, the mobile social network (MSN) service arises with a new era of mobile social big data, where users could conveniently converse and connect with others, and create new social data every second, everywhere through their mobile devices [1]. Besides, compared with the web-based social network, MSNs provide more opportunities on dedicated use in mobile and wireless networks, e.g., augmented reality, mobile communication, location-based services.

A. A taxonomy of mobile social network services

There are three broad categories of social network services.
1) **One-way following on MSN among strangers**: On Twitter or Weibo, following someone is not mutual. One can follow you without your approval (by default), and you do not have to follow him/her back. For example, a celebrity may have millions of Twitter or Weibo followers, most of whom he/she does not know in real life.

2) **Mutual-following on MSN between friends**: On Facebook, two friends mutually follow each other if one approves the other’s friend request. Meanwhile, Facebook also allows users to have one-way following of a page set up by businesses, organizations, brands that allow everybody to follow.

3) **Instant messenger**: In WhatsApp, WeChat\(^1\) or Line, two acquaintances can send instant messages to each other, if one approves the other’s friend request, instead of using the conventional SMS.

Recent reports indicate that many social network services have become a hybrid of the above three\(^2\).

**B. Messenger-based mobile social network: WeChat Moments**

The interpersonal tie in social networks, a.k.a. the social tie, is defined as the information-carrying connection among social circles \(^2\), within which they interact and exchange various kinds of information. The strength of the social-tie between two people can be evaluated by how frequently they communicate with each other via any online and/or offline channel.

People with strong social ties may want to communicate with each other at a higher frequency, via a messenger app. This is why many MSN services have incubated their own messenger app, such as Facebook Messenger, direct message on Twitter. Messenger apps have also started providing social-networking services. For instance, in WeChat, friends not only send instant messages to each other, but also have the access to each other’s page, i.e., the page of “WeChat Moments” (WM), or friend circle page. Therefore, we call this kind of social network as the **messenger-based mobile social network**. WM not only takes the advantage of web-based social network services, but also adds the following new features that may delight users.

**Keep strong social tie—no access to strangers’ page on WM**: The MSN and messenger apps have defined different access policies to one’s page, which affects the information diffusion

---

\(^1\) WeChat is a popular mobile messenger app, and the “WeChat Moments” (a.k.a. the friend circle) provides WeChat users with the social-networking services like viewing and posting texts/photos/links on their pages.

\(^2\)http://www.kpcb.com/blog/2016-internet-trends-report
process in various ways. On Twitter, one’s page is open for anyone’s access (by default). Besides, the path of retweeting is visible to everyone, and it is easy to follow anyone over the retweeting path. In WM, two users cannot see each other’s page if they are not connected as friends, and there is no visible path of retweeting (reposting) a post, and anyone has no idea about where the post is from.

Selected content display—private content displayed to selected friends: In certain scenarios over WM, there exists some content that would be better to be exposed only to selected close friends, which should not be viewed and reposted by other friends (note that there is no access to strangers’ page). This kind of private content is only shared among selected close friends, which further strengthens the social tie among these friends.

Group chat—a way of approaching the unfamiliars: As an instant messenger, WeChat provides the group chat function for strangers to converse, as every user can join a group by clicking a group link, scanning a QR code, or being invited by an existing group member, and two members in the same group do not have to be friends. Users could send the HTML5 pages into the group, which has become a popular way to propagandize online/offline activities.

C. New challenges in network applications

It is of vital importance for businesses, government, and institutes to understand how peoples’ online behaviors in the MSN can affect the underlying computer network, or their offline behaviors at large. However, new challenges may arise when we revisit the following problems, because the access control policy (e.g., no access to strangers’ page, private content) in WM confines the information diffusion among “acquaintances”, which reduces the chance that a post is exposed to strangers as well as the probability of reposting.

- **Information dissemination in mobile cellular networks**: Information dissemination from one user to others in mobile cellular networks depends on the reliability of connections among cellular users. It is a challenging problem as we have no idea about the quality of connections among them. Meanwhile, we observe that information diffusion over the MSN receives a high probability of success, as MSN services can easily stimulate users to share information, send messages with their loved ones, while on-the-go. Hence, it will be interesting to investigate whether there exist influential users that have high-quality connections to help disseminate information in cellular networks.
• **Backbone network traffic prediction**: Offline locations of MSN users can be predicted by mining their periodic behaviors. Human movement and mobility patterns have a high degree of freedom and variation, but they can still exhibit structural patterns due to geographical and social constraints [3]. Accordingly, user migration may cause the change in the backbone network traffic distribution—the location that have a high density of mobile users should be allocated with more backbone network resources. The traffic distribution of backbone network not only depends on the users’ movement, but also correlates to the frequency of they repost from friends. The online interaction frequency is an important indicator of the social tie among users in WM, which requires a fine-grained analysis for better allocating resources in the underlying backbone network.

• **Projecting mobile population distribution**: The online behaviors of users (e.g. reposting a link) may reflect certain geographical attributes. We could then conjecture that there lies the geo-homophily between an MSN and an offline mobile network. When large-scale migration (during Spring Festival in China, or winter/summer vacations) happens, some changes on the structure of the messenger-based social network could be monitored. Therefore, it is important for businesses to predict how the mobile population distribution varies, in order to deploy appropriate regional marketing strategies, or provide personalized recommendation for users at home, at workplace, or on travel.

In this article, we collect a dataset from WeChat Moments, called WeChatNet [4], which involves 25,133,330 WeChat users with 246,369,415 records of link reposting on their pages, from January 14 to February 27, 2016. This is the first released big dataset of WeChat Moments on users’ reposting behaviors. We revisit three network applications based on the data analytics over WeChatNet. We first present a voting strategy that finds the most influential users for information dissemination in mobile cellular networks. By observing the interaction between friends in time and spatial domains, we predict the traffic load in the underlying backbone network with a prediction accuracy rate of over 90%, which yields a near optimal resource allocation (i.e., the server placement in the underlying backbone network). Based on the location of users who view and repost a link in WM, we propose a model to project the distribution of the floating population. We further discuss the potential research opportunities for developing new applications using the WeChatNet dataset we release.
II. Dataset Collection and Relevant Network Applications

A. The WeChatNet dataset

As messenger-based social networks are usually developed on mobile devices, conventional web pages may not provide good visual experience. HTML5 is a promising way to adapt to different screen sizes of the mobile devices. WeChat provides several interfaces officially to help developers designing HTML5 pages. Users can then easily get access to the page content, and repost those interesting pages to their friends through the MSN like WM.

In the WM network, the links shared/posted by users usually lead to a post in HTML5 (H5). Such WM posts provide users with interactive operations such as an online greeting card, a lightweight online game (e.g., flappy birds), psychological test, etc. A WM post can be released by the WM service provider (Tencent), or a third-party web developer.

**WM data collection:** Our goal is to collect the statistics of WM post diffusions. We use the Application Programming Interface (API) provided by a business WM page creator platform, FIBODATA\(^3\), for crawling diffusion trails of pages created over the platform. Based on the collected data, we are able to construct a diffusion graph for each posted WM page. The dataset contains about 320,000 pages created by businesses from from January 14, 2016 to February 27, 2016, which involves 25,133,330 Wechat users with 246,369,415 link retweeting records\(^4\).

**Information diffusion process:** Suppose that user \(i\) and \(j\) are friends in the WM network. When user \(i\) shares the link of a WM post with his friends, user \(j\) may click the shared link to view the content of this post. If user \(j\) finds this post interesting, he may further repost the link to his friends, so that more users would have the chance to view this WM post. As this process is similar to the spread of infection,

- we call a user an *infected* user of a WM post if he views the post;
- we call a user an *infectious* user of a WM post if he views the post and reposts the post link.

**Post view record:** A post view record in our dataset is a 5-tuple in the following format:

\[
< U_1, U_2, PID, IP, t >,
\]

\(^3\)http://www.fibodata.com/

\(^4\)We release a WM dataset over https://github.com/pkumobile/WMdata. This dataset will be updated by crawling more recent data via FIBODATA.
where $U_1$ is the ID of the user whose post gets viewed; $U_2$ is the ID of the user who views the post by user $U_1$; $PID$ is the ID of the WM post that is assigned by the WM post creator platform; $IP$ the IP address of the post viewer, i.e., user $U_2$'s IP address; and $t$ is the time when the post view happens. The whole tuple records the post view event that user $U_2$ at the address $IP$ at time $t$ views post $PID$ of user $U_1$.

B. New features of WM

In Table I, we compare five most representative social network services and instant messengers. We can clearly see that Twitter and Weibo serve mostly as social networks, while WhatsApp serve mostly as instant messenger; Facebook (Messenger) and WeChat has combined both social network and instant messengers. WeChat Moments appears as a messenger-based social network with many noteworthy features.

First, WM takes the advantages of the influential users (account) in the social network, and proposes Subscription Accounts, where companies or individuals could edit web pages and push to the followers. This function has been proved to have great impact on attracting users’ attention.

Second, WM inherits the benefits of strong social tie of the instant messenger. The relationship between users in instant messenger (WeChat) is bidirectional rather than one-way following (Weibo). The relationship is also private, which means that strangers have no access to the information of certain users if they are not friends (i.e., no access to stranger’s pages). Therefore, the retweeting path of any content would not be recorded. This could be a double-edged sword for commercials, for instance, the advertisement may profit from the communication over strongly-tied friends, while the advertisement may have the difficulty of being spread widely and the business cannot obtain a long reposting path in most cases.

Third, WeChat exploits the superiority of group chat. The HTML5 page with abstract content instead of the URL address can be posted into the group chatting window, making it easier for users to acquire information.

Forth, WeChat is also unique on its connection to offline services, like ordering food delivery, calling a cab, obtaining coupons by scanning QR-code, etc.

C. Network applications using the WeChatNet dataset

In this article, we only study the impact of online reposting behaviors of users over the underlying computer networks, and we showcase three typical network applications using the
| Service category         | Service name                      | Twitter | Weibo | Facebook | WeChat | WhatsApp |
|-------------------------|-----------------------------------|---------|-------|----------|--------|----------|
| Social networking services | Video sharing                     | ✓       | ✓     | ✓        | ✓      | ×        |
|                         | Personal page                     | ✓       | ✓     | ✓        | ✓      | ×        |
|                         | Post search                       | ✓       | ✓     | ✓        | ✓      | ×        |
|                         | Favorite post save                | ×       | ×     | ✓        | ✓      | ×        |
|                         | Post App                          | ✓       | ✓     | ✓        | ✓      | ×        |
|                         | Access to pages of non-friends    | ✓       | ✓     | ✓        | ✓      | ×        |
| Messenger services      | Limit on # of followers/friends   | no limit| no limit| 5000     | 5000   | no limit |
|                         | Video/audio chat                  | ×       | ×     | ✓        | ✓      | ✓        |
|                         | Group messaging                   | ×       | ✓     | ✓        | ✓      | ✓        |
|                         | Sending location                  | ×       | ×     | ✓        | ✓      | ✓        |
|                         | Voice messaging                   | ×       | ×     | ✓        | ✓      | ✓        |
| Misc. services          | Mobile payment                    | ×       | ×     | ✓        | ✓      | ×        |
|                         | Video games                       | ×       | ✓     | ✓        | ✓      | ×        |
|                         | Offline services (taxi, ticket, etc.) | ×     | ×     | ✓        | ✓      | ×        |
|                         | Shopping                          | ×       | ✓     | ✓        | ✓      | ×        |

TABLE I: Comparison between different Twitter, Weibo, Facebook, WeChat, and WhatsApp. Statistics in the table are collected as of July 31, 2017.

WeChatNet dataset in Fig. 1. First, the reposting behaviors of users can form the information diffusion graph, where we could analyze the social influence over the network connections among users, which help the information dissemination in mobile cellular networks. Second, we have the access to the origin and the destination of every reposting record, which could be leveraged to predict the backbone network traffic. Third, statistics on the geographical information of users provide us with an opportunity of projecting the distribution of mobile population. We will discuss these three applications in details in the following sections.

III. INFORMATION DISSEMINATION IN MOBILE CELLULAR NETWORKS

A. Selection of the most influential user

The information dissemination process in mobile cellular network depends on the reliability of connections among cellular users. The more high-quality connections one has, the more-widely he/she can spread the information. Hence, we are interested in finding the most influential user that can best help disseminate the information in mobile cellular networks.

In the same geographical community, the connection graph of mobile cellular users can be built on basis of the information diffusion graph in mobile social network. Hence, we can transform
the problem of selecting the most influential user in the mobile cellular networks to that of finding key opinion leaders (KOLs) in mobile social networks.

B. Implementation by the voting strategy

The KOL with millions of followers (or a large number of friends) shows a strong social tie in the MSN, and he may have the power to help spread the information for online marketing/advertising. A user with millions of followers on Twitter or Weibo should be a KOL. However in WM, it is challenging to detect a KOL by counting the number of his/her friends, because most users in WM only have a limited number of friends, and WM has imposed a maximum number
of friends that one can connect to. Even a celebrity may only have approximately hundreds of friends in WM.

Instead of counting the number of his/her friends, we present a voting-based strategy that selects influential nodes by looking at the user’s local contribution to the information diffusion process [5], which is different from existing greedy solutions using Lazy Evaluation strategies that tend to seek nodes with a large marginal influence. The voting-based strategy works as follow. First, it tries to identify those major diffusion trees for the information diffusion process. Then, in each diffusion tree, each offspring node can vote for its father nodes as the WM page is diffused from the father nodes to the offspring node. After repeating these steps multiple times, the number of votes received by a node in the diffusion tree indicates the influence of the node in the network.

In WM, the size of the diffusion graph/tree could be very large, and we cannot enumerate all diffusion trees or nodes for executing the voting operations [5]. Hence, we leverage Gibbs Sampling technique to complete the above voting process within a limited time period. There are two parameters in this strategy, simply speaking, we need to sample a number of $R_1$ diffusion trees; for each selected tree, we execute a number of $R_2$ voting operations. Let $S$ denote the set of $K$ influential users that we find by using the voting strategy, and let $\sigma(S)$ denote the calculated influence of the set $S$.

We use the data on Jan. 14 2016 for performance evaluation. Figure 2a shows the influence value of $S$ under the voting strategy by varying $R_1$, while the other parameters are set as $K = 100, R_2 = 100000$. The green line is the performance of conventional greedy algorithm when $K = 100$. We observe that as $R_1$ grows, the variance of $\sigma(S)$ starts decreasing until it reaches a stable value. The influence value of $S$ under the voting strategy is greater than that under the greedy one, when $R_1$ is greater than 200.

Given $K = 100, R_1 = 500$, Figure 2b shows the influence value of $S$ under the voting strategy by varying $R_2$ from $10^9$ to $10^6$. The $\sigma(S)$ raises as $R_2$ increasing, while the variance of it implies that the enumeration is quite random if $R_2$ is small; otherwise a larger $R_2$ will make the $\sigma(S)$ more stable. Similar to results in Figure 2a, the influence value of $S$ reaches a stable value under the voting strategy, which is greater than that under the greedy one when $\log_{10}(R_2)$ is greater than four.

These results indicate that choosing the users with the greatest degree as KOLs will mis-determine the set of influential users in the WM network.
IV. PROFILING TRAFFIC DISTRIBUTION IN BACKBONE NETWORKS

By observing the interaction between friends in time and spatial domains, we are able to predict the traffic load in the underlying backbone network, and present a reverse greedy strategy to better place servers and reduce the traffic load in the network.

A. From human migration to traffic prediction

We model the human mobility pattern as a Markov random field and compute the communication distance between two users. With communication distance and communication frequency, we can predict the traffic generated by an online WM in the offline backbone network.

To validate the model, we utilize the data of first 19 days as the training set to train the model. The following five days are used for testing the performance of the prediction approach. Figures 3a and 3b show the real/predicted communication distance, and the real/predicted traffic load, where we observe that the model could do the prediction quite well, with an error rate lower than 10%.

B. Traffic optimization and server placement

We then design a heuristic approach, called the reverse greedy strategy, to ease the traffic load generated by the WM. The goal of traffic optimization is to reallocate the network resources (the servers) at proper locations to shorten the communication distance between users. Intuitively,
(a) The predicted average communication distance.

(b) The predicted traffic load.

(c) Location selection strategy performance on China Unicom backbone.

(d) Location selection strategy performance on China Telecom backbone.

(e) Results of the optimal strategy for placing 5 servers.

(f) Results of the reverse greedy strategy for placing 5 servers.

(g) Result of the naive greedy algorithm for placing 5 servers.

Fig. 3: Evaluation of the traffic optimization model.
we can select the location that can bring the most reduction in the generated traffic load in a greedy manner.

We use the backbone network graph of China Unicom and China Telecom to emulate the performance of the reverse greedy strategy. We enumerate all the possible combinations of server placement in the graph, calculate the resulting traffic load, and choose the case with the least load as the “optimal” solution. Figures 3c and 3d show that the reverse greedy strategy can achieve the near optimal result in the two backbone network graphs, while the naive greedy strategy performs inferiorly. In addition, the curves for the optimal and the reverse greedy strategy become flat gradually after placing 5 servers, which implies that replacing 5 servers are sufficient to reduce most of the traffic. We also visually display the results of server replacement of the two greedy strategies and the optimal solution in Figures 3e, 3f, and 3g, where the result of the reverse greedy strategy is quite similar to that of the optimal solution, and the naive greedy strategy fails to consider the nation-wide replacement of servers.

V. Projecting Mobile Population Distribution

Projecting the population distribution in geographical regions is important for many applications such as launching marketing campaigns or enhancing the public safety in certain densely-populated areas. Conventional studies require the collection of people’s trajectory data through offline means, which is limited in terms of cost and data availability. The wide use of MSN apps over smartphones has provided the opportunities of devising a lightweight approach of conducting the study using the online data of smartphone apps [6].

A. Modeling geo-homophily

A division of geographical regions is stable only if the MSN users in these divided regions show a strong geo-homophily—people in each region prefer communicating with others in the same region more than those in other regions. These inspire us to investigate the relationship between the online information diffusion, i.e., users’ communication in MSN, and the population distribution over a fixed division of offline regions. Intuitively, we can use the geo-location of messages among MSN users to derive the user distribution over the given regions. Then, the floating population across regions can be further inferred based on the derived distribution, which could be explained by the Dirichlet Process Mixture.
B. Geo-homophily in WM

The WM dataset records the page re-tweeting in 34 provinces in China, and we use these provinces as the geographical regions in this experiment. Every user in WM should have viewed a collection of pages, and each page view’s IP address corresponds to a province, among which the most frequently recorded one is set as the province where the user is located. We analyze the message diffusion process in two time periods: (1) Before Spring Festival, we monitor the message diffusion from Jan. 14 to Jan. 31, 2016, which are pre-holiday working and weekend days; (2) on the Spring Festival day, most people stay at home, and hence the structure of the message diffusion graph would be different.

Figure 4a shows the volume of message diffusion inside each province and that between every pair of provinces of China, where the amount of message diffusion inside a province is proportional to the size of the corresponding circle, and the amount between provinces are represented by the length of arcs.

The results indicate that most of the diffusions occur inside provinces, so the arcs are relatively sparse. In particular, there lies no arc between some pairs of circles in this figure, which does not mean that there is no message diffusion between the corresponding two provinces, but implies that the message diffusion between them is much weaker than that between those pairs of circles having arcs. For example, there were only hundreds of message diffusions between Tibet and
Taiwan in the dataset; in contrast, several millions of message diffusions occur between Beijing and Guangdong. This can be explained by the fact that those provinces are at distant locations, or they have little communication with most provinces in China mainland.

On the Spring Festival holiday, most people stay with their families in their home province. The graph structure changes, as most of the messages are sent for appointments and greetings, and these diffusions mainly took place between friends in the same vicinity. Thus, the proportion of the diffusion inside regions increases. The graph structure is illustrated in Figure 4b where some inter-province arcs disappear.

Then we obtain a chaotic segregation, which can hardly be said to have any geo-homophily. The amount of message diffusions inside and across these regions are shown in Figure 4c. Compared to Figure 4a under the same scale of plotting, the distribution of circles representing regions are very dense; and most of circles’ sizes are similarly small.

Fig. 5: (a) The viewing distribution of a hot message originated from Beijing; (b) Results of inferring the floating population (FP) that has excluded those whose home and remote regions are the same.

When considering the diffusion of a single page, we find that it will be reposted many times in the home region of the sender, while it may be sent to only a few non-home regions—those diffusions across regions only take up a small part. For example, we illustrate the distribution of views to a popular page with approximately one million views in Figure 5a, where the page is originally sent from the region of Beijing.

We then evaluate the performance of Dirichlet Process on WeChat dataset, and compare it
against the results of the latest national population census in China [5] which provides us the statistics of floating population in China. Here, the floating population (FP) in our experiments has excluded those whose home and remote regions are the same (e.g., those who rarely move out of the home region, as the work place and the home belong to the same region). As shown in Figure [5b] there lies a linear correlation between the real FP and the predicted population by the Dirichlet Process.

VI. Future Opportunities

Structure of new social networks: We have discussed that the structure of messenger-based social network (WM) is different from that of the blog-like social network. It would be interesting to investigate in more depth about the reasons why WM becomes different from other social networks. The data of WM provides new opportunities to explore the evolution of this new type of social network. For instance, the network representation of WM can recognize the similarity between users and predict the potential links [7] in the social network that may not be easily obtained in the diffusion graph.

Marketing in dynamic diffusion graph: Marketing is the most prevalent business model in online social network services. Our previous works have analyzed the results in the static diffusion graph. However, in a longer time period, the role or the influence of users in the social network may change, and it would be interesting to study the social tie between friends in a dynamic diffusion graph [8]. Moreover, the analysis on the effects of extrinsic rewards (e.g., the advertisement viewing duration), and on the friendsourcing [9] could contribute to the marketing strategy as well.

Understanding more demographic results: WM has been proven to have the ability to profile traffic distribution and to project population distribution, and indeed, the dataset can be used to understand more demographic results by observing the interaction among people, such as how to identify a user in the social network with incomplete knowledge about his friends [10].

Spam detection: The MSN facilitates the ubiquitous access to news/messages for everybody. However, rumours and cyber-violence are also easy to spread, breaking the right of legal citizens. We could analyze the network elements, i.e., the influence and vulnerability of users (by resembling mean opinion score) in the group and the confidence of links (by resembling stability), to help detect seditious news/messages [11], and prevent them from spreading.

5http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm
Promoting offline marketing activities: Businesses have seen WeChat as an important offline marketing tool, as HTML5 pages redirected by scanning QR-codes can provide more information to consumers. Businesses have to determine the qualification of treating certain potential consumers as a group by observing the stability and the activity of the group. This could help them decide whether to launch marketing in a certain area during a specific time period. Device-to-device and vehicular communication technologies are found to be efficient in sharing information among users [12], [13], which can greatly extends information diffusion in WM via either online or offline channels.

VII. Conclusion

In this article, we present a systematic study to understand how the online reposting behaviors of users in WeChat Moments will affect the computer networks and their offline behaviors. From the perspective of information diffusion, we present an online KOL detection method that is independent of the number of one’s friends. By observing the interaction between friends, we predict the traffic load in the underlying backbone network with a prediction accuracy rate of over 90%, which leads to a near optimal resource allocation. We also find that the Dirichlet Process Mixture can describe the distribution of the floating population, based on which we propose a model using the online message diffusion to project the offline population distribution. Moreover, we point out the future research opportunities of using the dataset we release in understanding the impact of online social tie over the offline social, business, and political lives.

REFERENCES

[1] S. Ioannidis, A. Chaintreau, and L. Massoulie, “Optimal and scalable distribution of content updates over a mobile social network,” in *INFOCOM 2009, IEEE*. IEEE, 2009, pp. 1422–1430.
[2] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, “Structure and tie strengths in mobile communication networks,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7332–7336, 2007.
[3] E. Cho, S. A. Myers, and J. Leskovec, “Friendship and mobility: user movement in location-based social networks,” in *KDD*. ACM, 2011.
[4] WeChatNet, “A Mobile Social Big Dataset of WeChat Moments.” [https://github.com/pkumobile/WMdata](https://github.com/pkumobile/WMdata)
[5] Y. Zhang, Y. Bai, L. Chen, K. Bian, and X. Li, “Influence maximization in messenger-based social networks,” in *Global Communications Conference (GLOBECOM), 2016 IEEE*. IEEE, 2016, pp. 1–6.
[6] Y. Zhang, Z. Li, K. Bian, Y. Bai, Z. Yang, and X. Li, “Population distribution projection by modeling geo homophily in online social networks,” in *Proceedings of the 2nd International Conference on Crowd Science and Engineering (ICCSE’17)*. ACM, 2017, pp. 1–8.
[7] Z. Wang, C. Chen, and W. Li, “Predictive network representation learning for link prediction,” in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2017, pp. 969–972.

[8] Y. Yang, X. Mao, J. Pei, and X. He, “Continuous influence maximization: What discounts should we offer to social network users?” in *Proceedings of the 2016 International Conference on Management of Data*. ACM, 2016, pp. 727–741.

[9] H. Zhu, S. Das, Y. Cao, S. Yu, A. Kittur, and R. Kraut, “A market in your social network: The effects of extrinsic rewards on crowdsourcing and relationships,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 598–609.

[10] M. Kearns, A. Roth, Z. S. Wu, and G. Yaroslavtsev, “Private algorithms for the protected in social network search,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 4, pp. 913–918, 2016.

[11] J. Sampson, F. Morstatter, L. Wu, and H. Liu, “Leveraging the implicit structure within social media for emergent rumor detection,” in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2016, pp. 2377–2382.

[12] J. Liu, S. Zhang, N. Kato, H. Ujikawa, and K. Suzuki, “Device-to-device communications for enhancing quality of experience in software defined multi-tier LTE-A networks,” *IEEE Network*, vol. 29, no. 4, pp. 46–52, 2015.

[13] H. Nishiyama, T. Ngo, S. Oiyama, and N. Kato, “Relay by smart device: Innovative communications for efficient information sharing among vehicles and pedestrians,” *IEEE Vehicular Technology Magazine*, vol. 10, no. 4, pp. 54–62, 2015.