Friend Network as Gatekeeper: A Study of WeChat Users’ Consumption of Friend-Curated Contents

Quan Li†, Zhenhui Peng‡, Haipeng Zeng†, Qiaoan Chen†, Lingling Yi†, Ziming Wu‡, Xiaojuan Ma‡, and Tianjian Chen†

† AI Group, WeBank, Shenzhen, Guangdong, China
‡ The Hong Kong University of Science and Technology, Hong Kong

{forrestli, tobychen}@webank.com, {zpengab, hzengac, zwual}@connect.ust.hk,
{kazechen, chrisyi}@tencent.com, mxj@cse.ust.hk

ABSTRACT
Social media enables users to publish, disseminate, and access information easily. The downside is that it has fewer gatekeepers of what content is allowed to enter public circulation than the traditional media. In this paper, we present preliminary empirical findings from WeChat, a popular messaging app of the Chinese, indicating that social media users leverage their friend networks collectively as latent, dynamic gatekeepers for content consumption. Taking a mixed-methods approach, we analyze over seven million users’ information consumption behaviors on WeChat and conduct an online survey of 216 users. Both quantitative and qualitative evidence suggests that friend network indeed acts as a gatekeeper in social media. Shifting from what should be produced that gatekeepers used to decide, friend network helps separate the worthy from the unworthy for individual information consumption, and its structure and dynamics that play an important role in gatekeeping may inspire the future design of socio-technical systems.

Author Keywords
Friend-curated content; information consumption; gatekeeper

CCS Concepts

- Human-centered computing → Empirical studies in collaborative and social computing; Social media;

INTRODUCTION
Social media services such as Twitter, Facebook, and WeChat† empower millions of users to consume content from and disseminate information to their social counterparts. For example, WeChat, one of the most popular friend-based social media services in China, generates and circulates over 1.5 million articles in the form of embedded posts or external links [23, 31]. Given the abundant context on social media, users face challenges of identifying authentic and high-quality information [38]. Take WeChat as an example, some public accounts on it mainly publish content that can be easily monetized to grab the audience’s eyeballs but lacks substance [49].

One way to safeguard the integrity of information for users is developing algorithms to remove fake news and promote high-quality ones [12, 40]. However, these algorithms could not present precisely what a specific user is interested in the current stage. As an another means, many social media services offer the “share” features for users to spread information in their social circles. These users are called “gatekeepers”, who pass along and comment on already available news items based on their interests [25, 35]. Previous research on social media has shown how the gatekeepers on Twitter affect the audiences’ information selection during a special event (e.g., 2009 Israel-Gaza conflict [25]), and how users play roles of the gatekeepers to control information flows on Reddit [26]. However, unlike Twitter and Reddit in which gatekeepers often have no close relationship with users (e.g., a famous star as the gatekeeper), WeChat builds friend-based social media in which a user ideally knows all members in his/her circle. Bakhshy et al. demonstrated that friends can expose individuals to cross-cutting content on Facebook [5]. Nevertheless, WeChat does not have algorithmically ranked News Feed as Facebook does, but present contents shared by friends in an ordered timeline manner. Moreover, most of the WeChat users are from China, and they have a different cultural background than the users of Facebook who mainly come from western countries. There is a lack of understanding of how the friend network acting as a gatekeeper on WeChat affects the users’ content curation behaviors. Such an understanding is important as it can not only help the content creators to learn how their works are spread in the friend-based social network but also facilitate such social media platforms to manage the information flow.

In this paper, we use WeChat as a lens to investigate how users leverage their friend networks as latent gatekeepers for content curations on the friend-based social media. Specifically, we reveal how WeChat users exploit the composition and tie strength of their friend network to safeguard the relevance, importance, popularity, and/or quality of information they consume. We further examine how users adopt the gatekeeping mechanism according to the changes of friend networks and interests over an extended period of time. We take a mixed-methods approach [45] to study the possible “friend network as a latent gatekeeper” phenomenon on WeChat. On the one hand, we quantitatively analyze over seven million WeChat

---

†http://www.wechat.com/en/
users’ reading behaviors and infer how these users accommodate and safeguard their varying information needs through different friend communities and social ties. On the other hand, we conduct an online survey with 216 participants to qualitatively understand how and why users view the gatekeepers in their networks. In general, we find that users tend to turn to the friend network for information consumption if there is overloading information. They tend to exploit weak-ties getting exposed to new domains and turn to strong-ties when demanding credible and reliable information. Elder users with shorter WeChat experiences and fewer friends depend primarily on their friend network for information consumption. Users leverage social circles to gatekeep information interests and the interests and attention paid to them curated by one social circle can shift to another circle. The major contributions of this paper are as follows:

- We qualitatively and quantitatively study the potential phenomenon of WeChat users leveraging their friend network as a collective and dynamic latent gatekeeper.
- We discuss the insights derived from our approach to inspire the future design of socio-technical systems.

BACKGROUND AND RELATED WORK

Gatekeeping in Social Media Era

Unlike traditional media, today’s social media can be indelibly remarked as We Media [10], i.e., user-operated media, in which there is no clear boundary between information producers, disseminators, and consumers, and the contents published are no longer constrained by length, timeliness, and the relevance to readers in a geographical and cultural sense [3, 14, 34]. Although social media users play an active role in shaping the online information landscape [41], they might not have the same level of professional qualities as the experts in the media industry for “gatekeeping”, i.e., scrutinizing content, safeguard its validity, veracity, and integrity before reaching the public [29]. Ira Basen [7] pointed out that digital media platforms have fewer filters and gates than traditional media, making it challenging for users to determine what is new and what is important. Keen [20] mentioned that Web 2.0 has a negative impact on gatekeeping because of the reduction in gates or official gatekeepers who are accountable and professional. He maintained that “gatekeepers are a necessity due to the flood of information coming digitally.” Clark [11] interviewed several news professionals and asked how social media plays roles in their daily professional lives, showing that the downsizing of newsrooms has made an impact on the traditional role of the editors as a gatekeeper. Besides, different from the definition of “gatekeeper” for traditional media that in a sense if something is “gatekept”, it won’t go public to anyone, many social media services offer the “share” features for users to spread information in their social circles. In other words, even if a user decides not to share the content, the content could still be seen by its friends through other friends (if they choose to share the content). The above studies mainly focus on studying the changes of gatekeeping from traditional media era to today’s social media era, under the context that social media consumers have to face a flood of fake news and information. Leavitt et al. [26] looked at Reddit to understand how the design of Reddit’s platform impacts the information visibility in response to ongoing events in the context of controlling information flows (through gatekeeping).

Algorithmic Content Curation on Social Media

People are increasingly relying on online socio-technical systems that employ algorithmic content curation to organize, select and present information. Several studies have addressed customers’ perception of automated curation [12, 40]. For example, Rader et al. [40] investigated user understanding of algorithmic curation in Facebook’s News Feed through an online survey. They found that over 60% of the respondents implicate that the algorithmic News Feed caused them to miss posts from friends yet they still believe the algorithm that prioritizes posts for display in the News Feed. Similarly, Eslami et al. [12] conducted a user study with 40 Facebook users to examine their perceptions of the Facebook news feed curation algorithm. In contrast with the above algorithmic curation, in this paper, we study social media users’ consumption of contents that come directly from their friend networks in the absence of any automated curation. Unlike algorithmic curation that arranges and ranks the news items based on designated features, consumption of friend-curated content on WeChat offers users complete controls over information selection. It would be interesting to have an in-depth analysis of the users’ internal ranking scheme of friend-curated content.

Factors that Affect Users’ Content Curation

Social media is an essential way for people to curate and present information and previous literature has studied several factors that affect users’ content curation behaviors. For example, Leskovec et al. [28] studied the spread of news across websites and found that blogs generally lag only a few hours behind mainstream news sites. Agrawal et al. [2] proposed a model to identify influential blog contributors. They found that the number of times a blog post is shared and the number of comments on it generates are positively related to the influence of its contributors. Khan [22] conducted an online survey that covered 1143 registered YouTube users and identified the factors that motivate user participation and consumption on YouTube through regression analysis. User-user relationship with various strengths is also an important factor that influence user’s information seeking experience [4, 8, 15, 16, 19, 36, 37, 46, 48, 52]. Gilbert et al. [16] bridged the gap between social theory and social practice by predicting the strength of interpersonal relationships in social media and conducting user study-based experiments on over 2000 social media relationships. Wu et al. [46] identified the two different types...
METHODS

Article Reading and Gatekeeping on WeChat

As a prevailing social media App, WeChat has its distinctive features. First, it owns the characteristics of traditional media. Users can read articles directly from the homepage of the subscription accounts from the WeChat Official Account Platform (Figure 1(1)), which serves as the main source of articles for publication. Subscription accounts are often used similarly to daily news feeds because they can push one or several new update(s) to their followers every day [31]. The update(s) could contain a single article or multiple articles bundled together. Users may subscribe to as many accounts as they like. All subscription accounts are placed together in a subscription accounts folder on the timeline of users. Similar to bloggers on Twitter, a WeChat subscription account also has its fixed author(s). Second, it owns the typical features of social media. The most intuitive feature is that users can read articles shared by and forward them to their friend network (Figure 1(2)). WeChat provides three channels, namely, Moments, private chatting, and group chatting for users to access and read articles curated by their friends. In this work, we focus on content curation through the Moments. Users can share articles through their Moments, an immensely popular feature used to share pictures, short videos, texts, and links with their friends. Users can scroll through this stream of contents, similarly to Facebook newsfeed but they appear in chronological order. Users can also share articles to a specified friend or a group of friends via direct private or group chatting (Figure 1(3)).

The gatekeeping process on WeChat is depicted in Figure 2. Given the large amount of information from the Internet, the hosts of subscription accounts act as the first level of gatekeepers. Users who want to curate content can directly read articles from these accounts, or they can read articles shared by their friends in the Moments. In the latter case, the user’s friend network is acting as the second level of gatekeepers. The subscription accounts and the friend network determine collectively which articles in the whole platform get to appear on individual users’ news feed.

Research Questions

To understand how users leverage their friend networks as latent gatekeepers for content curations, we first need to examine to what extent and in what way the subscription accounts and friend networks act collectively as gatekeepers for users. Therefore we have our first research question:

RQ1. How do WeChat users utilize friend networks and subscriptions collectively as latent gatekeepers for content consumption?

Li et al. revealed that WeChat users are often confronted by abundant friend-curated content from a wide variety of sources [31]. Users may need additional cues to reduce the cognitive burden of deciding what to read. One potentially useful and always available cue is the composition (e.g., classmate, relative) and tie strength (e.g., how close is the relationship) of their friend networks. If a close friend is believed to be highly knowledgeable or trustworthy about public affairs, these positive evaluations may transfer to the information curated by him/her. We have our second question regarding the composition and tie strength of the friend network:

RQ2. Any difference between a) social circles and b) social ties when acting as gatekeepers?

Noted that the social contacts and information interests can change noticeably over time. Understanding these dynamics allows us to leverage those facets to improve relevance, and better manage influence and different “gatekeepers” in information dissemination [43]. Therefore, we study the third research question regarding temporal dynamics in the gatekeeping process:

RQ3. How do WeChat users adapt gatekeeping for content consumption over time?

Quantitative Analysis Method

Archival data obtained through collaboration with WeChat reveal that messages coming from all WeChat channels, i.e., subscription accounts, private chatting, group chatting, and friend network (i.e., Moments), collectively create users’ information landscape on the platform, each taking up different proportions. On average, 57% of the articles consumed by a WeChat user come directly from subscription accounts. The
remaining 43% are shared by friends through private chatting (11%), group chatting (18%), and Moments (71%) under different social scenes [50]. Private chatting “digests” the articles exchanged in the communication [47]. Group chatting is a private conversation among a group of users pre-gathered for certain purposes. Note that members of a WeChat group may not necessarily be friends of one another. Therefore, it is a social environment with complicated and unpredictable factors [39]. Comparatively, Moments is like a public bulletin for one’s entire friend network, publishing all the contents posted by friends in a timeline manner. Theoretically, people can browse content on their Moments at will, similar to how they can treat posts curated in the subscription account folder (if we consider subscription accounts as a special type of “friend”). As the focus of this paper is on how users proactively leverage their friend network as a latent gatekeeper of their information landscape, we only consider voluntary reading behaviors related to the two broadcasting channels, i.e., subscription accounts and Moments.

Data collection and Description
The dataset in this work was collected by our collaboration colleagues from WeChat, Tencent. Particularly, the dataset used for RQ1 and RQ2 contains a one-week log of article-reading activities from March 12th - 18th, 2018 curated via the subscription accounts and friend network of 7,234,753 users, a stochastic sampling on all users. The dataset is anonymized with all identifiable information removed. It consists of three parts:

• A1. User Attributes include user information within the selected time frame, such as age, registration duration, the number of friend, and the list of official accounts subscribed.

• A2. User Social Relationship contains the list of friends of each user and their social relationship with the users. In this paper, we describe a social relationship through the following two dimensions:

  -- D1. Social Similarity calculates the number of common friend between two users, which is a common practice to indicate to what extent two users are similar in social network analysis [6]. We thereby adopt it to obtain the social similarity between two users.

  -- D2. Social Circle presents social community in this work. Our collaboration experts generate four types of labels for communities in one’s friend network: colleagues, family, schoolmates, and others (e.g., real estate agency and WeChat business) by adopting a community detection algorithm Fast Unfolding [9].

• A3. User Article Consumption includes (1) the list of articles published by all the subscription accounts that a user follows; (2) the list of articles consumed by the user from the subscription accounts; (3) the list of articles curated by the friend network; and (4) the list of articles consumed by the user from the friend network. Note that these data can be filtered based on the attributes D1 and D2.

To answer RQ3, we collect an additional one-year article-reading data (201709 - 201807) of 10,000 users via stochastic sampling on WeChat user pool. Compared with the previous one-week dataset for RQ1 and RQ2, this one-year sampling dataset only contains A1. User Attributes and D2. Social Circles. This dataset has an additional attribute that is not included in the one-week dataset:

• A4. Article Categories specify the aggregated number of articles each user consumes in terms of different article categories.

Preliminaries and Computational Metrics
In this subsection, we provide a brief overview of the definition of metrics and terms used in the quantitative analysis for RQ1 and RQ2 (note: metrics for RQ3 are described in RQ3 subsection in RESULT section). We define information consumption on WeChat as a behavior of user clicking on a certain article curated by friends that shows up in a user’s Moments. To the user, these friends are gatekeeper of his/her Moments, determining what can be circulated in it. We define:

• M1. Click-through Rate (CTR) is the ratio of the number of consumed articles to the number of exposed articles for a WeChat user:

\[ r_{i} = \frac{m_{i}}{n_{i}} \]

• M2. Influence Ratio (IR) is the influence ratio of user i to user j which is measured as:

\[ r_{i \rightarrow j} = \frac{m_{i \rightarrow j}}{n_{i}} \]
where \( m_{i \rightarrow j} \) is the number of times user \( j \) read articles shared by user \( i \) and \( n_j \) is the total number of articles that user \( i \) share. The IR quantifies the pairwise influence between the user and his/her friend. The larger the IR, the more attention user \( j \) pays to contents curated by user \( i \), and thus the greater influence user \( i \) has on user \( j \).

- **M3. Total influence of friends’ gatekeeping** indicates the total effect of friends’ gatekeeping on user \( j \) which is defined as:
  \[
  r_j = \frac{\sum_{i \in F_j} m_{i \rightarrow j}}{\sum_{i \in F_j} n_i}
  \]
  where \( F_j \) is the set of friends of users \( j \).

- **M4. Influence of a social circle** indicates the influence of a certain type of social circle \( E \) which is defined as:
  \[
  r_E = \frac{m_E}{n_E} = \frac{\sum_{i,j \in E} (m_{i \rightarrow j} + m_{j \rightarrow i})}{\sum_{i \in V(E)} n_i}
  \]
  where \( V(E) \) is the involved users in the set of friends \( E \).

- **M5. Influence of subscription accounts on user \( j \)** is defined as:
  \[
  s_j = \frac{k_j}{l_j}
  \]
  where \( l_j \) is the total number of articles published by all the subscription accounts that user \( j \) follow, and \( k_j \) is the total number of articles that user \( j \) reads directly from these subscription accounts.

- **M6. The ratio of the influence of subscription \( s \) over the influence of friends \( r \)**: \( \frac{s}{r} \) describes how users split the gatekeeping responsibilities between subscription accounts and the friend network.

**Qualitative Analysis Method**

To verify the potential quantitative results with the users and to understand why users have some specific behaviors for gatekeeping of friend networks, we conduct an additional qualitative analysis with WeChat users.

**Participants**

We recruited 216 participants (males: 53.7%, females: 46.3%; age (19-25): 24.0%, age (26-30): 28.2%, age (31-40): 24.5%, age (41-60): 20.8% and age (60+): 2.50%) via an online survey service to understand their information consumption experience on WeChat. All the survey participants have a good knowledge of WeChat, as well as information consumption on WeChat. Particularly, we choose the participants with good operation skills of WeChat, for which they could provide us more comprehensive insights. We further invited 10 survey respondents (P1-P10) (males: 60%; females: 40%; age (19-25): 20%, age (26-30): 20%, age (31-40): 30%, age (41-60): 20% and age (60+): 10%) for follow-up interviews about their choices in the survey. Each interview took 15 minutes and was audio recorded.

**Design of Questionnaires**

The questions used in the questionnaires take the form of multiple choices. As a supplement for RQ1, we ask participants to choose what kinds of articles are most welcome from their friend network and what kinds of articles they would further curate and repost. For RQ2, they are asked to indicate from which social circle(s) (options: family, colleague, schoolmate) do they curate information in their friend networks and choose the possible reasons we provide. For RQ3, we ask them whether and when would their social circles as “gatekeepers” experienced some changes.

**RESULTS**

In this section, we summarize the results for the research questions one by one, following the style of first presenting the quantitative results then listing the qualitative one if applicable.

**RQ1: Gatekeeping by Subscription Accounts and Friend Network Collectively**

Figure 3 shows the relationship between \( \frac{s}{r} \) (the ratio of the influence of subscription \( s \) over the influence of friends \( r \)) and \( \frac{l}{n} \) (the ratio of the number of articles \( l \) published by the subscription over the number of articles \( n \) shared by friends). We can see that \( \frac{s}{r} \) decreases with the increase in \( \frac{l}{n} \). Noted that in Figure 3, we adopt a logarithmic axis. One can see that the ratio of the influence of subscription accounts over the friend network has a power-law dependence on the ratio of the articles published by the subscription accounts over the ones by the friend network:

\[
\frac{s}{r} \approx \beta \left(\frac{l}{n}\right)^\alpha
\]

Through linear regression, we obtain \( \alpha \approx -0.48 \) and \( \beta \approx 0.75 \) with the significant level exceeding 99%. With a further deduction from Equation 5, we can infer that when \( \frac{l}{n} > 0.55 \), \( \frac{s}{r} < 1 \). That is: when the number of articles published by the
We also find that users of high or extra-high CTR tend to have attractive titles. Articles with finer "gates" to control the information flow.

(F1) (nourishing stories for one’s soul) (26.4%), and

The qualitative results for RQ2a: About Social Circle show the diversity of the curated and gatekeeping content. 79.6% of the survey respondents stated that they often consumed articles on WeChat. Articles with attractive titles (42.6%), news and events (38%), practical knowledge (34.7%), financial and investment knowledge (26.4%), and funny stories (25.5%) are most welcome from the friend network. Users would further curate and repost the articles about practical knowledge (56%), insightful stories (28.7%), industry trends (27.8%), “chicken-soup” articles (nourishing stories for one’s soul) (26.4%), and current events (21.8%), etc., a bit different from what they consume from the friend network. Respondents (P7, P9, females; P3, P5, males) stated that “sometimes, these articles represent what we thought,” and “are well responsive to our current status.” Responses from different participants indicate different media literacy, showing that contents curated by different friends can be quite diverse.

RQ2: Gatekeeping by Composition and Tie Strength of Friend Network

(CTR) also increases.

Figure 5. In overall, with the increase of x-axis: social similarity, y-axis: CTR also increases.

RQ2a: About Social Circle

To explore how the composition of friend network help WeChat users filter information in Moments, we analyze to what extent friends with different social attributes (D1. Social Similarity and D2. Social Circle) serve as users’ channel of choice for content consumption (measured by CTR). We divide the value range of social similarity (number of common friends) into seven intervals: 0, 1 - 3, 4 - 5, 6 - 10, 11 - 20, 21 - 40, and over 40 according to the input of domain experts from our collaborator. As shown in Fig 5, CTR goes up as social similarity increases. Among individuals having over 40 common friends with a user, those from the user’s schoolmates circle achieve the highest CTR. In the rest of the social similarity intervals, CTR of the family circle is the highest (noted as F3). When rendering the CTR values of friends from various social circles for users at different ages (Figure 6), we find that users across all age groups generally prefer to consume articles curated by the family circle, followed by colleagues and schoolmates. Interestingly, the influence of schoolmates first declines among users at age 20 and then rises again among users aged 32 and over, eventually surpassing the CTR of colleagues among users at age 54 and exceeding the family CTR for users aged 60 or older. Colleagues have about the same level of influence as (sometimes slightly higher than) family for users in their 20s and 30s, and gradually lose the leading position to family and then to schoolmates among users over 40. In a word, the proportion of information intake

Figure 4. Distribution of different age, registration duration, and number of friends over the four levels of click-through rate (CTR).
from different social circles varies across WeChat users in different age groups (notated as F4).

F3-4 manifests that people always care about family-curated content. However, 60.7% of the survey respondents do not consume information from the family circle, especially from the elder ones. Among them, 72.5% reported that they are not interested in the topics curated by the elder family members, and 29% indicated that their reading interests are not similar. "Family members like to share articles about inspiring stories, health maintenance, and festival-greetings, etc.," said P3 (male). 29.4% of those who consume family-curated articles "because of emotional support" and 74.1% of them "care about what my family members are interested in." Only 22.4% reported that they share similar interests with their family. "We may just take a glance at the title or quickly go through the contents," said P7 and P9 (females). In this case, it seems that people may already have a pre-assumption for the information curated by their families. Consuming information from them may be largely due to emotional support, rather than the information relevance or importance. To further verify this hypothesis, we need more data such as the average time of reading an article curated by different circles. Apart from the family circle, 74% and 76% of the survey participants like to consume information from colleagues and schoolmates because of the topics (71.5%) and similar hobbies (46.7%). 27.3% of users reading articles from colleagues stated that "these articles can be conversation-makers in the company." 23.6% of users (with 34.43% from the age of 26-30) who do not like to read articles from schoolmates stated that “our life becomes different.”

RQ2b: About Social Tie Strength

We then explore the use of tie strength of their friend network for filtering information in Moments. When Granovetter, M.S. first proposed the concepts of strong-/weak-ties, he did not provide a strict definition but a qualitative description: strong-ties refer to frequent connections and close relationships, and weak-ties are accidental connections with seldom communications [17]. In this study, we follow the approach proposed by Gupte et al. [18] and employ Jaccard Index\(^2\) of D1: Social Similarity to measure the tie strength quantitatively:

\[
J(F_i, F_j) = \frac{|F_i \cap F_j|}{|F_i \cup F_j|}
\]  

where \(F_i\) is user \(i\)’s friends and \(F_j\) is user \(j\)’s friends.\(^2\)

**Strong Ties Bring Trust.** Articles curated by friends on WeChat may be either from the subscription accounts that (Case 1) a user has already followed or from those (Case 2) the user has not yet followed. We utilize A2-3 to compute and compare M2. Influence Ratio in the two cases, and apply normalization as follows. For each case, we divide the influence ratio in each segment of tie strength by the minimum influence ratio among all segments in that case, and then obtain the normalized influence ratios of the corresponding cases in each segment. As shown in Figure 7, we find the influence of strong-ties among most segments in Case 2 is much higher than that in Case 1. For example, in Case 2, the influence ratio of the segment of [0.05, 0.1) is six times higher than that of the segment of [0, 0.001], whereas in Case 1, it is only two times. Case 2 confirms the “strong-ties theory” proposed by Krackhardt, D.: “consuming articles from unknown sources means making changes (e.g., exposure to new knowledge, cognitive changes.), and it is with discomfort; however, strong-ties can help overcome this discomfort.” [24] In Case 1, users have direct exposure to the articles from the subscription account. If they have determined to read the articles (or not), seeing the articles in their friends’ Moments may not change their decision. This is perhaps why the normalized influence ratios for Case 1 are pretty similar between the 2\(^{nd}\) and 7\(^{th}\) bar. However, there is a noticeable increase of Case 1 influence ratios in the last two bars, suggesting the likely persuasion effect of the strongest ties. This finding indicates that strong-ties bring a sense of trust as a gatekeeper. In other words, if an article comes from a subscription account that a user has not followed, the trust brought by this article is deficient. However, the friend curation behavior makes up for this lack of trust and thus this behavior can be considered as a trustful gatekeeper (noted as F5).

Weak Ties Bring Serendipity. We investigate the repetition rate of the articles curated by the friend network in each segment of tie strength (we calculate the repetition rate which indicates the percentage of friends ever curating the same articles based on A2-3) (the green curve in Figure 7). We find that with the increase of tie strength, the repetition rate rises gradually. This indicates that weak-ties are more likely to bring information that users have not seen before, and can act as “gates” leading to “unexpectedness” (noted as F6). This finding also corroborates the “weak-ties theory” [17] that “information curated by strong-ties is likely to be similar and redundant, whereas weak-ties can break the boundaries of people’s inherent social circles and bring new information.”
In the qualitative study, we asked the interviewees about how they treat the information curated by strong-/weak-ties. “I want to learn why my close friends read this article,” said P6 (female). “I will consume information from people I occasionally meet because of their comparatively fresh information.” “I pay special attention to some friends who have special ideas, or opinion leaders,” said P9 (female). “I’m interested in articles from strong-ties, but sometimes, weak relationships will also bring some current affairs-related articles which I am interested in,” said P5 (male). “Sometimes, I have a clear idea of what I want and go straight to appropriate friends for contents known to fulfill my consumption demand; other times, I am open to new information and just click around,” said P4 (male). When we further inquired him whether these friends could act as a “filter”, he said, “definitely, with so much information, I will choose the information curated by my close friends with trustworthiness.”

49.5% survey respondents stated that they would follow up a (female), “vance/types of articles within their reach at each time frame of one-month. We start with depicting each user \( u \) by its friend network composition \( \mathbf{D}(u) \), \( v_u = (r_c, r_f, r_s, r_o) \), where \( r_c, r_f \), \( r_s, r_o \) represent the ratio of the number of friends in colleagues, family, schoolmates, and others, respectively, to all friends. Based on its vector representation \( v_u \), we use K-Means to cluster all sampled users into four clusters. Each cluster indicates a different friend network composition. The number of clusters can be dynamically adjusted. As shown in Figure 8, when \( k = 4 \), we can achieve a balanced distribution among all clusters. Next, via regression analysis, we fit the friend network composition space to the article consuming space. The regression analysis has long been used to model the relationship between variables and to estimate how a dependent variable responds to changes [30]. Under the assumption that network composition can affect article consumption behaviors, we regard each social circle as an observed variable and use its combination to regress the consumption space for a certain article category \( (A4) \). For each cluster, we weigh each social circle by its contribution to the article consumption by using feature importance, i.e., we approximate the two spaces by \( f(n_p, n) \approx \text{reg}(w_c * r_c, w_f * r_f, w_s * r_s, w_o * r_o) \) where \( f \) computes the ratio of the number of consumed articles \( n_i \) of category \( k \) to all consumed articles \( n \) and \( w \) is the feature importance.

<table>
<thead>
<tr>
<th>Regression</th>
<th>( R^2_{k=0} )</th>
<th>( R^2_{k=1} )</th>
<th>( R^2_{k=2} )</th>
<th>( R^2_{k=3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>0.102</td>
<td>0.145</td>
<td>0.079</td>
<td>0.089</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.051</td>
<td>0.072</td>
<td>0.052</td>
<td>0.062</td>
</tr>
<tr>
<td>MLP</td>
<td>0.252</td>
<td>0.225</td>
<td>0.191</td>
<td>0.430</td>
</tr>
<tr>
<td>DT</td>
<td>0.761</td>
<td>0.742</td>
<td>0.543</td>
<td>0.801</td>
</tr>
<tr>
<td>RF</td>
<td>0.540</td>
<td>0.680</td>
<td>0.601</td>
<td>0.762</td>
</tr>
</tbody>
</table>

Table 1. Results for different regression models.

In this step, we apply five widely-used machine learning algorithms, including Linear Regression (LR), Lasso, Multiple-layer Perceptron (MLP), Decision Tree (DT), and Random Forest (RF) to conduct regression analysis. Among them, LR and Lasso are linear regressors and the rest fit data with non-linear kernels. We use the coefficient of determination \( (R^2) \) commonly employed in regression analysis to assess model performance (Table 1). The results indicate that DT performs the best with a sufficiently high \( R^2 \) score.

We then apply DT to extract the feature importance of each social circle for each cluster to reflect their contribution to the circulation of each type of article. Figure 9 gives an example. In Cluster 1, colleagues occupy about 60% of the friend network, followed by the family circle (25%). Color blocks
Figure 9. Color blocks indicate feature importance of four social circles for different article categories. Cluster 1 (colleagues occupying the most) and Cluster 2 (family occupying the most) are compared using data of 201709.

Figure 10. Word clouds indicate topics that social circles of Cluster 1 contribute to over time. The font size in word clouds encodes feature importance of the corresponding circle that contributes to the topic.

indicate the feature importance of the corresponding circles for different article categories. One can see that for Cluster 1, topics with high feature importance are practical knowledge from family as well as baby and child, holidays, and career events from colleagues. For Cluster 2 in which family circle dominates (80%), the distribution of topics with high feature importance (e.g., traditional culture, and traveling from family, games and shopping from schoolmates) is different from that in Cluster 1. For a specific cluster of users, different social circle curates different topics (noted as F7). For example, alcohol and tobacco-related articles come mostly from the schoolmate circle, and the schoolmate circle tends to circulate information about career events and entertainment. There is also the phenomenon that information about the same topic is curated by a different circle in different clusters. For example, in Cluster 1, the family is the main source of sports and food-related information, while such articles come mostly from “others” in Cluster 2. Another example is that users in Cluster 1 take in holiday-related content primarily from colleagues, while those in Cluster 2 read about holidays from their schoolmates’ posts. This is mainly due to different clusters of users who may have quite different media literacy [38].

Step 2: Visualizing feature importance over time. After calculating the feature importance for each month, we visualize the topics with feature importance encoded by font size in a word cloud. Take three social circles of users from Cluster 1 at four different time frames (i.e., 201709, 201712, 201803, and 201806) as an example (Figure 10). Apart from some topics that are always dominated by certain social circles, e.g., traveling for colleagues and plants for schoolmates, housing emerges on Dec. 2017 in the schoolmate circle and this circle then contributes significantly to housing, followed by the family circle contributing more to housing. Tea art-related articles shift between the family circle and the schoolmate circle. We also inspect other clusters and identify similar phenomena, i.e. although different social circles tend to curate some relatively stable topics (e.g., traveling) with strong characteristics, some interests can shift between circles and the attention paid to them has its own ups and downs (noted as F8).

In the qualitative results for RQ3, 50.5% of survey participants reported that leveraging different social circles as “gatekeepers” experienced some changes, with 32.4% of them after they entered colleges, 36.1% of them after they started to work, and 37.1% when their lives shift to a new stage, such as getting married, becoming parents, and getting retired. The follow-up interviews complement some detailed explanations. “now I prefer to read professional articles related to my major, and I will explore them by subscription accounts or my friend network,” reported by P1 (student, male); “I read more articles related to my field so I pay more attention to my colleagues,” said P4 (male). “After getting married, I consume more information about life, emotions, personal growth or the contents curated by my friends in similar circumstances. In fact, I have
DISCUSSION
The insights found in this study can provide implications for the future designs of socio-technical systems, e.g., social recommenders. Users tend to adjust their degree of reliance on subscription accounts or friend networks based on the quantity and quality of their content suppliers (F1). Meanwhile, elder users with fewer WeChat experience are more likely to consume friend-curated content (F2). For one thing, having more spare time, most of them enjoy socializing with old friends and classmates on social media (F4). As indicated in [32], “social media users are more likely to reconnect with people from their past, and these renewed connections provide a strong support network when people are near retirement.” For another, this group of users usually have strong information needs. Therefore, social recommender strategies for their contents, if applied, need to be tailored.

Regarding how gatekeeping gets reflected in different social circles and ties, survey participants indicate that the family circle cannot fully function as a gatekeeper, different from the quantitative analysis (F3). This is because of emotional supports, or because people may already have a pre-assumption of the contents curated by them. The qualitative study also finds that people’s reading interests may shift over time due to (1) changes of their information needs and tastes may have altered; and (2) changes of their social circles composition around the same time, which is in accordance with F7-8: although users with similar friend composition tend to get stable curation for some information, some articles and the attention paid to them can shift between different social circles. The performance of DT regression indicates that the article consumption space preserves different social circles in a non-linear way, i.e., different social circles can share common information interests. People with a clear idea of what they want to read will go straight to appropriate gatekeepers known to fulfill the consumption demands. These gatekeepers can be either close friends with strong-ties or trustful followed subscription accounts. In addition, the quantitative analysis indicates that the information curation behaviors of friends with strong-ties hold implications for the unfollowed subscription outlet trust (F5). With a little help from these friends, people may be able to connect with new subscription accounts and improve their readiness to participate in an informed democracy. People are also willing to acquire new knowledge from weak-ties (F6), which bring serendipity. In either case, users manage different gatekeepers as content curators. We can, therefore, model pairwise friend influence and apply to potential recommendation scenarios such as the social advertisements to give more exposure to users who are more likely to consume friend-curated content.

Limitation
There are several limitations to this research. First, in RQ3, since we only consider the relationship between social circles and article categories, regression models may fail to capture other factors with a possibly high correlation between the social circle and article consumption space. As we only have a one-year longitudinal data due to high maintenance cost of our collaborators, and users are clustered only once based on their social composition, we cannot conclude that changes of social circles would influence information gatekeeping, since dramatic changes in users’ social network or interests are unlikely to happen overnight. Second, in RQ2, we only use the number of common friends to measure tie strength. A more objective metric may be chatting frequency between two users. Third, in most cases, we just employ CTR to infer users’ reading interests and do not dig into deeper the motivations behind their clicks by using other metrics.

CONCLUSION
In this paper, we conduct a mixed-methods approach to studying “friend network as a latent gatekeeper” phenomenon on a friend-based social media WeChat. We analyze over seven million users to infer how they accommodate and safeguard their information through social circles and social ties. We also conduct a survey of 216 WeChat users about their reading activities on WeChat. Results indicate that WeChat users prefer the friend network when information is overloaded. They like to leverage weak-ties getting exposed to new domains, and turn to strong-ties when demanding credible and reliable information. Elder users with fewer experience using WeChat are more likely to consume friend-curated contents. Users leverage social circles to gatekeep information interests and the interests and attention paid to them can shift from one social circle to another.

ACKNOWLEDGMENTS
We are grateful for the valuable feedback and comments provided by the anonymous reviewers. This research was supported by WeChat-HKUST Joint Lab on AI Technology (WHATLAB) grant#1617170-0 and HKUST-WeBank Joint Laboratory Project Grant No.: WEB19EG01-d.

REFERENCES


