JOURNAL OF COMPUTER-MEDIATED COMMUNICATION

Journal of Computer-Mediated Communication

Information Overload, Similarity, and Redundancy: Unsubscribing Information Sources on Twitter

Hai Liang

School of Journalism and Communication, The Chinese University of Hong Kong, Room 424, Humanities Building, New Asia College, Shatin, N.T., Hong Kong

King-wa Fu

Journalism and Media Studies Centre, The University of Hong Kong, Room 206, Eliot Hall, Pokfulam Road, Hong Kong

The emergence of social media has changed individuals' information consumption patterns. The purpose of this study is to explore the role of information overload, similarity, and redundancy in unsubscribing information sources from users' information repertoires. In doing so, we randomly selected nearly 7,500 ego networks on Twitter and tracked their activities in 2 waves. A multilevel logistic regression model was deployed to test our hypotheses. Results revealed that individuals (egos) obtain information by following a group of stable users (alters). An ego's likelihood of unfollowing alters is negatively associated with their information similarity, but is positively associated with both information overload and redundancy. Furthermore, relational factors can modify the impact of information redundancy on unfollowing.

Keywords: Information Repertoire, Media Consumption, Information Redundancy, Unfollowing, Twitter.

doi:10.1111/jcc4.12178

Social media have changed the current media environment and can consequently influence individuals' selection of information sources. First, social media offer users a large number of information choices. All social media users can provide some kinds of content, and thus could be considered as information sources. However, the available human attention to consume information is always limited (Webster, 2010), leading to varying degrees of information overload on social media (Holton & Chyi, 2012). To cope with information overload, users will rely upon relatively small subsets or "repertoires" of their preferred channels (e.g., Kim, 2014; Taneja, Webster, Malthouse, & Ksiazek, 2012; Yuan, 2011).

Editorial Record: First manuscript received on November 12, 2015. Revision received on June 1, 2016. Accepted by Noshir Contractor on September 24, 2016. Final manuscript received on September 29, 2016. First published online on November 30, 2016.

Second, social media users are usually embedded in online social networks. They select information sources by "following" other users. Those users being followed are called followees. In social networks, users are inclined to follow other users who appear to be similar to themselves with respect to many attributes (McPherson, Smith-Lovin, & Cook, 2001). Given the increasing number of available choices on social media, this homophilous selectivity is reinforced in computer-mediated communication and potentially results in audience fragmentation (Sunstein, 2009). If users follow many content-similar followees, they are likely to receive many duplicated messages in their personal information streams, and thus lead to information redundancy (Harrigan, Achananuparp, & Lim, 2012).

Although the repertoire approach to media use provides an important framework to capture information consumption patterns under information overload (Kim, 2014; Taneja et al., 2012; Yuan, 2011), few studies have examined the dynamic process of repertoire formation and the role of content similarity and redundancy. In a networked environment on social media, information overload could increase the tendency of similarity-based selection and further leads to information redundancy, which in turn might increase information overload. Besides, little attention has been paid to the removal of information sources from personal repertoires. The deselection process helps reduce information overload (Webster, 2010) and stabilize personal information repertoires (Kwak, Chun, & Moon, 2011). In order to investigate the role of information overload, similarity, and redundancy in structuring information consumption patterns, this study extends the repertoire approach by incorporating media choice theories and social network analysis.

Followees as Information Repertoire on Social Media

Contemporary social media are usually conceived as a combination of information platforms and social network services, wherein "ordinary" users (as well as media organizations, journalists, and other "elite" users) create, share, and consume user-generated content (as well as professional news content) in social networks (e.g., Kwak, Lee, Park, & Moon, 2010; Murthy, 2012). This definition captures two essential aspects of the character of current social media platforms. First, ordinary people have turned themselves into media content providers on social media by producing and sharing news messages (Murthy, 2012). Second, people's reception of information is basically constrained by their personal social networks online. Social media users decide whose messages they wish to receive by following other users. By creating such following connections, users receive the content that their followees post. Thus, as users choose whom to follow, they also choose the information to which they will have access.

Twitter is one of the most popular social media platforms. Figure 1 illustrates the information consumption structure on Twitter. An individual (e.g., Ego 1) can subscribe to receive the tweets of another user (e.g., Alter 2). We say Alter 2 is a followee of Ego 1 while Ego 1 is a follower of Alter 2. On Twitter, information seeking takes the form of following relationships (Himelboim, Hansen, & Bowser, 2013). Egos receive messages from followees but do not receive them from followers. In Figure 1, Ego 1 is following Alters 1–3. Therefore, Ego 1 receives all tweets posted by these alters. The following relationships on Twitter may not be reciprocal. Users may rebroadcast a tweet by retweeting the message to their followers. They can also converse with any other users by replying to their tweets or mentioning other users using the "@" sign. All original tweets, retweets, and replies are displayed in users' timelines.

The emerging characteristics of social media have several implications for existing theories about the choice of information sources. First, all social media users, including ordinary people, journalists, and media organizations, can be their followers' information sources. Hermida, Fletcher, Korell, & Logan (2012) found that social media users are more likely to receive information from the individuals than from news organizations and journalists followed on social media. Even for media content, Twitter users get roughly half of their media referrals via intermediaries (Wu, Hofman, Mason, & Watts, 2011). In

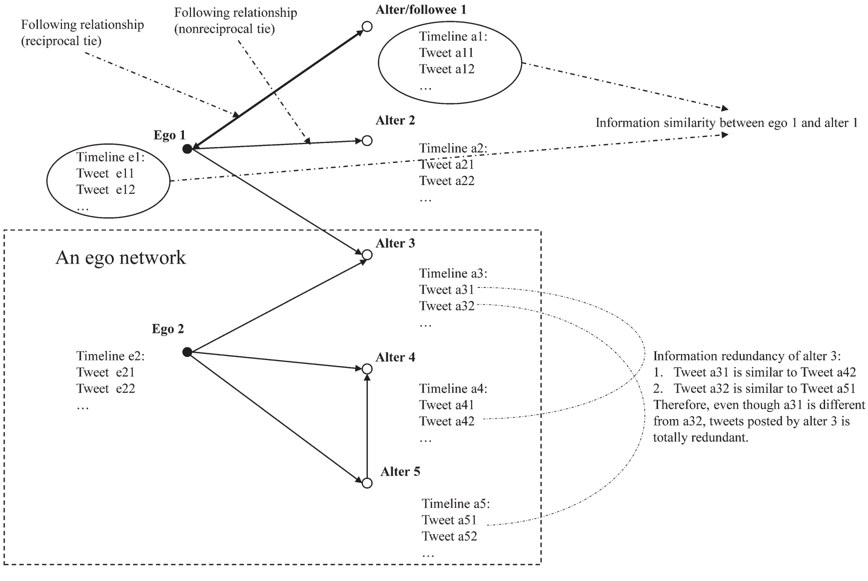

Figure 1 An ego network approach to conceiving followees as information repertoires. An ego network consists of a focal node ("ego") and the nodes to whom ego is directly connected to (alters) plus the ties (arrows in the figure) among the alters. The arrow pointing from ego 1 to alter 2 indicates that ego 1 follows alter 2. It also suggests information flows from alter 2 to ego 1.

this sense, the selection of information sources partly becomes the choice of followees on social media. Therefore, alters, followees, and information sources are interchangeable terms in this study.

Second, users' followees serve as their information repertoires to help cope with information overload on social media. The conventional notion about media choice is that people assess the available resources and choose among them in an effort to achieve their purposes rationally. However, that rationality is bounded by the overabundance of choice and the limited human attention available on social media. It is impossible to follow all available information sources or select relevant alters by examining every single message posted by these alters. One technique many people use to manage their choices is to limit the number of choices by paring down their options to a more manageable repertoire of preferred sources (Webster, 2010, 2014). On Twitter, that usually means following a small number of followees/alters.

Third, repertoire is a subset of available media that an individual uses regularly or frequently (e.g., Webster, 2014; Webster & Ksiazek, 2012), which actually reflects the habitual nature of media consumption. However, most studies analyze media repertoires in a static way. Building information repertoires on social media by following other users is a dynamic process. On Twitter, users can take two steps to establish and maintain their followees. When joining Twitter, users may subscribe to other users. Later on, users are free to unsubscribe and remove users from their following lists (i.e., unfollow). Kwak et al. (2011) have documented a relationship stabilization process on Twitter—users are less likely to unfollow those whom the users have been following for a long time. Finally, this process will result in stable repertories of information sources, which users rely on heavily for daily information consumption. It suggests that researchers should pay more attention to the unfollowing process in the dynamic analysis of repertoire formation.

Finally, previous studies focus on explaining the absolute size of repertoires, but often say little about their composition (see Webster & Ksiazek, 2012). An information repertoire is introduced to cope with information overload by selecting preferred sources. By taking the composition into consideration, similarity-based selection may create redundant messages in personal repertoires (Himelboim et al., 2013) and exacerbate information overload (Franz, 1999). It remains largely unknown to what extent users unfollow their information sources based on individual preference of certain content category, information overload, and information redundancy.

Building Repertoires: Informational Factors

The above discussion suggests three intertwined variables that are relevant to the formation of user repertoires on social media platforms. Unlike previous repertoire studies that focus on the structure of media use, the variables are related to the communication content. First, the primary purpose of building information repertoires is to cope with information overload. Information overload has traditionally been conceived as a subjective experience in which users are overwhelmed by a large supply of information in a given period of time (e.g., Savolainen, 2007). Although many reasons can cause information overload, the core component is the volume of incoming information (Franz, 1999). The growing number of information sources could have negative consequences for information seeking. Holton and Chyi (2012) found that 72.8% of respondents felt at least somewhat overloaded with the amount of news available today. In addition, the study reports that perceived overload depends on platforms: Use of Facebook is positively associated with information overload, whereas the use of Twitter is not significantly correlated with information overload. To cope with information overload, media users usually maintain a repertoire by limiting the number of sources. For instance, TV studies have demonstrated that people watch a small fraction of available channels in the US and across the world (see Webster, 2014).

However, previous studies merely focused on the implicit impact of information overload on the absolute size of personal information repertoires. When considering followees as a repertoire of information sources, information overload could influence the following or unfollowing of a specific user, because the volume of information delivered by each followee varies. Even when users have exactly the same repertoire size, the amount of incoming information they receive could be very different. Therefore, in addition to limiting the size of repertoire, ego users can control their incoming information by selecting or deselecting followees according to the followees' posting frequency. Kwak et al. (2011) found that burst posting is one of the major reasons of unfollowing on Twitter, irrespective of the topic of the tweets. If followees posted too many tweets in a short period of time, they were more likely to be unfollowed. Therefore,

H1: *Users are more likely to unfollow the alters who post more messages during a given period of time.*

Second, building information repertoires is closely related to the choice of specific information sources. Previous studies have used content preferences to explain cross-platform media repertoires (see Taneja et al., 2012). They found that information sources providing similar content are likely to be used together. For example, researchers have evidenced a news repertoire that combined Internet and television news (e.g., Dutta-Bergman, 2004; Yuan, 2011). This findings are consistent with the theory of media complementarity: Individuals who are interested in a particular content type expose themselves to various information channels that correspond with their area of interest (Dutta-Bergman, 2004).

Although the original purpose was to explain cross-platform media consumption, Himelboim et al. (2013) have extended the theory of channel complementarity by considering the complementary selection of information sources that occur within a single social media platform. They found that

Twitter users in different clusters follow different sets of sources that cover different areas of content (i.e., local versus national news). In a similar way, if users show consistent interest in posting a type of tweet, they are more likely to follow those users posting messages on a similar topic (Weng, Lim, Jiang, & He, 2010).

Even though empirical studies consistently suggest that users are likely to select sources providing content similar to their own tweets, it is hard to define and operationalize the content type on social media platforms. Instead of using the top-down approach to categorizing the tweets into politics, sports, entertainment, and so on, the repertoire studies have developed a user-defined approach by letting repertoires emerge from patterns of use, which is based on factor analysis (e.g., Kim, 2014; Taneja et al., 2012; Yuan, 2011). On Twitter, users can define their own content categories by using hashtags. Unlike the raw texts, hashtags are the keywords or topics prefixed by the # symbol in tweets to label and categorize tweets. Therefore,

H2: *Users are less likely to unfollow the alters who post similar hashtags.*

The effects of information overload and similarity might be mutually conditional. First, selections based on information similarity can influence the impact of information overload on information consumption. Information overload is a subjective feeling. For equal quantity of information, individuals may perceive different levels of overload. For sources providing similar hashtags, people may want to receive more messages from them, whereas for sources posting irrelevant hashtags, a few messages can cause information overload. Second, information overload can increase selectivity (Sunstein, 2009; Webster & Ksiazek, 2012) and thus the similarity-based selection of followees is expected to be reinforced on social media. Therefore,

H3: *The positive association between unfollowing and tweeting frequency is weaker for the alters who post similar hashtags to their egos.*

Third, selecting information sources based on information similarity in a networked communication environment can lead to information redundancy in one's personal repertoire. If users select the sources posting similar topics (using similar hashtags) consistently, they will be expected to receive lots of redundant messages. Information redundancy refers to message repetition in a series of received messages (Stephens, Barrett, & Mahometa, 2013). It does not mean posting repeated messages by a single alter user. As illustrated in Figure 1, information redundancy quantifies the extent to which an alter posted a type of content (i.e., hashtag) similar to other alters in an ego network, while information similarity is related to the similarity of hashtags between egos and alters. Even though there are no duplicated tweets within an alter's timeline, the tweets by alters could be totally redundant to their egos, because other alters may post similar tweets in the ego network.

The role of information redundancy in media choice has been only implicitly mentioned. For example, previous studies suggest that use of one medium will displace the use of functionally alternative media, because the time available in any day is fixed (e.g., Ferguson & Perse, 2000). It implies that people are less likely to choose media with redundant information relative to their current media repertoires under information overload conditions. Another relevant argument is based on the theory of media complementarity. It predicts that people who select one type of information from one channel will also select the same type of information from other channels (Dutta-Bergman, 2004; Himelboim et al., 2013). Eventually, that will cause duplicated messages received from different media platforms (Jenkins, 2006). It implies that redundancy might be acceptable if people are interested in only a few types of content. For example, users who are interested in pop music may follow many pop stars on social media. However, empirical studies found that people are actually interested in different types of content and are likely to possess a complementary architecture in their information repertoires (e.g., Chaffee, 1982; Webster, 2014; Yuan, 2011). That means if the users are interested in both pop music and sports, they may unfollow a few pop stars and then follow some sports stars to avoid information overload, even though the unfollowed pop stars were posting unique tweets concerning their own stories.

Twitter provides a common place where different information sources are available for users to choose from. Individuals now tend to search for information they are interested in by content type across multiple outlets (i.e., based on hashtag similarity). If users indeed have diverse interests and are inclined to maintain diversity in their information repertoires, they would tend to keep those followees who post hashtags that are less redundant relative to what they receive from all existing followees. This is a strategy to maintain diversity without causing additional information overload. Therefore,

> **H4:** *Users are more likely to unfollow those alters who post more redundant hashtags relative to what the users receive from all other alters.*

Furthermore, the redundancy effect on unfollowing might be conditional. According to previous studies, information redundancy can cause information overload (Farhoomand & Drury, 2002; Watson-Manheim & Belanger, 2007). Under information overload conditions, users might reduce information overload by unfollowing the followees who post redundant types of content. Otherwise, people could maintain diversity and redundancy simultaneously by following more users. Therefore,

> **H5:** *The positive association between information redundancy and unfollowing is stronger for the alters who post more messages during a given period of time.*

In addition, users have to balance the selection criteria based on content similarity and information redundancy. As we discussed, redundancy in personal information streams could be caused by the choice of sources based on hashtag similarity on Twitter. When followees create many redundant hashtags, users may be less likely to apply the similarity selection strategy in order to avoid unnecessary redundancy. Therefore,

> **H6:** *The negative association between hashtag similarity and unfollowing is stronger for the alters who post less redundant hashtags.*

The Role of Relational Factors

In addition to these informational factors, relational factors could be another set of factors structuring the dynamic process of repertoire building on social media. Relationship building (e.g., maintaining friendship online) and information seeking are the two major motivations of Twitter use (Kwak et al., 2010; Myers, Sharma, Gupta, & Lin, 2014; Wu et al., 2011). Information consumption on social media could be a byproduct of social networking behaviors. Many ordinary users socialize with their friends, family, and coworkers on Twitter. They establish ties for relational purposes, while the ties also serve as the conduits of information flow (Golder & Yardi, 2010; Myers & Leskovec, 2014). The network structure can influence what information users will receive. For example, users in densely connected communities are expected to receive more redundant information (Harrigan et al., 2012). That suggests that the relational factors might modify the relationships between the informational factors and unfollowing.

Following the literature on online friendship formation, we consider three variables that have been mentioned most frequently: popularity, reciprocity, and the number of common followees. First, Golder and Yardi (2010) reported that people who are already popular appear more likely to attract new followers due to the preferential attachment mechanism (Barabasi & Albert, 1999). Second, reciprocity refers to the tendency that two users follow each other to form a mutual connection. Kwak et al. (2011) argued that reciprocal following relationships on Twitter can bring emotional closeness (also see Golder & Yardi, 2010) and thus reduce the likelihood of being unfollowed. Reciprocal following relationships enable two users to follow each other's updates, which increases their interaction frequency and leads to more stable relations (e.g., Xu, Huang, Kwak, & Contractor, 2013). Finally, as two users share more followees, the likelihood of them being friends with one another is expected to increase (Golder & Yardi, 2010). Xu et al. (2013) argued that users with more common followees are more deeply embedded in Twitter networks. Studies have found that the number of common followees has a negative correlation with unfollowing on Facebook (Quercia, Bodaghi, & Crowcroft, 2012) and Twitter (Kivran-Swaine, Govindan, & Naaman, 2011; Kwak, Moon, & Lee, 2012; Xu et al., 2013). Given that all these variables have been demonstrated to be important predictors for unfollowing behavior on Twitter, we include them as control variables and further we will examine whether these variables will affect the informational factors involved in unfollowing.

RQ: *How will the relational factors influence the impacts of informational factors on unfollowing?*

Method

Data Collection

By using Twitter's REST APIs, we collected a two-wave panel dataset. To overcome the representativeness problem, this study sampled the panel users randomly from the population. First, we employed a method reported in Liang and Fu (2015) to generate random Twitter user IDs. We generated 90,000 random numbers. And then, we searched these numbers via the official API to check the existence of these Twitter IDs. Using this method, we obtained 34,006 valid Twitter user accounts (egos).

Second, we obtained the egos' user profiles, their followees' IDs, and up to 3,200 tweets and retweets (timeline) for each ego user. We collected the first wave of data in December 2014 and the second wave in March 2015. In the second wave, 33,774 egos still exist. Due to the privacy settings on Twitter, we could only get the tweets from the public accounts. In addition, since we are only interested in the unfollowing behavior, we exclude those users who are totally inactive during the period of data collection. Finally, we got 7,609 ego users who are both active and publicly available.

Third, we constructed ego networks in which nodes are users and ties are the following relationships between egos and followees. In the first wave, there are 1,314,156 nodes (including 7,360 ego users) and 1,766,269 ties in the ego networks. In the second wave, there are 1,403,291 nodes (including 7,464 ego users) and 1,888,039 ties in the ego networks. We further collected the followees' profiles and tweets, and their followees. We excluded the followees whose tweets and following relationships are kept private. The final dataset for our analyses include 7,449 ego networks with 1,180,903 nodes and 1,658,069 ties by combining the two waves.

Measures

Unfollowing was measured by comparing the ego networks between Wave 1 and Wave 2. If a followee in Wave 1 has not been observed in the followee list of Wave 2, we consider it was unfollowed during the two waves. Among the 1,658,069 ties, 2.89% (47,962) were removed during the two waves. Among the

7,449 egos, 38.44% (2,864) have deleted at least one followee. We note that most independent variables were extracted from the first wave only. Therefore, we only included the users (egos and followees) who appeared in Wave 1 (7,326 egos and 1,613,735 ties) for formal analyses.

In order to measure information similarity and information redundancy, we employed text mining techniques. First, we created a term-document matrix for each ego and its followees (i.e., 7,326 term-document matrices in total). In each term-document matrix, rows are the users (including an ego and their alters) and columns are the unique hashtags in the users' tweets (all available tweets). Given that our sample consists of active users, about 90% of the alters have posted at least one hashtag. Second, the hashtags used by user u were encoded into a feature vector of term frequency–inverse document frequency (tf-idf) $\phi(u)$. The ith element $\phi_i(u)$ represents the frequency of the hashtag indexed by i in all the hashtags used by u, scaled by the inverse document frequency (see Salton, Wong, & Yang, 1975). The tf-idf value increases proportionally to the number of times a word appears in the document, but is offset by the frequency of the word in the corpus.

Hashtag similarity was measured at the dyadic level by the semantic similarity between the hashtags used by the egos and those used by their alters. As a result, the score quantifying the similarity of information posted by two users u_1 and u_2 is given by a cosine similarity measure: $\langle \phi_i(u_1), \phi_i(u_2) \rangle / (||\phi(u_1)|| \cdot ||\phi(u_2)||)$. Theoretically, hashtag similarity ranges from 0 (completely dissimilar) to 1 (actually the same). The mean of the hashtag similarity score in our data is 0.009 ($SD = 0.047$). Since the hashtag similarity was calculated using the semantic distance based on the tf-idf values, if two hashtags are referring the same topic (e.g., #politics and #Obama), users are inclined to use them together, and thus the semantic similarity score between them will be high.

Hashtag redundancy was also measured based on the tf-idf values. Words are not equally important in terms of their uniqueness. In fact, some words have little or no discriminating power. For example, one's followees are all university researchers. It is likely that all followees may include "research" as a hashtag. This word is purely redundant. TF-IDF is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus (Robertson, 2004). It quantifies the word importance relative to all words used by alters in an ego network. If a followee has many hashtags with high tf-idf values, it means that the user included many unique hashtags and showed less redundancy relative to other alters in the same ego network. Therefore, we calculated the sum of all tf-idf values for each followee as an indicator of information uniqueness (IU). And then we subtract the value using the maximum value to measure information redundancy ($IR = \min(IR))/(\max(IR) - \min(IR))$. We further normalized the raw score to range from 0 to 1 by using the formula: $(IR = \min(IR))/(\max(IR) - \min(IR))$. As a result, the mean of information redundancy is in our data 0.368 ($SD = 0.283$).

Information overload was measured at the alter level by calculating the tweeting frequency—the number of messages posted by an alter during the two waves. On average, a followee posted 984 ($Mdn = 82$, $SD = 5{,}609$) messages. Although the information overload is a subjective feeling, we measured it in a more objective way. In order to control this individual (subjective) heterogeneity, we modeled its effect in a multilevel framework with other control variables.

Popularity of a followee was measured by the ratio of the number of followers to the number of followees at Wave 1. The numbers are directly provided by Twitter's profile API. The average of popularity is 19,880 ($Mdn = 2$, $SD = 332{,}295$). Similarly, we calculated the popularity score for each ego user ($M = 0.90$, $Mdn = 0.45$, $SD = 12.06$).

Reciprocity was measured by a binary variable. If a followee of an ego was also a follower of the ego at Wave 1, we said the tie was reciprocal at Wave 1. We calculated the reciprocity only at Wave 1 and use it as a time-lag predictor of the unfollowing behavior in Wave 2. In our data, 36.8% of the ties are reciprocal.

To calculate the number of *common followees* between the ego and their alters at Wave 1, we collected the followees of the egos and the followees of the alters. We then compared the followee lists of the egos and the followee lists of the alters to determine the number of shared followees. In Figure 1, Ego 2 and Alter 5 shared a common followee — Alter 4. On average, the egos and alters shared 62 ($Mdn = 12$, $SD = 208$) followees in our dataset.

We included two types of control variables that have been investigated in previous studies (Kivran-Swaine et al., 2011; Kwak et al., 2012; Xu et al., 2013). The ego specific predictors are the characteristics of ego users, whereas alter specific predictors are the characteristics of the followees. Ego specific variables include ego popularity, the number of tweets, and years since registration, all of which are directly provided by Twitter's profile API. Alter specific variables include the number of tweets, year since registration, hashtag rate (the proportion of the tweets containing at least a hashtag), and interaction frequency. Interaction frequency was measured by the sum of the frequency of the ego retweeting its followees' tweets, the frequency of the ego replying to its followees, and the frequency of mentioning.

Finally, we include *the order of follow* as an alter-specific control variable. Twitter does not offer information about the establishment time of each relationship. However, it does provide the temporal order of the establishment of relationships in the personal network (Kwak et al., 2011). Similarly, we constructed the relative order in relationship establishment of followees for each ego user by breaking the followees into 10 groups. The final score ranges from 10% (the most recent 10% of followees) to 100% (the oldest 10% of followees).

Data Analysis

We used multilevel logistic regression (Snijders & Bosker, 2012) to test our hypotheses. The multilevel framework has been successfully employed to model (ego-centric) network formation problems (e.g., Golder & Yardi, 2010; Kivran-Swaine et al., 2011). In our study, the unit of analysis is the tie between egos and alters. Each following relationship nested under the same ego user could be influenced by the unique characteristics of that particular ego. We choose logistic as the link function because our dependent variables are binary responses (i.e., unfollowing or not). All alter-specific measures are Level-1 variables. All ego-level predictors are Level-2 variables.

Results

Repertoire Stabilization

Table 1 presents the formal models to predict the unfollowing behavior on Twitter. We calculated two types of R^2 for multilevel models (Nakagawa & Schielzeth, 2013): Marginal R^2 is concerned with variance explained by fixed factors, and conditional R^2 is concerned with variance explained by both fixed and random factors. The full model (the second column in Table 1) can explain 69.7% of the variance. We should note that much of the variance comes from the ego level. The Intraclass Correlation Coefficient for a null model without any predictors is 95.5%. In our data, nearly two-thirds of users didn't unfollow any users during our observations. The marginal R^2 of the full model is 6.1%

If we only focused on the users who had unfollowed at least once, the same model could explain 10.4% of the variance by the fixed factors.

The order of follow shows a significant effect on unfollowing, suggesting that people are less likely to unfollow the users who have connected for a relatively long time. In this way, users re-examine their recent followees and decide whether to keep them in their information repertoires. Finally, the repertoire becomes more and more stable. An alternative explanation is that the older followees might imply strong

Table 1 Multilevel Logistic Regression Models Predicting Unfollowing

	Model 1		Full Model		Raw Text Measures	
	Estimate (SE)	Z	Estimate (SE)	Z	Estimate (SE)	Z
Informational Variables						
Similarity	−2.312**(.313)	−7.39	−1.700**(.339)	−5.02	−2.349**(.604)	−3.89
Overload#	0.102**(.023)	11.55	0.083**(.009)	9.18	0.038**(.013)	2.87
Redundancy	−0.019 (.002)	−0.80	0.120**(.024)	5.00	0.071*(.031)	2.26
Overload × Similarity	−0.353*(.141)	−2.50	−0.458**(.139)	−3.28	0.201**(0.77)	2.63
Overload × Redundancy	−0.068**(.018)	−3.82	−0.032 (.018)	−1.79	0.049*(.020)	2.47
Similarity × Redundancy	1.489**(.502)	2.97	0.953 (.562)	1.70	1.868 (1.056)	1.77
Relational Variables						
Alter popularity#			−0.036**(.009)	−4.14	−0.036**(.009)	−4.19
Reciprocity			−1.518**(.017)	−89.58	−1.513**(.0170)	−89.43
Shared followees#			−0.151**(.007)	−21.23	−0.151**(.007)	−21.16
Control Variables						
Alter-Specific						
Order of follow (old)	−1.555**(.021)	−72.95	−1.452**(.022)	−66.80	−1.456**(.022)	−66.95
Interaction frequency	−0.259**(.0130)	−19.95	−0.219**(.013)	−17.13	−0.219**(.013)	−17.13
# of tweets#	0.004 (.007)	0.57	0.013 (.007)	1.72	0.019**(.007)	2.59
Year since registration	0.005 (.004)	1.26	−0.060**(.004)	−13.48	−0.062**(.004)	−14.13
Hashtag rate	0.006**(.052)	0.11	0.106*(.052)	2.01	0.021 (.052)	0.41
Ego-Specific						
Ego popularity#$	0.034**(.012)	2.79	−0.036**(.013)	−2.85	0.034**(.011)	2.96
# of tweets#$	0.143**(.034)	4.19	0.192**(.036)	5.35	0.173**(.032)	5.41
Year since registration$	−0.031 (.026)	−1.20	0.003 (.027)	0.10	0.004 (.026)	0.15
Intercept	−5.357**(.096)	−55.83	−5.119**(.101)	−50.69	−4.746**(.093)	−50.82
Model Summary						
Var. of intercepts across users	8.138 (2.853)		9.021 (3.003)		7.177 (2.679)	
Log-Likelihood	−119,115.4		−113,763.1		−113,979.6	
Conditional R^2	67.8%		69.7%		65.2%	
Marginal R^2	2.6%		6.1%		7.0%	
# of ties			1,613,735			
# of users			7,326			

Note. $^{**}p < .01$, $^{*}p < .05$

#Variables were rescaled using Z-scores ($M = 0$, $SD = 1$) for multilevel analyses.

$Variables that were measured at the ego level (i.e., level-2 variables). All tie measures and alter-specific measures are Level-1 variables. All ego-level predictors are Level-2 variables. The first two models were based on the hashtag measures, while the last model was based on the raw text measures of similarity and redundancy.

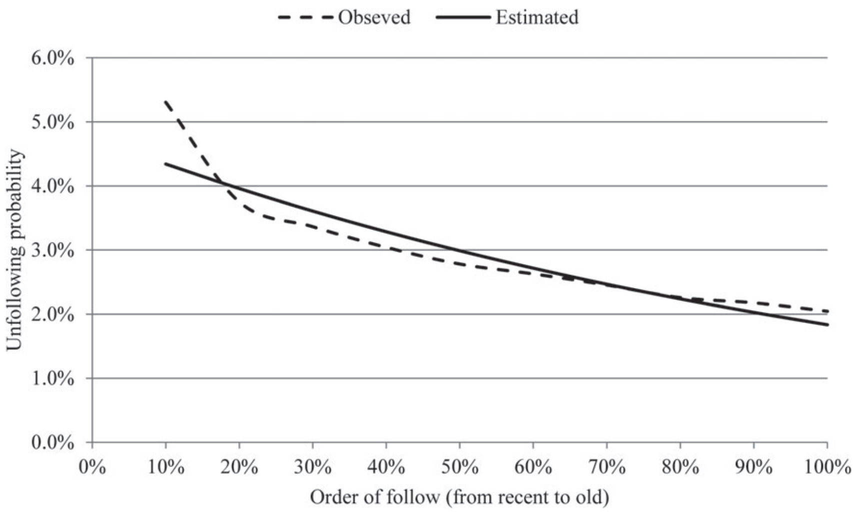

Figure 2 The observed and estimated probabilities of unfollowing as a function of the order of follow. The estimated probability was calculated based on the full model in Table 1.

ties, thus are less likely to be unfollowed. Figure 2 shows that even we controlled for the interaction frequency and other variables, the negative relationship between the order of follow and unfollowing probability still holds. The result implies users are intentionally stabilizing their information repertoires over time.

Informational Predictors

The major concern of the current study is to examine the role of informational factors in building personal repertoires of information sources. H1 stated that people are more likely to unfollow when they receive overloaded information. The full model in Table 1 shows that information overload is significantly associated with unfollowing, which means that users are more likely to unfollow the followees who posted too many tweets during the two waves ($B = 0.071$, $SE = .006$, $p < .01$). Therefore, H1 is supported.

H2 stated that users are less likely to unfollow the users sharing similar hashtags. The full model shows that hashtag similarity is negatively associated with unfollowing. That means users are more likely to keep the followees who tweeted similar hashtags ($B = -1.165$, $SE = .137$, $p < .01$). Therefore, H2 is supported.

H3 stated that hashtag similarity moderates the impact of information overload on unfollowing other users. Table 1suggests that the interaction effect of hashtag similarity and information overload on unfollowing is statistically significant ($B = -0.458$, $SE = .139$, $p < .01$). Figure 3 illustrates that the positive association between information overload and unfollowing is stronger when hashtag similarity between the ego and alter is low. From Figure 3, we also note that the difference of unfollowing probability between high and low similarity increases exponentially as information overload increases, indicating that information overload reinforces similarity-based selection. Therefore, H3 is supported.

H4 stated that users are more likely to unfollow the users whose tweets included many redundant hashtags. In the full model, hashtag redundancy is positively associated with unfollowing in the full model ($B = 0.120$, $SE = .036$, $p < .01$), suggesting that the increase of 1 hashtag redundancy will increase the probability of being unfollowed by 3%. Therefore, H4 is supported.

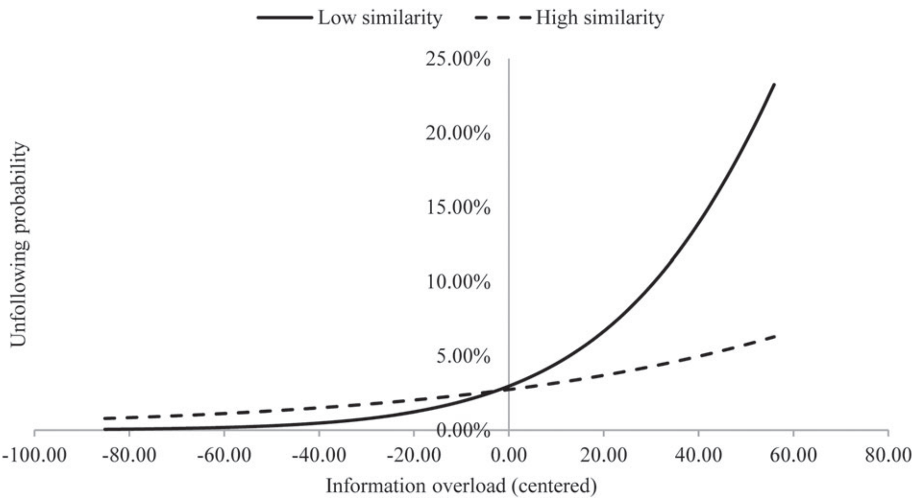

Figure 3 The interaction effect of information overload and similarity on unfollowing. The estimated probability was calculated based on the full model in Table 1.

Concerning H5 and H6, the full model suggests that both interaction effects are not significant at all. It suggests that the redundancy effect is not conditional on information overload and similarity when relational factors are controlled for. Therefore, H5 and H6 are not supported.

Relational Predictors

According to the full model in Table 1, all relational factors show significant impacts on unfollowing behavior. Consistent with previous studies, popularity is negatively associated with unfollowing ($B = -0.036$, $SE = .009$, $p < .01$). Ego users are less likely to unfollow users with more followers. Reciprocity is also negatively correlated with unfollowing ($B = -1.518$, $SE = .017$, $p < .01$). For followees with reciprocal ties with their egos at Wave 1, the probability of being unfollowed is 64% lower than the probability for the followees without reciprocal ties (82% versus 18%). Finally, the number of shared followees is negatively associated with unfollowing ($B = -0.151$, $SE = .017$, $p < .01$), indicating that the followees who share more followees with the egos are less likely to be unfollowed by the egos.

To answer the RQ, Model 1 in Table 1 excluded all relational factors. The inconsistency between Model 1 and the full model is caused by the exclusion of relational variables. First, the redundancy effect is no longer significant in Model 1. It indicates that relational factors are suppressors. In our data, information redundancy is positively correlated with reciprocity and the number of shared followees. For reciprocal ties, the followees' information redundancy is 0.43 on average, whereas the average information redundancy is 0.33 for nonreciprocal ties, $\chi^2(1, N = 1{,}613{,}733) = 50{,}676$, $p < .001$. The Spearman's rank correlation between the number of followees and information redundancy is 0.17 ($p < .001$). This means that high information redundancy implies dense connections (i.e., reciprocal and sharing more followees), which in turn decreases the unfollowing probability in Model 1. In addition, without considering the relational factors, the interaction effects with hashtag redundancy are significant in Model 1. The model suggests that users are less likely to unfollow the alters with redundant hashtags when information overload is high. It implies that users prefer information redundancy, which was produced by the relational factors.

Conclusion and Discussion

This study conceptualized social media followees as information source repertoires and examined the dynamics of repertoire formation using panel data from Twitter. First, this study suggests that users maintain relatively stable information repertoires to cope with information overload. During the 3 months of observation, only 5.56% of the following ties have been changed. Despite that, our findings suggest that some users actively and continuously adjust their information source repertoires over time. It is consistent with previous research (Kwak et al., 2011) that the new followees are most likely to be unfollowed, even when competing factors are controlled for. It implies that users are intentionally stabilizing their personal repertoires for daily information other than receiving it passively.

In our dataset, nearly two-thirds of users did not unfollow any users during our observations. This indicates that unfollowing actually is not a popular behavior on Twitter. However, it does not mean that unfollowing is a rare phenomenon or it lacks theoretical significance. We tracked the unfollowing behavior in a relatively short period of time. The number of users who have unfollowed other users should be much larger than 1/3. If we consider the frequency of unfollowing as an indicator for rational selection of information sources, the current study suggests that most users are not rational but habitual information consumers (Wood, Quinn, & Kashy, 2002). Instead of browsing all information channels, users would like to check information from a few sources repeatedly.

Second, this study extended the repertoire approach by examining the role of information overload, similarity, and redundancy in structuring information consumption patterns on a single social media platform. We found that seeking information similarity and reducing information redundancy could coexist in the process of optimizing information repertoires. One popular argument states that users are increasingly seeking content similar sources on social media. This is one of the important coping strategies people have for finding preferred content in an increasingly complex media environment (Webster & Ksiazek, 2012). Following this tendency, individuals would like to consume a steady diet of their preferred type of information sources. Finally, users with similar interests will cluster together (Himelboim et al., 2013) and cause information redundancy.

The current study indeed found that Twitter users are more inclined to keep those followees sharing similar hashtags. Under the information overload situations, the tendency of selecting content similar alters is reinforced (see Figure 3). However, the average hashtag similarity between egos and followees is only weakly associated with the average redundancy among the followees ($r = 0.038$, $t = 3.27$, $df = 7{,}267$, $p < .01$). The reason is that, as Table 1 suggests, people intentionally unfollowed the users with redundant information, even though they kept the similar alters at the same time. As a balance, their information repertoires contain the messages they are interested in and with very little redundancy. This also implies that people do have diverse interests and try to sample a diverse range of sources to build their information repertoires.

Third, we note that the formation process is significantly constrained by relational factors (i.e., popularity, reciprocity, and the number of common followees). In addition to their direct effects on unfollowing, the relational factors can alter the impacts of informational factors. We found that relational variables are suppressors of the redundancy effect. This implies that some users received unexpected and redundant information from their networked users. This relational constraint can also explain why previous research found that information overload is higher on Facebook than that on Twitter (Holton & Chyi, 2012), because relational constraint on Twitter is expected to be lower (e.g., Marwick & boyd, 2011). In addition, we hypothesized that the informational effects are conditional on each other. However, our results suggest that the redundancy effect is not dependent on information overload and similarity when the relational factors are controlled for.

Furthermore, although we focused on the information variables in building information repertoires, it does not mean that alternative explanations are impossible. On the contrary, our study is consistent with previous repertoire studies that the structural factors are more important that other factors (see Webster, 2014). The structural factors in the present study include the relational variables that characterize the online social networks and the control variables. For example, the low ratio of the number of followers to the number of followees (popularity) indicates that the users are inclined to keep more information sources. This finding is consistent with the idea of audience availability in television program choice (Webster & Wakshlag, 1983). Following many sources may suggest the users' availability in viewing new messages. However, these variables are at the microlevel or mesolevel in general. Future studies can explore the impacts of more macrolevel variables on the unfollowing behavior. As suggested by Webster (2014), the aggregate network level analysis would be beneficial to understand the bounded rationality of online user behaviors.

Limitations and Future Research

Several limitations can be associated with this study. First, when considering followees as information repertoire, we assume that users actually only read the messages posted by their followees. This assumption might not be accurate. Users can simply ignore the messages that they are not interested in to reduce information overload (Savolainen, 2007). In addition, users can receive messages beyond their immediate following networks. Social networks are not the only mechanisms through which users are directed to media. The recommender system and search engine are commonly used for direct audience attention on social media platforms (Webster, 2010). However, the following relationships do indicate awareness of the presence of the followees (Himelboim et al., 2013). Future studies can track users' browsing history on social websites to examine patterns of consuming specific messages other than sources.

Second, Twitter provides researchers with the unique opportunity to track patterns of individual selection of information sources. Although, the unobtrusive approach provides more objective measures, it lacks information on both demographic and psychological variables. Previous studies have found that demographic variables, such as gender and age, show significant impact on the composition of media repertoires (e.g., Yuan, 2011). In addition, users with different psychological characteristics may prefer different information-seeking approaches (Stefanone, Hurley, & Yang, 2013). Future studies need to further control these variables and examine the interaction effects between the self-reported and objective measures employed to build information repertoires.

In addition, the unobtrusive approach can cause potential measurement errors. For example, using tweeting frequency to measure information overload might be problematic. Even receiving the same amount of messages, some users may perceive more overload than would other users. We could not measure this subjective feeling directly. Instead, we employed the multilevel framework to control this individual heterogeneity carefully. First, the impact of tweeting frequency on the probability of being unfollowed by egos was considered separately for each ego. Furthermore, we included the potential compounding variables to control the individual differences. For example, Table 1 suggests that egos with more followees actually are less likely to unfollow other users, indicating that those users may have a higher threshold of information overload.

We measured information similarity and redundancy based on hashtags. This kind of operationalization was based on the repertoire approach to studying the user-defined channel types. Although the hashtag provides a convenient way to measure the content topics, 10% of followees did not post any hashtags in our sample. In the current study we considered them as "no preference" cases, i.e., the similarity and redundancy scores are zero. Another way to measure information similarity and redundancy is to calculate the variables based on the raw text. However, we think they are conceptually different things.

The purpose of the current study is to demonstrate that the user's choice of information sources is based on content topics other than using similar or unique words. As a robustness check, we conducted a post hoc analysis based on the raw text measures (see the last column in Table 1: Raw Text Measures). We found that the main effects are similar, whereas the interaction effects are slightly different. Future studies should use more advanced techniques to detect user-defined categories, such as the topic modeling approach, which is similar to factor analysis in media repertoire studies (see Weng et al., 2010).

Finally, social media platforms emphasize different technological characteristics. Our results rely on Twitter, which puts a greater emphasis on news sharing. For other social media platforms, like Facebook, studies may emphasize social networking. In this sense, users might be less susceptible to the information variables than was the case in our study. Furthermore, previous repertoire studies have demonstrated that people could build their personal repertoires across media platforms or rely on one of them. The choice of different repertoires is associated with user background characteristics (e.g., Kim, 2014). For studies based on a single platform, it is difficult to capture more general media use patterns. For example, watching TV news intensively may cause information overload or redundancy on Twitter. Therefore, future studies are encouraged to test our hypotheses across different social media platforms.

Note

1 This study was supported by the Public Policy Research Funding Scheme of the Central Policy Unit of the Government of the Hong Kong Special Administrative Region (2013.A8.009.14A) and the Small Project Funding from The University of Hong Kong (201409176011).

References

Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science,* 286(5439), 509–512. doi:10.1126/science.286.5439.509

Chaffee, S. H. (1982). Mass media and interpersonal channels: Competitive, convergent, or complementary. In G. Gumpert & R. Cathcart (Eds.), *Inter/media: Interpersonal communication in a media world* (3rd ed., pp. 62–80). New York, NY: Oxford University Press.

Dutta-Bergman, M. J. (2004). Complementarity in consumption of news types across traditional and new media. *Journal of Broadcasting & Electronic Media,* 48(1), 41–60. doi:10.1207/s15506878jobem4801_3

Farhoomand, A. F., & Drury, D. H. (2002). Managerial information overload. *Communications of the ACM,* 45(10), 127–131. doi:10.1145/570907.570909

Ferguson, D. A., & Perse, E. M. (2000). The World Wide Web as a functional alternative to television. *Journal of Broadcasting & Electronic Media,* 44(2), 155–174. doi:10.1207/s15506878jobem4402_1

Franz, H. (1999). The impact of computer mediated communication on information overload in distributed teams. *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences, 1999, HICSS-32, 6182030*. doi:10.1109/HICSS.1999.772712

Golder, S. A., & Yardi, S. (2010). *Structural predictors of tie formation in twitter: Transitivity and mutuality.* Paper presented at the 2010 IEEE Second International Conference on Social Computing (SocialCom).

Harrigan, N., Achananuparp, P., & Lim, E. P. (2012). Influentials, novelty, and social contagion: The viral power of average friends, close communities, and old news. *Social Networks,* 34(4), 470–480. doi:10.1016/j.socnet.2012.02.005

Hermida, A., Fletcher, F., Korell, D., & Logan, D. (2012). Share, like, recommend: Decoding the social media news consumer. *Journalism Studies,* 13(5–6), 815–824. doi:10.1080/1461670x.2012.664430

Himelboim, I., Hansen, D., & Bowser, A. (2013). Playing the same Twitter network: Political information seeking in the 2010 US gubernatorial elections. *Information Communication & Society*, 16(9), 1373–1396. doi:10.1080/1369118x.2012.706316

Holton, A. E., & Chyi, H. I. (2012). News and the overloaded consumer: Factors influencing information overload among news consumers. *Cyberpsychology Behavior and Social Networking*, 15(11), 619–624. doi:10.1089/cyber.2011.0610

Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York, NY: New York University Press.

Kim, S. J. (2014). A repertoire approach to cross-platform media use behavior. *New Media & Society*, 1–20. doi:10.1177/1461444814543162

Kivran-Swaine, F., Govindan, P., & Naaman, M. (2011). *The impact of network structure on breaking ties in online social networks*: *Unfollowing on Twitter*. Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems.

Kwak, H., Chun, H., & Moon, S. (2011). *Fragile online relationship*: *A first look at unfollow dynamics in twitter*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Paper presented at the Proceedings of the 19th international conference on World wide web*.

Kwak, H., Moon, S. B., & Lee, W. (2012). *More of a receiver than a giver: Why do people unfollow in Twitter?* Paper presented at the ICWSM.

Liang, H., & Fu, K. W. (2015). Testing propositions derived from Twitter studies: Generalization and replication in computational social science. *Plos One*. doi:10.1371/journal.pone.0134270

Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society,* 13(1), 114–133. doi:10.1177/1461444810365313

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology,* 27, 415–444. doi:10.1146/annurev.soc.27.1.415

Murthy, D. (2012). Towards a sociological understanding of social media: Theorizing Twitter. *Sociology-the Journal of the British Sociological Association,* 46(6), 1059–1073. doi:10.1177/0038038511422553

Myers, S. A., & Leskovec, J. (2014). *The bursty dynamics of the twitter information network.* Paper presented at the Proceedings of the 23rd International Conference on World Wide Web.

Myers, S. A., Sharma, A., Gupta, P., & Lin, J. (2014). *Information network or social network?: The structure of the twitter follow graph.* Paper presented at the Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution,* 4(2), 133–142. doi:10.1111/j.2041-210x.2012.00261.x

Quercia, D., Bodaghi, M., & Crowcroft, J. (2012). *Loosing friends on Facebook.* Paper presented at the Proceedings of the 4th Annual ACM Web Science Conference.

Robertson, S. (2004). Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation,* 60(5), 503–520. doi:10.1108/00220410560582

Salton, G., Wong, A., & Yang, C.-S. (1975). A vector space model for automatic indexing. *Communications of the ACM,* 18(11), 613–620. doi:10.1145/361219.361220

Savolainen, R. (2007). Filtering and withdrawing: Strategies for coping with information overload in everyday contexts. *Journal of Information Science,* 33(5), 611–621. doi:10.1177/0165551506077418

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.

Stefanone, M. A., Hurley, C. M., & Yang, Z. J. (2013). Antecedents of online information seeking. *Information Communication & Society,* 16(1), 61–81. doi:10.1080/1369118x.2012.656137

Stephens, K. K., Barrett, A. K., & Mahometa, M. J. (2013). Organizational communication in emergencies: Using multiple channels and sources to combat noise and capture attention. *Human Communication Research,* 39(2), 230–251. doi:10.1111/hcre.12002

Sunstein, C. R. (2009). *Republic.com 2.0.* Princeton, NJ: Princeton University Press.

Taneja, H., Webster, J. G., Malthouse, E. C., & Ksiazek, T. B. (2012). Media consumption across platforms: Identifying user-defined repertoires. *New Media & Society,* 14(6), 951–968. doi:10.1177/1461444811436146

Watson-Manheim, M. B., & Belanger, F. (2007). Communication media repertoires: Dealing with the multiplicity of media choices. *Mis Quarterly,* 31(2), 267–293.

Webster, J. G. (2010). User information regimes: How social media shape patterns of consumption. *Northwestern University Law Review,* 104(2), 593–612.

Webster, J. G. (2014). *The marketplace of attention: How audiences take shape in a digital age*: MIT Press.

Webster, J. G., & Ksiazek, T. B. (2012). The dynamics of audience fragmentation: Public attention in an age of digital media. *Journal of Communication,* 62(1), 39–56. doi:10.1111/j.1460-2466.2011.01616.x

Webster, J. G., & Wakshlag, J. J. (1983). A theory of television program choice. *Communication Research,* 10(4), 430–446. doi:10.1177/009365083010004002

Weng, J., Lim, E.-P., Jiang, J., & He, Q. (2010). *Twitterrank: finding topic-sensitive influential twitterers.* Paper presented at the Proceedings of the third ACM international conference on Web search and data mining.

Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology,* 83(6), 1281–1297. doi:10.1037//0022-3514.83.6.1281

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). *Who says what to whom on Twitter.* Paper presented at the Proceedings of the 20th international conference on World Wide Web.

Xu, B., Huang, Y., Kwak, H., & Contractor, N. (2013). *Structures of broken ties: Exploring unfollow behavior on twitter.* Paper presented at the Proceedings of the 2013 conference on Computer supported cooperative work.

Yuan, E. (2011). News consumption across multiple media platforms: A repertoire approach. *Information Communication & Society,* 14(7), 998–1016. doi:10.1080/1369118x.2010.549235

About the Authors

Hai Liang is Assistant Professor in the School of Journalism and Communication at the Chinese University of Hong Kong. His research interests include political communication, dynamic communication process, social media analytics, and computational social science. E-mail: hailiang@cuhk.edu.hk. Address: School of Journalism and Communication, Room 424, Humanities Building, New Asia College, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

King-wa Fu is Associate Professor at the Journalism and Media Studies Centre, The University of Hong Kong. His research interests cover political participation and media use, computational media studies, health and the media, and younger generation's Internet use. E-mail: kwfu@hku.hk. Address: Journalism and Media Studies Centre, Room 206, Eliot Hall, Pokfulam Road, The University of Hong Kong, Hong Kong.