

Xiao Wang:

You process content and I process context: Cross-platform divergence of retweetability between Twitter and Weibo

RESUMÉ

Artiklen undersøger forskelle i brugeres retweeting-adfærd på tværs af platformene Twitter og Sina Weibo. Med en heuristisk-analytisk informationsbehandlingsmodel for retweeting, som udspringer af dual-proces-teori, rationaliserer denne undersøgelse sammenligningen af brugeradfærd for de to platforme på basis af bruger-centrerede tværkulturelle kognitive forskelle. Resultaterne viser, at når der træffes beslutninger om, hvorvidt et indlæg skal retweetes, er brugere på Twitter mere tilbøjelige til at anvende en analytisk strategi baseret på informationsbehandling af indholds faktorer i sammenligning med brugere på Weibo, der er mere tilbøjelige til at benytte en heuristisk strategi baseret på informationsbehandling af kontekstuelle faktorer.

ABSTRACT

This article examines the cross-platform divergence between Twitter and Sina Weibo in users' retweeting behaviors. With a heuristic-analytic information-processing model of retweeting proposed on the basis of dual-process theory, this study rationalizes the cross-platform comparison by introducing user-centered cross-cultural cognitive differences. Results show that when making decisions about whether to retweet a post, users on Twitter are more likely to use an analytic strategy based on information processing of content factors compared to users on Weibo, who are more likely to adopt a heuristic strategy based on information processing of contextual factors.

EMNEORD

Microblogging, Twitter, Weibo, retweetability, heuristisk-analytisk informationsbehandling

KEYWORDS

Microblogging, Twitter, Weibo, retweetability, heuristic-analytic information processing

Introduction

As a member of the big family of social media that is straddling the borderline between traditional blogs and social networking sites, the microblog is a “group of Internet-based applications that build on the ideological foundations of Web 2.0, and that allow the creation and exchange of user-generated content” (Kaplan and Haenlein 2010, 61). Twitter, the first microblog website, made its debut in 2006 and started acting independently from April 2007. Twitter’s unique product concept and the favorable developmental momentum attracted not only a tremendous number of users but also a variety of ‘imitators’ around the world. As one of the undisputed giants in China’s internet universe, Sina Corporation also jumped into this market. It initiated the internal test of Sina Weibo in August 2009 and made this microblogging trial public in October.¹ The latest data shows that by March 2015, Twitter had 288 million monthly active registered accounts compared with 167 million for Sina Weibo.²

The most easily conceptualized information transitivity function afforded by microblogs is the *retweet*, which is literally equivalent to reposting but has its etymological roots in Twitter. In practice, if a microblogger finds an absorbing tweet when scanning his/her timeline, he/she is simultaneously offered a manual (i.e. using ‘RT@’ or ‘Via@’ as the input prefix of the retweeted post) and a one-click reweeting method of passing this post on from his/her followee (i.e. the original poster) to his/her followers.

Microbloggers retweet for various purposes, such as notification, entertainment, comment, agreement, and storage for future use (boyd et al. 2010). Users are believed to undergo a decision-making process when being exposed to tweets (Liu et al. 2012). That is to say, users’ decisions concerning whether to retweet an observed post are likely to correlate with numerous determinants contained in that post. This pragmatic idea is closely related to users’ information-processing behavior and forms the basis for a more down-to-earth question: How and why do certain posts, whether originals or

¹ <http://ir.weibo.com/phoenix.zhtml?c=253076&p=irol-irhome>

² <http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

retweets, receive more attention and popularity than others, i.e. garner greater “retweetability” (Suh et al. 2010)?

Retweetability can be defined either from a user-centered perspective (as a post’s *retweeting worthiness* perceived by its readers) or from a post-centered perspective (as the *objective probability* for a post to be re-posted by its readers). Attempts at understanding retweetability necessitate the examination of concealed but perceivable factors that can predict retweeting. A number of quantitative studies, despite their lack of a well-defined conceptual framework, mostly indicate a content-contextual categorization for the determinants of retweetability (e.g. Suh et al. 2010; Liu et al. 2012; Zhang et al. 2014). As far as content factors are concerned, a posts’ informativeness (e.g. type of topic; post length; post attractiveness; availability of multimedia; access to supplementary indexes or extensions such as hashtag, URL, and mention) plays a key role in impacting users’ evaluation of retweeting. As far as contextual factors are concerned, a poster’s credibility (e.g. trustworthiness, expertise, activity, authoritativeness, experience) shows significant influence on a microblogger’s retweeting judgment.

In terms of product architecture, Twitter and Sina Weibo stick to two strikingly different design and management philosophies. These product-based differences may prompt users on each platform to focus on a given series of post features when evaluating retweeting. However, in the context of the present study, a more noteworthy distinction between two platforms lies in the cultural identities of their user bases. We may ask whether cross-cultural cognitive differences are likely to persuade microbloggers on the different platforms to follow distinct principles (e.g. holistic vs. analytic; Nisbett et al. 2001) in the decision-making process with respect to whether to retweet a post. This study represents a systematic examination of the cross-platform divergence of retweeting behavior. First, a heuristic-analytic information-processing model of retweeting will be constructed on the basis of the dual-process theory. In light of the users’ geographic distribution on two microblog platforms, the holistic vs. analytic cognitive differences between Chinese and Western users will be addressed as the grounds for the proposed distinction between Weibo and Twitter in relation to the factors impacting retweets. With a rationalized hypothesis and a corresponding operational framework, this study employs the random digit search (RDS; Zhu et al. 2011) sampling

method. Sampled data is then incorporated into both linear regression and a general linear model for analysis.

Theoretical framework and research hypothesis

Although previous studies have explored retweetability on Twitter or Weibo, neither the data analysis design nor the categorization of determinants in these studies are conducive to comparability. In addition, these empirical contributions are mostly built upon either a superficial or nonexistent theoretical basis (e.g. Suh et al. 2010; Ma et al. 2013; Zhang et al. 2014), leading one to doubt the plausibility of their content-context scales. A more structured review of two distinct cognitive systems should thus be undertaken in order to provide a solid theoretical orientation for further empirical efforts.

Heuristic-analytic information processing model of retweeting

Microblog users are believed to undergo a judging and decision-making process when reading posts online (Liu et al. 2012). Retweetability is thus likely to correlate with various components of a particular post. If we employ dichotomous categorization to distinguish between a post's contextual and content factors, the dual-process theory – particularly notions that emphasize the interaction between two processing stages – becomes an appropriate guide for building an information-processing model of microbloggers' retweeting behavior.

Bearing in mind the variable terminologies used in the development of dual-process theory (e.g. *heuristic/systematic*, Chaiken 1980; *system1/system2*, Stanovich 1999), the current study adopts the terms *heuristic-analytic* (Evans 2010), hereafter making use of its typicality and understanding of convenience. As presented in Table 1, users' heuristic processing of contextual factors (e.g. posters' characteristics) is rapid, parallel, and automatic, with low cognitive effort and less involvement of consciousness. To reduce the information-seeking cost (Zhang et al. 2014), users take advantage of heuristic tactics to facilitate the recognition of contextualized properties and the synchronization between post information and existing knowledge in implicit memory. In contrast, users' analytic processing of content factors (e.g. information quality) is slow, sequential, and controlled with high cognitive effort and more involvement of consciousness. Users make use of normative rationality rather

than evolutionary rationality (Evans 2010) when processing content elements, which necessitate the engagement of working memory.

Heuristic processing of contextual factors	Analytic processing of content factors
Contextualized	Decontextualized
Fast	Slow
Parallel	Sequential
Automatic	Controlled
Low effort	High effort
Implicit knowledge	Explicit knowledge

Table 1. Major characteristics of heuristic and analytic processing of posts

As the processing of content and contextual factors in all probability takes place simultaneously while browsing an aggregation of posts, this study shows a preference for an interactive and concurrent rather than sequential model for interpreting information-processing behavior prior to retweeting. The heuristic-analytic information-processing model for retweeting proposed by this study is displayed in Figure 1.

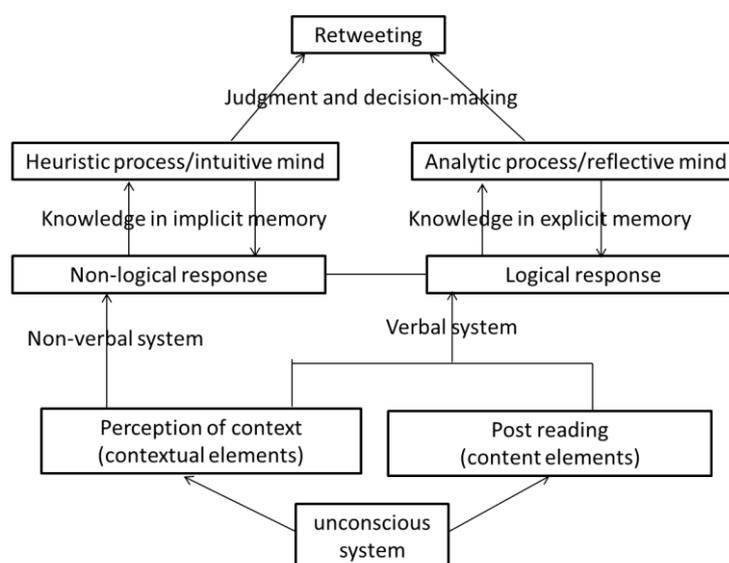


Figure 1. The information-processing model of retweeting (adapted from the two minds model, Evans, 2010)

On the bottom of the model, users' context perception and post reading are both supported by the unconscious system. At one end, most but not all contextual factors (e.g. activity, relation involvement, and visualized influence of the original poster) have activated users' non-verbal system to generate non-logical responses. As the information overload makes it impracticable for a user to manage each microblog post equally (Lang 2000), these responses are basically devoted to the tacit processing of focus and contextualized attributes of posts. They furthermore seek to synchronize the recognized information with existing knowledge in implicit memory without the engagement of the explicit system. This synchronization enables a microblogger to use his/her intuitive mind to heuristically process a post and eventually decide whether to retweet. At the other end, all content factors (e.g. information interactivity, vividness, and affectiveness) and a few text-based contextual factors (e.g. self-description details of the original poster) have activated users' verbal system to generate logical responses. With the engagement of the explicit system, especially the working memory, a microblogger makes use of his/her reflective mind to analytically process a post before deciding whether to retweet.

Grounds for the cross-platform divergence

The upsurge of social media is changing the world's communicative landscape in that physical proximity no longer plays a decisive role for users expanding their social networks (Leetaru et al. 2013). In other words, the classic sociological consideration of strong-weak ties (Granovetter 1973; Lin et al. 1981) is digitally weakened when microbloggers make decisions on the target they are inclined to follow. However, the microblogger's dual identity as both an internet user and a real-world individual makes it essential to assess cultural properties before judging platform properties. Microblogging platforms across various cultural contexts might thus differ from one another in the determinants for retweetability, largely because of the differing cultural properties to which these platforms are subject.

a. Users' geographic distribution: Western majority vs. Chinese majority

Leetaru et al. (2013) mapped the global Twitter trend in their 'Global Twitter Heartbeat' project. They streamed 1,535,929,521 tweets from 71,273,997 users over the course of 40 days and coded each geo-located tweet by both the Place

Location and the Exact Location. Results of both the city rank by the percentage of geo-referenced tweets or retweets and the global distribution map of most retweeted cities showed that users in areas influenced by Western society, particularly the United States and Western Europe, perform the most intensive tweeting activities on Twitter.

The 2012 annual report of Sina Weibo conducted a two-week survey with a sample size of approximately 101,000.³ As the backstage data revealed, of 368,000,000 accounts registered till July 2012, more than 300,000,000 accounts were registered locally in China (including Taiwan, Hong Kong, and Macau). Although Weibo has long-term plans to extend its overseas business, the number of overseas users can hardly exceed 10% of the total.⁴ Together with the leading role of China-based Weibo, consolidated after its appearance on NASDAQ,⁵ and Twitter's inaccessibility in mainland China,⁶ it is straightforward to conclude that Weibo is a Chinese-majority social platform.

b. Holistic vs. analytic cognitive differences

Social psychologists have long argued and experimentally demonstrated that cognitive patterns differ from culture to culture. Among these cross-cultural contributions, a *holistic vs. analytic* perspective stands out as universally accepted (Peng and Nisbett 1999). Based on the social orientation hypothesis that the cognitive differences are embedded in a long tradition of science, mathematics, philosophy, and social systems, *holistic thought* is defined as "involving an orientation to the context or field as a whole, including attention to relationships between a focal object and the field, and a preference for explaining and predicting events on the basis of such relationships." This is compared to *analytic thought*, which is defined as "involving the detachment of the object from its context, a tendency to focus on attributes of the object to assign it to categories, and a preference for using rules about the categories to explain and predict the object's behavior" (Nisbett et al. 2001, 293).

³ <http://wenku.baidu.com/view/94a2bb8d83d049649b6658be.html>

⁴ <http://web2.iresearch.cn/weibo/20110607/141253.shtml>

⁵ <http://www.chinabgao.com/info/74803.html>

⁶ <http://money.cnn.com/gallery/news/2014/05/07/banned-china/index.html>

Although each culture on this globe might take up a relatively stable position within the holistic-analytic range, an Eastern-Western antithesis has attracted the most research interest. As for the visual perception, Chinese Americans might find it much more difficult than European Americans to detach an object from its embedded environment (Ji et al. 2000). Specifically, Chinese Americans have offered more “whole card” description when responding to Rorschach cards (Abel and Hsu 1949) and have performed significantly better on the background-involved information than their counterparts in both the classical Rod & Frame Test (Witkin et al. 1954) and the free recall task (Park et al. 1999). This attentional distinction marked by field independence vs. field dependence can be experimentally achieved by the application of not only the quiescent physical stimuli mentioned above but sometimes also dynamic or even realistic scenes (Masuda and Nisbett 2001).

The holistic-analytic cognitive differences may also affect how people from different cultures make judgments and undertake reasoning with regard to a particular piece of information (Norenzayan and Nisbett 2000). On the basis of the hypothesis that the long-term attitude towards formal logic in ancient sciences and philosophy might reverberate in the cognitive processes of people living in modern society, Nisbett *et al.* (2001) suggest that when judging the convincingness or the soundness of formal arguments, East Asians might be more reliant than Americans on prior beliefs and experience-based tactics. For instance, Chinese and Korean university students have been shown to take advantage of more intuitive strategies than do European American students, who are apt to use formal logic in reasoning (Norenzayan et al. 2002a). Conversely, European American students are more inclined to disregard prior beliefs and experiential knowledge in favor of formal logic when the logical structure of an argument (e.g. a syllogism) conflicted with common sense (Norenzayan et al. 2002b).

The influence of holistic-analytic cognitive differences might be felt throughout the entire information-processing model of retweeting. At the bottom of the model, differences in attentional pattern tend to influence users’ unconscious system in guiding them to attend to different sets of visual elements. Accordingly, users on the Western-majority Twitter tend to detach the post’s content from its context and to focus on content elements rarely intercepted by context elements, whereas users on the Chinese-majority Weibo are inclined to

attend to the web page as a whole and are less capable of detaching the post's content from its context. In other words, field-independent attention might lead microbloggers on Twitter to concentrate on reading the post, while field-dependent attention of users on Weibo probably sets aside a larger proportion of their cognitive capacities for perception of context.

At the top of the model, differences in reasoning might affect the principles by which users abide when judging whether a post is worth retweeting. With more logical responses activated by the verbal system, Twitter users are more likely to draw upon the explicit system (e.g. working memory) and seek suggestions from the formal logic of the post content in the subsequent judgment. In contrast, Weibo users are more likely to synchronize their non-logical responses, which might be predominantly activated by the context perception via the non-verbal system, with their knowledge of the implicit system (i.e. intuitive strategies represented by prior belief and experiential knowledge) before deciding whether to retweet a post.

As a result, the holistic vs. analytic thought systems are likely to be reflected in users' information-processing behaviors in the microblogging world. The cross-platform divergence of factors influencing retweetability might be largely shaped by cross-cultural cognitive differences with respect to issues such as attention and reasoning.

H: When judging a post's retweetability, Twitter users make greater use of analytic processes/reflective strategies than do Sina Weibo users; whereas Sina Weibo users employ more heuristic processes/intuitive strategies than do Twitter users.

In other words: When deciding whether to retweet a post, Twitter users are more likely to make a judgment based on information processing of content factors compared to Weibo users, who are more likely to make a judgment based on information processing of contextual factors.

Operational framework

This theoretically oriented empirical study devotes the next step to building a structured operational framework to underpin further analysis. The similarity and comparability of matched pairs of variables are particularly important for ensuring the validity of measurements when testing cross-platform differences.

It becomes, therefore, necessary to exclude non-comparable factors from this framework. In this case, a post's content factors are operationalized as various aspects of its information quality, whereas a post's contextual factors are operationalized as various dimensions of information credibility. Two confounding factors are also controlled to further validate the cross-platform comparison.

Content factors: information quality

As this study plans to adopt computerized large-scale data rather than manual data-coding (Naaman et al. 2010) in testing the hypothesis, the assessment of content factors must go beyond semantic and syntactic constraints, such as a post's topical categorization. A framework characterized by 'information quality', consisting of five operational dimensions, is thus proposed as follows.

a. Information load (post length)

When users are processing posts on microblog pages, the information load functions as an important benchmark "beyond accuracy" (Wang and Strong 1996, 5) for their perception of information quality. In this study, a post's information load is operationalized as the post length in characters. The post length is defined as the number of characters instead of words. The APIs of both platforms set the character as the default counting mode, and characters seem to provide a well-rounded data presentation, taking emoticons, punctuations, and URLs into account.

b. Information vividness (extended rich-media information)

As the input limit is set as 140 characters on both Twitter and Weibo, users tend to visualize their text messages by attaching an extended piece of rich-media information. Information vividness is operationalized as a dichotomous variable that represents whether or not a post contains one or more attached pictures (including animated GIFs). Video and other forms of rich media information are excluded due to the incommensurability of the two platforms.

c. *Topic highlighting (hashtag)*

Posts with distinctive topics, such as world events, enjoy more popularity among microbloggers (Zhao et al. 2011). A typical Twitter hashtag takes the form of a character '#' seamlessly followed by the topical focus. On Weibo, however, the hashtag is presented as two '#', between which the topical focus is inserted, for instance, '#BrazilWorldCup#'. The hashtag feature functions as a tool for categorizing trending topics. The contents of a post involving a hashtag normally have a topical focus initiated by a user or a background operation. The topical highlighting is thus operationalized as a dichotomous variable that represents whether or not a post involves a hashtag.

d. *Information indexing and extension (URL)*

In order to clarify news sources and improve information completeness (Zhang et al., 2014), microblog users tend to insert URLs into their newsworthy posts (Castillo et al. 2011). Once clicked upon, the URL will direct users to an 'enlarged' version of the original post, helping them obtain a clear view of causes and effects. As such, information indexing and extension are operationalized as a dichotomous variable that measures whether one or more URLs are included in a post.

e. *Information interactivity (mention)*

Online chatting and news reporting are just two of the diverse usages identified from microbloggers' public timelines (Java et al. 2007). To promote mutual communication and strengthen online relationships, users often take the initiative to start a conversation rather than remain restrained by the tweet-retweet mode. Mentions are a feature that make point-to-point communication available for microbloggers. People use '@' to notify or simply draw the attention of their targets. Information interactivity is thus operationalized as a dichotomous variable that indicates whether one or more mentions are included in a post.

Contextual factors: information credibility

A social media user must shoulder the responsibility of acting as both a content provider and an acquisition gatekeeper (Haas & Wearden, 2003). This dual

identity has prompted concern among researchers with regards to social media's content credibility (e.g. Flanagin & Metzger, 2000). In a microblogging context, the perception of microblog credibility may play a crucial role in users' heuristic processing of posts (Yang et al. 2013) based on the reflective mind. Additionally, a poster's user attributes (e.g. physical appearance, trustworthiness, reputation, and personal experience) have been found to exert a significant influence on the evaluation of the perceived credibility of web-based information (Flanagin and Metzger 2000; 2007). As a consequence, a contextual framework consisting of six operational dimensions closely related to the information credibility was established.

a. Poster's degree of expressive activity (number of statuses)

Statuses consist of tweets and retweets a user has posted on his/her timeline. The number of statuses acts as a suitable indicator of the user's devotion to and activity in the microblogging community. A study concerning the social awareness streams on microblogs found that approximately four-fifths of the original message content on Twitter can be categorized as self-expression (Naaman et al., 2010). People frequently provide updates on their up-to-the-minute situations as well as personal feelings, partly giving rise to the fact that the average status on Twitter consists of just 9.4 words (Leetaru et al., 2013). The poster's degree of expressive is operationalized as a scale variable that represents the number of statuses a user has posted on the timeline.

b. Poster's degree of evaluational activity (number of favorites)

As the personal accumulation of saved records (possibly for further reference), the number of favorites quantifies the efforts a user has made to evaluate others' content (e.g. agreement, disagreement, disputation, controversy). The poster's degree of evaluational activity is operationalized as a scale variable that represents the number of favorites a user has selected since his/her microblogging debut.

c. Poster's degree of relationship involvement (number of followees)

A poster's initiative to connect with others represents a rarely explored dimension in the assessment of information credibility. A user's following behavior implies acceptance that the followee's acts will be incorporated

into the user's personal timeline. As a measurement of a user's following size, the number of followees is an out-degree indicator that measures the user's willingness to be involved in microblogging relationships. The poster's degree of relationship involvement is operationalized as a scale variable that represents a user's number of followees.

d. Poster's degree of self-disclosure (self-description)

Microbloggers are empowered to tailor the visibility of their private lives to their own needs. The more private information a user has disclosed, the more trust and information credibility he/she earns (Zhang et al. 2014). As an aspect of self-disclosure, the self-description has a prominent appearance in a user's profile. Users are offered freedom to provide and update that information which they would like others to know about themselves. Self-introduction, personal motto, and life insights are among the various common forms that the self-description may take. The poster's degree of self-disclosure is thus operationalized as a dichotomous variable that weighs whether or not a user has provided a self-description in his/her profile.

e. Poster's visualized influence (number of followers)

Although a microblogger's networked power is a multidimensional concept that cannot be quantified by investigating a single factor, number of followers appears to be the value that affords the most direct visualization for users. This in-degree measurement renders a user's audience size overt to visitors, thereby providing an intuitive cue for all to perceive a user's popularity in this cyber community. Moreover, the number of followers also has a strong linear relationship with the retweet rate on Twitter (Suh et al. 2010). In this study, the poster's visualized influence is operationalized as a scale variable that represents a user's number of followers.

f. Poster's experience (duration of account)

The online experience in a given UGC community performs as a relatively stable user attribute attached to a microblogger. Moreover, internet experience positively correlates to the perception of web-based information credibility (Flanagin and Metzger 2000; 2007). The author argues that the richness of microblogging experience cannot simply be reflected by the

intensity of online behaviors (e.g. posting frequency) but is typically also represented by the age of a user's account. As such, the poster's experience is operationalized as a scale variable that measures the account duration since its first registration. Duration is calculated in months, rounded up.

Confounding factors

a. Account verification

Despite the important role played in users' heuristic perception of a poster's authoritativeness and expertise (Zhang et al. 2014), account verification does not fit into the current contextual framework due to the lack of comparability between the two platforms. The marketing strategy and product orientation advocated by Sina Corporation have made it much easier for users, especially small-scale or even virtual organizations, to create authenticated and verified accounts on Weibo than on Twitter. As a result, this dichotomous variable (i.e. whether or not a user has been systematically verified) is controlled in further analysis.

b. Posting time

The microblogging cycle is a chain of peaks and troughs. Krishnamurthy *et al.* (2008) examined the local time of day when statuses were updated and found that the Twitter workload exhibited an ascent during late-morning hours, a descent in the small hours of the morning, and a relatively smooth frequency of use at other times throughout the day. Following this suggestion, the posting time is also regarded as a confounding factor since posts published during the online rush hours are apt to have a higher retweet rate. The posting time is dichotomized into rush hour (from 8 AM to 12 AM) and non-rush hour (from 12 AM to 8 AM).

Figure 2 provides a graphical summary of the operational framework:

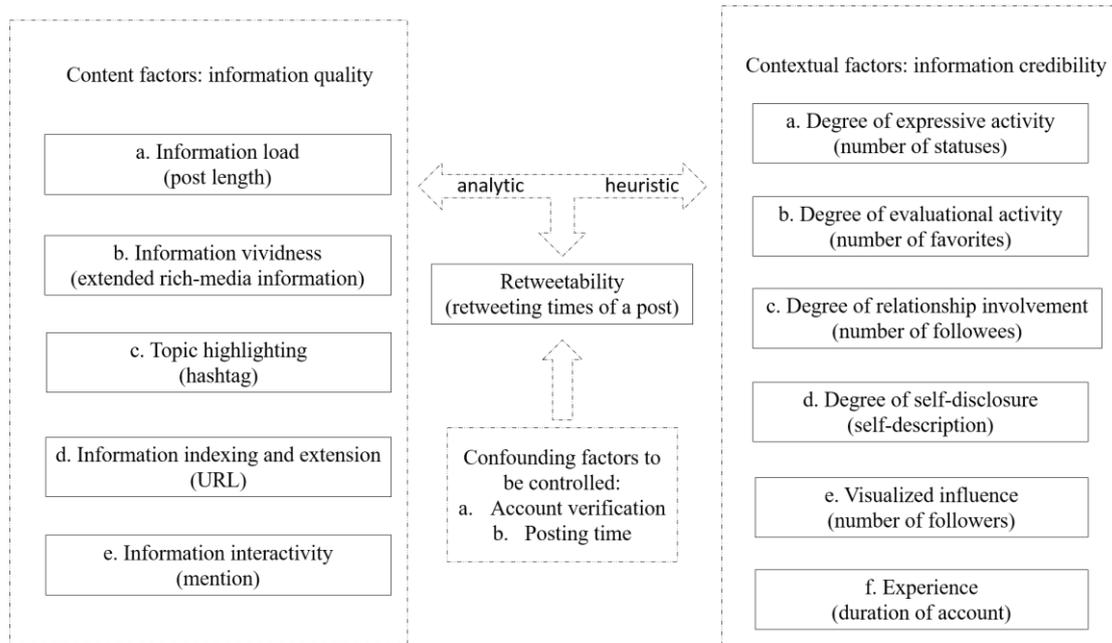


Figure 2. Graphical summary of the operational framework

Method

As neither the manually collected nor the randomly crawled data is capable of ensuring a probability sample for the user-generated content, the random digit search method (RDS, Zhu et al. 2011) was applied to the sampling process. RDS makes full use of the numeric user ID to generate probability samples that consistently exhibit an acceptable approximation of the population parameters of a certain platform. In other words, the probability sampling of users functions as the basis of the probability sampling of target data. Bearing in mind the computerized procedures running throughout RDS, a professional software developer was consulted for practical advice and provided technical support during the sampling process.

APIs provided by both platforms⁷ were constantly utilized during this probability sampling of both users and posts. Consequently, the number of valid sampled IDs amounted to 23,881 for Weibo and 22,081 for Twitter. The time frame was set as 1 July 2014 to 31 July 2014, and all the original tweets posted by sampled users within this time span were entered into the database for further analysis. Retweets were excluded to guarantee the model's

⁷ <https://dev.twitter.com/streaming/overview> & [http://open.weibo.com/wiki/微博 API](http://open.weibo.com/wiki/微博_API)

applicability and reduce the inherent uncertainty of both heuristic and analytic cues presented to users when judging the retweetability of a retweet *per se*. As a time difference might exist for both in-platform and between-platform users, the time frame was executed in accordance with the posting time (i.e. the local time in the poster's time zone), thereby avoiding incomparability of variables such as the rush/non-rush time on the respective platforms. The author began carrying out the data crawling at approximately 8:00 AM (Beijing Time) on 1 August 2014 to achieve the greatest possible reduction in numeric changes to the variables, such as the number of retweets. In this way, 178,288 posts were collected from 22,081 users on Twitter and 190,815 posts were collected from 23,881 users on Sina Weibo. As the coding unit is each individual post collected from the sampled users' personal homepages during a certain period of time, this study utilized MySQL⁸ (MySQL Workbench 6.1 CE), one of the most popular relational database management systems, to retrieve and store relevant data from all of the collected posts.

All of the 14 variables contained in the operational framework were extracted from the 369,103 posts. For the data analysis, the author first conducted a data reduction using the factor analysis to verify the construct validity of the operational framework. Variables passing through the factor analysis were entered into the linear regression and the general linear model (GLM) analyses, which were adopted to manage the major focus areas of this study. The entire process of data analysis was performed separately for Twitter and Weibo.

Results

In order to evaluate the construct validity of the operational framework, two independent factor analyses were performed for the data collected from Twitter and Weibo respectively. Coordinate transformation represents the mathematical principle of factor analysis. It aims to extract a relatively small number of disparate comprehensive indicators, namely factors from a group of original variables observed in practice. This data reduction tactic can thus achieve both dimension reduction and variable categorization for the operational framework. All 11 of the variables with respect to content and contextual factors were incorporated into the analysis. Results of the factor

⁸ <http://www.mysql.com/products/workbench/>

analysis provided an acceptable discriminant validity for the operational framework, which formed the basis for the subsequent rationalization of the linear regression and GLM.

Linear regression

Linear regression is usually adopted to analyze the linear relationship between multiple independent variables and one dependent variable. To distinguish three sets of independent variables (i.e. content variables, contextual variables, and confounding variables) in the operational framework, this study entered them into the regression model in a stepwise manner. As such, retweeting time was set as the dependent variable, and blocks for content factors, contextual factors, and control variables were included in the model sequentially.

For Twitter (Table 2), the retweetability shows significant Pearson correlation with all of the independent variables except for duration of account and posting time. 10 out of 13 independent variables included in the model show significant standard coefficients with retweetability. These 10 variables are ranked by their influence on retweetability from high to low as follows: extended rich media (.032), hashtag (.027), number of followers (.023), number of followees (-.016), URL (-.014), number of favorites (.014), length (.013), mention (-.008), account verification (.007), and self-description (.005). This sequence demonstrates that content factors afford more probabilities than do contextual factors in influencing Twitter users' retweeting decisions.

	Retweeting times			
	Pearson Correlation	Std. Coefficients (Beta)	Sig.	Std. Error
a_Length	.015***	.013	.000	.000
a_Extended_rich_media	.035***	.032	.000	.064
a_Hashtag	.025***	.027	.000	.047
a_URL	-.011***	-.014	.000	.043
a_Mention	.012***	-.008	.001	.049
b_Number_of_statuses	.009***	.000	.903	.000
b_Number_of_favorites	.017***	.014	.000	.000
b_Number_of_followees	.012***	-.016	.000	.000
b_Self_description	.007***	.005	.035	.049
b_Number_of_followers	.023***	.023	.000	.000
b_Duration_of_account	.001	-.002	.530	.001
c_Account_verification	.019***	.007	.008	.514
c_Posting_time	-.001	.000	.932	.058

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 2. Linear regression analysis of Twitter

For Sina Weibo (Table 3), the retweetability shows significant Pearson correlation with all of the independent variables except for information interactivity (mention) and posting time. 9 out of 13 independent variables included in the regression model show significant standard coefficients with retweetability. These 9 variables are ranked by their influence on retweetability from high to low as follows: number of followers (.058), number of followees (-.031), length (.027), extended rich media (.024), URL (-.016), hashtag (.015), self-description (.012), account verification (.012), number of favorites (.006). In contrast to the above regression model for Twitter, this sequence explicitly shows that contextual factors have greater influence on Weibo users' retweeting behavior than do content factors.

	Retweeting times			
	Pearson Correlation	Std. Coefficients (Beta)	Sig.	Std. Error
a_Length	.026***	.027	.000	.000
a_Extended_rich_media	.026***	.024	.000	.064
a_Hashtag	.017***	.015	.000	.047
a_URL	-.009***	-.016	.000	.043
a_Mention	.002	-.002	.340	.049
b_Number_of_statuses	.017***	.004	.147	.000
b_Number_of_favorites	.009***	.006	.014	.000
b_Number_of_followees	.011***	-.031	.000	.000
b_Self_description	-.119***	.012	.000	.049
b_Number_of_followers	.042***	.058	.000	.000
b_Duration_of_account	.008***	.005	.074	.001
c_Account_verification	.035***	.012	.000	.514
c_Posting_time	.001	.002	.466	.058

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 3. Linear regression analysis of Sina Weibo

General linear model (GLM)

The linear regression analysis has revealed some preliminary associations between retweeting totals and two categories of factors. However, the standard coefficients obtained above still seem incapable of encompassing the explanatory power of both categories of factors for predicting users' retweeting behavior. The general linear model (GLM) was adopted to further examine the main hypothesis. GLM not only goes a step further than the multivariate regression model by allowing for linear transformations or linear combinations of dependent variables; it also provides a solution for the normal equations when variables are not linearly independent.⁹ As in the linear regression, independent variables were entered into the model in a stepwise manner. Once again, retweeting total was set as the dependent variable, and content factors, contextual factors, and control variables were included in the GLM sequentially. R squared change (ΔR^2) is the variation of R squared once an independent variable is included in or excluded from the model. In this sense, ΔR^2 serves as a proper indicator of the statistical contribution each category of

⁹ http://www.uta.edu/faculty/sawasthi/Statistics/stglm.html#reg_extension

factors makes in explaining a post's retweetability. Results of the GLM are shown in Table 4.

Step-wise sequence	Retweeting times					
	Twitter			Sina Weibo		
	R ²	Adjusted R ²	ΔR^2	R ²	Adjusted R ²	ΔR^2
Content factors included	8.0%	7.8%	8.0%	6.3%	6.0%	6.3%
Contextual factors included	10.3%	8.4%	2.3%	12.8%	9.7%	6.5%
Control variables included	16.3%	13.7%	6.0%	21.1%	18.3%	8.3%
Total R ²		16.3%			21.1%	

Table 4. Results of GLM analysis

Overall, 13 independent variables (control variables included) carry more weight in explaining the retweeting behavior of Weibo users (21.1%) than that of Twitter users (16.3%). More importantly, the hypothesis proposed in the current study has once again been strongly validated by the contrast between the two microblogging platforms in terms of how the variance in retweeting is explained by two categories of factors. For Twitter, content factors account for 8.0% of the variance in retweeting totals, which overwhelmingly outstrips the 2.3% that is accounted by contextual factors. For Sina Weibo, content factors account for 6.3% of the variance in retweeting times, whereas contextual factors account for 6.5% of the variance in a post's retweetability. The ΔR^2 afforded by contextual factors is even slightly higher than that afforded by content factors. In short, content factors carry more weight than contextual factors in explaining a Twitter post's retweetability, while contextual factors carry more weight in explaining a Weibo post's retweetability. Results achieved from the linear regression analysis are thus further authenticated by the GLM. That is, when making decisions about whether to retweet a post, users on Twitter are more likely to make a judgment based on the information processing of content factors/analytic cues when compared to users on Weibo, who are more likely to make a judgment based on the information processing of contextual factors/heuristic cues.

Conclusion, limitations, and expectations

This study aims to examine the cross-platform divergence between Twitter and Sina Weibo users' retweeting behaviors. With a heuristic-analytic information-processing model of retweeting established on the basis of dual-process theory, this study rationalized the comparison by introducing the cognitive differences between Chinese and Western users. Results of both linear regression and GLM proved that when evaluating the retweetability of microblog posts, users on Twitter make more use of an analytic process/reflective tactics than do users on Weibo. In contrast, users on Weibo employ more a heuristic process/intuitive tactics than do users on Twitter.

This study makes significant scholarly contributions. The information-processing model of retweeting has implications for further explorations into information-processing and sharing behaviors on social media. Moreover, the ready-made operational framework can be adjusted or refined to make it compatible with further efforts to model post popularity. The practical value of this study, however, extends beyond this. For ordinary users, who tend to maintain multiplatform usage of social media, this study offers a clear-cut processing model of information sharing, which can help users adopt different usage rules on different sites and minimize potential impacts brought about by undesirable posts such as rumors, spam, and unwanted advertising. For those public opinion and business accounts, a clear recognition of the underlying mechanism for cross-platform differences can assist in the targeted adjustment of content features and social marketing strategies. For government agencies, situational awareness may be strengthened to make a better use of microblogs as a channel for delivering political information, thereby avoiding communicational obscurity.

Limitations and expectations of this study also go hand-in-hand. First, the sampled data is all from original tweets. Retweets were excluded from the sample to guarantee the model's applicability. Future studies could fit tweets and retweets into the model as a whole to test the applicability of the contextual-content scale. Second, an inevitable deficiency in the use of big data is a lack of self-reported data, with the result that subjective measurements of users' information-processing and sharing behaviors on social media should be considered in the future research design. Moreover, a much more comprehensive explanatory framework for the cross-platform divergence in

retweetability can be expected in future studies. With respect to the user-centered causes, Western-Eastern differences – either in social cognition, such as causal attributions and social inferences (Morris et al. 1995), or in behavioral norms, such as societal tightness (Gelfand et al. 2006) and face-work uniquely attached to the high-context Chinese society (Oetzel & Ting-Toomey 2003) – are of specific value and must be examined to explain the results of the current study. Conversely, we must ask whether the cross-cultural differences have been attenuated simply because we are increasingly converging on similar websites or applications with similar programming norms.

Author's note

I would like to thank all of the individuals who helped me throughout the process of organizing and writing this article, especially my advisor Dr. Liu Jun and censor Dr. Lisbeth Kjastrup for consistently providing valuable advice and motivating me to get my paper published, as well as Mr. Wang Jinghua for the programming support during the sampling process.

References

- Abel, Theodora M., and Francis LK Hsu. 1949. "Some aspects of personality of Chinese as revealed by the Rorschach Test." *Journal of Projective Techniques* 13(3): 285-301.
- Boyd, Danah, Scott Golder, and Gilad Lotan. 2010. "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter." Paper presented at HICSS, Kauai, HI, US, January 6.
- Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. 2011. "Information credibility on twitter." Paper presented at the 20th international conference on World Wide Web, Hyderabad, India, March 28-April 1.
- Chaiken, Shelly. 1980. "Heuristic versus systematic information processing and the use of source versus message cues in persuasion." *Journal of Personality and Social Psychology* 39(5): 752-766.

- Evans, Jonathan St BT. 2010. *Thinking twice: Two minds in one brain*. Oxford University Press.
- Flanagin, Andrew J., and Miriam J. Metzger. 2000. "Perceptions of Internet information credibility." *Journalism & Mass Communication Quarterly* 77(3): 515-540.
- Flanagin, Andrew J., and Miriam J. Metzger. 2007. "The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information." *New Media & Society* 9(2): 319-342.
- Gelfand, Michele J., Lisa H. Nishii, and Jana L. Raver. 2006. "On the nature and importance of cultural tightness-looseness." *Journal of Applied Psychology* 91(6): 1225-1244.
- Granovetter, Mark S. 1973. "The strength of weak ties." *American Journal of Sociology* 81(5): 1287-1303.
- Haas, Christina, and Stanley T. Wearden. 2003. "E-credibility: Building common ground in web environments." *L1-Educational Studies in Language and Literature* 3(1-2): 169-184.
- Java, Akshay, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. "Why we twitter: understanding microblogging usage and communities." Paper presented at the 9th WebKDD and 1st SNA-KDD, San Jose, CA, US, August 12.
- Ji, Li-Jun, Kaiping Peng, and Richard E. Nisbett. 2000. "Culture, control, and perception of relationships in the environment." *Journal of Personality and Social Psychology* 78(5): 943-955.
- Kaplan, Andreas M., and Michael Haenlein. 2010. "Users of the world, unite! The challenges and opportunities of social media." *Business Horizons* 53(1): 59-68.
- Krishnamurthy, Balachander, Phillipa Gill, and Martin Arlitt. 2008. "A few chirps about twitter." Paper presented at the 1st workshop on Online Social Networks, New York, NY, US.

- Lang, Annie. 2000. "The limited capacity model of mediated message processing." *Journal of Communication* 50(1): 46-70.
- Leetaru, Kalev, Shaowen Wang, Guofeng Cao, Anand Padmanabhan, and Eric Shook. 2013. "Mapping the global Twitter heartbeat: The geography of Twitter." *First Monday* 18(5-6).
- Lin, Nan, Walter M. Ensel, and John C. Vaughn. 1981. "Social resources and strength of ties: Structural factors in occupational status attainment." *American Sociological Review* 46(4): 393-405.
- Liu, Zhiming, Lu Liu, and Hong Li. 2012. "Determinants of information retweeting in microblogging." *Internet Research* 22(4): 443-466.
- Ma, Haixin, Weining Qian, Fan Xia, Xiaofeng He, Jun Xu, and Aoying Zhou. 2013. "Towards modeling popularity of microblogs." *Frontiers of Computer Science* 7(2): 171-184.
- Masuda, Takahiko, and Richard E. Nisbett. 2001. "Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans." *Journal of Personality and Social Psychology* 81(5): 922-934.
- Morris, Michael W., Richard E. Nisbett, and Kaiping Peng. 1995. "Causal attribution across domains and cultures." In D. Sperber, D. Premack, and A. J. Premack (Eds.), *Causal cognition*, 577-614, New York: Oxford University Press.
- Naaman, Mor, Jeffrey Boase, and Chih-Hui Lai. 2010. "Is it really about me? message content in social awareness streams." Paper presented at the conference on Computer Supported Cooperative Work, Savannah, GA, US, February 6-10.
- Nisbett, Richard E., Kaiping Peng, Incheol Choi, and Ara Norenzayan. 2001. "Culture and systems of thought: holistic versus analytic cognition." *Psychological Review* 108(2): 291-310.
- Norenzayan, Ara, Incheol Choi, and Richard E. Nisbett. 2002. "Cultural similarities and differences in social inference: Evidence from behavioral predictions and lay theories of behavior." *Personality and Social Psychology Bulletin* 28(1): 109-120.

- Norenzayan, Ara, and Richard E. Nisbett. 2000. "Culture and causal cognition." *Current Directions in Psychological Science* 9(4): 132-135.
- Norenzayan, Ara, Edward E. Smith, Beom Jun Kim, and Richard E. Nisbett. 2002. "Cultural preferences for formal versus intuitive reasoning." *Cognitive Science* 26(5): 653-684.
- Oetzel, John G., and Stella Ting-Toomey. 2003. "Face concerns in interpersonal conflict a cross-cultural empirical test of the face negotiation theory." *Communication Research* 30(6): 599-624.
- Park, Denise C., Richard Nisbett, and Trey Hedden. 1999. "Aging, culture, and cognition." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 54(2): 75-84.
- Peng, Kaiping, and Richard E. Nisbett. 1999. "Culture, dialectics, and reasoning about contradiction." *American Psychologist* 54(9): 741-754.
- Stanovich, Keith E. 1999. *Who is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Erlbaum.
- Suh, Bongwon, Lichan Hong, Peter Pirolli, and Ed H. Chi. 2010. "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network." Paper presented at the International Conference on Social Computing, Minneapolis, MN, US, August 20-22.
- Wang, Richard Y., and Diane M. Strong. 1996. "Beyond accuracy: What data quality means to data consumers." *Journal of Management Information Systems* 12(4): 5-33.
- Witkin, Herman A., Helen B. Lewis, Max Hertzman, Karen Machover, P. Bretnall Meissner, and Seymour Wapner. 1954. *Personality through Perception*. New York: Harper.
- Yang, Jiang, Scott Counts, Meredith Ringel Morris, and Aaron Hoff. 2013. "Microblog credibility perceptions: comparing the USA and China." Paper presented at the conference on Computer Supported Cooperative Work, San Antonio, TX, US, February 23-27.
- Zhang, Lun, Tai-Quan Peng, Ya-Peng Zhang, Xiao-Hong Wang, and Jonathan JH Zhu. 2014. "Content or context: Which matters more in

information processing on microblogging sites." *Computers in Human Behavior* 31(2): 242-249.

Zhang, Yang, Zhiheng Xu, and Qing Yang. In press. "Predicting Popularity of Messages in Twitter using a Feature-weighted Model." *International Journal of Advanced Intelligence*.

Zhao, Wayne Xin, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. 2011. "Comparing twitter and traditional media using topic models." Paper presented at the European Conference on Information Retrieval, Dublin, Ireland, April 18-21.

Zhu, Jonathan JH, Qian Mo, Fang Wang, and Heng Lu. 2011. "A random digit search (RDS) method for sampling of blogs and other user-generated content." *Social Science Computer Review* 29(3): 327-339.