More spreaders but less influence: An analysis of indirect influence on Twitter

Xin Shuai[a], Ying Ding[b], Jerome Busemeyer[c], Shanshan Chen[c], Yuyin Sun[b], Jie Tang[d]

[a]*School of Informatics and Computing, Indiana University Bloomington, IN, USA*
[b]*School of Library and Information Science, IN, USA*
[c]*Dept. of Psychological and Brain Science, Indiana University Bloomington, IN, USA*
[d]*Dept. of Computer Science and Technology, Tsinghua University, Beijing, China*

Abstract

Social influence in social networks has been extensively researched. Most studies have focused on direct influence, while another interesting question can be raised as *whether indirect influence exists between two users who're not directly connected in the network and what affects such influence.* In addition, the theory of *complex contagion* tells us that more spreaders will enhance the indirect influence between two users. Our observation of intensity of indirect influence, propagated by *n* parallel spreaders and quantified by retweeting probability in two Twitter social networks, shows that complex contagion is validated globally but is violated locally. In other words, the retweeting probability increases non-monotonically with some local drops. A *quantum cognition* based probabilistic model is proposed to account for these local drops.

Keywords: Twitter, Complex Contagion, Quantum Cognition

1. Introduction

Thanks to the fast development of Web2.0, many online social networks have emerged, where the observation of information diffusion, or social influence, in large-scale data becomes possible. Social influence has been studied by many researchers, including the validation of influence (Anagnostopoulos et al., 2008; Crandall et al., 2008; Gomez Rodriguez et al., 2010), the propagation of influence among multiple types of social media (Gruhl et al., 2004; Cha et al., 2009; Hong et al., 2011), the maximization of influence spread in the whole network (Kempe et al., 2003; Chen et al., 2009), and the probabilistic modeling of direct influence (Xiang et al.,

2010; Tang et al., 2009; De Choudhury et al., 2007). However, most relevant studies focused on direct influence, while another relevant question regarding social influence may be raised as to *whether a user can exert indirect influence on his/her friends' friends and what affects such influence.* Normally, multiple intermediate persons called spreaders are involved in the indirect communication between two persons, i.e., the sender and the receiver. Those spreaders may have a combinational effect on the indirect influence propagated from the sender to the receiver.

Some have studied the problem of *serial indirect influence* as shown in Figure 1(a), where the node *A* exerts indirect influence on the

1

node *B* through the *n* serial intermediate nodes, C_1 from C_n. For example, Fowler & Christakis (2008) found that happiness spreads among people as far as 3 hops. Liu et al. (2010) tried to quantify the indirect influence of n-degree friends at the topical level. However, very few studies have focused on another type of indirect influence, which we call *parallel indirect influence*, as shown in Figure 1(b). $C_1, ..., C_n$ nodes are located in parallel between node *A* and node *B* to propagate influence from the former to the latter.

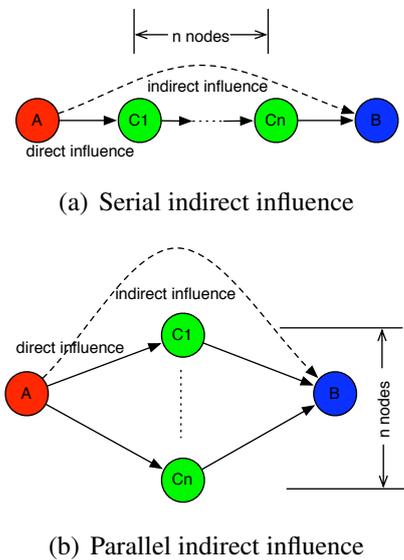(a) Serial indirect influence

(b) Parallel indirect influence

Figure 1: Two types of indirect influence

A concept closely related to parallel indirect influence is *complex contagion*. Unlike *simple contagion*, which can spread in social networks after just one contact with a single infected neighbor like a disease, complex contagion is a phenomenon where multiple sources of exposure to a new idea are required before an individual adopts the idea (Centola & Macy, 2005). That is to say, repeated exposures of an individual to an idea recommended by his/her multiple neighbors positively affect the probability he/she will eventually follow that idea. Romero et al. (2011) studied the

spread of hashtags in Twitter and quantified the probability of a user adopting a new hashtag as the function of the number of his/her neighbors who have already adopted it. They found that the spread of political hashtags validates the complex contagion, where the adoption probability increases monotonically as the number of neighbors who had already adopted the same hashtags increases, until a plateau is finally reached. By contrast, for idiom hashtags, complex contagion does not take effect, and the adoption probability decays rapidly when more neighbors have adopted the same hashtags.

The problem we are studying is similar to Romero et al. (2011), but we focus on message spread behavior and indirect influence on Twitter. A concrete example of this is shown in Figure 2, where Alice sends out original mes-
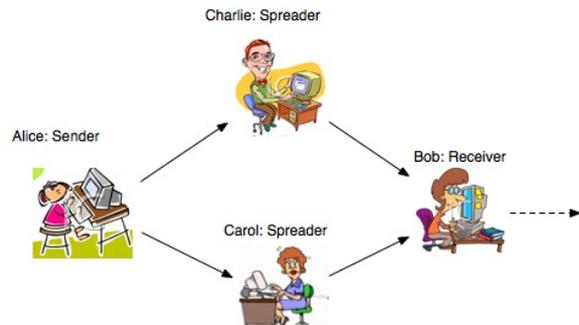

Figure 2: Typical information spread in a social network

sages, Charlie and Carol further spread Alice's messages (i.e., by retweeting) and Bob finally receives them. After that, Bob may choose to further spread Alice's messages to others, just like his two neighbors Charlie and Carol have done, or not. Here, the intent of Bob to further spread Alice's messages would reflect the intensity of the indirect influence of Alice on Bob, which can be measured as the probability that Bob will further spread Alice's messages, given that Charlie and Carol have already spread these messages. If complex con-

2

tagion takes effect, the influence intensity will be higher when both Charlie and Carol spread Alice's messages than when either or none of the two spread them.

In this paper, we examine the intensity of indirect influence as the function of the number of parallel spreaders between two users on Twitter who don't have direct following relations. We found that *complex contagion* is observed globally but is violated locally. Especially, when the number of spreaders increases from one to two, there's an obvious drop in the intensity. The newly emerging field of *quantum cognition* is applied to interpret the local drops in terms of interference effect on the process of decision-making. Recently, an article from NewScientist indicated how humans may actually think in a "quantum" manner (Buchanan, 2011). Research from cognitive science has also provided some initial evidence of quantum-like cognitive interference in human decision-making (Khennikov, 2010; Busemeyer et al., 2009). These cognitive experiments showed that the classical law of total probability was violated. Instead, quantum probability (Gudder, 1988) was applied to explain the experimental results. In addition, quantum cognition has been employed to further advance the theory of information retrieval (IR) (Piwowarski et al., 2010; Zuccon et al., 2009; Zhang et al., 2010).

Our main contributions are:

- Examined the change of parallel indirect influence between the sender and the receiver, quantified by retweeting probability, with the number of spreaders and found that such probability increases non-monotonically with some local drops;

- Verified the existence of complex contagion in the indirect influence on the global scale;

- Proposed a probabilistic model based on quantum cognition to explain local drops in retweeting probability.

This paper is organized as follows: Section 2 defines the problem. Section 3 shows the experimental results. Section 4 proposes quantum cognition model. Section 5 lists related literature and Section 6 concludes the study.

2. Problem Definition

Twitter users send and read messages called tweets, which contain no more than 140 characters. One user can read another user's messages by following them. In addition, one user's message can be re-sent by his followers via retweeting. A retweeting message starts with the identifier "*RT @username*". Such following/follower relationships connect Twitter users and form the social network where information flows through retweeting. Given a collection of tweets $C = \{t\}$, V represents all Twitter users while $E = \{(u, v) \mid u, v \in V\}$ represents all following relations where u follows v. We provide several formal definitions as follows:

- DEFINITION 1. [Following Triple] $\forall t$ starting with "*RT @y: RT @x*" posted by z, we build a following triple $T_{xyz} = (x, y, z)$, $x, y, z \in V$ and claim that $(z, y) \in E$ and $(y, x) \in E$. We also define $C(T_{xyz})$ as the total count of tweets that belongs to T_{xyz} and $C(v), v \in V$ as the total count of tweets v posted.

- DEFINITION 2. [Spreaders] $\forall a, b \in V$, we define spreaders between a and b as $S_{ab} = \{y \mid T_{ayb} \neq NULL, y \in V\}$.

- DEFINITION 3. [N-spreader Retweeting Pattern] $\forall a, b \in V$ we define a retweeting pattern $P_{ab} = \{T_{ayb} \mid y \in S_{ab}\}$ and $|S_{ab}| =$

n. Consequently, we define a n-spreader retweeting pattern as $P_n = \{P_{ab} \| S_{ab}| = n\}$, and P_{ab} is an *instance* of P_n.

- DEFINITION 4. [Retweeting Probability] $\forall P_{ab} \neq \varnothing$, we define the probability of *b* retweeting from *a* as $Pr(b|a; S_{ab}) = \sum_{y \in S_{ab}} C(T_{ayb})/C(a)$. Consequently, we define the retweeting probability of n-spreader retweeting pattern as $Pr(n) = \sum_{P_{ab} \in P_n} Pr\{b|a; S_{ab}\}/|P_n|$.

- DEFINITION 5. [Indirect Influence] $\forall x, z \in V \cap P_{xz} \neq \varnothing$, we think *x* exerts indirect influence on *z*. $Pr(n)$ indicates the average intensity of indirect influence in n-spreader retweeting pattern.

Starting from 1-spreader retweeting pattern to illustrate how information spreads in Twitter (Figure 3(a)), where *B* follows *C* and *C* follows *A*. We assume that *B* does not directly follow *A* so it can only read *A*'s messages through *C*. *A* posts a tweet *T* and *C* reads and retweets it by adding the sign *RT @A*. Such information flow is represented by solid arrays because it is actually observed. Then *B* reads *T* and decides whether to further retweet it by adding a sign *RT @C*. The information flow from *B* is represented by a dashed array because it is not an actual but a potential flow. We can measure the intent of *B* to further propagate *A*'s messages transferred by *C* using the retweeting probability $Pr(B|A; C)$. For instance, *A* posts a total of 100 tweets and 20 of them are retweeted by *C*. Then *B* further retweets 5 of the 20 tweets, which makes $Pr(B|A; C) = 5/100 = 0.05$.

In Figure 3(b), two spreaders C_1 and C_2 are involved in transferring messages from *A* to *B*. Assume that *A* posts 100 tweets, C_1 retweets 15 of them and C_2 retweets 30 of them. Then *B* retweets 5 of C_1's 15 tweets and 10 of C_2's 30 tweets. As a result, $Pr(B|A; C_1, C_2) = (5 + 10)$
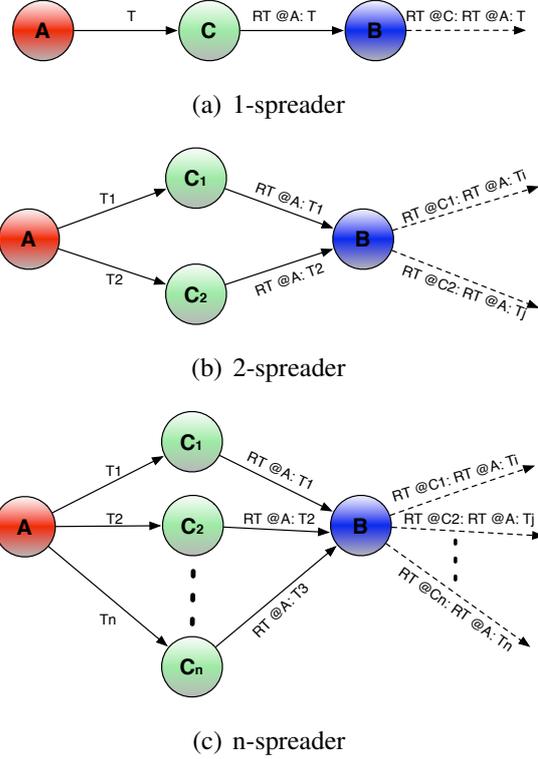

(a) 1-spreader

(b) 2-spreader

(c) n-spreader

Figure 3: Retweeting patterns with different number of spreaders

$/100 = 0.15$. The general n-spreader retweeting pattern is shown in Figure 3(c), and our research question can be formulated as: *Given n spreaders, how does the curve Pr(n) change with n?*

According to complex contagion, $Pr(n)$, which represents the intensity of indirect influence, increases monotonically with *n*, since more spreaders involved in transferring the sender's messages leads to higher probability of the receiver to further propagate the sender's messages. We examined the probability in the real Twitter data to see whether this is true.

3. Results

In this section, we show the results of retweeting probability in two different Twit-

4

ter datasets and testify the existence of complex contagion. One dataset consists of tweets from the public timeline that can be seen as a whole social network, while the other consists of tweets from a certain user's ego network.

3.1. Dataset1

The first dataset[1] contains 467 million tweets from 20 million Twitter users from June to December 2009 , which covers 20%-30% of total public tweets during this period. Figure 4 shows the distribution of n-spreader retweeting pattern. As the number of spreaders increases, the corresponding number of instances drops, indicating that the situation where too many spreaders are involved is infrequent. The maximum number of spreaders we found is 29.
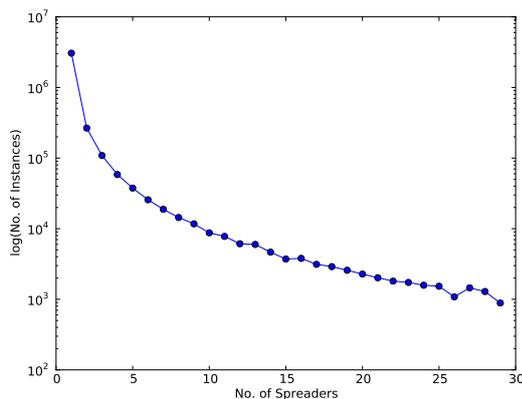
Figure 5: The curve of $Pr(n)$ in dataset1

Figure 4: Distribution of the number of instances in all n-spreader retweeting patterns in dataset1

Figure 6: Distribution of the number of instances in all n-spreader retweeting patterns in dataset2

We plotted the curve $Pr(n)$ for $n = 1$ to 10 in Figure 5. The reason that we didn't show the part of the curve for $n > 10$ is that the relatively rare occurrence of patterns with a large number of spreaders makes the result subject to random disturbance and yields unreliable observations. From Figure 5 we can see that
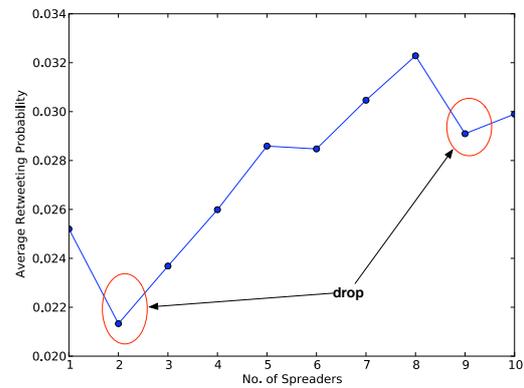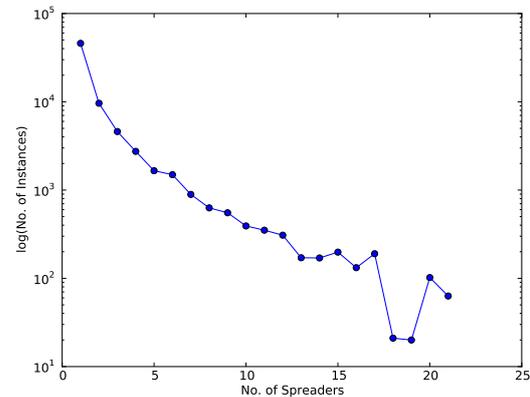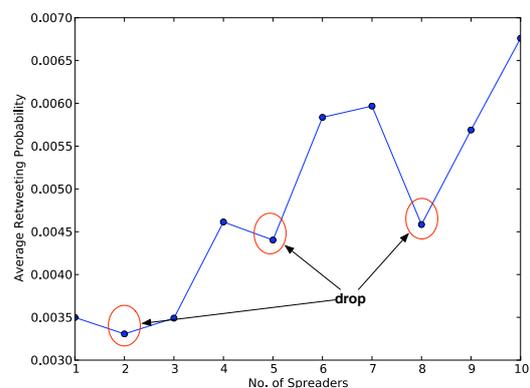
Figure 7: The curve of $Pr(n)$ in dataset2

the global trend of $Pr(n)$ is increasing as n increases. That is to say, overall, the intensity of indirect influence tends to become higher, or at least persists, as more spreaders are included, which validates the phenomenon of complex contagion in the global level. However, there are two drops spotted in $Pr(n)$, i.e. from n =1 to 2 and 8 to 9. To testify the statistical significance of the two decreased probability values we observed, we used t-test (Gosset, 1970) of *difference between two means*[2] since retweeting probability is a type of mean value according to DEFINITION 4. Specifically, we tested two hypothesis: $Pr(1) > Pr(2)$ and $Pr(8) > Pr(9)$. Two indices, t-score and p-value were calculated, with higher t-score and lower p-value indicating more statistically significant difference. Generally, we consider the difference (decrease) as statistically significant if p-value is less than 0.05; otherwise, the difference (decrease) is very likely to be caused by random noise. The results are shown in Table 1: We can see both p-values and t-

decrease	from 1 to 2	from 8 to 9
t-score	88.44	8.38
p-value	0.00	0.00

Table 1: T-test results of decrease of retweeting probability in dataset1

scores show that the two drops of retweeting probability are statistically significant, especially when one spreader becomes two spreaders. This implies that the decrease is indeed caused by some other reason rather than random noise. In other words, the indirect influence decays occasionally as the number of spreaders increases, wherein the complex contagion is violated locally.

---

[2]http://www.rossmanchance.com/applets/ TOSCalculations/TOSCalculations.html

3.2. Dataset2

The second dataset comes from Knowledge Engineering Lab at Tsinghua University and covers the time span from August to December in 2009. The process of tweets crawling started from @yanglicai, a popular Twitter user in the Chinese community, then extended to all contacts of @yanglicai identified from his replying and retweeting messages. Subsequently, tweets from the contacts of his contacts are crawled as well. Finally, 192,999 tweets from 8254 users were obtained, and 25.5% of all crawled tweets were retweeting messages. If dataset1 represents the whole network, we view dataset2 as an ego network with @yanglicai located at the center with all other users having direct/indirect interactions with @yanglicai.

decrease	from 1 to 2	from 4 to 5	from 7 to 8
t-score	5.15	1.51	5.11
p-value	0.00	0.06	0.00

Table 2: T-test results of decrease of retweeting probability in dataset2

We did the same investigation for the second dataset as the first dataset. Figure 6 shows the distribution of n-spreader retweeting patterns in this ego-network. Again, the number of instances decreases as n increases. We plotted the curve of $Pr(n)$ in Figure 7. Similar to dataset1, several drops are spotted, i.e. from n =1 to 2, 4 to 5 and 8 to 9. We also did the same t-test with results shown in Table 2. The p-value of the test for the decreased probability occurring when n changes from 4 to 5 is a little larger than 0.05, implying that such decrease is only marginally trustable and is possibly caused by random noise. However, the other two drops still show high statistical significance because both p-values are close to 0. In summary, the phenomenon that retweeting probability increases non-monotonically with

6

some local drops is observed in both the whole network (dataset1) and certain ego-network (dataset2), verifying that complex contagion is globally validated but locally violated.

4. Quantum Cognition

Although the effect of complex contagion on the indirect influence is observed on a global scale, the local scale decreasing in influence needs further interpretation. Actually, a reason for this may be related to human psychology and cognition, a topic too complex to explain. The emerging field of *quantum cognition*, however, might be able to provide a potential interpretation for the decreased influence phenomenon. This field applies the formalism of quantum theory to model cognitive phenomena such as memory, judgment and decision making. Notably, in the process of decision making where a decision depends on multiple factors, quantum cognition assumes that these factors are not independent but have quantum-like interference effects on the final decision in a manner similar to the explanation for results from double-slit experiments (Khennikov, 2010; Pothos & Busemeyer, 2009; Yukalov & Sornette, 2011).
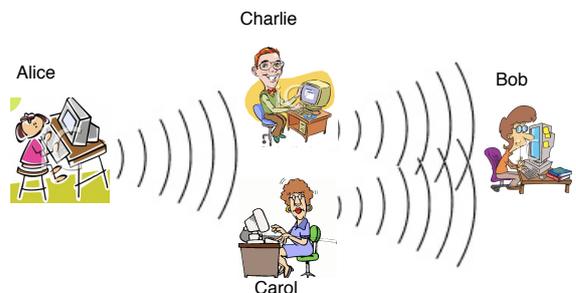
4.1. A Physical Metaphor

From the perspective of quantum cognition, the information spread is compared here to the physical wave quantum phenomenon. In the wave theory, the phase and magnitude of a wave at a certain point is the linear superposition of all waves from all sources. This superposition can result in constructive (when the wave phases align) or destructive (when the waves are out of phase) interference. Such quantum interference was first observed in the well-known double-slit experiment Young (1804), where quantum waves that pass through two slits interfered with each

other and generated a pattern of bright and dark bands on the screen. (Figure 8(a))
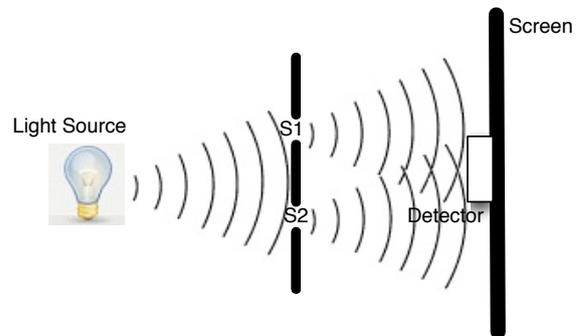
(a) Double-slit experiment in quantum physics

(b) Double-slit phenomenon in information spread

Figure 8: Double-slit phenomenon in both physics and information

We use our example in Figure 2 to illustrate the physical metaphor for information spread in Figure 8(b). Alice's sending out an initial message to Charlie and Carol creates an information 'wave' in the form of a primary wavefront. Then Charlie and Carol forward Alice's messages by creating two secondary wavefronts. Finally, the two waves arriving at Bob may have constructive/destructive interference effects on his intent to further propagate the information flow. If constructive interference occurs, the intensity of Bob's intent to propagate Alice's messages will increase, which is very similar to the effect of complex contagion. If destructive interference occurs, the intensity of Bob's intent will decrease, which goes against

7

complex contagion.

We give an intuitive illustration to show how Bob's intent decreases due to a destructive interference effect. Assume that initially only Charlie spreads Alice's messages while Carol does not. Bob receives Alice's messages through Charlie and becomes interested in these messages and further spreads them, because Bob has obtained relevant and fresh information. Later on, Carol also begins to spread Alice's messages but they largely overlap with those already spread by Charlie. Bob therefore becomes less interested in Alice's messages because he's overwhelmed with redundant information. Thus Bob's intent to further spread Alice's messages decreases, and so does the indirect influence of Alice on Bob. Here, the interference between two spreaders leads to destructive effects on the indirect influence from the sender to the receiver, which may explain the local decrease spotted in Figure 5. If a third spreader joins in and provides the receiver with new and relevant information about the sender, the indirect influence may rise again.
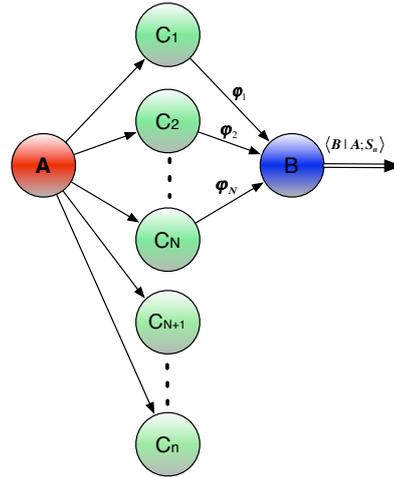
4.2. Cognition Model

Although an intuitive interpretation of locally decreased influence is given in terms of quantum cognition, formal mathematical modeling is needed. We propose a cognitive model called a *q-attention* model. This model originates from Batchelder & Riefer (1990), whose work presents a family of processing models for the source-monitoring paradigm in human cognition but is especially tailored to our social influence research in Twitter. Note that most concepts and notations used in our quantum modeling come from the theory of *quantum probability* (Gudder, 1988).

The q-attention model is designed for the situation where a group of Twitter users $S_n = \{C_j\}$, $j = 1, 2, ..., n$, follow user A and are also

followed by user B. S_n receive tweets from A and also retweet them to B. In addition, we assume that there exists an upper limit on the total number of users B can pay attention to among all B's followees, due to human's cognitive processing capacity. We set the number as $N \leq n$. In order to formulate the probability $Pr(B|A; S_n)$ as a function of n, we propose two versions of q-attention model: one is based on classical probability theory, and the other is based on quantum probability theory.
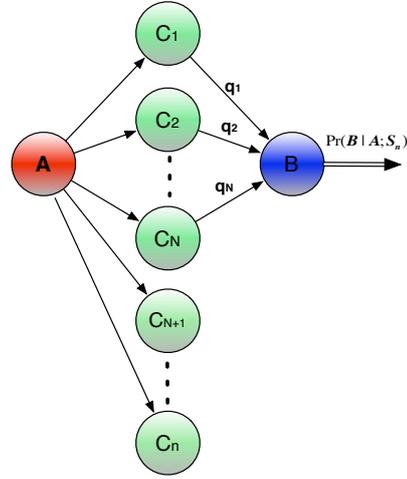
(a) Classical q-attention model

(b) Quantum q-attention model

Figure 9: Q-attention model

8

4.2.1. Classical q-attention model

We model the cognitive process of B retweeting from A's messages through $C_j \in S_n$ by dividing it into two phases: First, B pays attention to $C_j's$ messages; Second, B reads those messages retweeted by C_j from A and decides whether to further retweet them (Figure 9(a)). In the first phase, we define q_j as the probability that B pays attention to C_j and require $\sum_{q=1}^{N} q_j = 1$. In the second phase, we have already defined $Pr(B|A; C_j)$ to measure B's retweeting intent through C_j. Finally, the probability that B retweets from A through C_j can be computed by the classical law of total probability:

$$Pr(B|A; S_n) = \sum_{C_j \in S_n} q_j \cdot Pr(B|A; C_j) \qquad (1)$$

We replace $Pr(B|A; S_n)$ with $Pr(n)$ to represent the general case and two implications can be drawn:

- The total effect of all spreaders on the receiver is just the linear superposition of an individual effect of each spreader, which is represented by an independent single term without interfering with other spreaders' effects;

- $Pr(n)$ increases monotonically as n increases, indicating that adding spreaders only increases the final intent of B to further retweet A's messages.

Actually, the classical q-attention model matches the sense of complex contagion. Inspired by the description of influence curve of political hashtag in Romero et al. (2011), we can represent $Pr(n)$ as a monotone increasing function. In the simplest case, we can assume that $Pr(n)$ is just a linear increasing function of n from 1 until N . When $n \geq N, Pr(n) \equiv Pr(N)$, because the upper limit of human's information processing probability is reached and the intensity of influence persists at a relatively stable level. To summarize, $Pr(n)$ can be approximately represented by the following piecewise defined function:

$$\begin{cases} Pr(n) = Pr(1) + \frac{Pr(N)-Pr(1)}{N-1}(n-1), n = 2, ..., N-1 \\ Pr(N+k) = P(N), k \in \mathbb{N}^+ \end{cases}$$
$$(2)$$

4.2.2. Quantum q-attention model

To take into account the cognitive interference effect that is ignored in the classical q-attention model, we now formulate a quantum version of the q-attention model. According to quantum probability theory Gudder (1988), a probability p of an event is not primitive but derived from something more primitive called a probability amplitude φ (a complex number) — the probability is the square of magnitude of the amplitude, i.e. $p = |\varphi|^2$. As is shown in Figure 9(b), the structure of quantum q-attention model is almost the same as the classical version, except for the following three differences:

- The classical probability of paying attention to C_j, previously denoted by q_j , is replaced by a probability amplitude denoted by φ_j . B is assumed to be in a superposition state, denoted by:

$$|S_n\rangle = \sum_{j=1}^{N} \varphi_j \cdot |C_j\rangle \qquad (3)$$

where φ_j represents the "potential" of B to consider the tweets from C_j, but the potential is represented by a complex number. The probability of paying attention to the tweets from C_j is obtained from the squared magnitude $q_j = |\varphi_j|^2$ and again we require $\sum_{j=1}^{N} |\varphi_j|^2 = 1$. Overall, Equation 3 represents B's cognitive state of selecting spreaders C_j to pay attention to in a quantum manner;

- The classical probability that B retweets from A through C_j, previously denoted by $Pr(B|A; C_j)$, is replaced by the probability amplitude $\langle B|A; C_j \rangle$, which represents the B's "potential" to retweet but such potential is represented by a complex number. The probability of retweeting is obtained from the squared magnitude $Pr(B; A, C_j) = |\langle B; A, C_j \rangle|^2$;

- The quantum probability obeys *the law of total amplitude* rather than *the law of total probability*. The probability that B retweets from A passed along by $C_j \in S_n$ is represented by $Pr(B; A, S_n) = |\langle B; A, S_n \rangle|^2$. To determine the amplitude $\langle B; A, S_n \rangle$, we replace the sum of probabilities shown in Equation 1 with the sum of amplitudes given below:

$$\langle B|A; S_n \rangle = \sum_{C_j \in S_n} \varphi_j \cdot \langle B|A; C_j \rangle \qquad (4)$$

To make it clear, we show some examples of how quantum q-attention model naturally encompass interference effects. First, we consider $n = 1$ in which case only one spreader C_1 exists. Then $\langle B|A; S_1 \rangle = |\varphi_1 \cdot \langle B|A; C_1 \rangle|^2 = |\varphi_1|^2 \cdot |\langle B|A; C_1 \rangle|^2 = q_1 \cdot Pr(B|A; C_1)$. Apparently, the quantum version is exactly the same as the classical version because no interference occurs.

Next we consider $n = 2$ in which case $S_2 = \{S_1, S_2\}$. From Equation 4 we have:

$$
\begin{aligned}
\langle B|A; S_2 \rangle =& \quad |\varphi_1 \cdot \langle B|A; C_1 \rangle + \varphi_2 \cdot \langle B|A; C_2 \rangle|^2 \\
=& \quad |\varphi_1 \cdot \langle B|A; C_1 \rangle|^2 + |\varphi_2 \cdot \langle B|A; C_2 \rangle|^2 \\
& + (\varphi_1^* \cdot \varphi_2) \cdot \langle B|A; C_1 \rangle^* \cdot \langle B|A; C_2 \rangle \\
& + (\varphi_1 \cdot \varphi_2^*) \cdot \langle B|A; C_1 \rangle \cdot \langle B|A; C_2 \rangle^*.
\end{aligned}
$$

The first two terms corresponds to the same probability that we obtain from the classical

model, while the last terms form a conjugate pair:

$$
\begin{aligned}
(\varphi_1^* \cdot \varphi_2) \cdot \langle B|A; C_1 \rangle^* \cdot \langle B|A; C_2 \rangle =& \\
|\varphi_1 \cdot \varphi_2 \cdot \langle B|A; C_1 \rangle \cdot \langle B|A; C_2 \rangle| \cdot& \\
(\cos(\theta_{12}) + i \cdot \sin(\theta_{12}))& \\
(\varphi_1 \cdot \varphi_2^*) \cdot \langle B|A; C_1 \rangle \cdot \langle B|A; C_2 \rangle^* =& \\
|\varphi_1 \cdot \varphi_2 \cdot \langle B|A; C_1 \rangle \cdot \langle B|A; C_2 \rangle| \cdot& \\
(\cos(\theta_{12}) - i \cdot \sin(\theta_{12}))&
\end{aligned}
$$

where θ_{12} is the phase difference of the two complex number $\varphi_1 \cdot \langle B|A; C_1 \rangle$ and $\varphi_2 \cdot \langle B|A; C_2 \rangle$, which is dependent on the relation between C_1 and C_2. The sum of the conjugate pair produces a real number that we call an *interference* term:

$$I_{12} = 2 \cdot |\varphi_1 \cdot \varphi_2 \cdot \langle B|A; C_1 \rangle \cdot \langle B|A; C_2 \rangle| \cdot cos(\theta_{12})$$

Note that the cosine term can be positive (producing constructive interference), negative (producing destructive interference) or zero (no interference). Notably, if the cosine value is sufficiently negative, then the probability of B retweeting from A through two spreaders C_1 and C_2 can be smaller than the probability given only one spreader C_1, i.e. $|\varphi_2 \cdot \langle B|A; C_2 \rangle|^2 + I_{12} < 0$.

Finally we consider the case $n = 3$ and $S_3 = \{S_1, S_2, S_3\}$. Similarly, we obtain:

$$
\begin{aligned}
\langle B|A; S_3 \rangle =& |\varphi_1 \cdot \langle B|A; C_1 \rangle|^2 + |\varphi_2 \cdot \langle B|A; C_2 \rangle|^2 \\
& + |\varphi_3 \cdot \langle B|A; C_3 \rangle|^2 + I_{12} + I_{13} + I_{23},
\end{aligned}
$$

with the interference term defined as:

$$I_{ij} = 2 \cdot |\varphi_i \cdot \varphi_j \cdot \langle B|A; C_i \rangle \cdot \langle B|A; C_j \rangle| \cdot cos(\theta_{ij}) \qquad (5)$$

Again, each interference term can be positive or negative. If they are all large negative numbers, then we may then find a decrease in the probability for three spreaders compared with one or two spreaders. There is also a possibility that all the interference terms will cancel

out each other in which case the combinational interference would be equal to none. In other words, the quantum version in this case is reduced to the classical version.

4.2.3. Comparison between two q-attention models

We now systematically compare the classical and quantum q-attention models and mathematically interpret the locally decreased influence based on the quantum q-attention model. The architecture of the two models is actually the same yet the main conceptual difference between the two lies in the representation of probability. Based on above equations, we can obtain the mathematical relation between the two models as:

$$Pr_q(n) = Pr_c(n) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I_{ij} \qquad (6)$$

where $Pr_q(n)$ represents the average result of the quantum model while $Pr_c(n)$ represents the average result of the classical model. The only difference between the two models is the sum of interference terms, which represents the interference effect. If we consider the delta value, we obtain:

$$\triangle Pr_c(n) = Pr_c(n) - Pr_c(n-1) = q_n \cdot Pr(B|A; C_n) \qquad (7)$$

$$\triangle Pr_q(n) = Pr_q(n) - Pr_q(n-1)$$
$$= q_n \cdot Pr(B|A; C_n) + \sum_{j=1}^{n-1} I_{jn} \qquad (8)$$

From Equation 7, we can easily obtain $\triangle Pr_c(n) > 0$, i.e., $Pr_c(n)$ increases monotonously with n. In other words, $B's$ intent to retweet $A's$ messages always becomes stronger as more spreader C_j are involved in relaying $A's$ messages. However, the situation becomes more complex in Equation 8, due to the interference items. It

becomes very difficult to predict the changing tendency of $B's$ retweeting intent. Especially, if $\sum_{j=1}^{n-1} I_{ij} > -|q_n \cdot Pr(B|A; C_n)|$, we have $\triangle Pr_q(n) < 0$. In other words, sometimes the destructive effect will become so strong that it decreases $B's$ intent.

Compared with the classical q-attention model, the quantum q-attention model fully captures the trend of indirect influence shown in Figures 5 and 7. When constructive cognition interference occurs or interference cancel out (i.e., the classical model), the influence increases and complex contagion takes effect; when destructive cognition interference occurs, the influence may decrease and complex contagion is violated. In addition, the constructive/null effect occurs more frequently than the destructive effect, corresponding to the global increase of indirect influence with occasional local drops. Although our proposed quantum cognition model shows promise, it is a purely theoretical hypothesis based on current research in the field and needs further empirical verification.

5. Discussion

Equation 6 shows that the core part of quantum q-attention model is the interference terms, which can be modeled as a linear combination of different $cos(\theta_{ij})$ terms. Then the remain problem is how to evaluate the capability of quantum model (i.e. those θ_{ij}) in characterizing the dynamic of indirect influence regarding retweeting behavior in Twitter. It is difficult to solve because those interference terms are affected by complex human psychological or other factors, which needs further investigation and experimentation. Moreover, the quantum cognition model is more like a conceptual and theoretical prototypical model, rather than traditional probabilistic models, like Markov random filed or Bayesian

network, which already have mature frameworks for parameters learning and prediction evaluation. Therefore, the complete evaluation of the quantum model is beyond our work at the moment and we will leave it for future work.

In spite of the difficulty of model evaluation illustrated above, we still attempt to justify the mathematical capability of the quantum q-attention model in interpreting the co-existence of global increase and local drops found in the empirical data, by answering two questions: First, if we have already observed change of the retweeting probability with the number of spreaders in one empirical dataset, can we find the valid mathematical solutions of $cos(\theta_{ij})$? Second, can we apply some simple rule to generate simulated/fake $cos(\theta_{ij})$ terms, which can lead to global increased but locally dropped retweeting probabilities against different different number of spreaders?

To answer the first question, we rely on the simple classic linear model described in Equation 2 and 1. In addition, we assume that the empirical values of retweeting probabilities represent the output of quantum model and the difference between the quantum and classical model are fully captured by interference terms, as is shown in Equation 6. Then by solving the linear equations composed of Equation 1,2,6, we can obtain θ_{ij} for different values of n. There're two points that worth attention. First, if $|cos(\theta_{ij})| > 1$, then it is an invalid solution and the quantum q-attention model fails. Second, most of the time, the solutions are not unique and multiple combinative solutions of $cos(\theta_{ij})$ can be found.

We take dataset1 as the example to illustrate how to find those $cos(\theta_{ij})$. We set N as half of the total number of retweeting patterns found in dataset1, i.e. N=14. When $n = 2$, we solve the linear equations above and obtain $cos(\theta_{12}) = -0.524$, which is a valid value since

$-1 \leq cos(\theta_{12}) \leq 1$. It also indicates that the interference between C_1 and C_2 is destructive, which leads to a local drop. When $n = 3$, we solve the linear equations and use the already known value of $cos(\theta_{12})$. We obtain

$$44.4 \cdot cos(\theta_{13}) + 7.8 \cdot cos(\theta_{12}) = 7.9 \cdot \quad (9)$$

Obviously, $cos(\theta_{13})$ and $cos(\theta_{12})$ have more than one solutions and they may have opposite signs. In other words, when C_3 is added, the interference between C_1 and C_3 and the interference between C_2 and C_3 may behave in different directions, leading to the rise of the global indirect influence again. To be general, those cosine terms satisfy the following equation when $n = k > 2$:

$$\sum_{i}^{k-1} Diff(i) + 2(\alpha cos(\theta_{1k}) + \sum_{j=2}^{k} \beta cos(\theta_{jk})) = Diff(k)$$
$$(10)$$

where $Diff(i) = Pr_q(i) - Pr_c(i)$, $\alpha = \sqrt{Pr(1) \cdot \beta}$ and $\beta = \frac{Pr(N) - Pr(1)}{N - 1}$. We also try to find values of $cos(\theta_{ij})$ for other n values and the solutions are all valid, which indicates the mathematical plausibility of proposed quantum model. However, It is worth note that the $cos(\theta_{ij})$ items calculated from one dataset do not apply to other dataset, because different sets of users have different interactions.

To answer the second question, we run a simple simulation to generate those $cos(\theta_{ij})$ without any prior knowledge by assuming that $cos(\theta_{ij})$ is just uniformly distributed from [-0.5, 0.5] and can be randomly sampled. Here we examine three types of retweeting probabilities: classical model value from Equation 2, emprical value from the dataset1, and simulated quantum model value from Equation 10, 6 and randomly sampled $cos(\theta_{ij})$. We plot two groups of values by using different random number generator seeds. The simulated curves in both Figure 10(a) and Figure 10(b) show the trend of global increase

plus local drop, which matches our motivation of building quantum q-attention model. However, the change of retweeting probability in Figure 10(b) shows large deviation from the empirical curve, implying that $cos(\theta_{ij})$ is not randomly distributed but follows some rule that needs further examination in order to fit the real data.
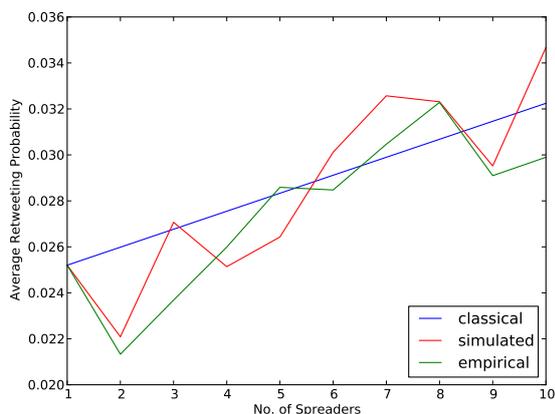
(a) simulation with random seed 1
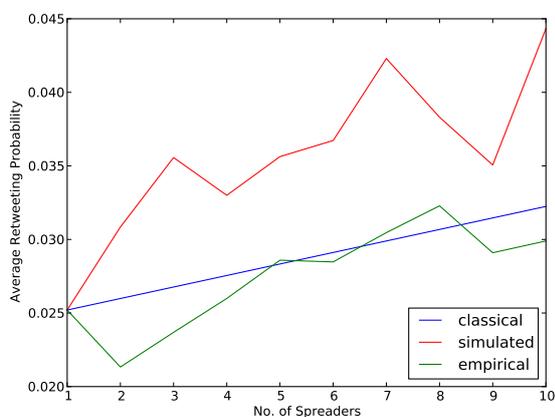
(b) simulation with random seed 2

Figure 10: Simulation results of quantum model using randomly sampled interference items

Although we demonstrate the mathematical plausibility of quantum model above, our work just initiates a new direction of applying quantum cognition to social influence study, which

is not seen in any other current related researches yet. We will leave the study of complete parameter estimation and model evaluation of quantum q-attention models for future work, possibly by combining existing machine learning techniques and psychological experiments.

6. Related Work

6.1. Information and Influence Propagation

The origin of information propagation in social networks comes from social influence, which occurs when an individual's thoughts, feelings or actions are affected by other people. Information propagation characterizes the way that a node in social networks can spread an information meme to its neighbor nodes via exerting social influence on them. The existence of social influence has been validated by many researchers (Anagnostopoulos et al., 2008; Crandall et al., 2008; Gomez Rodriguez et al., 2010). Different types of influence across various fields has been studied, including social media like Blog (Gruhl et al., 2004), Flickr (Cha et al., 2009) and Twitter (Hong et al., 2011; Yang et al., 2010), as well as academia (Ding, 2011; Ding & Cronin, 2011; Yan et al., 2011). The maximization of influence spread in the whole network was investigated (Kempe et al., 2003; Chen et al., 2009) as well. In addition, different methods of probabilistic modeling of social influence were also proposed. Tang et al. (2009) used Topical Affinity Propagation to model the topic-level social influence on large social network; De Choudhury et al. (2007) found three contextual factors that affect the influence propagation between two friends in MySpace; Xiang et al. (2010) developed an unsupervised model to estimate relationship strength from interaction activity and user similarity. Although many interesting phenomena

have been observed from the above researches, most of the time only direct communication between two adjacent persons in a social network is considered in modeling and analyzing the information propagation. Normally, multiple intermediate persons called spreaders are involved in the indirect communication between two persons, i.e., the sender and the receiver. Those spreaders may not be independent but have combinational interference on the information flow from the sender to the receiver. Some have studied the problem of serial indirect influence where indirect influence is propagated through n serial nodes. Fowler & Christakis (2008) found that happiness spreads among people as far as 3 hops. Liu et al. (2010) tried to quantify the indirect influence of n-degree friends at the topical level. However, very few studies have focused on the parallel indirect influence, where indirect influence is propagated through n parallel nodes, the focus of our paper.

6.2. Quantum cognition

A recent article from NewScientist pointed out the strong connection between quantum theory and human thought processes, where human cognition is found to follow fuzzy logic more than classical logic (Buchanan, 2011). Cognitive science researches have also identified the existence of quantum-like cognitive interference in human decision-making (Busemeyer et al., 2009; Khennikov, 2010), where the experiments showed that the classic law of total probability was violated. Instead, quantum probability (Gudder, 1988) was introduced to explain the experimental results. Quantum cognition has been introduced in the IR field with the hope to build more user-centric methods in searching and ranking. Arafat & Rijsbergen (2007) applied QT to address some fundamental issues in searches, which investigated and modeled user cogni-

tion and their interactions during the search process. Hou & Song (2009) proposed an extended vector space model (EVSM) to model context-sensitive high-order information. Zuccon et al. (2009) suggestted ranking the document relevance using QT as judgment of relevance that is not independent from other documents, and the interference of other documents play an important role in judging the relevance. They proposed a novel quantum probability ranking principle (QPRP) to model a situation, where a document relevance assessment is influenced by other documents. Zhang et al. (2010) proposed the use of quantum finite automation (QFA) to represent the transition of the measurement states (the relevance degrees of the document judged by users) and dynamically model the cognitive interference of users when they are influenced by other related documents. Although quantum cognition has been extensively applied in IR, our paper takes the first lead in applying quantum cognition to analyze the indirect social influence in social networks.

6.3. Complex Contagion

Complex contagion (Centola & Macy, 2005) refers to the phenomenon where multiple sources of exposure to a new idea are required before an individual adopts the idea. Simple contagion, however, can spread in social networks with just one contact with a single infected neighbor, as a disease may function. Centola (2010) found that individuals are more likely to acquire new health practices while living in networks with dense clusters of connections, that is, when in close contact with people they already know well. For instance, people are more likely to participate regularly in the health forum if they had more health buddies who registered for it. Romero et al. (2011) studied the spread of hashtags in Twitter and quantified the probability of a

user adopting a new hashtag as the function of the number of his/her neighbors who have already adopted it. They found that the spread of political hashtags validates the complex contagion, where the adoption probability increases monotonically as the number of neighbors who have already adopted the same hashtags increases, until finally reaching a plateau. By contrast, for idiom hashtags, complex contagion does not take effect and the adoption probability decays rapidly when more neighbors have adopted the same hashtags.

7. Conclusions

In this paper, we investigated the propagation of parallel indirect influence on Twitter with a focus on how the intensity of influence changes with the number of spreaders. We constructed two social networks formed by the following relations (confirmed by RT in messages) in Twitter, quantified the intensity of indirect influence with the retweeting probability, and plotted the curve of retweeting probability with the number of spreads. We found that the phenomenon of complex contagion is validated globally since the overall trend of the retweeting probability is gradual increasing against the number of spreaders. However, the probability decreases locally as well. We applied quantum cognition theory in an attempt to interpret the local anomaly.

Although some interesting findings about indirect influence on Twitter are observed in this study, a great deal of future work remains to be done on the topic. First, the phenomenon of locally decreased influence needs further study. Quantum cognition itself may not be sufficient to explain it. Some theories from psychology, sociology, and probability will help address these complexities. Second, we plan to compare the indirect influence curves for different topics. We think that different pat-

terns in retweeting behaviors exist for different topics on Twitter. It will be very meaningful to compare them and see whether complex contagion works or not in different contexts. Third, we hope to testify our quantum cognition model on other social networks including Facebook. For instance, "like" is a common action in Facebook. We can model the probability of a use who "like" some item on Facebook as the number of his/her social neighbors who already "liked" the same item, and see whether complex contagion with local drops reemerges. Finally, our current work is more descriptive and observational than mathematical formulations. We hope to build a probabilistic model to represent and predict indirect influence in the future.

References

Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008). Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* KDD '08 (pp. 7–15). Las Vegas, USA.

Arafat, S., & Rijsbergen, C. J. V. (2007). Quantum theory and the nature of search 2007. In *AAAI Symposium on Quantum Interaction* (pp. 114–221). California.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548–564.

Buchanan, M. (2011). Your quantum minds. *New Scientist*, *3*, 34–37.

Busemeyer, J. R., Wang, Z., & Lampert-Mogiliansky, A. (2009). Empirical comparison of markov and quantum models of decision making. *Journal of Mathematical Psychology*, *53*, 423–433.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, *329*, 1194–1197.

Centola, D., & Macy, M. (2005). *Complex contagion and the weakness of long ties*. Technical Report.

Cha, M., Mislove, A., & Gummadi, K. P. (2009). A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th international conference on World wide web* (pp. 721–730). Madrid, Spain.

Chen, W., Wang, Y., & Yang, S. (2009). Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* KDD '09 (pp. 199–208). Paris, France.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., & Suri, S. (2008). Feedback effects between similarity and social influence in online communities. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* KDD '08 (pp. 160–168). Las Vegas, USA.

De Choudhury, M., Sundaram, H., John, A., & Seligmann, D. D. (2007). Contextual prediction of communication flow in social networks. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence* WI '07 (pp. 57–65). Silicon Valley, USA: IEEE Computer Society.

Ding, Y. (2011). Topic-based pagerank on author cocitation networks. *J. Am. Soc. Inf. Sci. Technol.*, *62*, 449–466.

Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? measures of scholarly esteem. *Inf. Process. Manage.*, *47*, 80–96.

Fowler, J. H., & Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *British Medical Journal*, *337*.

Gomez Rodriguez, M., Leskovec, J., & Krause, A. (2010). Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* KDD '10 (pp. 1019–1028). Washington DC, USA.

Gosset, W. S. (1970). *Studies in the History of Statistics and Probability*. London.

Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web* (pp. 491–501). New York.

Gudder, S. P. (1988). *Quantum Probability*. Academic Press.

Hong, L., Dan, O., & Davison, B. D. (2011). Predicting popular messages in twitter. In *Proceedings of the 20th international conference companion on World wide web* (pp. 57–58). Hyderabad, India.

Hou, Y., & Song, D. (2009). Characterizing pure high-order entanglements in lexical semantic spaces via information geometry. In *Proceedings of the 3rd International Symposium on Quantum Interaction* (pp. 237–250). Germany.

Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* KDD '03 (pp. 137–146). Washing ton DC, USA.

Khennikov, A. Y. (2010). *Ubiquitous Quantum Structure: From Psychology to Finance*. Springer.

Liu, L., Tang, J., Han, J., Jiang, M., & Yang, S. (2010). Mining topic-level influence in heterogeneous networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management* CIKM '10 (pp. 199–208). Toronto, Canada.

Piwowarski, B., Frommholz, I., Lalmas, M., & Rijsbergen, K. (2010). What can quantum theory bring to information retrieval. In *Proceedings of CIKM'2010* (pp. 59–68).

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of rational decision theory. In *Proceedings of the Royal Society B Biological Sciences* (pp. 2171–2178). The Royal Society volume 276.

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web* WWW '11 (pp. 695–704). Hyderabad, India.

Tang, J., Sun, J., Wang, C., & Yang, Z. (2009). Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* KDD '09 (pp. 807–816). Paris, France.

Xiang, R., Neville, J., & Rogati, M. (2010). Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web* WWW '10 (pp. 981–990). Hong Kong, China.

Yan, E., Ding, Y., & Sugimoto, C. R. (2011). P-rank: An indicator measuring prestige in heterogeneous scholarly networks. *J. Am. Soc. Inf. Sci. Technol.*, *62*, 467–477.

Yang, Z., Guo, J., Cai, K., Tang, J., Li, J., Zhang, L., & Su, Z. (2010). Understanding retweeting behaviors in social networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 1633–1636). Toronto, ON, Canada.

Young, T. (1804). The bakerian lecture: Experiments and calculations relative to physical optics. *Philosophical Transactions of the Royal Society of London*, *94*, 1–16.

Yukalov, V. I., & Sornette, D. (2011). Decision theory with prospect interference and entanglement. *Theory and Decision*, *70*, 283–328.

Zhang, P., Song, D., Hou, Y., Wang, J., & Bruza, P. (2010). Automata modeling for cognitive interference in users' relevance judgment. In *AAAI-Fall 2010 Symposium on Quantum Informatics for Cognitive, Social, and Semantic Processes (AAAI-QI)*. Washington DC.

Zuccon, G., Azzopardi, L., & van Rijsbergen, K. (2009). The quantum probability ranking principle for information retrieval. In *In Proceedings of ICTIR'2009*.