Fast Information Cascade Prediction Through Spatiotemporal Decompositions

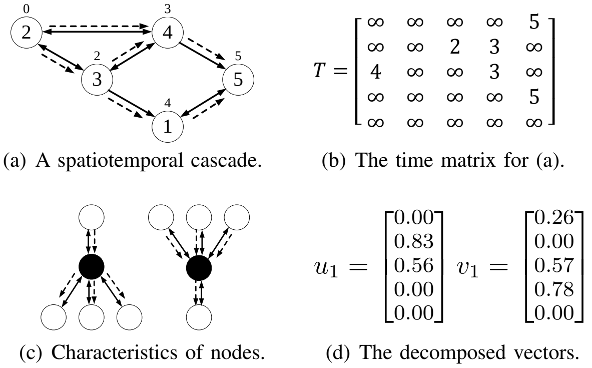
Huanyang Zheng and Jie Wu
Department of Computer and Information Sciences, Temple University, USA
Email: {huanyang.zheng, jiewu}@temple.edu

***Abstract*—In online social networks, information cascades occur when people observe the actions of others (followees) and then make the same choices that the others have made (followers). Cascade predictions are important, since they can detect and help resist bad cascades. We focus on photo cascade predictions in Flickr: given the current cascade and social topology, we want to predict the number of propagated users at a future time slot. Information cascades include a large amount of data that crosses both space and time. To reduce prediction time complexities, our main idea is to decompose the spatiotemporal cascade information (a larger size of data) to user characteristics (a smaller size of data) for subsequent predictions. Space and time matrices are introduced to record the cascade information. We introduce a set of new notions, persuasiveness and receptiveness (represented as two vectors for complexity reduction), to capture characteristics of followees and followers. In this case, persuasiveness includes followees' abilities to propagate information, while receptiveness includes followers' willingness to accept information. Then, we propose a three-stage parallel prediction scheme as follows. (1) We map the spatiotemporal cascade information to a weighted matrix, in which the weights of space and time information are tuned. (2) Singular value decomposition is used to extract nodes' persuasiveness and receptiveness (two vectors) from the weighted matrix. (3) Predictions are conducted based on nodes' persuasiveness and receptiveness. Finally, extensive evaluations on the Flickr dataset are conducted to verify the competitive performance of the proposed scheme.**

Keywords*—*Cascade prediction, online social network, parallel, persuasiveness and receptiveness, spatiotemporal decomposition.

I. Introduction

Nowadays, online social networks (OSNs), which belong to typical large distributed systems, are a fundamental medium for spreading information, such as sharing startling news, creative ideas, and interesting stories. An information cascade may occur if a user follows another user: if Alice (a followee) shares a photo, Bob (a follower) may scan this photo and then share it to his/her followers later. This type of iterative information propagation is called an *information cascade*. Meanwhile, cascade predictions are important in various aspects of human lives, such as in the control of computer viruses, prevention of infectious diseases, inhibition of terrible rumors, estimation of economic products, and the forecast of marketing strategies. However, the cascade prediction is very difficult, due to its intrinsic complexities: when will a user further propagate the information (called propagation boundaries)? In this paper, we capture propagation boundaries spatiotemporally, i.e., through both social topological information and time information. More specifically, given a cascade before a time τ_1 and the social topology, we want to predict the number of propagated users (called the cascade size) at a future time slot τ_2 (assuming that the information source appears at $\tau_0 = 0$). To reduce prediction time complexities, our main idea is to decompose the spatiotemporal cascade information (a larger size of data) to user characteristics (a smaller size of data) with bounded information loss; then, predictions are conducted based on the decomposed information, as to have a low time complexity.



$$T = \begin{bmatrix} \infty & \infty & \infty & \infty & 5 \\ \infty & \infty & 2 & 3 & \infty \\ 4 & \infty & \infty & 3 & \infty \\ \infty & \infty & \infty & \infty & 5 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}$$

$$u_1 = \begin{bmatrix} 0.00 \\ 0.83 \\ 0.56 \\ 0.00 \\ 0.00 \end{bmatrix} \; v_1 = \begin{bmatrix} 0.26 \\ 0.00 \\ 0.57 \\ 0.78 \\ 0.00 \end{bmatrix}$$

(a) A spatiotemporal cascade.
(b) The time matrix for (a).
(c) Characteristics of nodes.
(d) The decomposed vectors.

Fig. 1. Illustrations of information cascades. In (a) and (c), solid directional edges among nodes (numbers inside nodes are user IDs) represent follower-followee relationships (the pointed node is the follower). Dashed directional edges indicate the cascade. The label on the top of a node indicates the time when this user starts to propagate information after having been influenced. Node 2 is the information source. In (c), the left dark node has high persuasiveness and receptiveness (the right one is the opposite). The decomposition result for the cascade of the first four time slots is shown in (d).

In a macro view, information cascades of OSNs include a large amount of data that crosses both time and space, i.e. spatiotemporal information. Therefore, we use matrices S and T to respectively capture the space and time dimensions of cascades. Here, S is the network adjacency matrix, which shows the social topology. Then, the time matrix T indicates the propagated nodes (i.e., users) in terms of time sequences. The time matrix T, which corresponds to the cascade of all five time slots in Fig. 1(a), is shown in Fig. 1(b). The element t_{ij} of T is the time when user j starts to propagate information after having been influenced by user i. We assume that a propagated node influences its followers immediately without a delay, while time durations of influences can be deducted from T. Note that T includes complete time information and partial space information: nodes that are closer within the social topology are more likely to be propagated at closer times. Although S and T can be used for predictions directly, the prediction time complexity is unacceptable due to matrix operations. For example, in the Flickr dataset [1], S and T involve 2,302,925 users with 11,267,320 photos, which are unacceptable for matrix representations.

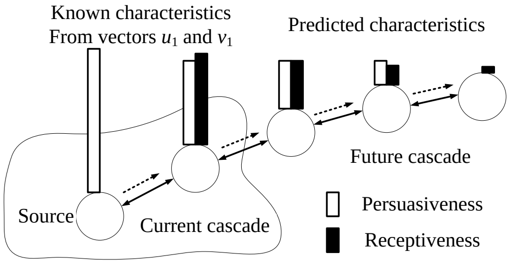


Fig. 2. The decay pattern of nodes' persuasiveness and receptiveness.

The micro view of cascades is that followees iteratively propagate information to their followers. Therefore, we introduce a set of new notions, *persuasiveness* and *receptiveness*, to capture characteristics of followees and followers: the persuasiveness is defined as followees' abilities to propagate information; the receptiveness represents followers' willingness to accept information. As shown in Fig. 1(c), the left dark node has high persuasiveness and receptiveness (the right dark node is the opposite). Vectors u_1 and v_1 are used to record nodes' persuasiveness and receptiveness, where the i^{th} elements in the vectors u_1 and v_1 show node i's persuasiveness and receptiveness, respectively. We further consider these two characteristics to be *spatiotemporally-sensitive*: if a node with a high out-degree is spatially far away from the information source, it may not be propagated, and thus it cannot positively propagate the information further (i.e., low persuasiveness). In the case of a temporal remote node, it also has low persuasiveness, since its followers may have been propagated by other nodes. The same rule works for the receptiveness. Therefore, in terms of the distribution, nodes' persuasiveness and receptiveness should decay with respect to their spatiotemporal distances to the information source. Moreover, their decay patterns indicate propagation boundaries: the cascade terminates when nodes have low persuasiveness and receptiveness.

Our prediction scheme is based on both the macro and micro properties of cascades. This scheme has three stages as follows. (1) In the first stage (Section IV), we map the time matrix, T, to a weighted matrix M. As previously mentioned, T includes spatiotemporal information. The mapping objective is to tune the weights of space and time information. We also highlight earlier propagations in the mapping process, since they are more important than later ones. (2) In the second stage (Section V), we introduce the singular value decomposition (SVD [2]) to extract nodes' persuasiveness and receptiveness (two vectors) from the weighted matrix M, with bounded information loss. This is because the element m_{ij} of M represents a joint result of followee i's persuasiveness and follower j's receptiveness. Fig. 1(d) shows the result for the cascade of the first four time slots ($\tau_1 = 4$, and node 5 is waiting for the prediction) in Fig. 1(a). u_1 shows that nodes 2 and 3 are followees, while v_1 shows that nodes 1, 3 and 4 are followers. Now, the spatiotemporal cascade information (matrices) is compressed into nodes' persuasiveness and receptiveness (vectors), resulting in a reduced prediction time complexity. (3) In the third stage (Section VI), we conduct predictions based on the decomposed information. The decay pattern of nodes' persuasiveness and receptiveness along shortest paths are focused, as shown in Fig. 2. Then, the persuasiveness and receptiveness of currently unpropagated nodes are predicted. For example, in Fig. 1, node 5's persuasiveness and receptiveness are predicted according to vectors u_1 and v_1. Based on the prediction result, $\hat{u}_1$ and $\hat{v}_1$, we can reconstruct the predicted weighted matrix, $\hat{M}$. The predicted number of propagated users can be obtained by mapping $\hat{M}$ back to the predicted time matrix.

Our contributions are manifold: (1) we consider cascades spatiotemporally, and propose a parallel prediction scheme to deal with the large amount of cascade information. (2) We introduce persuasiveness and receptiveness to capture characteristics of followees and followers, which are completely novel. Persuasiveness and receptiveness can be decomposed from the spatiotemporal cascade information, i.e., the complete cascade information is compressed efficiently with bounds. (3) User personalities (e.g., gender and age) can be incorporated into our model. (4) Prediction methods, based on nodes' persuasiveness and receptiveness, are proposed, the performance of which are verified by real-data driven evaluations.

The remainder of this paper is organized as follows: In Section II, we survey the related work; in Section III, basic concepts are shown with the dataset description; in Section IV, we show the mapping process; in Section V, the spatiotemporal decomposition is introduced to extract nodes' persuasiveness and receptiveness; in Section VI, we show the whole prediction process; in Section VII, extensive real data-driven evaluations are shown; and finally, in Section VIII, we conclude the paper.

II. Related Work

An information cascade occurs when people observe the actions of others and then make the same choices that the others have made. The most popular cascade models include the linear threshold model [3, 4], and the independent cascade model [3–6]. In the linear threshold model, each person has a weight and a threshold. A person starts to spread information further, only if the weight summation of propagated persons that he/she follows is larger than his/her own threshold. Instead of the deterministic model, the independent cascade model introduces probabilities: once propagated, each node has a certain likelihood of further spreading the information to its followers. More models are derived from these two models. For example, Ghasemiesfeh et al. [7] considers a k-complex model, where a node is further propagated if no less than k neighbors of this node are propagated. Sadikov et al. [8] considers a k-tree model. However, these models mainly focus on the spatial cascade information.

The study on spatiotemporal cascade has been proposed in [9], where the time dimension also matters. Differing from former studies, we compress the spatiotemporal cascade information into nodes' persuasiveness and receptiveness, which are completely novel. This compression also sheds light on the big data processings [10], since it reduces the dimensions for describing cascades. Rather than using statistical approaches, our method reserves insights on cascades. Our model can also be extended by considering user personalities.

Another branch of cascade studies focuses on the data mining of real datasets, such as Facebook [11], Flickr [1] and Twitter [4, 12]. These studies observe real cascades and then match real cascade properties to theoretical models. Our study is based on the Flickr dataset [1].

TABLE I. FLICKR DATASET SUMMARY

Time period (two periods)	11/02/2006 to 12/03/2006 02/03/2007 to 05/18/2007
# Links	17,034,807 to 33,140,018
# Users	1,487,058 to 2,302,925
# Photos	11,267,320
# Favorite marks	34,734,221
# Popular photos	14,002
Most popular photo	Marked by 2,998 times
Largest in / out-degree	21,001 / 26,367

TABLE II. NOTATIONS

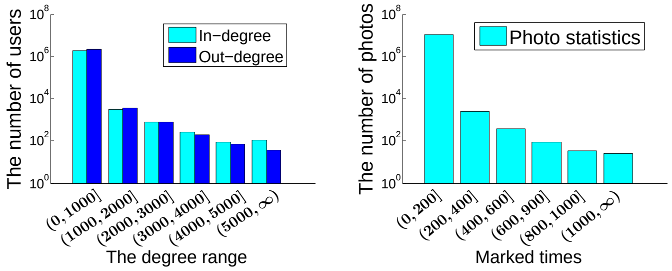
Notation	Description
τ_1 / τ_2	The current / future cascade time ($\tau_2 > \tau_1$).
τ_0	The appearance time of the information source.
S / T	The space / time matrix with elements s_{ij} / t_{ij}.
N_i / N_p / N	The set of influenced / propagated / total users.
E_i / E_p / E	The edge set corresponding to V_s / V_t / V.
M	A matrix mapped from the time matrix T.
U / Σ / V	The SVD result of M ($M = U\Sigma V^*$).
σ_i	The i^{th} largest singular value of the matrix M.
u_i / v_i	The vector in U / V corresponding to σ_i, and u_1 / v_1 shows persuasiveness / receptiveness.
$\hat{u}_1$ / $\hat{v}_1$	The predicted u_1 / v_1 in the future cascade.
$\hat{M}$ / $\hat{T}$	The predicted M / T in the future cascade.
$f(t) = e^{-ct}$	The mapping function, which maps t_{ij} to m_{ij}.

III. BASIC CONCEPTS AND DATASET DESCRIPTION

A. Basic Concepts

Flickr is an online social network for sharing photos (i.e., the information to propagate) among users, the relationships of which are directional: a directional edge from Bob to Alice means that Bob follows Alice. Users share photos among each other by labeling a “favorite-mark” to a photo. We refer to users who label photos with a “favorite-mark” as *propagated* users in the cascade of that photo. Meanwhile, users are called *influenced* if they have seen this photo. Note that an influenced user may not be a propagated user, since he/she may not mark the photo as a favorite for further sharing. Then, a photo cascade process can be formally defined as a spatiotemporal photo spreading process on all influenced users, rather than on all propagated users. Information cascades include a large amount of data that crosses both time and space, i.e. spatiotemporal information. Then, the space matrix, S, is defined as the adjacency matrix of the social topology among all the users (including the users that need to be predicted). Theoretically, S should include the complete social topology (i.e., all users on Flickr), since a cascade may propagate over the whole network. However, for practical usage, S can be a large enough subgraph.

Once a user shares a photo, we consider that this user is influenced by all the propagated users that he/she follows. The element t_{ij} of matrix T is the time when user j starts to propagate information after having been influenced by user i. Here, T is called the time matrix, which includes all users corresponding to S. The elements in T that represent currently unpropagated users are set to be infinite. The users in T are corresponding to the users in S. We will further discuss the size of S and T in Section VI.D, since including all users is redundant, and is not feasible for practical usage. A large enough subgraph can be used for the prediction. Note that a user j may have been influenced by multiple users before his/her own propagation at the time t_{ij}. For example, in Fig. 1(a), user 5 has been influenced by user 4 since time 3, but he/she finally decides to propagate (i.e., label a “favorite-mark” to the photo) at time 5. Note that time durations of influences can be deducted from T. Therefore, complete information of a photo cascade has been reserved in both S and T. In addition, let τ_0, τ_1 and τ_2, respectively, denote the appearance time of the information source, the current time (i.e., we know the whole cascade process between τ_0 and τ_1), and the future time at which we want to predict the cascade size. In the following cascade examples of this paper, we set $\tau_0 = 0$, $\tau_1 = 4$, and $\tau_2 = 5$ as a default setting.



(a) User degree distribution. (b) Favorite mark distribution.

Fig. 3. Statistics of the Flickr dataset.

B. Dataset Description

The Flickr dataset is collected by Cha et al. [1] through Flickr APIs. It was collected during the time periods from November 2^{nd} to December 3^{rd}, 2006, and February 3^{rd} to May 18^{th}, 2007. The number of users and their links are growing with respect to time. Note that a user, on average, has less than 15 links: this network is definitely *sparse* (i.e., matrices S and T are sparse). The degree distribution is shown in Fig. 3(a), indicating that a few users have very high degrees. 11,267,320 photos are shared during this period, with 34,734,221 favorite marks in total. 34,484 photos are not marked, but are recorded in the system. Most photos (11,218,834 photos) are marked no more than 100 times, while only 25 photos are marked more than 1,000 times. The distribution of photos, in terms of times marked, is shown in Fig. 3(b). Since photos of different popularity stand for cascades of different types, we choose *popular photos* (defined as the photos that are shared more than 100 times) for further analysis in the following part. We consider that popular photos have similar cascade properties. The other dataset statistics are shown in Table I, and all the notations are shown in Table II.

IV. TUNING THE SPATIOTEMPORAL INFORMATION

The first stage of our prediction scheme is introduced in this section, where we show the mapping process that tunes the weights of space and time information. The guiding rules of the mapping are shown, with their insights following.

A. Mapping Process

In this paper, we independently map each element in T to the element in M of the same position. Then, the mapping function is defined as $f : t_{ij} \rightarrow m_{ij}$, or $m_{ij} = f(t_{ij})$, over real positive numbers. Since earlier propagations are more

$$T(\tau_1 = 4) = \begin{bmatrix} \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & 2 & 3 & \infty \\ 4 & \infty & \infty & 3 & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}$$

(a) The time matrix at $\tau_1 = 4$.

$$M(\tau_1 = 4) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.67 & 0.55 & 0 \\ 0.45 & 0 & 0 & 0.55 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(b) The corresponding mapping result.

Fig. 4. Mapping T to M through $f(t) = e^{-t/5}$, where $c = \frac{1}{5} = \frac{1}{\tau_2}$.

important (explained later in the next subsection), we have the following mapping rule:

***Guiding Rule** 1:* The function $f(t)$ is strictly decreasing with respect to t. When $t \to \infty$, we have $f(t) \to 0$.

Another concern is that the starting time of the cascade is not important: the cascade in Fig. 1(a) can be viewed as starting at $\tau_0 = 0$ and finishing at $\tau_2 = 5$; however, it can also be viewed as starting at $\tau_0 = 1$ and finishing at $\tau_2 = 6$ (i.e., a position translation of 1 on the time domain). Obviously, this translation should not influence mapping results (relationships among elements m_{ij}). Therefore, we have:

***Guiding Rule** 2:* The function $f(t)$ satisfies $\frac{f(t+\tau)}{f(\tau)} = f(t)$, i.e., $f(t+\tau) = f(t)f(\tau)$. Here, τ is a parameter for tuning the starting time of the cascade.

An interesting phenomenon is that Guiding Rules 1 and 2 have determined the function form of $f(t)$, as follows:

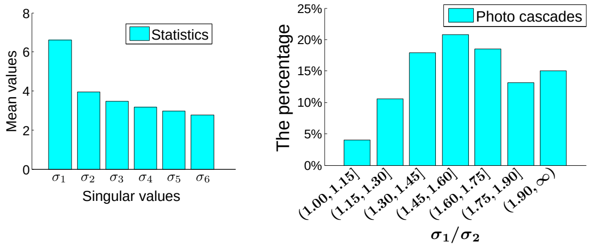
***Theorem** 1:* The only feasible family of solutions for the above guiding rules are exponential functions, i.e., $f(t) = e^{-ct}$ where c is an arbitrary positive number.

Proof: Let us start with Guiding Rule 2, where $f(t+\tau) = f(t)f(\tau)$. If $\tau = t$, then $f(t+t) = f(2t) = f(t)^2$. If we do this iteratively, then we can have $f(Ct) = f(t)^C$, where C is a parameter. Exchanging t and C, we have $f(Ct) = f(t)^C = f(C)^t$. Let $C = 1$, and then we have $f(t) = f(1)^t$. Obviously, $f(1)$ is an arbitrary constant. If we replace $f(1)$ with e^c, then the result is $f(t) = e^{ct}$. Here, c is an arbitrary real number. According to Guiding Rule 1, the function $f(t)$ is strictly decreasing. Therefore, we change the result $f(t) = e^{ct}$ to be $f(t) = e^{-ct}$, and restrict c to be a positive number. In addition, the result can also be proved through (1) operating a logarithm on $f(t+\tau) = f(t)f(\tau)$ to be $\ln f(t+\tau) = \ln f(t) + \ln f(\tau)$, and then (2) using Cauchy's functional equation. □

Here, parameter c's insight is its functionality for tuning time scales (e.g., 1 hour is equivalent to 60 minutes): time scales should not change the mapping result. Generally speaking, the value of c is determined empirically. The value c can be set in the range of $[\frac{1}{\tau_2}, \frac{1}{\tau_1}]$. In addition, the corresponding mapping process of the cascade (only the first four time slots) in Fig. 1(a) is shown in Fig. 4.

B. Mapping Insights

As previously mentioned, we consider cascades spatiotemporally, which includes a large amount of data that crosses both space and time in a macro view. Meanwhile, in a micro view, the persuasiveness and receptiveness is used to capture characteristics of followees and followers: the persuasiveness includes followees' capacities to propagate information; the receptiveness includes followers' willingness to accept information. Nodes' persuasiveness and receptiveness are considered to be spatiotemporally-sensitive: if a node with a high out-going degree is spatially far away from the information source, it may not be propagated, and thus it cannot positively propagate the information further. In the case of a temporal remote node, it also has low persuasiveness, since its followers may have been propagated by other nodes. The same rule works for the receptiveness: propagations that fail to reach the sources' neighbors may lead to a premature abortion of further information propagations; a successful propagation of a remote node does not change the overall cascade trend. Therefore, in terms of the distribution, nodes' persuasiveness and receptiveness should decay with respect to their spatiotemporal distances to the information source, as previously mentioned in Fig. 2. That is the reason why we highlight earlier propagations in the Guiding Rule 1. Moreover, the decay pattern of nodes' persuasiveness and receptiveness reveals boundaries for further propagations: the cascade terminates when nodes have low persuasiveness and receptiveness.



(a) Statistics on σ_i.

(b) σ_1/σ_2 of photo cascades.

Fig. 5. Statistics on singular values of photo cascades.

Let us go back to the element m_{ij} in M. Obviously, m_{ij} is a joint result of followee i's persuasiveness and follower j's receptiveness. Note that a larger value of m_{ij} means an earlier propagation, i.e., a larger persuasiveness of followee i, and a larger receptiveness of follower j. Meanwhile, matrices T and M have included complete time information and partial space information of the cascade: nodes that are closer within the social topology are more likely to be propagated at closer times. Now, the parameter c in the mapping function $f(t) = e^{-ct}$ has another insight meaning: it balances the weights of space and time information. If $c \to 0$, then M is composed of zeros and ones: we only focus on the space information, regardless of time sequences. On the other hand, if c is large, the time information is highlighted. Therefore, M can be viewed as a tuned spatiotemporal information matrix. In the next section, we show the decomposition process through SVD operations, where we extract nodes' persuasiveness and receptiveness from the tuned spatiotemporal information matrix.

V. Spatiotemporal Decomposition

The second stage of our prediction scheme is shown in this section, where we introduce the SVD [13] operation on the weighted matrix M to extract information on nodes' persuasiveness and receptiveness. This is because the element m_{ij} of M represents a joint result of followee i's persuasiveness and follower j's receptiveness.

$$U = \begin{bmatrix} \mathbf{0.00} & 0.00 & 1.00 & 0.00 & 0.00 \\ \mathbf{0.83} & -0.56 & 0.00 & 0.00 & 0.00 \\ \mathbf{0.56} & 0.83 & 0.00 & 0.00 & 0.00 \\ \mathbf{0.00} & 0.00 & 0.00 & -1.00 & 0.00 \\ \mathbf{0.00} & 0.00 & 0.00 & 0.00 & -1.00 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \mathbf{0.98} & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.55 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix} \quad V = \begin{bmatrix} \mathbf{0.26} & 0.68 & 0.69 & 0.00 & 0.00 \\ \mathbf{0.00} & 0.00 & 0.00 & -1.00 & 0.00 \\ \mathbf{0.57} & -0.68 & 0.46 & 0.00 & 0.00 \\ \mathbf{0.78} & 0.27 & -0.56 & 0.00 & 0.00 \\ \mathbf{0.00} & 0.00 & 0.00 & 0.00 & -1.00 \end{bmatrix}$$

Fig. 6. The corresponding SVD result (U, Σ, and V) for the mapped matrix M in Fig. 4(b).

$$M = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.67 & 0.55 & 0.00 \\ 0.45 & 0.00 & 0.00 & 0.55 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix} \qquad M_1 = \sigma_1 u_1 v_1^* = \mathbf{0.98} \cdot \begin{bmatrix} \mathbf{0.00} \\ \mathbf{0.83} \\ \mathbf{0.56} \\ \mathbf{0.00} \\ \mathbf{0.00} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{0.26} \\ \mathbf{0.00} \\ \mathbf{0.57} \\ \mathbf{0.78} \\ \mathbf{0.00} \end{bmatrix}^* = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.21 & 0.00 & 0.46 & 0.63 & 0.00 \\ 0.14 & 0.00 & 0.31 & 0.42 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}$$

Fig. 7. The rank-1 approximation of the matrix M. There is a bounded information loss from M to M_1.

A. SVD Preliminaries and Dataset Verification

In the SVD, M is factorized to a product of three matrices: U, Σ, and V ($M = U\Sigma V^*$, where $*$ is a transpose). The matrix Σ is a diagonal matrix with nonnegative real numbers on the diagonal (generally in descending order), while the diagonal entries σ_i of Σ are known as the singular values of M. The number of singular values equal the matrix rank of M. In this paper, we focus on the SVD's functionality of low-rank approximations, i.e., the matrix M is approximated by vectors. Let u_i and v_i denote the i^{th} columns of matrices U and V, respectively. Assuming r is the rank of M, then we can select the largest k ($k < r$) singular values to approximate M:

$$M_k = \sum_{i=1}^{k} \sigma_i u_i v_i^* \tag{1}$$

where M_k is the approximated M through the k largest singular values. Moreover, the difference between matrices M_k and M is bounded by $||M_k - M||_2 = \sigma_{k+1}$, where $||\cdot||_2$ denotes the 2^{nd} order Frobenius norm. In addition, M can also be accurately represented as $\sum_{i=1}^{r} \sigma_i u_i v_i^*$.

We then conduct experiments on the Flickr dataset, as to verify the effectiveness of this decomposition. The corresponding time matrices of popular photos (i.e., photos that are shared more than 100 times) are mapped by $f(t) = e^{-ct}$ with the parameter c as the reciprocal of the cascade duration, i.e., $c = 1/(\tau_2 - \tau_0)$. Singular values are averaged with respect to different photos, and the result is shown in Fig. 5(a). It can be seen that the difference of σ_1 and σ_2 is much larger than the differences of other consecutive singular values (such as σ_2 and σ_3). The distribution of σ_1/σ_2 is shown in Fig. 5(b). It means that the main pattern of M is highlighted (note that $||M_k - M||_2 = \sigma_{k+1}$), i.e., the cascade information is greatly concentrated in σ_1 and its corresponding vectors (u_1 and v_1). This observation is explained later in subsection C. In addition, for further analysis, the corresponding SVD for the mapped matrix M in Fig. 4(b) is shown in Fig. 6. The low rank ($k = 1$) approximation of M is shown in Fig. 7.

B. Information Decomposition

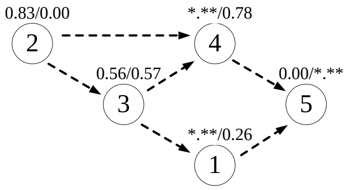
As discussed in Section III, the matrices T and M are sparse. Generally speaking, a sparse matrix has a relatively-low rank, i.e., M should have a few singular values with respect to its size. The relationship between matrix sparsity and rank has been studied in [14]. Moreover, experiments in Fig. 5 show that the largest singular value is a concentration of the cascade information. Therefore, we use $\sigma_1 u_1 v_1^*$ (i.e., M_1) as the compressed cascade information for further processing: predictions are based on u_1 and v_1. Note that this information compression has limited information loss. We can also use more singular values (rather than only using σ_1), which can bring more accurate predictions at the cost of higher time complexities (a tradeoff between accuracy and complexity). Let $\hat{u}_1$ and $\hat{v}_1$ denote the predicted vectors in the future cascade (described later in Section VI), then M can be reconstructed through $\hat{M} = \sigma_1 \hat{u}_1 \hat{v}_1^*$. The predicted number of propagated users can be obtained by mapping M back to the predicted time matrix (i.e., reconstruction).

The decomposition can reduce the difficulties of cascade predictions, since it reduces dimensions for describing cascades. Spatiotemporal cascades are compressed. Instead of matrix operations, vector operations are used to reduce prediction time complexities. According to [13], a centralized SVD of an r-rank $n \times n$ matrix takes a time complexity of $O(rn^2)$.

Moreover, vectors u_1 and v_1 have their insights: u_1 shows nodes' persuasiveness; v_1 represents nodes' receptiveness. As previously mentioned, m_{ij} is a joint result of followee i's persuasiveness and follower j's receptiveness. Meanwhile, the element corresponding to m_{ij} in $\sigma_1 u_1 v_1^*$ (i.e., M_1) is the product of the i^{th} element in u_1 (persuasiveness) and the j^{th} element in v_1 (receptiveness). Note that a larger value in u_1 and v_1 means a larger persuasiveness and receptiveness, respectively, since they would lead to an earlier propagation, i.e., a larger corresponding element in M. The example in Fig. 7 shows the SVD for the cascade in Figs. 1 and 4, while u_1 and v_1 have been shown in Fig. 1(d). Note that only the information on the first four time slots is available now, and we want to predict the cascade of the following time slots. As mentioned in Fig. 1(d), u_1 shows that nodes 2 and 3 are key spreaders, which conforms to Fig. 1(a). v_1 shows that nodes 3 and 4 are more important receivers than node 1, which also meets Fig. 1(a). Meanwhile, u_1 and v_1 also show the decay of nodes' persuasiveness and receptiveness, with respect to their spatiotemporal distances to the information source.

C. SVD Insights and Personalities

As mentioned in subsection A, the cascade information is greatly concentrated with respect to the largest singular value, while σ_1 is almost twice that of σ_2 in Fig. 5. A reasonable explanation for this phenomenon is that *each singular value*



$$\hat{u}_1 = \begin{bmatrix} \mathbf{0.37} \\ 0.83 \\ 0.56 \\ \mathbf{0.83} \\ 0.00 \end{bmatrix} \quad \hat{v}_1 = \begin{bmatrix} 0.26 \\ 0.00 \\ 0.57 \\ 0.78 \\ \mathbf{0.78} \end{bmatrix}$$

(a) Persuasiveness and receptiveness. (b) Predicting u_1 and v_1 in (a).

Fig. 8. The corresponding nodes' characteristics of Fig. 7. In (a), dashed directional edges show the cascade process, while numbers within nodes are their IDs. Labels on top of the nodes are persuasiveness/receptiveness, which are extracted from u_1 and v_1 in Fig. 7. The symbol *.** means needs to be predicted, the results of which are shown in bold font in (b).

represents a cascade mode: σ_1 shows a general global mode, e.g., almost all users enjoy beautiful high-definition photos rather than normal low-definition ones; σ_2 shows a popular mode, e.g., lots of users share beautiful high-definition photos on landscapes; σ_3 shows a comparatively local mode, e.g., a small group of users like landscape photos on mountains; so on so forth. SVD extracts global and common photo cascade modes into larger singular values, while it leaves local and personal photo cascade modes as smaller singular values.

Therefore, our scheme can utilize the information on user personalities to pursue better performances. Let $\bar{\sigma}$, $\bar{u}$, and $\bar{v}$ respectively denote the weight, the additional persuasiveness, and the additional receptiveness brought by user personalities. $\bar{\sigma}$, $\bar{u}$, and $\bar{v}$ of each user can be concluded from the gender, the age, the total number of shared photos, the total online time, and so on. Then, we can revise our prediction through $\hat{M} = \sigma_1 \hat{u}_1 \hat{v}_1^* + \bar{\sigma}\bar{u}\bar{v}^*$. Therefore, our model can easily be extended by considering user personalities.

D. Parallel SVD

Another advantage of our scheme is its parallelism. First, mapping matrices T to M can be done in parallel, since mapping elements in T are independent of each other. Meanwhile, SVD also has parallel methods [15, 16]. Given p processors, SVD of a $n \times n$ matrix can be done [15] within a time complexity of $O(n^3/p)$. The centralized method takes $O(rn^2)$, where r is the rank of the matrix. Note that both the centralized and distributed methods target the complete SVD, while we only need the largest singular value σ_1 and its corresponding vectors. Therefore, there exist possibilities to further reduce time complexities. Since SVD is a standard tool, we do not focus on further improving its efficiency.

VI. Information Cascade Prediction

The third stage of our prediction scheme is described in this section, where we conduct predictions through extracting patterns of u_1 and v_1. Then, we construct the spatiotemporal cascade information as the final prediction.

A. Non-historical Prediction

In this subsection, we predict $\hat{u}_1$ and $\hat{v}_1$ based on the current cascade (called non-historical prediction), i.e., the historical data of former cascades is not utilized. The persuasiveness and receptiveness of unpropagated nodes are predicted based on their *shortest path* to the information source. Here, nodes' persuasiveness and receptiveness are considered as node weights in the shortest path algorithm, while all edge weights are 0. For persuasiveness predictions, nodes with known persuasiveness (non-zero elements in u_1) use their persuasiveness as node weights, while nodes with unknown persuasiveness (i.e., need to be predicted) use constant units as their weights. The receptiveness predictions are similar. The reason for the shortest path is that it has a relatively-high probability (among all paths) of gradually propagating the information from the source to the node. Another reason for using the shortest path is to *complement the space information*, since M only includes partial space information. Along a shortest path, nodes' persuasiveness and receptiveness should decay because of increased spatiotemporal distance to the information source. Note that, an information source's receptiveness is 0 (it only spreads the information out), and the persuasiveness of the end user of a propagation is also 0 (it only receives the information without further propagations).

A simple but effective method is to use the decay of propagated nodes' persuasiveness and receptiveness for predicting that of unpropagated nodes, and an example is shown in Fig. 8, which corresponds to the example in Fig. 7. In Fig. 8(a), the labels on top of nodes represent their persuasiveness and receptiveness (extracted from u_1 and v_1), where the symbol *.** means needs to be predicted. Let us start with the persuasiveness of node 4. Its shortest path to the information source is from node 4 to node 2 directly; therefore, 0.83 is predicted as the persuasiveness of node 4, since no decay pattern exists on this path. Then, the persuasiveness of node 1 can be calculated through the path of nodes 2, 3, and 1. The persuasiveness decay from node 2 to node 3 is $\frac{0.56}{0.83}$, therefore, node 1's persuasiveness is predicted as $\frac{0.56}{0.83} \times 0.56 = 0.37$. Note that, node 5's persuasiveness is predicted to be 0, since it is the end of a propagation path (i.e., it cannot further propagate the information). As for node 5's receptiveness, it is predicted through the path of nodes 2, 4, and 5. Since the source only spreads information, node 5's receptiveness is predicted to be the same as node 4's receptiveness. The prediction results of $\hat{u}_1$ and $\hat{v}_1$ are shown in Fig. 8(b). Then, $\hat{u}_1$ and $\hat{v}_1$ are normalized to be [0.27,0.61,0.41,0.61,0.00]* and [0.20,0.00,0.45,0.62,0.62]*, respectively. Then, we reconstruct $\hat{M} = \sigma_1 \hat{u}_1 \hat{v}_1^*$. The predicted time matrix, $\hat{T}$, can then be obtained through mapping $\hat{M}$ back. Elements $\hat{t}_{15} = 5$ and $\hat{t}_{45} = 9$ in $\hat{T}$ show two predicted propagation times of node 5. We use the minimum values of 5 and 9 as the final prediction (i.e., node 5 will be propagated at time 5), which is the same as the actual cascade in Fig. 1(a).

In the above example, we have not considered the case where the length of the shortest path is longer than three. Instead of using the averages of former decays, we use *the decay of the most similar pair of former nodes* for predictions in a shortest path with larger length. The similarity is defined as *the summation of squared social topological degree differences of followees and followers*. Here, the degree can be either in-degree, out-degree, or both. If the path length is too short to extract the decay information, the followee's persuasiveness and receptiveness are used directly, as shown for predicting node 4's persuasiveness in Fig. 8. The pairwise similarities enable different followers of the same followee to have d-ifferent predictions. To further reveal *directions* of cascades,

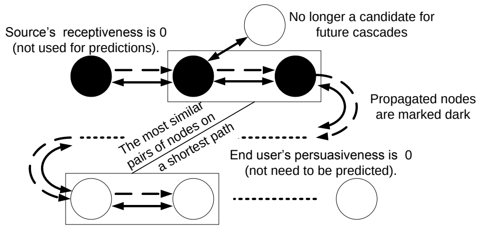

Source's receptiveness is 0
(not used for predictions).
No longer a candidate for
future cascades
Propagated nodes
are marked dark
The most similar
pairs of nodes on
a shortest path
End user's persuasiveness is 0
(not need to be predicted).

Fig. 9. Three rules for non-historical predictions.

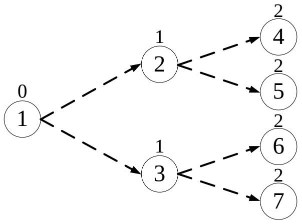


Fig. 10. A case study (the same notation with Fig. 1).

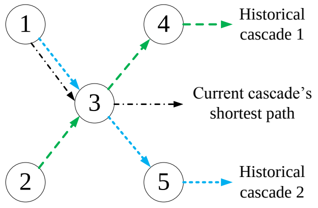


Fig. 11. Historical predictions.

unpropagated nodes within a certain number of hops to the information source are kicked out for considerations of being propagated in the future. In other words, unpropagated nodes near to the source are not receptive, and thus they are no longer candidates for future cascades. This hop-count threshold is empirically determined based on the number of nodes currently propagated. Rules for non-historical predictions are shown in Fig. 9 (currently propagated nodes are marked dark while the remaining nodes are unpropagated at present) and are summarized as follows:

- The information source's receptiveness is 0, and is not used for receptiveness predictions along the shortest path. Meanwhile, the persuasiveness of the user at the end of the shortest path is fixed to be 0.
- Predictions are based on shortest paths. Along a shortest path, the persuasiveness and receptiveness decay between a pair of nodes are predicted as the corresponding decay between the most similar (in terms of degree differences) pair of currently propagated nodes. Nodes' persuasiveness and receptiveness can be derived from decays.
- Unpropagated nodes within a certain number of hops to the information source are kicked out for being propagated in the future. They are not receptive, and thus, are no longer candidates for future cascades, i.e., they are influenced by the cascade without further propagations (influenced but not propagated).

A supervised learning process on the pattern of nodes' persuasiveness and receptiveness should bring a better prediction. However, it also has a higher time complexity as a tradeoff. Since the current method has obtained a good result, we do not further explore learning-based methods. To better understand decay patterns, a case study on "branching" cascades is conducted, which is shown in Fig. 10. This type of cascade spreads without resistances, where the number of propagated nodes increases exponentially. Assuming the usage of $c = \frac{1}{2}$ for the mapping process, the decomposition result for the cascade in Fig. 10 is $u_1 =$ [1, 0, 0, 0, 0, 0, 0]* and $v_1 =$ [0, 0.71, 0.71, 0, 0, 0, 0]*. In other words, the cascade is compressed into relationships among nodes 1, 2, and 3, since the later cascade repeats their propagation mode. Therefore, the pattern of this "branching" cascade can be captured.

B. Historical Prediction

In the previous subsection, we predicted a cascade without historical information. Now, we study predictions based on former cascades, i.e., using decay patterns of nodes' persuasiveness and receptiveness in former cascades to help predict the current cascade (called historical prediction). The prerequisite of historical predictions is that former cascades are homogenous with the current one: cascades of popular photos are different than unpopular ones; therefore, we should not use the historical data on cascades of popular photos to predict cascades of unpopular ones.

Shortest paths of the current cascade are cooperatively used with the historical data. Instead of calculating decays of nodes' persuasiveness and receptiveness based on currently propagated nodes, we use decays of former cascades as predictions. An example is shown in Fig. 11: the black dashed directional path indicates a shortest path of the current cascade. The historical decay of cascade 2 is used to predict the decay from node 1 to 3, while cascade 1 is not used, since it does not have an intersection with the decay from nodes 1 to 3. In the case of multiple available historical cascades, the decay of the current cascade is predicted to be their average decay.

C. Algorithm Complexities

As previously mentioned, S and T include all users in the network. However, this is unnecessary, since most cascades only influence a very small portion of users in the network. Therefore, for practical usage, we can have a subgraph just large enough for predictions (e.g., all users that are within 5 hops of the information source and their relationships).

We have used shortest paths with node weights in predictions; however, this can be solved by slightly modifying Dijkstra's algorithm (use the node weight instead of the edge weight when greedily adding a new node). Therefore, it has the same time complexity with the normal Dijkstra's algorithm. Let N_p and N ($N_p \ll N$), respectively, denote the number of currently propagated nodes and total nodes (E_p and E to represent the number of corresponding edges). Then, the centralized Dijkstra's algorithm takes $O(E + N \log N)$ through a Fibonacci heap. Calculating decays (and nodes' persuasiveness or receptiveness) can follow the same order of the shortest path. Since the path length is bounded by the network diameter D ($D \leq N$), the decay calculation takes at most $O((E + N \log N)D)$. The mapping and its reversion (mapping $\hat{M}$ back to $\hat{T}$) takes $O(N^2)$. The centralized SVD takes $O(rN_p^2)$, where r is the rank of M. Here, we do not need $O(rN_p^2)$ for the SVD, since the decomposition results for currently unpropagated nodes are useless; the persuasiveness and receptiveness corresponding to unpropagated nodes are 0, and we only need to decompose the cascade information among currently propagated nodes. Therefore, the total time complexity is $O(N^2 + DN \log N + rN_p^2)$ in a centralized calculation method, when considering a sparse graph.

According to [17], Dijkstra's algorithm can be done in parallel. The idea is to divide the graph into pieces for each processor. Given p processors, the time complexity can be brought down to $O(N^3/p)$ (a more accurate description is given in [17]). The SVD takes $O(N_p^3/p)$. Mapping and its reversion can be solved in parallel, since each element's map is independent from the others. So the total time complexity of our scheme is $O((N^3+N_p^3)/p)$ in parallel.

If we conduct predictions with whole information S and T directly, then the time complexity will not be acceptable. For each unpropagated node, we need to scan and process S and T, which takes at least $O(N^2)$. Therefore, at least $O(N^3)$ is needed for a centralized method. Even if direct predictions can be implemented in parallel, they should have a higher time complexity than $O(N^3/p)$ due to the overhead. Meanwhile, our decomposition method has compressed all cascade information (S and T) into nodes' persuasiveness and receptiveness with limited information loss, resulting in a reduced time complexity.

VII. Evaluation

In this section, extensive evaluations are conducted. After presenting the basic settings, we show baseline algorithms and evaluation metrics. Finally, the evaluation results are shown from different perspectives to provide insightful conclusions.

A. System Settings

Our evaluations focus on cascades of popular photos that are marked "favorite" more than 100 times, since photos of different levels of popularity stand for cascades of different types. However, each photo may be involved in multiple cascades: unconnected users (in terms of social topology) may share the same photo coincidentally, leading to different cascades in the network. Therefore, for each popular photo, we select its earliest cascade (in the sense of the earliest appearance time of the information source) for predictions. Generally speaking, the earliest cascade is also the largest one.

For a photo cascade predition, it is almost impossible to take all unpropagated nodes into consideration, since we have 2,302,925 users in total. Meanwhile, only 25 photos are propagated over more than 1,000 users. Therefore, a subgraph is expected to improve the prediction efficiency. This subgraph is constructed as a combination of (1) all propagated users of the earliest cascade of the photo, (2) three random out-going neighbors for each of these propagated users, and (3) all social topology relationships between the above users.

The input parameters of our prediction scheme include the complete cascade information (from the time τ_0 to τ_1), while we will predict the size of the cascade at its finishing time τ_2. In the following experiments, we will tune the ratio of τ_1/τ_2 to observe the performance of our scheme. This value stands for the amount of prior knowledge, i.e., a larger τ_1/τ_2 should bring a better prediction. Note that, our prediction scheme will not achieve 100% accuracy even if $\tau_1/\tau_2 = 1$, since the information compression from M to u_1 and v_1 is lossy. In addition, the parameter c is set to be $1/\tau_1$, while the hot count threshold is empirically set to be 3. As for user personalities, we use normalized out-degree and in-degree to respectively describe $\bar{u}$ and $\bar{v}$, while we set $\bar{\sigma}$ to be $0.1 \times \sigma_1$.

B. Baseline Algorithms and Evaluation Metrics

Three baseline algorithms (the first two prediction schemes are non-historical schemes, while the last scheme is a historical scheme) are used for comparison as follows. (1) Largest in-degree: among all unpropagated nodes, the node with the largest in-degree (in terms of the social topology) is considered to be the next propagated node. The propagation time delay is considered as the largest propagation time delay of the current cascade. This scheme is based on the observation that a user with a larger in-degree is more likely to accept new information. (2) Most influenced: among all unpropagated nodes, the node that has the largest number of incoming propagated neighbors is considered to be the next propagated node. The propagation time delay is also the largest propagation time delay of the current cascade. This scheme is based on the observation that a user with more in-neighbors in the cascade is more likely to be influenced. (3) Most active: among all unpropagated nodes that are outgoing neighbors of propagated nodes, the node that is the most active, in terms of having been propagated by former cascades for the most number of times, is considered to be the next propagated node. The propagation time delay is calculated as the historical delay.

As for the evaluation metrics, the standard Receiver Operating Characteristic (ROC) metrics [18] are employed, including the detection rate (the higher the better), the false positive rate (the lower the better), and the accuracy (the higher the better). More details can be found in [18].

C. Evaluation Result

The evaluation result is shown in Fig. 12, in terms of non-historical (top row) and historical (bottom row) prediction schemes. Each column corresponds to one of the three ROC metrics. For the non-historical schemes, the proposed algorithm outperforms the two naive baselines, among all three metrics. This is beacuase our algorithm considers spatiotemporal information, while the two naive algorithms mainly focus on the space information. Overall, our algorithms get about 20% higher accuracy than the two baselines. Another observation is that all these schemes have diminishing return effects: the increasing rate of the accuracy decreases with respect to τ_1/τ_2. This is because the early propagations are more important and more deterministic for the future trend of the cascade, and thus the amount of information contributed by a early propagation is larger than that by a late propagation. The initial information helps predict the cascade framework, while the following information just fulfills predicting details of the cascade. The prediction gain is marginal when $\tau_1/\tau_2 \geq 0.1$.

As for the historical predictions (bottom line), it can be seen that they perform better than non-historical schemes, since additional information is utilized. Meanwhile, the baseline algorithm (i.e., most active) does not have a very good performance, since it relies on the user histories too much, without considerations of the spatiotemporal propagations of the current cascade. Our algorithm extracts users' persuasiveness and receptiveness from former cascades, and then combines that information with the spatiotemporal information of the current cascade to obtain a better result. It can been seen that the historical prediction has an accuracy of about 0.9 when $\tau_1/\tau_2 = 0.1$. The corresponding detection rate and false positive rate is more than 0.8, and less than 0.1, respectively.

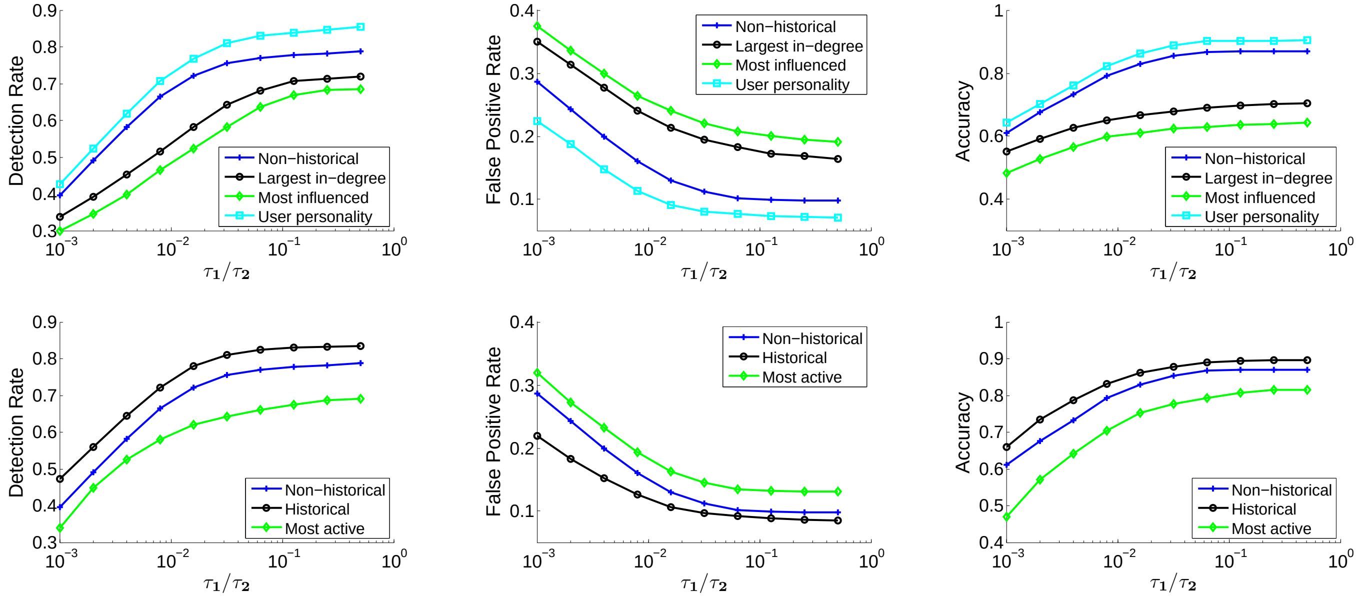


Fig. 12. The evaluation results. The top row shows non-historical prediction schemes (The algorithm "User personality" is the proposed non-historical scheme with additional considerations on user personalities), while the bottom row consists of historical prediction schemes. Note that the history information has included the information on user personalities. Each of the three columns indicates one of the three metrics (detection rate, false positive rate, accuracy).

VIII. Conclusion

Information cascade predictions are important, due to their functionalities of detecting bad cascades. Given the current cascade and the social topology, we want to predict the cascade size at a future time slot. In a macro view, a cascade is described by space and time dimensions: the time information also includes partial space information, since closer nodes in the social topology are more likely to propagate information at closer times. In a micro view, we use the spatiotemporally-sensitive persuasiveness and receptiveness to respectively describe followees and followers. The SVD operation is used to decompose the spatiotemporal cascade information (matrices) into vectors u_1 and v_1, which stand for nodes' persuasiveness and receptiveness, respectively. Predictions are conducted based on these vectors, as to have a low time complexity. User personalities can also be incorporated into our scheme. Furthermore, our prediction scheme can be implemented in parallel. Finally, extensive real-data driven evaluations verify the competitive performance of the proposed scheme.

References

[1] M. Cha, A. Mislove, and K. P. Gummadi, "A Measurement-driven Analysis of Information Propagation in the Flickr Social Network," in *Proc. of WWW 2009*, pp. 721–730.

[2] E. Henry and J. Hofrichter, "Singular value decomposition: application to analysis of experimental data," *Essential Numerical Computer Methods*, vol. 210, pp. 81–138, 2010.

[3] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. of SIGKDD 2003*, pp. 137–146.

[4] W. Galuba, K. Aberer, D. Chakraborty, Z. Despotovic, and W. Kellerer, "Outtweeting the twitterers - predicting information cascades in microblogs," in *Proc. of WOSN 2010*, pp. 3–3.

[5] V. Gómez, H. J. Kappen, and A. Kaltenbrunner, "Modeling the structure and evolution of discussion cascades," in *Proc. of HT 2011*, pp. 181–190.

[6] K. Saito, R. Nakano, and M. Kimura, "Prediction of information diffusion probabilities for independent cascade model," *Knowledge-Based Intelligent Information and Engineering Systems*, vol. 5179, pp. 67–75, 2008.

[7] G. Ghasemiesfeh, R. Ebrahimi, and J. Gao, "Complex contagion and the weakness of long ties in social networks: revisited," in *Proc. of EC 2013*, pp. 507–524.

[8] E. Sadikov, M. Medina, J. Leskovec, and H. Garcia-Molina, "Correcting for missing data in information cascades," in *Proc. of WSDM 2011*, pp. 55–64.

[9] P. Mohan, S. Shekhar, J. A. Shine, and J. P. Rogers, "Cascading spatio-temporal pattern discovery: A summary of results," DTIC Document, Tech. Rep., 2010.

[10] Y. Zhao and J. Wu, "Dache: A data aware caching for big-data applications using the mapreduce framework," *Proc. of INFOCOM 2013*, pp. 35–39.

[11] P. A. Dow, L. A. Adamic, and A. Friggeri, "The anatomy of large facebook cascades," in *Proc. of ICWSM 2013*.

[12] C. Hui, Y. Tyshchuk, W. A. Wallace, M. Magdon-Ismail, and M. Goldberg, "Information cascades in social media in response to a crisis: a preliminary model and a case study," in *Proc. of WWW 2012*, pp. 653–656.

[13] M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra and its Applications*, vol. 415, no. 1, pp. 20–30, 2006.

[14] J.-G. Dumas and G. Villard, "Computing the rank of large sparse matrices over finite fields," *Computer Algebra in Scientific Computing CASC, Technische Universität München*, 2002.

[15] P. Shah, C. Wieser, and F. Bry, "Parallel higher-order SVD for tag-recommendations," in *Proc. of IADIS International Conference WWW/Internet 2012*.

[16] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2002.

[17] http://www.mcs.anl.gov/~itf/dbpp/text/node35.html.

[18] D. L. Streiner and J. Cairney, "What's under the ROC? an introduction to receiver operating characteristic curves," *A Guide to the Statistically Perplexed*, vol. 52, no. 4, p. 304, 2013.