Dynamic Shaker Detection from Evolving Entities

Xiaoxiao Shi* Wei Fan† Philip S. Yu*‡
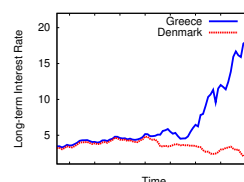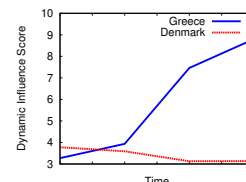
Abstract

Finding the most influential entities as well as conducting causality analysis is an important topic in economics, healthcare, sensor networks, etc. One famous example in economics is the bankruptcy of Lehman Brothers that triggered the 2008 global financial crisis. In recent years, some works were proposed to infer the causal relationships among several entities, and subsequently find the most influential ones (*i.e.*, shakers). However, most of the previous works assume that the causal relationships and the shakers are static. In other words, they are assumed to be stable over time. This assumption may not necessarily be true, especially when we study volatile entities or long-term time series. In this paper, we propose a dynamic model called "DShaker" to capture the evolving causal relationships and dynamic shakers. The intuition is to model the causality propagation into a graph called "dynamic cascading graph". We then find the optimal cascading graphs, by maximizing their likelihoods in a non-convex multi-objective optimization formulation. We solve it by mapping into a trace norm minimization problem. Experiments included three datasets in social sciences. The proposed method can effectively capture those entities with increasing impacts, while existing methods missed most of them. For example, in the experiment of studying the banks' statistics from 1998 to 2007 (before the financial crisis), the proposed method successfully captures Lehman Brothers as one of the most precarious banks in subprime loans.

1 Introduction

In an interconnected world, it is an important topic to infer the causal relationships among evolving entities, and detect the most influential ones (*i.e.*, shakers [19]). For example, the bankruptcy of Lehman Brothers triggered the 2008 global financial crisis; the surge of Greek debts leads to the recent European sovereign debt crisis. These shakers are having significant impacts to the world. In recent years, two lines of works had been proposed to find the shakers. The first set of works assumes that the entities have binary states (active and inactive), and the states can be turned from

*Computer Science Department, University of Illinois at Chicago, USA. {xshi9, psyu}@uic.edu.

†Huawei Noahs Ark Lab, China. david.fanwei@huawei.com.

‡Computer Science Department, King Abdulaziz University, Jeddah, Saudi Arabia

(a) Increasing influence of Greece reflected by long-term interest rate

(b) Increasing influence of Greek debts reflected by the proposed model

Figure 1: Increasing Influence of Greece from Year 2005 to Year 2011

inactive to active if the entities are affected by some shakers. Models like independent cascade (IC [13]) and influence maximization model (*e.g.*, [9, 15, 22]) are built on the binary-state assumption. As an important improvement, recently, [19] proposes a framework that allows entities to have states with continuous values. The authors in [19] propose to use "cascading graph" to capture the causal relationships among the entities, and use a static model to capture the shakers. The method in [19] can thus be used in quantitative areas such as finding the financial market shakers.

However, the approach proposed in [19] assumes that the underlying causal relationships and the shakers are static. In other words, given a set of evolving entities over time T, it is assumed that the causal relationships never change during the given period, neither are the shakers. We thus refer to the model in [19] as "static shaker model" in the rest of the paper. The "static shaker" assumption may not necessarily be true, especially when (1) either the time period is relatively long, or (2) the evolving data is quite volatile in the given period. An example is illustrated in Fig. 1, which shows the changing influence of Greece. Fig. 1(a) plots the changing long-term interest rate of Greece as compared to Denmark. As can be observed, the two countries have similar interest rates at the beginning. However, as time goes by, the interest rate of Greece surges to over 20%, which is a strong signal that finally triggers the recent European sovereign debt crisis. Under the assumption of a static causal relationship and a stable influence, the IC model as well as the static shaker model [19] identify that Denmark has a comparable influence to Greece. This result is contrary to the reality. More details of the results are studied in the experiment

section (Section 4).

In this paper, we propose a dynamic model called "DShaker" to capture evolving causal relationships and detect dynamic influential entities (such as Greece in the above example). However, there are at least three challenges.

- First, it is not clear on how to capture the evolving causal relationships, especially given that the traditional static shaker model cannot be easily adapted to solve the problem.

- Second, it is not obvious on how to filter the noisy time series signals when capturing the causal relationships.

- Third, suppose we can successfully capture the evolving causal relationship. It is not clear on how to detect the most influential entities (dynamic shakers).

We propose a multi-objective optimization formulation to solve the above problems. The basic idea is to find a series of evolving adjacency matrices (cascading graphs) by maximizing their likelihood given the observed time series. Furthermore, low rank minimization technique is adopted to reduce noise and amplify the signals emitted from the underlying shakers. With no directly applicable methods to solve the non-convex multi-objective optimization problem, we propose a strategy derived from combining gradient descent with trace norm minimization. After solving the optimization, we obtain a set of evolving cascading graphs. Three metrics are proposed to capture the dynamic shakers on the set of cascading graphs. They are used to evaluate (1) average influence, (2) time-decay influence, and (3) potential influence, respectively.

We conducted three sets of experiments. These experiments focused on the following problems in social sciences: finding the European economy shakers, finding the most precarious banks based on their subprime loan records, and finding the NASDAQ stock market shakers. Lacking directly applicable dynamic methods, the independent cascade model as well as the static shaker model ([19]) is used as comparison methods. It is observed that the proposed model can effectively capture the dynamic shakers. For example, in the experiment of finding stock market shakers over 3 years of data, the DShaker model outperforms the baselines by as much as 40% in accuracy, with the ground truth coming from Yahoo! Finance.

2 Preliminary

Given a set of evolving entities, the objective is to find shakers whose value changes can significantly affect the values of many other entities. We next formally define the related concepts.

DEFINITION 1. (EVOLVING ENTITY) *Denote $v_i \in \mathcal{V}$ as an entity where $i \in \{1, 2, \cdots, |\mathcal{V}|\}$. Let $a_t(i) \in \mathbb{R}$ be the*

attribute of v_i that evolves over the time variable t. In this paper, the time variable is considered to be discrete, and we use $a_t(i)$ to denote the attribute value of entity v_i at time step t.

As in [19], it is assumed that the value of the entity v_i can be influenced by the other entities under the Markovian assumption. In other words, $a_t(i)$ depends on the values of all entities at time $t - 1$:

$$(2.1) \qquad a_t(i) = \sum_{j=1}^{|\mathcal{V}|} w_{(j,i)} a_{t-1}(j)$$

where $w_{(j,i)}$ is the weight of influence from v_j to v_i. It has the following property:

- The value $|w_{(j,i)}|$ is large if v_j has significant influence on v_i.

- $w_{(j,i)} < 0$ if v_j has a negative influence on v_i; that is, the rising of $a_{t-1}(j)$ will cause $a_t(i)$ to decrease.

- $w_{(j,i)} > 0$ if v_j has a positive influence on v_i.

In matrix forms, Eq. 2.1 can be written as:

$$(2.2) \qquad \mathbf{a}_t = \mathbf{a}_{t-1} \mathbf{W}$$

where $\mathbf{a}_t = [a_t(1), a_t(2), \cdots, a_t(|\mathcal{V}|)] \in \mathbb{R}^{1 \times |\mathcal{V}|}$ and $\mathbf{W} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ where the (j, i)-th entry is $w_{(j,i)}$, or the influence from node v_j to node v_i. In [19], the matrix $\mathbf{W}$ is also called the adjacency matrix of a *cascading graph*. With these definitions, a shaker is a node that causes significant value changes of the other nodes. Several quantitative measurements are proposed to capture shakers in Section 3.3.

3 Proposed Algorithm: DShaker

In this section, we explore the strategy to construct dynamic cascading graphs, and further propose three metrics to capture the dynamic shakers. The general flow of the proposed model is summarized in Fig. 2.

3.1 Cascading Graph Construction In [19], the cascading graph is constructed via evaluating the pairwise correlation between each pair of entities. There are several drawbacks of this strategy. First, it is very time-consuming to perform the pairwise computation. Second, it neglects the fact that the influence can be accumulated from a combination of many entities, instead of a pairwise relationship. In addition, it is not clear how to adapt this strategy to capture dynamic causal relationships. In this paper, we propose a new formulation to construct the cascading graph $\mathbf{W}$. The intuition behind is to solve $\mathbf{W}$ given a series of equations derived from Eq. 2.2:

$$(3.3) \qquad \begin{cases} \mathbf{a}_2 = \mathbf{a}_1 \mathbf{W} \\ \cdots \\ \mathbf{a}_{S+1} = \mathbf{a}_S \mathbf{W} \end{cases}$$
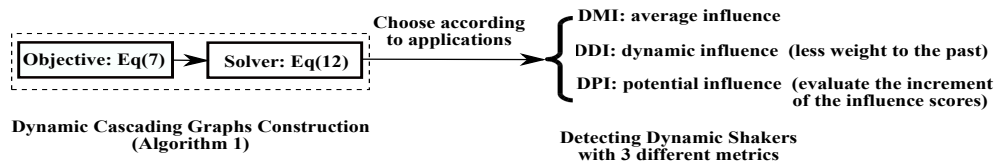
Figure 2: The DShaker algorithm flow

It can be written as

$$(3.4) \qquad \mathbf{A}^{S+1} = \mathbf{A}^S \mathbf{W}$$

where $\mathbf{A}^N = [\mathbf{a}'_{N-S+1}, \cdots, \mathbf{a}'_{N-1}, \mathbf{a}'_N]' \in \mathbb{R}^{S \times |\mathcal{V}|}$ and S is the length of the time interval to construct the cascading graph. It is important to note that if we have sufficient data and $\mathbf{A}^S$ is nonsingular, the optimal cascading graph $\mathbf{W}$ can be obtained directly by $\mathbf{W} = (\mathbf{A}^S)^{-1} \mathbf{A}^{S+1}$. However, there is no guarantee that the matrix $\mathbf{A}^S$ is nonsingular. Furthermore, intuitively, if $S < |\mathcal{V}|$, there are fewer equations than the unknowns, and hence there will be many solutions to Eq. 3.4. Hence, we should have extra constraints in solving the cascading graph $\mathbf{W}$.

Static Cascading Graph Construction In the following paragraphs, we first introduce how to construct static cascading graph, and then extend it to construct dynamic graphs. We first use least square estimator to solve Eq. 3.4, and also impose a low rank constraint on the cascading graph $\mathbf{W}$ as follows

$$(3.5) \qquad \begin{aligned} \min_{\mathbf{W}} &\|\mathbf{A}^{S+1} - \mathbf{A}^S \mathbf{W}\|_F^2 \\ &\text{w.r.t. } \mathbf{rank}(\mathbf{W}) < r \end{aligned}$$

where $\|\mathbf{A}\|_F$ is the Frobenious norm, and $\mathbf{rank}(\mathbf{W})$ is the rank of the matrix $\mathbf{W}$, and r is the upper bound of the rank. The intuition of the low rank constraint is as follows. First, there should be only a limited number of shakers that affect the entities. This assumption is widely used in the financial market, where only 30 stocks are selected to construct the DOW stock index, and only 5 currency pairs are chosen to construct the US dollar strength index, etc. Second, the adjacency matrix $\mathbf{W}$ in solving Eq. 3.4 may contain significant noise, which is the basic characteristic of most time series data [6]. The low rank constraint acts as a regularization term to reduce the impact of noise. Note that Eq. 3.5 is not a convex function and it is not easy to solve. We follow the idea of trace norm minimization [7] to cast Eq. 3.5 into a relaxed optimization problem as follows:

$$(3.6) \qquad \min_{\mathbf{W}} \|\mathbf{A}^{S+1} - \mathbf{A}^S \mathbf{W}\|_F^2 + \lambda \|\mathbf{W}\|_*$$

where $\|\mathbf{W}\|_*$ is the trace norm (nuclear norm) of the matrix $\mathbf{W}$, defined as the sum of the singular values of the matrix. Moreover, λ is the regularization parameter, which is studied

in the experiment section. It is shown in [7] that Eq. 3.6 is the convex envelope of Eq. 3.5 over the unit ball of spectral norm.

Dynamic Cascading Graph Construction With the static construction formulation in Eq. 3.6, we next introduce how to allow the dynamic evolution of the cascading graph $\mathbf{W}$. We first denote $\mathbf{W}^{(t-1)}$ as the cascading graph at time $t-1$. We then show how to construct the new cascading graph $\mathbf{W}^{(t)}$ at time t, which allows a smooth evolution from $\mathbf{W}^{(t-1)}$ as follows

$$(3.7)$$
$$\begin{aligned} \min_{\mathbf{W}^{(t)}} F(\mathbf{W}^{(t)}) = &\|\mathbf{A}^{S+1} - \mathbf{A}^S \mathbf{W}^{(t)}\|_F^2 \\ &+ \gamma \|\mathbf{W}^{(t)} - \mathbf{W}^{(t-1)}\|_F^2 + \lambda \|\mathbf{W}^{(t)}\|_* \end{aligned}$$

where the first term is to minimize the modeling error as in Eq. 3.6. The second term is to allow a smooth evolution of $\mathbf{W}^{(t)}$ from $\mathbf{W}^{(t-1)}$. In other words, although new cascading graph $\mathbf{W}^{(t)}$ is allowed, it should not be too different from the previous one $\mathbf{W}^{(t-1)}$. The last term is the trace norm term. The sensitivities of the regularized parameters γ and λ are studied in the experiment section.

We summarize the challenges of the problem as follows:

1. The proposed cascading graph construction formulation in Eq. 3.4 is not directly solvable because of the nonunique solutions and noise. Even after imposing a low rank constraint as in Eq. 3.5, the optimization objective is non-convex, which is still difficult to solve.

2. After we cast the non-convex problem into a convex trace norm minimization in Eq. 3.7, it contains multiple objectives (minimize $\|\mathbf{A}^{S+1} - \mathbf{A}^S \mathbf{W}^{(t)}\|_F^2$ and $\|\mathbf{W}^{(t)} - \mathbf{W}^{(t-1)}\|_F^2$). One cannot use traditional trace norm minimization solver (*e.g.*, singular value thresholding or SVT) to solve it.

3. When applying a gradient descent approach to solve the problem, we face the difficulty that the derivative cannot be easily calculated. Hence we adopt a novel approach using the singular value decomposition techniques and then applying soft-thresholding on the singular values. More details are introduced in the next section.

We describe the framework of the algorithm as in Algorithm 1. The first line of the code calculates the minimal time

Input: $\mathbf{D}$ is the dataset; M is the maximal number of cascading graphs; T is the maximal time step; α is the ratio of the maximal error, default value is 1.5.

Output: A series of cascading graphs $\mathbf{W^{(1)}}, \mathbf{W^{(2)}}, \mathbf{W^{(3)}}, \cdots, \mathbf{W^{(M)}}$

1 Construct the minimal time interval to construct the cascading graph: $\tilde{t} = T/M$;

2 Construct the matrix $\mathbf{A}^{\tilde{t}}$ and $\mathbf{A}^{\tilde{t}+1}$;

3 Solve $\mathbf{W^{(1)}}$ by solving Eq. 3.6, let
$\epsilon \leftarrow \|\mathbf{A}^{\tilde{t}+1} - \mathbf{A}^{\tilde{t}}\mathbf{W}^{(1)}\|_F^2$

4 **for** $t = 2$ *to* M **do**

5   Construct the matrix $\mathbf{A}^{\tilde{t}+t\tilde{t}}$ and $\mathbf{A}^{\tilde{t}+t\tilde{t}+1}$ accordingly;
    /* Test whether the cascading graph should be updated. */

6   **if** $\|\mathbf{A}^{\tilde{t}+t\tilde{t}+1} - \mathbf{A}^{\tilde{t}+t\tilde{t}}\mathbf{W}^{t-1}\|_F^2 > \alpha \times \epsilon$ **then**

7   Solve $W^{(t)}$ by solving Eq. 3.7;
        $\epsilon \leftarrow \|\mathbf{A}^{\tilde{t}+t\tilde{t}+1} - \mathbf{A}^{\tilde{t}+t\tilde{t}}\mathbf{W}^{(t)}\|_F^2$;

8   **else**

9         $\mathbf{W^{(t)}} \leftarrow \mathbf{W^{(t-1)}}$;

10   **end**

11 **end**

12 Return the series of cascading graphs
$\mathbf{W^{(1)}}, \mathbf{W^{(2)}}, \mathbf{W^{(3)}}, \cdots, \mathbf{W^{(M)}}$.

Algorithm 1: Dynamic Cascading Graph Construction

interval $\tilde{t}$ to update the dynamic cascading graph. The algorithm will then automatically update the cascading graphs at each time interval. The second line constructs the matrices $\mathbf{A}^{\tilde{t}}$ and $\mathbf{A}^{\tilde{t}+1}$, and the third line calculates the first cascading graph by using the formula in Eq. 3.6. Step 4 to Step 12 continuously construct new cascading graphs as the time series proceeds. It first checks whether we can directly use the previous cascading graph in Step 6. If the error by using old cascading graph is tolerable, the cascading graph is not updated; otherwise, it solves Eq. 3.7 to update a new cascading graph. As such, the output of the algorithm is a set of temporal cascading graphs $W^{(1)}, W^{(2)}, W^{(3)}, \cdots, W^{(M)}$. We next introduce how to solve Eq. 3.7, followed by the details of capturing the shakers, given the set of evolving temporal cascading graphs.

3.2 Optimization Note that the optimization in Eq. 3.7 is convex. Hence we can approach the global optimal solution via gradient descent. In other words, the optimal solution will be found by an iterative process with the following updating rule:

$$(3.8) \qquad \mathbf{W}_i^{(t)} = \mathbf{W}_{i-1}^{(t)} - g(\mathbf{W}_{i-1}^{(t)})\nabla F(\mathbf{W}_{i-1}^{(t)})$$

where $\mathbf{W}_{i-1}^{(t)}$ and $\mathbf{W}_i^{(t)}$ are the $\mathbf{W}^{(t)}$ at the $(i-1)$-th iteration and the i-th iteration respectively, $F(\mathbf{W}_{i-1}^{(t)})$ is the objective function, and $g(\mathbf{W}_{i-1}^{(t)})$ is the step size of the gradient descent. Usually, $g(\mathbf{W}_{i-1}^{(t)})$ is set to be a constant. In later section, we show that we can choose a better step size $g(\mathbf{W}_{i-1}^{(t)})$ to boost the convergence.

It is important to note that the objective function $F(\mathbf{W}_{i-1}^{(t)})$ in Eq. 3.7 contains a trace norm term $\|\mathbf{W}\|_*$. It is thus not obvious on how to calculate the derivative $\nabla F(\mathbf{W}_{i-1}^{(t)})$. We follow the idea of [11] to update $\mathbf{W}^{(t)}$ in two steps in each iteration. We first denote

$$Z(\mathbf{W}^{(t)}) = \|\mathbf{A}^{S+1} - \mathbf{A}^S\mathbf{W}^{(t)}\|_F^2 + \gamma\|\mathbf{W}^{(t)} - \mathbf{W}^{(t-1)}\|_F^2$$

Hence, $F(\mathbf{W}^{(t)}) = Z(\mathbf{W}^{(t)}) + \lambda\|\mathbf{W}^{(t)}\|_*$. We next calculate
(3.9)
$$\nabla Z(\mathbf{W}_{i-1}^{(t)}) = 2(\mathbf{A}^S)^T(\mathbf{A}^{S+1} - \mathbf{A}^S\mathbf{W}^{(t)}) + 2\gamma(\mathbf{W}^{(t)} - \mathbf{W}^{(t-1)}).$$

We then update $\mathbf{W}_i^{(t)}$ with the following equation
(3.10)
$$\min_{\mathbf{W}_i^{(t)}} \|\mathbf{W}_i^{(t)} - \left(\mathbf{W}_{i-1}^{(t)} - g(\mathbf{W}_{i-1}^{(t)})\nabla Z(\mathbf{W}_{i-1}^{(t)})\right)\| + \lambda\|\mathbf{W}_i^{(t)}\|_*$$

It is proven in [3] that the minimization with the form like Eq. 3.10 can be solved by first computing the singular value decomposition (SVD) of the matrix $\mathbf{W}_{i-1}^{(t)} - g(\mathbf{W}_{i-1}^{(t)})\nabla Z(\mathbf{W}_{i-1}^{(t)})$, and then applying some soft-thresholding on the singular values. This process is summarized as follows [3].

THEOREM 3.1. *Let $\mathbf{M} \in \mathbb{R}^{m \times k}$ and let $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^T$ be the singular value decomposition of $\mathbf{M}$ where $\mathbf{U} \in \mathbb{R}^{m \times r}$ and $\mathbf{V} \in \mathbb{R}^{k \times r}$ have orthonormal columns, $\Sigma \in \mathbb{R}^{r \times r}$ is diagonal, and r is the rank of the matrix $\mathbf{M}$. Then*

$$(3.11) \qquad \arg\min_{\mathbf{W}} \frac{1}{2}\|\mathbf{W} - \mathbf{M}\| + \lambda\|\mathbf{W}\|_*$$

can be solved by $\mathbf{W} = \mathbf{U}\Sigma_\lambda\mathbf{V}^T$, where Σ_λ is diagonal with $(\Sigma_\lambda)_{ii} = \max\{0, (\Sigma)_{ii} - \lambda\}$.

With Theorem 3.1, we can update the matrix $\mathbf{W}^{(t)}$ at each iteration by

1. Calculate $\mathbf{M} = \left(\mathbf{W}_{i-1}^{(t)} - g(\mathbf{W}_{i-1}^{(t)})\nabla Z(\mathbf{W}_{i-1}^{(t)})\right)$;

2. Use singular thresholding as in Theorem 3.1 to update the value of $\mathbf{W}_i^{(t)}$.

In the updating rule, we can choose different step size $g(\mathbf{W}_{i-1}^{(t)})$ in each iteration. We next introduce a strategy to choose the optimal $g(\mathbf{W}_{i-1}^{(t)})$ to boost the optimization. We first introduce the following theorem.

THEOREM 3.2. *By following Theorem 3.2 in [11], if $g(\mathbf{W}_i)$ satisfies the following equation*

(3.12)
$$F(\mathbf{W}_{i+1}) \leq Z(\mathbf{W}_i)$$
$$+ \mathbf{Tr}\Big((\mathbf{W}_{i+1} - \mathbf{W}_i)\nabla F\Big) + g(\mathbf{W}_i)\|\mathbf{W}_{i+1} - \mathbf{W}_i\|_F^2$$

we have

(3.13) $$F(\mathbf{W}_n) - F(\hat{\mathbf{W}}) \leq \frac{g(\mathbf{W}_n)\|\mathbf{W}_0 - \hat{\mathbf{W}}\|_F^2}{2n}$$

where $\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} F(\mathbf{W})$.

Theorem 3.2 can be directly derived from Theorem 3.2 in [11] by doing some algebra manipulations. Because of limited space, the proof is omitted here. With Theorem 3.2, we can improve the convergence of the optimization by judiciously choosing the value of $g(\mathbf{W}_i)$ to satisfy Eq. 3.12. In the experiment, we starts with a small value of $g(\mathbf{W}_i)$, and keep double its value until Eq. 3.12 is satisfied.

3.3 Evaluation Metrics for Dynamic Shakers Note that the above algorithm generates a series of cascading graphs evolving over time. We thus cannot use the algorithm proposed in [19] to find shakers because it only works on a single static graph. In this section, we derive new metrics to capture the dynamic shakers in temporal cascading graphs. The general flow is described in Fig. 2. We first introduce the static accumulated influence proposed in [19].

DEFINITION 2. (STATIC INFLUENCE) *Given an entity v_i, the static accumulated influence $\mathbf{AI}(v_i)$ is defined as*

(3.14) $$\mathbf{AI}(v_i) = \sum_{t=1}^{T}\sum_{j=1}^{|\mathcal{V}|} \mathbf{W}^t(i,j)^2$$

where T is the duration of the influence, $\mathbf{W}^t$ is the t-th power of the matrix $\mathbf{W}$, and it is also the cascading graph after t step (as transition probability matrix), and $\mathbf{W}^t(i,j)$ is the value of the (i,j)-th entry of the matrix $\mathbf{W}^t$.

The intuition behind is to summarize the influence of the target entity v_i to all the other entities, as reflected by $\mathbf{W}^t(i,j)$. More theoretical explanations can be found in [19]. In the proposed algorithm, we obtain a series of evolving cascading graph, but the static influence above can only be used in a single static graph. Hence, we first use the above formula to evaluate the influence score of each entities on each single graph in $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \cdots, \mathbf{W}^{(M)}$. In this way, for each entity, we get a series of influence scores $\mathbf{AI}^{(1)}(v_i), \cdots, \mathbf{AI}^{(M)}(v_i)$ where $\mathbf{AI}^{(t)}(v_i)$ is the influence score of the entity v_i at the t-th cascading graph. We next introduce three metrics to measure the dynamic influence.

DEFINITION 3. (DYNAMIC MEAN INFLUENCE (DMEAN)) *Given an entity v_i, and its influence score over time $\mathbf{AI}^{(t)}(v_i)$ where $t = 1, 2, \cdots, M$, we define the dynamic mean influence as*

(3.15) $$\mathbf{DMI}(v_i) = \frac{1}{M}\sum_{t=1}^{M} \mathbf{AI}^{(t)}(v_i)$$

The Dynamic Mean Influence (DMean) score is straightforward. It is basically the average influence score across the duration. A further improvement is to consider the time decay factor as follows:

DEFINITION 4. (DYNAMIC DECADE INFLUENCE (DDECADE)) *Given an entity v_i, and its influence score over time $\mathbf{AI}^{(t)}(v_i)$ where $t = 1, 2, \cdots, M$, we define the dynamic decay influence as*
(3.16)
$$\mathbf{DDI}(v_i) = \frac{1}{\sum_{t=1}^{M}\exp(t - M)}\sum_{t=1}^{M}\exp(t - M)\mathbf{AI}^{(t)}(v_i)$$

where $\exp(t - M)$ is the decay factor that puts higher weights on recent influences (the time step close to M).

Note that the previous two influence scores both evaluate the dynamic influence in an accumulative manner. In other words, they are models to summarize the influence scores across time. However, in some cases, we may be more interested in the entities that have increasing influence although the influence is still weak at the current time step. For example, in economy, "BRIC" [2] are the four countries (Brazil, Russia, India, and China) that have more and more economic impact to the world. Their potential influences are more attractive to the economist. We next introduce a different metric that captures the potential influence of an entity.

DEFINITION 5. (DYNAMIC POTENTIAL INFLUENCE) *Given an entity v_i, and its influence score over time $\mathbf{AI}^{(t)}(v_i)$ where $t = 1, 2, \cdots, M$, we define the dynamic potential influence as*
(3.17)
$$\mathbf{DPI}(v_i) = \frac{d\mathbf{AI}(v_i)}{dt} \approx \frac{1}{k}\sum_{t=M-k}^{M-1}\Big(\mathbf{AI}^{(t+1)}(vi) - \mathbf{AI}^{(t)}(vi)\Big)$$

where k is the duration used to calculate a smooth slope.

In essence, Dynamic Potential Influence (DPotential) is a metric to calculate the slope, or direction of the evolving influence score. As a result, it will give higher weights to the entities that have a significant potential influences to the others, regardless of its current influence score. Note that the three metrics are developed for different purposes and they have different interpretations. One should choose among them based on their applications.

4 Experiments

We evaluate the proposed DShaker model on (1) US banks' subordinated debt datasets, (2) European countries' loan datasets, as well as (3) stock market datasets.

4.1 Finding the shaking banks with subordinated debts

The first dataset is from the FDIC bank database[5]. We downloaded the dataset including 3412 banks, and we looked at the quarterly statistics of each bank from year 1998 to 2007, before the 2008 subprime crisis. More specifically, we looked at the subordinated loans of the banks (105 of the banks have records of subordinated loans), and the objective is to capture the banks that have significant influences on the subordinated loan market. We include two comparison methods in the experiment. The first one is the IC (independent cascading) model, which usually works on static graph to capture the most influential nodes. The second one is the static shaker proposed in [19], and we denote it as "SShaker". Note that these two models only work on static graph. We thus first constructed a static cascading graph by using the pairwise local construction approach proposed by [19]. Then, the most influential nodes were captured based on the static cascading graph by IC model and the algorithm in [19]. Furthermore, we report the shakers found by the DShaker model using the three metrics respectively (Dynamic Mean Influence, Dynamic Decade Influence, and Dynamic Potential Influence). The top 10 shakers are presented in Table 1.

Among the results, it can be clearly observed that the independent cascading (IC) model performs the worst, and it seems to provide a random list of influential banks. It misses big banks like Bank of America, Citigroup which are among the leaders of the subprime morgage problem. The static shaker model has a better performance than IC model. For instance, it successfully captures big banks like Fifth Third, Citigroup, HSBC, *etc.*. However, it also generates banks like "Valley National", which do not seem to have a big influence on the subprime market. Furthermore, we cannot find big banks like Chase, Wells Fargo in the list. The reason is that the dataset records subordinated debt data over 9 years, during which the causal relationships among the banks may have already changed several times. If we force a single static cascading graph over the 9 year period of the data, it may produce poor results. The proposed Dynamic Mean Influence and Dynamic Decade Influence in the DShaker model successfully captures most of the big banks at that time. These banks are very active in producing subordinated debts where one of them is the notorious subprime mortgage. More interestingly, the Dynamic Potential Influence captures Lehman Brothers in its top-ten list. Note that the amount of subordinated debts of Lehman Brothers is not comparable to the big banks. For example, in the last quarter of 2007, the amount of subordinated debts of Citigroup is about 200

times of that of Lehman Brothers. Moreover, the amount of subordinated debts of Bank of America is 10 times larger than that of Citigroup. Hence, it seems that as compared to these big banks, the problem of Lehman Brothers should be the last thing to be considered. However, as captured by the Dynamic Potential Influence, the influence of Lehman Brothers is growing rapidly although its spot influence score is not as high as those big banks. Its bankruptcy in 2008 finally triggers the crash of the subordinated debt market, which leads to the latest financial crisis.

4.2 Finding the most precarious EU countries

The second dataset we are looking at is the loan dataset of 25 European countries. It is downloaded from the European Central Bank [4], and it contains the quarterly statistics of the countries from year 2000 to year 2011. It is important to emphasize that the dataset is about the loans of the countries over time. The shakers detected in this dataset are thus those countries that may deteriorate the debt situation of European countries. Among the 25 countries, we present the top 5 most precarious EU countries in Table 2.

Among the results, the IC model does not perform well. It generates a list of countries that do not seem to have essential influence on the EU debt problem. The static shaker model works better. It captures countries like Spain and Italy, and they are among the countries that lead EU into debt trouble as reported in [17]. However, the static shaker model regards Greece to have a similar influence score as Denmark. As such, neither Greece nor Denmark is considered to be one of the precarious countries. The reason is also analyzed in Fig. 1 of the introduction section. The two countries do look similar if we neglect the dynamic changes of the causal relationships and look at the data starting from year 2000. The Dynamic Mean Influence and the Dynamic Decade Influence perform better than the static shaker model. They find more key countries to this problem, such as Greece and Belgium. Note that the Dynamic Mean Influence reports Estonia and Finland as among the precarious countries, due to their financial crisis during 2008-2009. Although their economies are now recovering, Dynamic Mean Influence is based upon an averaging strategy. Dynamic Decade Influence is a better strategy because it puts higher weights to recent influence scores. Furthermore, Dynamic Potential Influence ranks Greece as the first precarious country. It successfully captures the growing influence of Greece in the EU debt problem.

4.3 Finding the stock movers and shakers

The last dataset is on the stock market. We downloaded the historical daily price data for 2000 stocks from Yahoo! Finance. The dataset records 2000 NASDAQ stocks, and their daily closing prices in 1000 business days (last date is Dec 16th, 2011). One can imagine that the dataset may be extremely

Table 1: Top 10 Precarious Banks in Subprime Debts

IC	SShaker	DMean	DDecade	DPotential
LNB	FIRST FINANCIAL	CITIGROUP INC.	CITIGROUP INC.	GOLDMAN SACHS
FIDELITY	FIDELITY	BANK OF AMERICA	WELLS FARGO	PNC
FIRST FINANCIAL	FIFTH THIRD	WELLS FARGO	BANK OF AMERICA	BANCO SANTANDER
SUSQUEHANNA	CITIGROUP INC.	CHASE	CHASE	BANCO BILBAO
BOK FINANCIAL	BANK OF AMERICA	U.S. BANCORP	U.S. BANCORP	TORONTO BANK
BANQUE NATIONALE	HSBC	HSBC	HSBC	**LEHMAN BROTHERS**
BANK OF MONTREAL	VALLEY NATIONAL	SUNTRUST BANKS	SUNTRUST BANKS	STATE STREET CORP.
CITY NATIONAL	NORTHERN BANK	KEYCORP	PNC	BB&T CORPORATION
BANK LEUMI	BANQUE NATIONALE	PNC	KEYCORP	TCF FINANCIAL CORP.
FARMERS BANCORP	FNB CORP	MORGAN STANLEY	MORGAN STANLEY	MORGAN STANLEY

Table 2: Top 5 Precarious EU Countries in Debts

IC	SShaker	DMean	DDecade	DPotential
Estonia	Spain	Finland	Italy	Greece
Poland	Hungary	Greece	Neth.	Italy
Hungary	Italy	Italy	Greece	Spain
Ireland	Luxembourg	Estonia	Spain	Neth.
Denmark	Neth.	Belgium	Belgium	Belgium

Table 3: Top 10 Stock Market Shakers

IC	SShaker	DMean	DDecade	DPotential
ACMR ✓	ZOLT ✓	PCBC ✓	PCBC ✓	PCLN ✓
ININ ✗	ZLCS ✗	PCLN ✓	NFLX ✓	AAPL ✓
URRE ✗	ZION ✓	CME ✓	PCLN ✓	ATRI ✓
AAON ✓	ZHNE ✗	OCNF ✗	AAPL ✓	GMCR ✓
INSM ✗	YRCW ✗	FNBN ✗	CME ✓	AMZN ✓
MAGS ✓	ZINC ✓	AAPL ✓	AMZN ✓	DNEX ✗
ATAX ✗	XNPT ✗	NFLX ✓	RIMM ✓	HANS ✓
PROV ✗	ZOLL ✓	TIBB ✗	FFIV ✓	GOLD ✓
SPRT ✗	YHOO ✓	AMZN ✓	GOLD ✓	TZOO ✓
MLHR ✗	WYNN ✓	RIMM ✓	BIDU ✓	DNDN ✓
30%	60%	70%	**100**%	90%

volatile (over 3 years), and it may contain complicated evolving cascading relationships among the stocks. The objective is to find the stock market "movers and shakers" who affect the market the most. In order to evaluate the results, we follow the procedure used by [19] to download the ground truth from Yahoo! Finance. More specifically, we first crawl the stock industry leadership rankings [8] with various criteria such as market capitalization, PEG ratio, *etc.* For example, Google Inc. (GOOG) ranks the first among the 85 stocks of the same industry according to its market capitalization, and ranks 15th according to its PEG ratio. We then re-rank the stocks according to their average rankings across all criteria in ascending order. For example, the average ranking score of GOOG is $(1 + 15)/2 = 8$ if we only consider capitalization and PEG ratio, and it ranks first if the average scores of other stocks are larger than 8. The top 20% of stocks in each industry are considered to be the true market shakers. The top 10 most influential stocks captured by the cascading graphs are presented in Table 3.

In Table 3, we place checkmarks besides the stocks that are real shakers, crossmarks besides the false positive. As can be observed, the simple IC model has the worst performance, which just finds 3 real shakers among 10. The static shaker model improves the IC model by finding 3 more real shakers. However, owing to the long term (over 3 years), the static shaker model cannot capture the dynamics of the causal relationships, and the resulting accuracy is 60%. Among the three proposed models, the Dynamic Decade Influence has the best performance. It has a 100% accuracy for the top 10 shakers, as compared to 70% accuracy of the

DMean model, and 90% accuracy of the Dynamic Potential Influence model. It is important to mention again that the Dynamic Mean Influence model just takes the average of the influences as the final score without considering the time factor. This is the main reason that it cannot beat the Dynamic Decade Influence. In addition to the top 10 most influential stocks, we also performed evaluation on the top 20, top 30, and up to top 100 influential stocks found by the 5 approaches. In this time, we directly reported their accuracies under different settings, and the results are summarized in Fig. 3. It can be clearly observed that Dynamic Decade Influence and Dynamic Potential Influence beat the other three approaches in accuracy. Between the two, Dynamic Decade Influence performs notably better. However, it is important to mention again that two metrics serve in different purposes. Dynamic Potential Influence is used to capture the entities that have increasing influences, while Dynamic Decade Influence focuses on the current influence scores. One should choose between them based on their own necessities.

Importance of Capturing Evolving Influence Given the dynamic nature of the stock dataset, we also study (1) the evolution of causal relationships, and (2) the dynamic of the influence scores, or equivalently, the dynamic of shakers. We
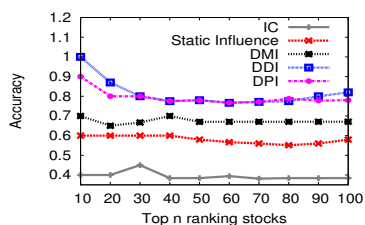
Figure 3: Accuracy of shaker detection in NASDAQ stocks over 3 years' data

(a) Evolving Closing Prices over 3 Years
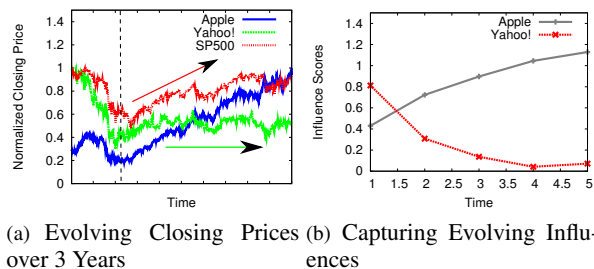
(b) Capturing Evolving Influences

Figure 4: Example

take two technology stocks Apple and Yahoo! as examples, and use S&P500 index as the market indicators. We first plot the daily closing prices of the three stocks over three years as in Fig. 4(a). As can be observed, it is a volatile market with lots of uncertainties. However, we can clearly observe that all three stocks have a clear slump (the 2008 financial crisis), and then a turning point as highlighted by the dotted line. One can also observe that before the turning point, Yahoo! has stronger correlation with the S&P500 than Apple. During this time, we can observe that the trend of Yahoo!'s stock almost overlaps with the market. However, after the dotted line, Yahoo!'s stock goes flat with some fluctuations, while the S&P500 is rising continuously. During this period, the stock of Apple has stronger correlation with the S&P500 than Yahoo!. We also plot the changing of influence scores of the two stocks over 3 years as in Fig. 4(b). One can also clearly observe the influence scores of Apple is soaring while the influence of Yahoo! keeps decreasing. If we stick to use the static cascading graph, we may mistakenly take Yahoo! as more influential than Apple especially looking at a long period. Hence, we should use dynamic cascading graph and dynamic shaker detection technique, when we are dealing with a volatile dataset during a long period.

Parameter Sensitivity There are two regularized parameters γ and λ in the optimization objective in Eq. 3.7. We next analyze the sensitivity of the parameters on the stock dataset. The accuracies on the top-100 stock shakers are plotted in Fig. 5. First, the parameter γ determines how smooth of the

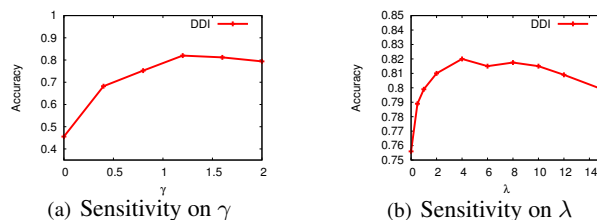(a) Sensitivity on γ

(b) Sensitivity on λ

Figure 5: Parameter Sensitivity Analysis

temporal cascading graphs are. In other words, if γ is large, the evolving cascading graphs have strong similarities with the graphs at previous time step. Fig. 5(a) is the accuracy of the dynamic decade influence model (DDecade) with different γ while we set $\lambda = 4$. As can be observed, when γ is small, the result is not good. This is because in this case, the cascading graphs overfit to the data at each time interval, and they are not required to evolve smoothly. It then causes prediction error. When $\gamma = 1.2$, the DMean model has the best performance. In practice, one can choose the best γ with cross-validation if some prior knowledge is available; otherwise, $\gamma = 1.2$ should be a good choice. Fig. 5(b) shows the result of DDecade while we set $\gamma = 1.2$. Note that λ is the parameter to force the cascading graphs to be of low rank. It shows that when λ increases from 0 to 4, the accuracy keeps increasing. This is because λ is a very important component to reduce noise. However, if λ is too large, the accuracy drops since it will force the rank of the cascading graphs (adjacency matrix) to decrease significantly. In this case, the modeling error $\|\mathbf{A}^{(S+1)} - \mathbf{A}^{(S)}\mathbf{W}\|_F^2$ will go up and more mistakes are made. The choice of λ can be determined by cross-validation if one can use some prior knowledge; otherwise, it can be set as the default value 4. In the experiment mentioned before, we set $\gamma = 1.2$ and $\lambda = 4$.

5 Related Work

There are several areas of work that we build upon. First, information diffusion and influence maximization (*e.g.*, [9, 15, 22]) is one area of related work. The general idea is to trace the information propagation on a given network by various models, and infer the top influential nodes or effectors that can best explain the observations (*e.g.*, [1, 16, 14, 13]). Usually, the entities have binary states (active or inactive) and the discovered influential nodes usually have wide spreads. Different from these works, the static shaker model [19] is proposed to capture influential nodes on evolving entities with numeric change. It can thus work on a wider range of applications. However, it assumes that the underlying causal relationships among the evolving entities do not change over time, and so do the shakers. In this paper, we propose a dynamic model to capture the evolving causal relationships. Note that it will be cast into the static

shaker model if the underlying causal relationships do not change over time. However, it will automatically change the captured causal relationships if the underlying causality has been changed. The second line of related work is about the study of causality. It has attracted intensive interest in various fields. A major assumption of our research is to find causalities via temporal correlations (time series data [12]). Note that although the two concepts are not exactly the same, yet the existence of some correlations can lead us to induce, under some fairly broad conditions, the existence of certain causal relations. This assumption is broadly used in other fields such as Philosophy [20], Bioinformatics [21, 18] and Economics (*e.g.*, the 2003 Nobel Prize winning work in economic time series analysis [10]).

6 Conclusion

In recent years, several models (*e.g.*, IC, static shaker, *etc.*) were proposed to capture the most influential evolving entities (*i.e.*, shakers). However, most of them assume that the underlying causal relationships among entities do not change over time, and hence the shakers should also be static. In reality, this assumption may not necessarily be true, especially for volatile data, or for long-term time series (*e.g.*, financial market data, long-term economy indicies, sensor data, *etc.*). In this paper, we propose to capture the evolving causal relationships, and discover the dynamic shakers. The model is formulated as a non-convex multi-objective optimization problem, which we solve by combing the idea of gradient descent and trace norm minimization. After solving the optimization problem, we obtain a set of temporal cascading graphs. Three metrics are proposed to capture the dynamic shakers on the evolving cascading graphs. They evaluate the entities' average influence, time-decay influence, and potential influence, respectively. Three sets of experiments were conducted. It shows that the proposed DShaker model is more accurate than the static shaker model in capturing real shakers. For instance, in the experiment to discover stock market shakers, the dynamic shaker improves the static shaker model by as much as 40%. As a future work, the proposed model will be adopted to study the causal effects in medical science, intrusion detection and cyber-security tasks.

References

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In *WSDM*, pages 207–218, 2008.

[2] BRIC. http://en.wikipedia.org/wiki/bric.

[3] J. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[4] European Central Bank dataset. `http://www.ecb.int/stats/html/index.en.html`.

[5] FDIC dataset. `http://www.fdic.gov/bank/statistical/`.

[6] S. Ellner and P. Turchin. Chaos in a noisy world: new methods and evidence from time series analysis. *American Naturalist*, 145:343–375, 1995.

[7] M. Fazel, H. Hindi, and S. P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *In Proceedings of the 2001 American Control Conference*, pages 4734–4739, 2001.

[8] Yahoo! finance industry ranking. http://finance.yahoo.com /q/in?s=aapl+industry.

[9] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *KDD*, pages 1019–1028, 2010.

[10] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[11] S. Ji and J. Ye. An accelerated gradient method for trace norm minimization. In *ICML*, page 58, 2009.

[12] E. J. Keogh and S. Kasetty. On the need for time series data mining benchmarks: a survey and empirical demonstration. In *KDD*, pages 102–111, 2002.

[13] M. Kimura, K. Saito, and R. Nakano. Extracting influential nodes for information diffusion on a social network. In *AAAI*, pages 1371–1376, 2007.

[14] T. Lappas, E. Terzi, D. Gunopulos, and H. Mannila. Finding effectors in social networks. In *KDD*, pages 1059–1068, 2010.

[15] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *EC*, pages 228–237, 2006.

[16] J. Leskovec, M. McGlohon, C. Faloutsos, N. S. Glance, and M. Hurst. Patterns of cascading behavior in large blog graphs. In *SDM*, pages 551–556, 2007.

[17] PIGS. `www.enotes.com/topic/PIGS_(economics)`.

[18] C. A. Ratanamahatana, J. Lin, D. Gunopulos, E. J. Keogh, M. Vlachos, and G. Das. Mining time series data. In *Data Mining and Knowledge Discovery Handbook*, pages 1049–1077. 2010.

[19] X. Shi, W. Fan, J. Zhang, and P. S. Yu. Discovering shakers from evolving entities via cascading graph inference. In *KDD*, pages 1001–1009, 2011.

[20] Q. Smith. Causation and the logical impossibility of a divine cause. *Philosophical Topics*, 21(1):169–191, 1996.

[21] I. Tsamardinos. Causal data mining in bioinformatics. *ERCIM News*, 2007(69), 2007.

[22] D. Wu, Y. Ke, J. X. Yu, P. S. Yu, and L. Chen. Detecting leaders from correlated time series. In *DASFAA*, pages 352–367, 2010.