Discovering Shakers from Evolving Entities via Cascading Graph Inference

Xiaoxiao Shi[†], Wei Fan[‡], Jianping Zhang[*], and Philip S. Yu[†]
[†]Department of Computer Science, University of Illinois at Chicago
[‡]IBM T.J.Watson Research Center
[*]Analytic & Forensic Technology , Deloitte Financial Advisory Services LLP
{xiaoxiao,psyu}@cs.uic.edu, weifan@us.ibm.com, jianpzhang@deloitte.com

ABSTRACT

In an interconnected and dynamic world, the evolution of one entity may cause a series of significant value changes for some others. For example, the currency inflation of Thailand caused the currency slump of other Asian countries, which eventually led to the financial crisis of 1997. We call such high impact entities *shakers*. To discover shakers, we first introduce the concept of a *cascading graph* to capture the causality relationships among evolving entities over some period of time, and then infer shakers from the graph. In a cascading graph, nodes represent entities and weighted links represent the causality effects. In order to find hidden shakers in such a graph, two scoring functions are proposed, each of which estimates how much the target entity can affect the values of some others. The idea is to artificially inject a significant change on the target entity, and estimate its direct and indirect influence on the others, by following an inference rule under the Markovian assumption. Both scoring functions are proven to be only dependent on the structure of a cascading graph and can be calculated in polynomial time. Experiments included three datasets in social sciences. Without directly applicable previous methods, we modified three graphical models as baselines. The two proposed scoring functions can effectively capture those high impact entities. For example, in the experiment to discover stock market shakers, the proposed models outperform the three baselines by as much as 50% in accuracy with the ground truth obtained from Yahoo! Finance.

Category and and Subject Descriptors H.2.8 [Database Management]: Database applications-Data mining
General Terms Algorithms.
Keywords Cascading Graph, Shakers, Dynamic Entity

1. INTRODUCTION

In an interconnected and dynamic world, the evolution of one entity may cause a series of significant value changes for some others. Examples include the volatile and intercorrelated financial markets, or the climate changes of neighboring areas. This is typically called the domino or butterfly effect. A famous case is the currency inflation of Thailand, which caused the financial crisis of 1997. We call such high impact entities *shakers* and our objective is to find the shakers in a set of evolving entities. It is important to mention that shakers are different from the influential nodes studied in [12, 33] and the effectors in [15]. One of the key differences is that the former two studies investigate entities with only one of two possible states (active and inactive), while shakers allow entities with numerical attributes and continuous changes (*e.g.*, stock prices), which is more general and more complicated. In addition, shaker formulation allows for different types of effects (*e.g.*, positive effect and negative effect introduced in Section 2). With the general problem setting, the following two challenges need to be addressed:

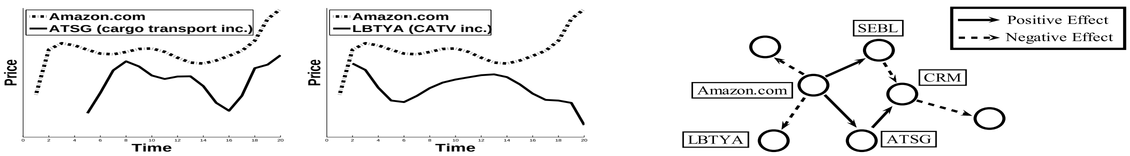
C1: Given the set of evolving entities with continuously changing numerical attributes, how to determine who affects whom?

C2: After determining the causalities among the entities (by solving C1), how to detect hidden shakers?

To solve C1, we use evolving data (*e.g.*, time series, streaming, sensory) to model the entity evolutions and detect the latent entity causalities by measuring and capturing the similarity of the evolution trends. The intuition is that if an entity A affects entity B, then: (1) the changes in A occurs prior to the changes in B, and (2) they will have similar trends. For example, Fig. 1(a) plots the price evolutions of two stocks over time. One is Amazon.com and the other is its business partner, a cargo transportation company. It is clear that the evolutions of the two stocks are almost the same but the change in Amazon.com takes the lead. Hence, we say that Amazon.com affects its business partner in the stock market. This is the major assumption of our research to find causalities via these temporal correlations. 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 [29], Bioinformatics [32] and Economics (*e.g.*, the 2003 Nobel Prize winning work in economic time series analysis [9]). By detecting such causalities between each entity pair, we construct a cascading graph as follows:

[*]This research was conducted at MITRE Corporation prior to the author's joining Deloitte Financial Advisory Services LLP.

KDD'11, August 21–24, 2011, San Diego, California, USA.



(a) Positive effect (b) Negative effect (c) Fraction of the cascading graph with positive and negative edges (effects)

Figure 1: Example of the cascading graph on stock data learned from the algorithm

Definition 1 (Cascading Graph). *A cascading graph is a weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathbf{E}, \mathcal{A})$ where*

- *$\mathcal{V}$ represents the set of entities. $\forall v \in \mathcal{V}$, $a_v(\tau) \in \mathcal{A}$ is its attribute that evolves over time τ. For simplicity, $\mathcal{A} \subseteq \mathbb{R}$ is a set of real numbers.*
- *$\mathbf{E} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$ is called the* cascading matrix *such that $\mathbf{E}(i,j) = e$ is the causality effect from entity v_i to entity v_j. And $|e|$ determines the strength or confidence of the causality. Furthermore, $e < 0$ if it is a* negative effect *such that the value increase of one entity causes the value decrease of another as in Fig. 1(b); $e > 0$ if it is a* positive effect *as in Fig. 1(a).*

An example of a cascading graph on stock data is illustrated in Fig. 1(c) where solid lines represent edges with positive effects ($e > 0$, or as in Fig. 1(a)) and dotted lines represent negative effects ($e < 0$, or as in Fig. 1(b)). It is important to note that the introduction of negative effect provides an advantage not found in other graphical models (*e.g.*, [10, 4]). It extends the expressive power to model a wide range of networks that naturally contain opposite correlations among the nodes. Examples include the cooperation and competition networks of the business institutions, stock correlation networks, *etc.* However, with previously proposed methods, it is not obvious how to discover latent shakers in such a graph:

1. Graphical Models: In a cascading graph, since it allows for negative effects (edges), the adjacency matrix, or the cascading matrix $\mathbf{E}$ is not stochastic. Hence, as far as we know, there is no graphical model (*e.g.*, [10, 4, 14]) that can be directly applied to mine the latent shakers.
2. Influence Propagation Models: Most of the previous methods (*e.g.*, [1, 33, 12] and the IC model [13]) investigate the graph with positive edges. However, given the cascading graph with negative edges, the influence propagation becomes more complicated, because some of them may be reduced or canceled. As in Fig. 1(c), "Amazon.com" has two inference propagation paths to "CRM". One is "Amazon $\rightarrow$ ATSG$\rightarrow$CRM" that can increase the value of CRM, but the other is the negative effect "Amazon $\rightarrow$SEBL $\rightarrow$CRM" that decreases its value. Hence, with the negative effect canceling the positive effect, "Amazon.com" may finally cause no value change (no influence) on "CRM". Traditional influence propagation models do not consider such situations.

Since previous models are not directly applicable, in this paper, two scoring functions are proposed to find the shakers by evaluating each node's accumulated influence and maximal influence. The key idea is to artificially inject a significant value change in the target node, and estimate its direct and indirect influence on other nodes by following a graph inference rule under the Markovian assumption. Using linear algebra to perform a transformation, the two scoring functions are proven to be (a) only dependent on the structure of the cascading graph, and (b) they can be calculated in polynomial time.

We conducted three sets of experiments. These experiments focused on the following problems in social sciences: finding the economy shakers of countries, finding the most influential economy indicators, and finding the NASDAQ stock market shakers. Without directly applicable methods, three previous graphical models are modified as baselines. It is observed that the two proposed scoring functions can effectively capture the shakers. For example, in the experiment of finding stock market shakers, the proposed models outperform the three baselines by as much as 50% in accuracy, with the ground truth coming from Yahoo! Finance. Moreover, one of the shakers, AAPL (Apple Inc.), is successfully captured; it later leads to a big drop of the NASDAQ stocks in January 2011 despite the good news of most other stocks.

2. PROBLEM FORMULATION

Given a set of evolving entities, the objective is to find shakers whose value changes can significantly affect the values of many other entities. The following provides a more formal definition of the related concepts.

Definition 2 (Evolving entity). *Denote $v \in \mathcal{V}$ as an entity, $a_v(\tau) \in \mathcal{A} \subseteq \mathbb{R}$ as the attribute of v that evolves over the time variable τ. In this paper, the time variable is considered to be discrete, and we use $a_v(t)$ to denote the attribute value of entity v at time step t.*

The evolving attributes are used to detect causality relationships among the entities. The general idea is that $\forall v, w \in \mathcal{V}$, if the trend of $a_v(\tau)$ is similar to that of $a_w(\tau)$ and one is prior to the other as in Fig. 1(a), then v and w have a causality relationship. More details can be found in Section 3. Note that once the causality relationships are determined, the present values of the entities (*i.e.*, $a_v(\tau), a_w(\tau)$, *etc.*) are no longer important. Instead, to find the shakers, we are more interested in their value changes and the entities causing those changes. As such, we use a $1\times|\mathcal{V}|$ vector $\mathbf{C}_t$ to denote the value changes of the entities at time t such that

$$\mathbf{C}_t(i) = a_{v_i}(t) - a_{v_i}(t-1)$$

is the value change of node v_i at time t. Then, shaker is defined as follows

Definition 3 (Shaker). *Given an entity $v \in \mathcal{V}$, assume its value significantly changes, say, at the beginning where $t = 0$. By following the influence propagation path, if the values of other entities significantly change afterwards where $t = 1, 2, \cdots$, we call the entity v a* shaker.

This is a general definition of shaker and the key intuition here is to estimate whether the target entity v can cause significant changes of many others afterwards ($t > 0$).

Cascading Graph

The key step to find shakers is to capture and summarize the complicated causality relationships among the entities, and unveil the shakers by inferring how the causality effect is propagated. To do so, we propose to use cascading graph $G = (\mathcal{V}, \mathbf{E}, \mathcal{A})$. A cascading graph is a graph, in which nodes in $\mathcal{V}$ are entities and weighted directed edges in $\mathbf{E}$ represent causalities. In Section 3, we show how to extract a cascading matrix $\mathbf{E}$ from evolving entities. An example of a cascading graph is shown in Fig. 2, where v_1, v_2 and v_3 have some effects on the entity v_4. The main advantage of cascading graph, not found in traditional graph models, is its flexibility to model negative effects. These are formulated as negative edges, *e.g.*, edge (v_1, v_4) in Fig. 2. One challenge is that negative effects may cancel the influence of positive effects in later influence propagation. Thus, the temporal correlation detected at the initial state may disappear in later inferences. More details can be found in Eq. 1. Hence, in addition to shaker discovery, cascading graphs can be used in a wide range of applications where entities have opposite relationships. Traditional stochastic graphical models, however, are not able to model such complicated relationships because the adjacency matrix is non-stochastic. We next define an inference rule for cascading graphs.

Inference on Cascading Graph. Similar to first order Markov process, we make the Markovian assumption that the attribute value changes $\mathbf{C}_t$ at time t only depend on those at time $t-1$:

$$\mathbf{C}_t(i) = \sum\nolimits_{j=1}^{|\mathcal{V}|} \mathbf{C}_{t-1}(j)\mathbf{E}(j,i) \tag{1}$$

where $\mathbf{E}(j,i)$ is the causality effect from entity v_j to entity v_i such that $\sum_{j=1}^{|\mathcal{V}|} |\mathbf{E}(j,i)| = 1$. Hence, $\mathbf{E}(j,i)$ can also be interpreted as the weight of the "ingredient" of the mixture effects to v_i. For instance, in Fig. 2, 50% of the effects to v_4 are from v_1, 20% are from the entity v_2 and the rest are from the entity v_3 such that:

$$\mathbf{C}_t(4) = -0.5\mathbf{C}_{t-1}(1) + 0.2\mathbf{C}_{t-1}(2) + 0.3\mathbf{C}_{t-1}(3)$$

Note that $\mathbf{E}(1,4) = -0.5 < 0$ is a negative effect. Hence, when the node v_1's value increases at time $t-1$, v_4's value tends to decrease at time t, since $\mathbf{C}_t(4) \propto -0.5\mathbf{C}_{t-1}(1)$. As such, negative effect (edge) tends to lead the value to change in an opposite direction, while positive effect makes the value moving in the same direction. In matrix form, the inference rule in Eq. 1 can be written as:

$$\mathbf{C}_t = \mathbf{C}_{t-1}\mathbf{E} = \mathbf{C}_{t-2}\mathbf{E}^2 = \cdots = \mathbf{C}_0\mathbf{E}^t \tag{2}$$

It is interesting to mention that the cascading matrix $\mathbf{E}$ is similar to a transition probability matrix in Markov process, and $\mathbf{C}_t$ is similar to a state vector at time t. However, unlike a transition probability matrix, the cascading matrix $\mathbf{E}$

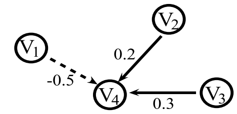


Figure 2: Inference on cascading graph

allows for negative values and hence it is not stochastic. In Section 4, we provide an analysis on how traditional graphical models do not perform well in this case.

Model Formulation for Shaker Discovery

To evaluate the influence of the target node v_i, we artificially change its value at time $t = 0$, and then infer its effect on other entities during $0 < t < T$, where T is the time span we are interested in. In other words, we do not care about the changes after time T, which is a user-specified parameter studied in the experiment section. To find shakers, the following steps are performed:

1. Artificially change the value: Define a new vector $\mathbf{C}_0^{(v_i)}$ such that $\forall j \neq i, \mathbf{C}_0^{(v_i)}(j) = \mathbf{C}_0(j)$, but $|\mathbf{C}_0^{(v_i)}(i)| \gg |\mathbf{C}_0(i)|$. In other words, $\mathbf{C}_0^{(v_i)}$ is different from $\mathbf{C}_0$ only in the target node v_i.

2. Influence inference: Define vector $\mathbf{C}_t^{(v_i)}$ as the estimated value changes at time t with $\mathbf{C}_0^{(v_i)}$ as the initial state, or $\mathbf{C}_t^{(v_i)} = \mathbf{C}_0^{(v_i)}\mathbf{E}^t$ with the influence rule in Eq. 2.

Note that the vector $\mathbf{C}_t^{(v_i)}$ is derived after the value of the target entity v_i suffers a big jump ($|\mathbf{C}_0^{(v_i)}(i)| \gg |\mathbf{C}_0(i)|$). Hence, we want to study the difference between the two vectors $\mathbf{C}_t^{(v_i)}$ and $\mathbf{C}_t$ to investigate how the jumping value affects others. Two scoring functions are defined.

Definition 4 (Accumulated Influence). *Given an entity v_i, the accumulated influence $\mathbf{AI}(v_i)$ is defined as the accumulated relative difference between the two vectors*

$$\mathbf{AI}(v_i) = \sum_{t=1}^{T} \frac{\partial \mathbf{C}_t^{(v_i)}}{\partial \mathbf{C}_0^{(v_i)}} \tag{3}$$

where $\partial \mathbf{C}_t^{(v_i)} = \|\mathbf{C}_t^{(v_i)} - \mathbf{C}_t\|$ and $\| * \|$ is the L-2 norm that summarizes the difference.

It is interesting to note that $\partial \mathbf{C}_t^{(v_i)} / \partial \mathbf{C}_0^{(v_i)}$ has a similar form as function derivative $\frac{\partial f(x)}{\partial x}$ that evaluates how fast a function changes at a certain direction. We use it to capture how effectively the influence can be propagated through node v_i. However, it is not easy to obtain $\partial \mathbf{C}_t^{(v_i)} / \partial \mathbf{C}_0^{(v_i)}$. It is not known how to determine $\mathbf{C}_0^{(v_i)}$, since we only know $|\mathbf{C}_0^{(v_i)}(i)| \gg |\mathbf{C}_0(i)|$ without any specific value. In the next section, we prove that $\mathbf{AI}(v_i)$ is actually independent of $\mathbf{C}_t$, $\mathbf{C}_t^{(v_i)}$ and $\mathbf{C}_0^{(v_i)}$, and it can be efficiently calculated only with a cascading matrix $\mathbf{E}$, or the topology of the cascading graph. Similarly, we define maximal influence as follows:

Definition 5 (Maximal Influence). *Given an entity v_i, we define the maximal influence as*

$$\mathbf{MI}(v_i) = \max_{t=1,2,\ldots,T} \frac{\partial \mathbf{C}_t^{(v_i)}}{\partial \mathbf{C}_0^{(v_i)}} \tag{4}$$

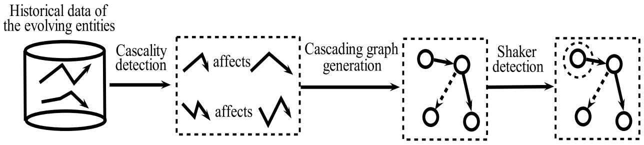


Figure 3: Algorithm flow

Unlike accumulated influence, maximal influence only considers the moment when the target entity exerts the most influence on others despite the effect potentially disappearing later. In the next section, we introduce how to construct the cascading graph and how to calculate the two scoring functions in polynomial time.

3. PROPOSED ALGORITHM

The algorithm flow is sketched in Fig. 3, which contains two main parts. The first is to examine the entity causalities and generate the cascading graph by evaluating the similarities of their evolution trends. The second is to infer shakers in the static cascading graph.

Cascading graph construction

To construct a cascading graph, or equivalently, to fill out all the entries of the cascading matrix $\mathbf{E}$, we need to determine the causality $\mathbf{E}(i,j)$ between each entity pair $\forall v_i, v_j \in \mathcal{V}$. If they have similar evolution trends and one follows another, we say that they have strong causality relationship. To do so, we first fit each evolution trend of $\forall v_i \in \mathcal{V}$ by a function $f_i(t)$. Then we utilize basic function transformations (such as scaling, shifting, *etc.*) to match the two functions and find their similarity. In particular, the following optimization problem is devised

$$\min_{\hat{h},\hat{u},\hat{t},\hat{s}} \epsilon(i,j) = \sum_{t=0}^{L} \Big(f_i(t) - \big(\hat{h} f_j(\hat{u}t + \hat{t}) + \hat{s}\big)\Big)^2 \tag{5}$$

where L is the time span determined by the data itself, $\hat{t}$ is the shifting along the time variable whereas $\hat{s}$ is the shifting along the output, $\hat{h}$ is the constant to change the scale of the output and $\hat{u}$ is the constant to allow dynamic time warping [25]. The optimization function thus matches the two evolution functions with the above four types of algebra transformations. If the two entities have very similar evolving trends, they will be almost perfectly matched by Eq. 5. In this case, $\epsilon(i,j)$ will be small and we say that they have strong causality relationship; otherwise we say they have weak causality relationship. Hence, we define the strength of a causality as $e^{-\frac{\epsilon(i,j)}{\sigma}}$, which is similar to the similarity function in normalized cut [28]. Furthermore, the solved values of $\hat{h}$ and $\hat{t}$ encode the properties of the causality

- if $\hat{h} > 0$, it is a positive effect as in Fig. 1(a); if $\hat{h} < 0$, it means the value increase of one entity causes the value decrease of another; we call it a negative effect.

- if $\hat{t} > 0$, it means $f_i(t)$ leads the trend, or v_i affects v_j. We define

$$\mathbf{E}(i,j) = \frac{\hat{h}}{|\hat{h}|} e^{-\frac{\epsilon(i,j)}{\sigma}}, \ \mathbf{E}(j,i) = 0 \tag{6}$$

If $\hat{t} < 0$, it means v_j affects v_i

$$\mathbf{E}(i,j) = 0, \ \mathbf{E}(j,i) = \frac{\hat{h}}{|\hat{h}|} e^{-\frac{\epsilon(j,i)}{\sigma}} \tag{7}$$

If $\hat{t} = 0$, neither v_i nor v_j takes the lead. Hence $\mathbf{E}(i,j) = \mathbf{E}(j,i) = 0$.

In the above formula, the constant σ is a normalization term to ensure $\sum_{i=1}^{|\mathcal{V}|} |\mathbf{E}(i,j)| = 1$, and the sign of $\mathbf{E}$ is determined by $\hat{h}$. As such, the cascading matrix $\mathbf{E}$ is determined and we assume that it is static and that it does not change over the duration of the study. To solve Eq. 5, we assume that the evolution of each entity v_i is determined by an n-degree polynomial function such that $f_i(t) = \sum_{r=0}^{n} k_r t^r$ where k_rs are the polynomial coefficients. Choosing a polynomial function has several advantages: (1) It can fit both linear and nonlinear (*e.g.*, [30, 18]) evolution functions perfectly. (2) It can effectively capture the turns of the evolving trends, which is an important clue in determining causality. Hence, we denote the evolution functions of v_i and v_j as $f_i(t) = \sum_{r=0}^{n} k_r t^r$ and $f_j(t) = \sum_{r=0}^{n} \hat{k}_r t^r$ where n is the degree discussed in Section 4. We then substitute them into Eq. 5 and use coordinate descent to solve the optimization problem. The general idea is to solve one variable at a time by fixing the values of the others. To be specific, we first solve $\hat{h}$ by setting $\frac{\partial \epsilon(v,w)}{\partial \hat{h}} = 0$ with $\hat{u}$, $\hat{t}$ and $\hat{s}$ set as the initial values.

$$\hat{h} = \sum_{t=0}^{L} (\sum_{r=0}^{n} k_r t^r - \hat{s}) / \sum_{t=0}^{L} \sum_{r=0}^{n} \hat{k}_r (\hat{u}t + \hat{t})^r \tag{8}$$

We solve $\hat{t}$, $\hat{u}$ and $\hat{s}$ similarly, and then solve $\hat{h}$ again with the new values of $\hat{t}$, $\hat{u}$ and $\hat{s}$. The process is performed iteratively C times to approach the optimal solution; the convergence rate is discussed in Section 4. We then use the optimal solution $\epsilon(i,j)$ of each entity pair v_i and v_j to determine each entry of the cascading matrix $\mathbf{E}$.

Shaker detection by computing accumulated influence and maximal influence

Recall that we define the accumulated influence (Eq. 3) and maximal influence (Eq. 4) to capture the shakers. However, both scoring functions involve some undetermined variables such as $\mathbf{C}_t^{(v_i)}$. As a result, the two functions seem to be not computable with the unknown variables. We next prove that both scoring functions actually only depend on the cascading matrix $\mathbf{E}$ or the topology of the graph.

THEOREM 1. *The value of the accumulated influence* $\mathbf{AI}(v_i)$ *only depends on the cascading matrix* $\mathbf{E}$*.*

PROOF. We can rewrite the accumulated influence as

$$\begin{aligned}\mathbf{AI}(v_i) &= \sum_{t=1}^{T} \frac{\partial \mathbf{C}_t^{(v_i)}}{\partial \mathbf{C}_0^{(v_i)}} = \sum_{t=1}^{T} \frac{\|\mathbf{C}_t^{(v_i)} - \mathbf{C}_t\|}{\|\mathbf{C}_0^{(v_i)} - \mathbf{C}_0\|} \\ &= \sum_{t=1}^{T} \frac{tr\Big((\mathbf{E}^t)^T \cdot (\mathbf{C}_0^{(v_i)} - \mathbf{C}_0)^T (\mathbf{C}_0^{(v_i)} - \mathbf{C}_0) \cdot \mathbf{E}^t)\Big)}{tr\Big((\mathbf{C}_0^{(v_i)} - \mathbf{C}_0)^T (\mathbf{C}_0^{(v_i)} - \mathbf{C}_0)\Big)}\end{aligned} \tag{9}$$

Denote $\Delta\mathbf{C} = \mathbf{C}_0^{(v_i)} - \mathbf{C}_0$, and it is important to note that $\Delta\mathbf{C}(i) \neq 0$ and $\forall j \neq i, \Delta\mathbf{C}(j) = 0$ since only the target entity v_i is changed in $\mathbf{C}_0^{(v_i)}$. As such, $(\Delta\mathbf{C})^T \Delta\mathbf{C}$ is a $|\mathcal{V}| \times$

Input: $\mathcal{V}$:the set of evolving entities; n (default 9) : the degree of the polynomial functions used to fit the data; C (default 3) : the number of iterations in the optimization; T (default 6) : parameter to control the time span;

Output: R_A=Ranking according to $\mathbf{AI}(v)$ and R_M=Ranking according to $\mathbf{MI}(v)$ for $\forall v \in \mathcal{V}$

/* Construct the cascading graph */

1 **for** $\forall v_i \in \mathcal{V}$ **do**

2   **for** $\forall v_j \in \mathcal{V}$ **do**

3   Solve $\min_{\hat{h},\hat{u},\hat{t}} \epsilon(i,j)$ in Eq. 5 with n and C;

4   Determine $\mathbf{E}(i,j)$ and $\mathbf{E}(j,i)$ as Eq. 6 and 7.

5   **end**

6 **end**

/* Shaker detection */

7 Calculate $\mathbf{E}^t$ for $t = 1, \cdots, T$

8 **for** $\forall v_i \in \mathcal{V}$ **do**

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

10     $\mathbf{MI}(v_i) = \max\{\sum_{j=1}^{|\mathcal{V}|} \mathbf{E}^t(i,j)^2\}$

11 **end**

12 $R_A \leftarrow$Rank $v \in \mathcal{V}$ according to $\mathbf{AI}(v)$ (descending)

13 $R_M \leftarrow$Rank $v \in \mathcal{V}$ according to $\mathbf{MI}(v)$ (descending)

Algorithm 1: Shaker Detection via Cascading Graph

$|\mathcal{V}|$ matrix with only the (i,i)-th entry being non-zero. We denote the value as q. Hence, it is not difficult to get that

$$tr\Big((\mathbf{E}^t)^T \cdot (\mathbf{C}_0^{(v_i)} - \mathbf{C}_0)^T(\mathbf{C}_0^{(v_i)} - \mathbf{C}_0) \cdot \mathbf{E}^t)\Big) = \sum_{j=1}^{|\mathcal{V}|} q\mathbf{E}^t(i,j)^2 \tag{10}$$

And therefore

$$\mathbf{AI}(v_i) = \sum_{t=1}^{T} \frac{\sum_{j=1}^{|\mathcal{V}|} q\mathbf{E}^t(i,j)^2}{q} = \sum_{t=1}^{T}\sum_{j=1}^{|\mathcal{V}|} \mathbf{E}^t(i,j)^2 \tag{11}$$

Hence, $\mathbf{AI}(v_i)$ only depends on the cascading matrix $\mathbf{E}$. $\square$

Similarly, we have the following theorem.

THEOREM 2. *The maximal influence* $\mathbf{MI}(v_i)$ *only depends on the cascading matrix* $\mathbf{E}$ *where*

$$\mathbf{MI}(v_i) = \max_{t=1,\cdots,T}\{\sum_{j=1}^{|\mathcal{V}|} \mathbf{E}^t(i,j)^2\} \tag{12}$$

It is necessary to mention again that both Theorem 1 and Theorem 2 are very important, since they provide methods to calculate the two scoring functions that are superficially not computable. Furthermore, it is not difficult to see that the computations of the two scoring functions are now linear to the cardinality of the entity set $|\mathcal{V}|$ with T as a relatively small number (discussed in Section 4).

Algorithm complexity analysis

The complete algorithm is presented in Algorithm 1, which contains two main parts: cascading graph generation and shaker detection. The algorithm complexity from Step 1 to Step 6 is $O(C \cdot n \cdot L \cdot |\mathcal{V}|^2)$ where n is the degree of the polynomial function, L is the time span of the data, and C is the number of iterations of the coordinate descent in Step 3. With n and C being relatively small numbers (discussed in Section 4), the complexity is $O(L|\mathcal{V}|^2)$. The most expensive step of the second part is to calculate the matrix power (Step 7), whose complexity is $O(T|\mathcal{V}|^{2.4})$ with divide and conquer strategy. As a result, the complexity of the whole algorithm is $O(|\mathcal{V}|^{2.4})$ and the space complexity is $O(|\mathcal{V}|^2)$, given that T is a small constant as discussed in Section 4.4.

4. EXPERIMENTS

Recall that the proposed model contains two main parts: one is to generate the cascading graph from the evolving entities; the second is to find the shakers in the graph. We consider the steps together and focus on the evaluation of the shakers discovered at the end. Also note that the output (Algorithm 1) is the ranking of the entities according to their estimated influence scores; the entities rank at the top are the shakers. Hence, we evaluate the methods with both classification and ranking related criteria. Three methods are chosen for comparison. It is important to note that since these comparison methods are not designed to handle negative edges and mine shakers, some modifications (as outlined below) are necessary:

1. The first method to compare is the weighted PageRank [4], a state-of-the-art algorithm to rank influential websites. It is used in [34] to find influential nodes in a network, and we adopt the same mechanism to find the shakers. Note that PageRank does not work on the network with negative edges. Hence, we treat all negative and positive edges the same by setting the edge weights as their absolute values, in order to run PageRank on the dataset without getting errors.

2. The second comparison method is based on the Independent Cascade (IC) model [15, 11] that is widely used as an information propagation framework on network. We use the weighted IC model in [33] as a comparison model to find the top-k nodes that maximize the influence. Similar to PageRank, we modify the cascading graph by setting the edge weights as their absolute values in order to make the model work.

3. The last comparison method is the Linear Threshold (LT) model used in [11, 13]. It is also a popular network influence propagation model, and we use the influential node selection algorithm in [11] as the third comparison model. As the previous two, in order to make the model work on the dataset, we set the edge weights as their absolute values.

It is important to emphasize again that it is a necessary step to treat all positive and negative edges the same by setting edge weights as their absolute values. This strategy is considered to induce the least information loss in order to make the comparison models work, without getting a software error.

Three datasets (Table 1) from social sciences are used to evaluate the algorithms. The economic dataset contains the historical data of 967 economic indicators of the United States covering a 30 year period. The GDP dataset is the historical GDP data for 150 countries, and the Stock dataset contains the historical daily prices of 2730 stocks. We treat

Table 1: Summary of experimental datasets and the number of edges of their cascading graphs.

	# Entities	Time Span	# Edges	Meaning of the Entities	Meaning of the Shakers
Economic dataset	967	30	80,195	Economic indicators	Key economic indicators
GDP dataset	150	50	5,823	Countries with GDP	High impact countries in economy
Stock dataset	2730	260 (52 weeks)	2,793,323	Stocks	Stock leaders

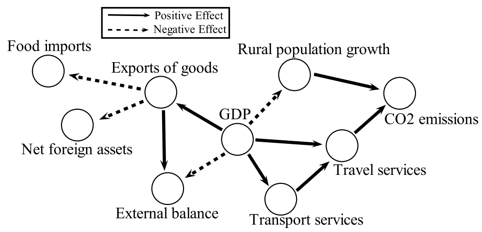


Figure 4: A fraction of the cascading graph of economic indicators

the economic indicators, countries and the stocks as entities respectively; their historical data are used as the evolving attributes, as described in Table 1 . The physical meanings of the shakers in the three datasets are also shown in Table 1. Note that we adopt a pruning step to set $\mathbf{E}(i,j) = 0$ if its value is less than 10^{-4} to reduce noise.

4.1 Finding the most important economic indicators

The first experiment is to find the shakers in 967 economic indicators (*e.g.*, consumer expenditure, interest rate, *etc.*). The data was downloaded from the World Bank [2]. Each of the economic indicators is regarded as an entity and the objective is to find the shakers who contribute the most to the value changes of other indicators. Recall that we use the evolution trend similarity between each entity pair to detect the causalities, which are then summarized into the cascading graph. A fraction of the generated cascading graph is shown in Fig. 4. As it can be observed, it is a complicated network. For example, the increase of GDP has a positive effect on the number of travel services, but negative effect on the rural population growth rate. Moreover, both the rural population and travel services will then increase the CO_2 emissions. Hence, it is not obvious to derive whether GDP has a positive, negative or even no effect (effect cancellation) on CO_2 emissions. With such a complicated network, we aim at finding the shakers who will contribute the most to the value changes of other entities. The top-5 shakers found by the proposed models as well as the three comparison methods are presented in Table 2. As pointed out in [31, 19], some of the widely accepted important economic indicators are the GDP, Gross National Income (GNI), Purchasing Power Parity (PPP), the GDP derived from PPP, and the GNI derived from PPP. Note that all of these indicators are discovered by the two proposed scoring functions. But in the comparison methods, only the Independent Cascade (IC) model discovers one such indicator (GNI), while missing the rest. The most important reason for the poor performances is that the comparison models are not designed to handle a cascading graph that has a more general but more complicated influence propagation schema. Hence, the two scoring functions offer more advantages to find the shakers.

4.2 Finding country-level economy shakers

We downloaded the gross domestic product (GDP) historical data for 150 countries covering a 50 year period (1960-2009) from the World Bank [2]. Each country is modeled as an entity. The objective is to find the top-10 economy shakers according to the development of their GDPs. We use the economy ranking of countries in 2010 as the ground truth [24]. Note that we only consider the countries with complete GDP data for all 50 years in the World Bank and hence some important countries such as Russia, Germany, *etc.*, do not appear in the data and ground truth. The top-10 countries from the ground truth are the United States, China, Japan, France, United Kingdom, Italy, Brazil, Canada, India and Spain. If the found shaker belongs to the top-10 countries regardless of its rank, then we consider it as a correct prediction; otherwise, it is a false prediction.

The results are presented in Table 3, with the accuracy summarized in the last row. It can be observed that the two proposed scoring functions effectively capture the high impact countries in economy, while the comparison models have poor performances. For example, in the economics textbook [31], it is not surprising that United States, United Kingdom and Japan have had strong impacts on the world during the past 50 years. It is also reasonable that countries like Brazil and China are becoming more and more important in recent years [3]. All of these countries were found by the two scoring functions just by tracing the GDP value changes but without any other information. On the contrary, the comparison model LT misses "United States" and IC misses "United Kingdom". Note that between the two proposed scoring functions, the accumulated influence criterion performs a little better (90% accuracy) on this dataset. But it is important to mention again that the two scoring functions have their own statistical meanings and one should choose from them based on different needs of interpretations. For instance, if one only cares about the entities' maximal impact regardless of the impact disappearing later, then the maximal influence criterion should be chosen.

4.3 Stock market shaker detection

We downloaded the historical daily price data for 2730 stocks from Yahoo! Finance [6]. All the 2730 stocks are from NASDAQ in a single year period (from 11/30/2009 to 11/30/2010) and we aimed at finding the stock market shakers who affect the market the most. It is important to mention that we also generated the ground truth from Yahoo! Finance. More specifically, we can find the stock industry leadership rankings [7] with various criteria such as market capitalization, PEG ratio, *etc.* For example, Google Inc. (GOOG) ranks 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 the two criteria, and it ranks first if the average scores of other stocks are

Table 2: Top 5 Shakers of Economic Indicators

PageRank	IC	LT	Accumulated	Maximal
Net current transfers	GNI	Arable land	GNI	GNI
Merchandise imports	Workers' remittances	Agricultural machinery	GDP	GDP
Current account balance	Export volume index	Net income from abroad	GDP, PPP	GNI, PPP
Export volume index	Changes in inventories	Net bilateral aid flows	GNI, PPP	GDP, PPP
Net trade in goods and services	Net current transfers	Changes in inventories	PPP	PPP

Table 3: Top 10 Economy Shakers (only consider the countries with complete GDP data in World Bank)

PageRank	IC	LT	Accumulated Influence	Maximal Influence
Zimbabwe	Japan ✓	Austria	United States ✓	United States ✓
United States ✓	United States ✓	Andorra	Japan ✓	United Kingdom ✓
United Kingdom ✓	Zimbabwe	Israel	United Kingdom ✓	Japan✓
Belgium	Andorra	Serbia	France ✓	France✓
Switzerland	Namibia	Timor-Leste	China ✓	Brazil ✓
Japan ✓	Vietnam	Guinea-Bissau	Brazil ✓	Italy ✓
France✓	Italy ✓	Malta	Italy ✓	Spain ✓
Sweden	Angola	Hungary	Spain ✓	Sweden
Brazil ✓	Lithuania	Australia	Mexico	China ✓
Italy ✓	Fiji	Chile	Canada ✓	Mexico
Accuracy: 60%	Accuracy: 30%	Accuracy: 0%	**Accuracy: 90%**	Accuracy: 80%

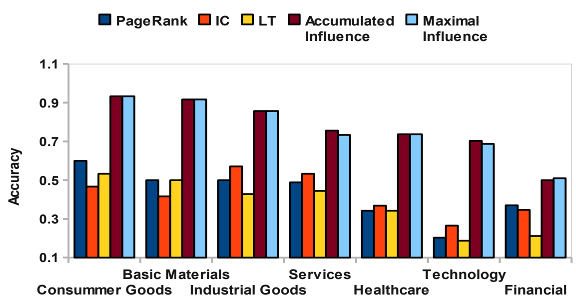


Figure 5: Accuracy of the top 10% of rankings

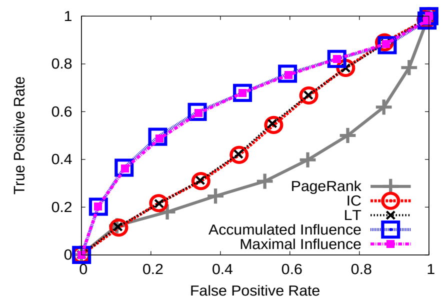


Figure 6: Lift curves of shaker detection in 2730 NASDAQ stocks

larger than 8. As such, the top 20% of stocks in each industry are considered to be the true market shakers. We then conduct three sets of experiments in this dataset to evaluate the models

1. Finding stock market shakers in each of the 7 sectors classified in Yahoo! Finance: consumer goods, basic materials, industrial goods, services, healthcare, technology and financial.

2. Finding stock market shakers in the overall market.

3. Case study to see how the stock market shakers affect the market after the studied period (after 11/30/2010).

Result analysis By tracing the evolutions of stock prices, we generate 7 cascading graphs for the 7 different sectors and analyze their stock market shakers respectively. Recall that the final output (Algorithm 1) is the rankings of the entities with respect to their influence scores, with the top ranked ones being the shakers. We evaluated the accuracies of the top 10% of the rankings (50 such stocks in each sector on average) and the results are presented in Fig. 5. It is clear that the two proposed scoring functions outperform the comparison models. For example, in the healthcare sector, the accuracies of the three comparison methods are just around 35%, while those of the proposed models are over 70%. Again, the reason is that the comparison methods are not suitable to handle networks such as a cascading graph with opposite correlations. The aim of the second experiment was to find the stock market shakers in all of the 2730 stocks. In this case, only one cascading graph is generated to summarize the causality relationships among all stocks to find the global market shakers. The lift curves are plotted in Fig. 6, where we can clearly observe that the two proposed models beat the comparison methods. We also present the top 10 stock market shakers in Table 4, in which the two proposed scoring functions clearly outperform the comparison models by as much as 50% in accuracy. Note that the Apple Inc. (AAPL) is identified by both proposed functions. More interestingly, AAPL suffers a big slide on January 18th, 2011 following the news release of its CEO Steve Jobs taking a medical leave. The NASDAQ 100 index ($NDX) follows its big loss the next day [26] even though there are strong earning reports from companies like IBM and despite the rally of the Asian market. Note that the internal market mechanism may be very complicated and there may be other reasons for the market slide in addition to AAPL, but the market shaker does reveal the risk ahead.

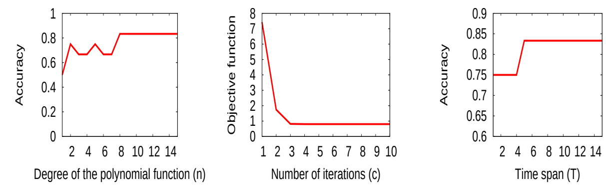
4.4 Parameter Sensitivity

There are three parameters in the proposed model. We

Table 4: **Top 10 Stock Shakers in NASDAQ**

PageRank	IC	LT	Accumulated Influence	Maximal Influence
AACOU ✓	VCLT ✓	AAPL ✓	ACGL ×	ACGL ×
CAPS ✓	FITBP ✓	RDWR ×	MICC ✓	MICC ✓
TWER ×	WGOV ×	ABBC ×	HBANP ✓	HBANP ✓
JSDA ×	TLCR ✓	LCUT ✓	SHPGY ✓	SHPGY ✓
EVBS ×	VYFC ✓	AKAM ×	VGIT ✓	VGIT ✓
ZANE ×	BTFG ×	ADTN ✓	ATRI ×	ATRI ×
AUTH ×	ALOG ✓	AFSI ×	VCLT ✓	VCLT ✓
PFBC ×	ABCD ×	ACXM ✓	COST ✓	COST ✓
IBNK ×	ZUMZ ×	AAWW ✓	AAPL ✓	AAPL ✓
IMOS ×	AACC ×	DNDN ×	PNRA ✓	PNRA ✓
Accuracy: 20%	Accuracy: 50%	Accuracy: 50%	**Accuracy: 80%**	**Accuracy: 80%**

choose the "Basic materials" sector of the stock data to analyze the three parameters, and choose the accumulated influence as the scoring function since the maximal influence criterion has similar performances. The first parameter is the degree n of the polynomial function, which is used to fit the data; the result is shown in Fig. 7(a). It can be observed that the result tends to be stable when the degree is larger than 8. This is because even if we set the degree larger, the coefficients of the high degree terms tend to be zero in order to minimize the ridge loss with regularization term. In practice, one can choose the degree n to be a median number (*e.g.*, $n = 10$), or adaptively choose n according to the fitting loss with regularization. The second parameter is the number of iterations of the coordinate descent to solve the optimization in Eq. 5. As in Fig. 7(b), the objective function actually converges very fast since the four variables have closed-form solutions in each iteration. The last parameter is the time span T across which the influence is estimated. In other words, we only consider the influence of the target nodes within T step. It can be observed from Fig. 7(c) that the performance is stable after $T > 5$. It means that the effects are propagated efficiently. In the experiment, we set $n = 9, C = 3$ and $T = 6$.



(a) Degree (b) Convergence (c) Time span

Figure 7: Parameter sensitivity analysis

5. RELATED WORK

There are several areas of work that we build upon. First of all, causality has attracted intensive interest in artificial intelligence. An increasing number of works are proposed to model the causality among discrete and countable events by using their frequency of co-occurrences (*e.g.*, [27, 10, 23]). In reality, however, on the one hand, there may be a huge number of observed events; on the other hand, there is usually a time latency between the cause and the effect. Hence, it is generally very difficult to detect the underlying causality of the entities just by observing their co-occurrence. The present work investigates the transformations of the entity evolution trends instead of the co-occurrence to detect causality. In one of the steps of the proposed model, we borrow the idea of time series (*e.g.*, [22, 35]), although it is not the focus of the paper. We fit the evolving data with polynomial functions and detect the similarity among the function transformations. There may be other time series techniques that can be used to find the causalities, such as using some discrete transformations ([20, 21]). But the most important advantage of polynomial fitting is that it can give us closed form solutions in each iteration that leads to the efficient computation (Fig. 7(b)).

Information diffusion and influence maximization is another area of related work. It has attracted considerable attention (*e.g.*, [8, 16, 34]) in recent years. The general idea is to trace the information propagation on a given network by various models, and find the top influential nodes or effectors that can best explain the phenomenon (*e.g.*, [1, 17, 15, 13]). Usually, the entities have only one of only two states (active or inactive) and the found influential nodes usually have wide spreads. With a different objective, the proposed model allows numerical attributes of the entities and considers whether the target nodes cause attribute value changes for some others, which are not necessarily the ones with large spreads. Furthermore, the proposed model allows for negative edges in the network where the influence diffusion becomes more complicated. Extensive experiments show that it is better to handle the entities that naturally contain opposite correlations. Another area of related work is data mining in economics. For example, in [5], a network based approach is applied to find similarities/communities among time-series foreign exchanges, but it does not consider influence cancellation between negative effects and positive effects.

6. CONCLUSION

Given a set of evolving entities with complex, latent and dynamic relationships, this paper investigates the problem of mining shakers or entities that can cause significant value changes for others. An example is the stock market shakers that significantly affect the prices of other stocks. We address the problem by proposing a cascading graph that summarizes the causalities among the entities by capturing the similarities of their evolving trends. More importantly, two scoring functions (accumulated and maximal effects) are formulated in order to capture the degree or efficiency on how the target entity affects others. To do so, we artificially inject a significant change on the target entity where its effects propagate throughout the network, and then estimate its influence on other nodes for a given period of time. Although the two scoring functions seem unsolvable because of several undetermined variables, we formally prove that the unknown variables can be canceled with the inference rule derived from Markovian assumption, and the two scoring functions are actually only dependent on the structure of the cascading graph. The proposed model is evaluated on three real world datasets from social sciences: an economic indicator dataset, a country GDP development dataset and the stock market dataset. It is observed that the proposed model outperforms the three comparison methods substantially. For example, in the stock market dataset, the accuracies of the three comparison methods are just around 35% while those of the proposed models are over 70% in the healthcare sector.

Acknowledgement The authors would like to thank the MITRE Corporation for their support to this work. They would also like to express their gratitude to Dr. Ali Hadjarian for his proofreading of the paper.

7. 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] World bank dataset. http://data.worldbank.org/.

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

[4] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *WWW*, pages 107–117, 1998.

[5] D.J. Fenn, M.A. Porter, P.J. Mucha, M. McDonald, S.Williams, N.F. Johnson, and N.S. Jones. Dynamical clustering of exchange rates. *preprint (2010) arXiv:0905.4912v2*, 2010.

[6] Yahoo! finance. http://finance.yahoo.com/.

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

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

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

[10] D. Heckerman, D. Geiger, and D. M. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995.

[11] D. Kempe, J. Kleinberg, and É. Tardos. Influential nodes in a diffusion model for social networks. In *IN ICALP*, pages 1127–1138, 2005.

[12] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, pages 137–146, 2003.

[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] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins. The web as a graph: measurements, models, and methods. In *COCOON*, pages 1–17, 1999.

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

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

[17] 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.

[18] D. Li, J. Shen, and J. Xie. Study on representation of time series based on subsection polynomial fitting. In *FSKD*, pages 1371–1376, 2007.

[19] D. Miles and A. Scott. *Macroeconomics: understanding the wealth of nations*. John Wiley and Sons., 2nd edition, 2005.

[20] A. Mueen and E. J. Keogh. Online discovery and maintenance of time series motifs. In *KDD*, pages 1089–1098, 2010.

[21] A. Mueen, E. J. Keogh, Q. Zhu, S. Cash, and M. B. Westover. Exact discovery of time series motifs. In *SDM*, pages 473–484, 2009.

[22] S. Papadimitriou, J. Sun, and C. Faloutsos. Streaming pattern discovery in multiple time-series. In *VLDB*, pages 697–708, 2005.

[23] J. Pearl. Causal inference in statistics: an overview. *Statistics Surveys*, 3:96–146, 2009.

[24] Country ranking in GDP. http://en.wikipedia.org/wiki/list-of-countries-by-gdp-(nominal).

[25] 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.

[26] Market reaction on AAPL. http://www.finanzen.net/nachricht/aktien/market-snapshot-stocks-mostly-down-after-apple-goldman-results-1012909.

[27] G. Shafer. Causal logic. In *ECAI*, pages 711–720, 1998.

[28] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000.

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

[30] Y. Sun, Q. Meng, Z. Chen, and Y. Zhang. Normalizing the polynomial match for the nonlinear signal in transducers. *Sensors and Actuators*, 125:76–83, 2005.

[31] T. D. Tregarthen and L. Rittenberg. *Macroeconomics*. Worth Publishers, 2nd edition, 2000.

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

[33] Y. Wang, G. Cong, G. Song, and K. Xie. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In *KDD*, pages 1039–1048, 2010.

[34] 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.

[35] B. Yi, N. Sidiropoulos, T. Johnson, H. V. Jagadish, C. Faloutsos, and A. Biliris. Online data mining for co-evolving time sequences. In *ICDE*, pages 13–22, 2000.