Mining Causality from Co-location Patterns by Approximating Direct Causes from Granger Causes

SU Hong, LU Weiqi

September 2, 2022

Abstract

Co-location patterns and Causality are two well studied fields nowadays. The problem of mining causalities among co-location patterns is an important primitive in knowledge discovery, with wide-ranging applications from ecosystem to praxiology and epidemiology. However, there are still some limitations in solving the above problem. Many causality mining models lack full utilization of the temporal dimension. And one major challenge of Granger causality test is that it only guarantees the cause precedes the effect in time. To tackle these limitations, this report proposes a novel and scalable model for mining causality from co-location patterns based on Granger causality and designs an algorithm for approximating direct causes from Granger causes based on unique causal information. And extensive experiments are conducted on real-world datasets to evaluate our result and compare it with others.

1 Introduction

1.1 Motivation

In the age of big data, the large amount of location-based data obtained from electronic devices equipped with GPS gives people the opportunity to mine rules of interest. The spatial-temporal dataset is one of the most common location-based datasets of great value to study, with dimensions of time, location, and other attributes of interest. One hot area of study of spatial-temporal data is association rule mining, which can be applied in many areas, such as ecology, environmental study, and location-based recommendation.

However, one drawback of the association rule is its lack of use of the dimension of time, which would lead to the loss of information on cause and effect hidden in the spatial-temporal dataset. This motivates us to develop the algorithm for the causality rule mining spatial-temporal dataset, in particular, object-based. There are several benefits of causal rules over association rules. Firstly, causal rules can give us a better understanding of the relationships between different types of objects, which can further inspire researchers to take a deeper study of the true causes of some events, for example, the locust plague. Moreover, given the causal rules, people can make a better prediction of future events with the up-to-date dataset, while avoiding the situation where a cause is predicted to occur in the future given its effect observed at present. There are many applications of causality rule mining from a spatial-temporal dataset, for example, recommending places of interest to a user given the user's location and recent behavior, finding the food chains of an ecosystem, and investigating the root cause of a natural disaster.

Our approach first finds the prevalent co-location patterns in the spatial-temporal dataset, then counts the number of instances for each prevalent co-location pattern at each timestamp, and summarizes the counted numbers as a time series dataset. Then, we apply the Granger causality test to the time series dataset to find the Granger causes of some prevalent co-location patterns. Finally, we use partial correlation to prune the indirect causes from the Granger causes. Our approach can be applied to both moving-object and event-based spatial-temporal datasets to approximate the direct causes of a co-location pattern.

1.2 Applications

- **Ecosystem**: The evolution of an ecosystem is the outcome of interactions between animals, plants, microbes, and the natural environment. Finding the causal relationships between these factors helps analyze and predict the evolution of the ecosystem over time, e.g., figuring out the reasons why the crops do not grow well.
- **Praxiology**: Praxiology is the theory of human action. People, items, and venues are different spatial features that can be correlated with each other. The causal relationships between these factors can be a reference for spatial-temporal activity preference of people [11] with numerous applications ranging from personalized location-based recommendations to group-oriented advertisements.
- **Epidemiology**: During a global pandemic, implemented policies, people, entertainment places, and confirmed cases can be regarded as different spatial feature types. Building the causal chains that eventually lead to confirmed cases can help locate the root causes of the pandemic and perform appropriate actions to break the causal chains.

2 Related Work

- **Co-location Pattern**: co-location pattern [4, 1] is a well studied field, which studies the co-occurrence of spatial features in close geographic proximity. Our research related to co-location pattern is based on the classical and discrete version of co-location pattern study [4] and then extends its idea to the spatial-temporal scenario. [4] selects prevalent co-location patterns by the minimum participation ratio of the involved spatial features. However, there exists some situations where a spatial feature has much larger number of instances than others, which results in the low participation ratio of this feature in related co-location patterns. In this case, co-location patterns containing this spatial feature would be pruned and the related causalities are overlooked. To avoid over-pruning, we select co-location patterns based on the maximum participation ratio of involved features, which is more reasonable in the context of causality mining.
- **Causality Mining without Time**: Causality mining is a research hot topic. There are many causality mining algorithms or models which do not involve temporal dimension. For instance, some traditional algorithms such as Wermuth-Lauritzen algorithm and SGS algorithm [9]. Recently, [12] introduces a non-combinatorial method NO TEARS to learn the directed acyclic causal graph from a dataset with multivariate samples. NO TEARS algorithm works well if no temporal dimension is involved or the effects immediately follow its causes. However, the loss of temporal information makes the estimated causal graph unreliable as the time lag between causes and effects becomes significant.
- **Causality Mining among Co-location Pattern**: Though causality mining and co-location patterns are two well studied fields, there is inadequate research in the combination of both. To the best of our knowledge, the only existing relevant research is [5], which proposes an algorithm CRDA to discover causal rules by analyzing the newly adding/disappearing co-locations. However, CRDA only considers three timestamps to mine causality, which is very likely to produce unstable results due to the presence of noises. What is more, if we directly extend this idea into n timestamps by analyzing three consecutive timestamps at a time and summarize the results, there still exist two major limitations. 1) CRDA requires the cause and the effect to form a co-location. For instance, if co-location pattern {A,B} → {C,D}, then the appearance of {A,B} is assumed to be co-located with the appearance of {C,D}. However, in the general scenario where moving objects may exist, the union of the cause and effect is not necessarily a co-location. 2) It is hard or inefficient to determine the time interval between t_1, t_2, t_3 as different pairs of cause and effect may only be discovered on different time intervals. However, CRDA fails to handle this situation unless a naive enumeration is performed, which is not efficient.
- **Granger Causality**: Granger causality model is a popular causality model which fully utilizes the temporal dimension.Two major research fields of Granger causality are: Granger causality test and Hawkes process.

Granger causality test [13, 3] is a statistical hypothesis testing for determining whether one variable is helpful to forecast another one. [13] introduces the general framework of Granger causality testing. But, Granger causality test only guarantees the cause precedes the effect in time. As a result, Granger causes without too much unique causal information may also be included. [3] is a non-spatial Granger causality test paper with an algorithm to select Granger causes. But, the stepwise selection algorithm is based on precision instead of unique causal information, which is more useful in prediction but not applicable in approximating direct causes.

Hawkes process studies the impact of history events on the probability of some event occurs at present [7, 10, 8]. As proposed by [10], the Granger causality A → B exists if and only if the impact function of A on B is not constant zero. Two major algorithms to estimate the parameters of the Hawkes process model are MLE-SGLP [10] and VI-SG [8]. MLE-SGLP applies Group-Lasso to enforce the sparsity of Granger causes. However, the time complexity of MLE-SGLP is $\mathcal{O}(N^3)$, where N is the size of the dataset. In this case, fine-tuning the hyper-parameters such as coefficients for regularization terms becomes unfeasible as the size of the dataset grows significantly. To address this problem, VI-SG uses variational expectation-maximization algorithm to optimize both parameters and hyper-parameters of the Hawkes process model, and returns the probability distribution of model parameters. It assumes that the parameters have log normal distribution because the parameters of the Hawkes process model are non-negative by definition. This implies that the estimated parameters have non-zero means, which makes it difficult to define a threshold to extract direct causes from Granger causes.

Table 1: Comparison between causality mining models

	Time Domain	**Cyclic Causality**	**Scalable**	**Compatible with M.O.***
NO TEARS			✓	✓
VI-SG	✓	✓		✓
CRDA	✓	✓	✓	
ours	✓	✓	✓	✓

* denotes Moving Objects.

3 Contributions

1. Slightly modify the popular but static co-location pattern metric **Participation Index** to fit the dynamic time scenario such that we can determine a co-location pattern over a period of time. Furthermore, this modification avoids over-pruning of co-location patterns caused by the unbalanced populations and participation ratios of different spatial features, so the downstream causality mining from co-location patterns is more comprehensive. More details are given in Section 4.1.2.
2. Use Granger causality model to mine causal relationships from co-location pattern, fully utilizing the temporal dimension.
3. Granger causality test requires the input time series to be stationary. We use ΔN_t (or generally $(1-B)^k N_t$), where N_t is the number of instances of certain co-location pattern, as the variables in Granger causality test to fulfill the stationarity assumption.
4. A select and prune algorithm is proposed to approximate the direct causes of c based on whether unique causal information is contained.
5. The time complexity of our algorithm is $\mathcal{O}(n^2U^2)$, where n is the number of timestamps, U is the number of spatial features. In practice, $n \ll N$, where N is the total number of co-locations formed in all timestamps. Compared with the related work, our algorithm is much more efficient without loss of accuracy when the dataset is large, since it focuses more on the global changes of co-location numbers with time to discover causalities.

4 Technical Approach

4.1 Technical details

4.1.1 Co-location

According to [4], a **co-location** c is defined as a subset of spatial features $\mathcal{T} = \{f_1, f_2, \cdots, f_K\}$, say $\{f_{n_{c1}}, f_{n_{c1}}, \cdots, f_{n_{ck}}\}$. Let θ be a user defined threshold in a sense that $\mathcal{R}(x, y)$ if and only if $d(x, y) \le \theta$, where d is a metric, $\mathcal{R}$ is a binary

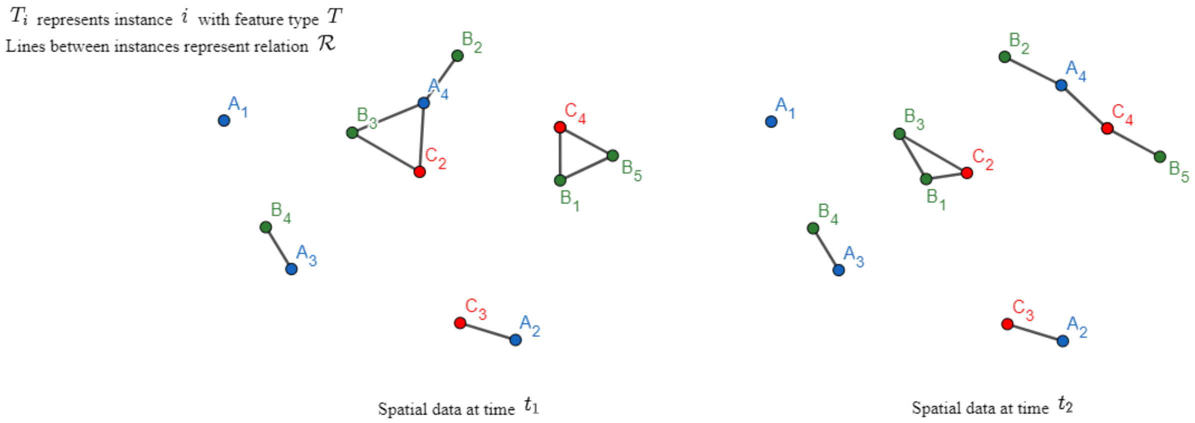

Figure 1: Sample data for illustration

relation, x, y are instances of spatial feature types. We define a subset of spatial feature instances $c_i^t = \{i_1, i_2, \cdots, i_k\}$ at time t as **co-location instance** if it contains all spatial features in c and $\mathcal{R}(i_p, i_q)$, $\forall p \neq q$.
Take Figure 1 as an example. $\mathcal{R}(A_3, B_4)$ at time t_1 since $d(A_3, B_4) \leq \theta$ for some user defined threshold. $\{A_4, B_3, C_2\}$ at time t_1 is a co-location instance of $\{A, B, C\}$ since all pairwise distances are within the threshold θ.

4.1.2 Modified Participation Index(PI)

$$\mathbf{PI} = \max_{f_i \in c}\{\Pr(c, f_i)\} \tag{1}$$

where

$$\Pr(c, f_i) = \frac{\pi_{f_i}(|\text{table\_instance}(c)|)}{|\text{table\_instance}(f_i)|} \tag{2}$$

Take Figure 1 as an example. Consider the Modified Participation Index of the co-location $c := \{A, B\}$. At time t_1, the co-location instances include $\{A_3, B_4\}, \{A_4, B_2\}, \{A_4, B_3\}$. Thus, $Pr(c, A) = \frac{2}{4}$, and $Pr(c, B) = \frac{3}{5}$. Then, $PI = \max_{f \in c}\{Pr(c, f)\} = \frac{3}{5}$.
Remark: In [4], PI is defined to be $\mathbf{PI} = \min_{f_i \in c}\{\Pr(c, f_i)\}$. We use max here because max can help to mine more causality. For example, assume there is a causal relationship $A \Rightarrow \{A, B\}$, if we use min definition to determine whether $\{A, B\}$ is a co-location pattern (The determining procedure is shown in Section 4.1.3, then $\{A, B\}$ is a co-location pattern if and only if the majority of A objects form a co-location with B objects $(*)$ and the majority of B objects also form a co-location with A objects $(**)$. But when we mine the causal relationship $A \Rightarrow \{A, B\}$, it is not necessary that both $(*)$ and $(**)$ hold. Either $(*)$ or $(*)$ holds is sufficient.

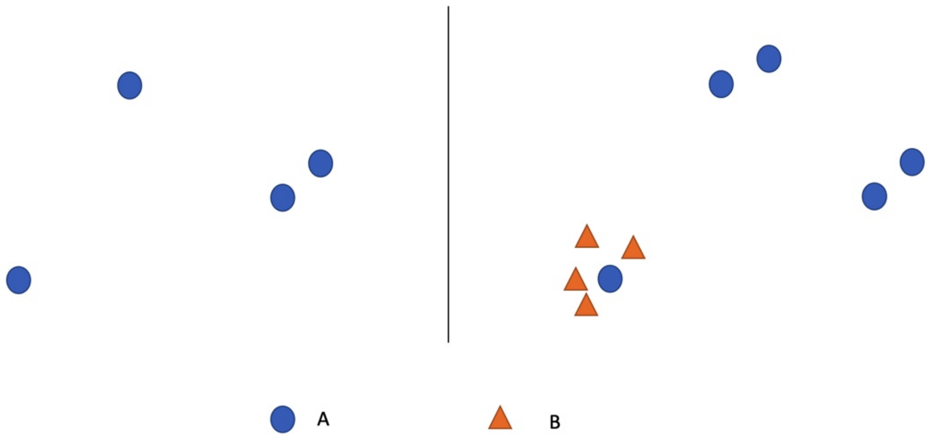

Figure 2: Sample data for illustration

See Figure 2 for more details. We only consider two timestamps for simplicity, left part is t_1, right part is t_2 By definition, **A** itself is a co-location pattern. After a period a time, the instances of objects **A** and **B** are shown in the right part. Intuitively, object**A** will "attract" object **B**, i.e. there should exist a causality rule $A \Rightarrow \{A, B\}$. By definition, consider $c = \{A, B\}$, then $Pr(c, A) = \frac{1}{5}, Pr(c, B) = 1$. But if we use min for **PI**, we can see **PI** for $\{A, B\}$ is $\frac{1}{5}$, which indicates that $c = \{A, B\}$ is not a co-location pattern if we set the threshold to be $> \frac{1}{5}$. Then we will miss this causality when we use min, while max can fix this problem.

4.1.3 Co-location Pattern

The idea of determining co-location pattern is the same as in [4], if the set of co-locations c satisfying the condition that average of its corresponding **PI** over $t_1, \cdots, t_N \geq \xi$, i.e. $\frac{1}{N} \sum_{t=t_1}^{t_N} PI_t \geq \xi$, where ξ is some user defined threshold between 0 and 1, then we claim c is a prevalent co-location pattern.
Take Figure 1 as an example. Consider the co-location $c := \{A, B\}$. Set the threshold $\xi = 0.5$. Then, we can calculate that $PI_{t_1} = 0.6$, $PI_{t_2} = 0.5$, $\bar{PI} = \frac{1}{2}(PI_{t_1} + PI_{t_2}) = 0.55 > \xi$. c is a prevalent co-location pattern by the definition.

4.1.4 Granger Causality

Let us discuss two time series first. Consider two stationary time series x and y. To test whether x is a Granger cause of y, we conduct autoregression of y with appropriate time lagged values L:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_L y_{t-L} + \varepsilon_{1t} \tag{3}$$

Next, we augment the autoregression by adding lagged values of x:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_L y_{t-L} + b_1 x_{t-1} + \cdots + b_L x_{t-L} + \varepsilon_{2t} \tag{4}$$

where $\{a_i\}_{i=0}^L, \{b_i\}_{i=1}^L$ are regression coefficients, and ε_{1t}, ε_{2t} are independent Gaussian white noises. Or in most of the cases, the sample size of time series data is large enough for these residuals to approximate to the normal distribution by central limit theorem.
If augmenting x helps to add explanatory power to the regression according to the F-test, i.e. using F-statistics to test the null hypothesis $H_0 : b_1 = b_2 = \cdots = b_L = 0$. If we reject H_0, then we claim x is a Granger cause of y.

4.1.5 Formulating Number of Co-locations

Lemma 1. *If co-location pattern c only contains single type of object and is the root cause of the causality chain, then its number of co-locations $N_t \sim ARIMA(0, 1, 0)$.*

Proof. Consider a co-location pattern c which is the root cause of the causality chain. Denote N_t as the number of co-locations of c, then

$$N_t = N_{t-1} + a_t \tag{5}$$

where $a_t \sim N(0, \sigma^2)$. Equation 5 can be easily verified as c does not have any cause, then N_t actually follows random walk, i.e. a Markov Process whose future value only depends on the current value. So $N_t \sim \text{ARIMA}(0, 1, 0)$. □

Definition 4.1. The variables used in Granger Causality test are k-time difference of number of co-locations, i.e. $(1 - B)^k N_t,\ k \geq 1$. Generally speaking, $k = 1$ is enough. When $k = 1$, the variable used for Granger Causality test is ΔN_t. And ΔN_t is used as default unless exception.

Remark. The reasons why we use $(1 - B)^k N_t$ (most of the time, ΔN_t) are as follows:
1)Since for root cause at least one time difference is needed, and if we intend to use Granger Causality test, all N_t should have the same times of difference. As a result, one time difference is needed if one of the co-location patterns is root cause.
2)ΔN_t is of high level of interpretation. Because it makes sense that if c_i is the cause of c_j, then the increment of number of co-locations of c_i is closely related to that of c_j. For example, if there is a positive increment of number of co-locations of c_j, there is very likely to be a positive increment of number of co-locations of c_j.
3)Based on our experience, $k = 1$ is enough for most of the time.

4.1.6 Partial Correlation

For partial correlation between X and Y conditioning on N controlling variables $\mathbf{Z} = \{Z_1, Z_2, \cdots, Z_N\}$, denoted as $\rho_{XY\cdot\mathbf{Z}}$:

$$\begin{aligned} \rho_{XY\cdot\mathbf{Z}} &= \text{corr}(e_X, e_Y) \\ &= \text{corr}(X - L(X|\mathbf{Z}), Y - L(Y|\mathbf{Z})) \end{aligned} \tag{6}$$

where $L(\cdot|\mathbf{Z})$ is the linear regression whose response is $\cdot$ and predictors are $\mathbf{Z}$.
The intuition behind is that one compute the correlation between X and Y are removing the linear dependencies of $\mathbf{Z}$ to get the partial correlation.

4.2 Selecting Best Approximation of Direct Cause

Without loss of generality, consider a co-location pattern c and a set of all its Granger causes $\mathcal{G} = \{c_{i_1}, \cdots, c_{i_k}\}$, our goal is to select a subset of $\mathcal{G}$, denoted as $\mathcal{G}'$ such that $\mathcal{G}'$ is the best approximation of the direct cause of c. See Figure 3 for more intuition.

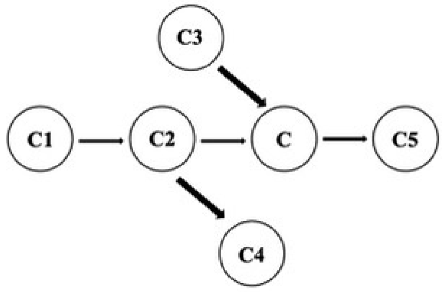

Figure 3: Arrows here imply causal relationship. By definition of cause and Granger cause, $C5$ is not Granger cause since C is prior to $C5$, $C4$ is not a cause of c but may be a Granger cause of c. $C1$ is an indirect cause of c. Then by Granger Causality Test, we will obtain $C1, C2, C3$ (or $C1, C2, C3, C4$) as the set of Granger Causes of C. But not all $C1, C2, C3$ (or $C1, C2, C3, C4$) are direct causes of C.

4.2.1 Selecting Best Time Lag L for Granger cause

In Section 4.1.4, assume x is a Granger cause of y, then we need to choose a suitable time lag L such that the residual error is minimized. The idea is computing partial correlation between x_t and x_{t-L} conditioned on $x_{t-1}, \cdots, x_{t-L+1}$ for $L = 1, 2, 3, \cdots$, and we choose a user defined threshold ϵ such that we choose the minimal L with the partial correlation between x_t and $x_{t-L} < \epsilon$. Then we can get the time lag that is the most representative of the trend of x.

4.2.2 Goodness of Granger cause

Suppose X and Y are both Granger causes of Z, if X and Y both contain unique information about the effect Z, then the partial correlation between X and Z conditioned on Y (removing the linear dependency of Y) should be significantly greater than 0. The same argument holds for correlation between Y and Z conditioned on X.

However, if X is the direct cause but Y only contains some information originated from X, then the correlation between Y and Z conditioned on X should be insignificantly greater than 0. As a result, we can prune Y from the set of Granger causality of Z.

As a result, a "good" Granger cause that is the direct cause or part of the direct cause should contain some unique information about the effect. To formally introduce the goodness of a Granger cause, with respect to how close it is to the direct cause. Define:
Let c be the effect and $\mathcal{G}$ be the set of Granger causes of c. Denote $\mathcal{G}'$ as the subset of $\mathcal{G}$ such that $\mathcal{G}'$ is the best approximation of the direct cause of c.
$c_i \in \mathcal{G}'$ if and only if

$$\rho_{c_i c \cdot \mathcal{G}^{-c_i}} > \delta \tag{7}$$

where ρ stands for partial correlation, $\mathcal{G}^{-c_i}$ is a subset of $\mathcal{G}$ excluding c_i and δ is some user defined threshold to ensure the partial correlation between c_i and c is significantly greater than 0.

4.2.3 Algorithm for Selecting and Pruning

The algorithm for finding causes is proposed as follows: Algorithm 1 first computes the training accuracy of autoregression for each Granger cause c_i of c. We use Mean Square Error (MSE) to measure the training accuracy for each predictor c_i. We assume that the direct cause can predict the future of c most accurately. Therefore, we choose c_j that has the least MSE. After that, we remove predictors from Granger causes $\mathcal{G}$ that have small partial correlation with c conditioned on c_j. The intuition is that if c_i has information for prediction of c similar to that of c_j, then, the partial correlation between c and c_i conditioned on c_j should be very small since the linear relationship between c and c_i has been well represented by c_j. We repeat the process until all Granger causes have been chosen or pruned. More concrete details of the algorithm are shown in Section 5.4.

5 Experiment

In this section we detail the experiments and the results for our Granger Causality model and algorithm. Firstly, we briefly describe our dataset and data pre-processing in Section 5.1. The first dataset is a check-in dataset, which records users' visits to places of interest. A co-location instance is formed between a user and a venue if the user checks in at the venue. All the co-location patterns discovered from the check-in dataset are shown in Section 5.2. Next, tests for checking stationarity of ΔN_t are introduced in Section 5.3. Finally in Section 5.5, Algorithm 1 is applied to the Granger causes of certain co-location pattern to get the approximation of direct causes.

5.1 Data Description and Pre-processing

The first dataset is NYC Check-in Dataset obtained from [11], which is available in the web link (https://sites.google.com/site/yangdingqi/home/foursquare-dataset). The dataset contains nearly 230,000 check-in records

---

Algorithm 1 Algorithm for selecting and pruning cause.

Given a set of co-location pattern c*'s Granger causes* $\mathcal{G} = \{c_{i_1}, \cdots, c_{i_k}\}$
Initialize $\mathcal{G}' = \{\}$
for all c_i *in* $\mathcal{G}$ **do**
    Compute the best time lag L for c_i
    Compute R-squared ${R^2}_{(i)}$ *for autoregression* $c^t \sim c^{t-1}, \cdots, c^{t-L}, {c_i}^{t-1}, \cdots, {c_i}^{t-L}$
end for
while $\mathcal{G}$ *is not empty*: **do**
    Choose $c_j = \arg\max_{c_i \in \mathcal{G}} {R^2}_{(i)}$
    $\mathcal{G}' \leftarrow \mathcal{G}' \cup \{c_j\}$
    Prune c_j *from* $\mathcal{G}$
    for all c_i *in* $\mathcal{G}$ **do**
        if $|\mathrm{corr}(c_j, c_i)| >$ *user defined threshold* τ **then**
            $\mathcal{G}' \leftarrow \mathcal{G}' \cup \{c_i\}$
        else
            Compute the partial correlation ρ_{ij} *between c and* c_i *conditioned on* $\mathcal{G}'$
            if $|\rho_{ij}| <$ *user defined threshold* θ **then**
                Prune c_i *from* $\mathcal{G}$
            end if
        end if
    end for
end while
return $\mathcal{G}'$

---

collected from 1,083 users for about 10 months. Each check-in record is of the form (User ID, Venue ID, Venue category ID, Venue category name, Latitude, Longitude, Timezone offset in minutes, UTC time).

The dataset originally contains 251 venue categories, which is too fine-grained to mine causalities. Therefore, we merge venue categories with similar purposes, for example, various restaurants, different schools and colleges, many indoor places for entertainment. With this approach, we classify all venues into 49 coarse-grained venue categories. The advantage of combining venue categories is that we can gain more observations of each coarse-grained category for a better analysis of user check-in behaviors. Even so, there are only 15 samples on average per coarse-grained category per day, which is not sufficient to analyze daily behaviors of users.

To address the time sparsity of the dataset, we make an assumption that user behaviors on different days are independent since we are more interested in the behavioral patterns of users from morning to night, and the causalities across a day or more are usually dependent on personal weekly schedules, which are not easy to be generalized to all users. Under this assumption, users on different days can be considered as different persons. Therefore, we take (User ID, Date of UTC time) as a unique identifier for users, and only use the time without date for causal analysis. After that, all records are concentrated in one day to avoid the time sparsity of observations, which ensures the statistical reliability of the result obtained by our algorithms.

5.2 Mining Co-location Pattern

Besides coarse-grained venue categories, we also regard users as an object type. The main purpose of our experiment on the check-in dataset is to explore user check-in behaviors, so we ignore the co-locations between venues and define that a co-location instance is formed only when a user checks in at a venue. Then, the causalities we mine are of the form (user, venue category A) → (user, venue category B).

With the above definition, we evaluate the modified PI for each co-location pattern (user, venue category), and select those with relatively large modified PI as prevalent co-location patterns for causal analysis.

5.3 Stationarity Testing

There exist many methods to check the stationarity of a time series such as Augmented Dickey & Fuller test (ADF) [2] and Ljung-Box test [6]. As Ljung-Box test required more manual effort, we mainly apply ADF to check the stationarity of ΔN_t mentioned in Section 4.1.5 and use Ljung-Box test as a supplementary test for stationarity. And the results of ADF suggest all ΔN_t stationary at 99% confidence level.

5.4 Selecting & Pruning

After using Granger causality test to mine all Granger causes of certain co-location patterns, Algorithm 1 is performed to select and prune some of the Granger causes discovered. For example, the set of all Granger causalities of {People, Factory} is {Convention Center, Food & Snack, School, Store, Market & Fair, Transportation, Travel-Related Place}. By human common sense, some pairs such as School & Factory should not have causal relationship. And the later analysis will show that School indeed does not contain adequate unique causal information of Factory.

As in Algorithm 1, from line 3 to line 6, all corresponding R^2 are computed, e.g. R^2 of Factory ~ History of Factory + History of School. Next, based on the R^2 computed, at the first iteration Transportation is selected to set $\mathcal{G}'$. Then by line 11 to line 19, all the remaining co-location patterns are pruned based on the partial correlation metric. The select & prune result is consistent with the overall causal chain. For instance, School is the approximated direct cause of transportation according to Figure 4, which forms the causal chain School → Transportation → Factory. Hence, the causal information of School to Factory is actually contained via Transportation. Consequently, if we select & prune based on unique causal information, School is pruned.

5.5 Experimental Results

We have run three models, ours, VI-SG, and NO TEARS on the dataset. The result is shown in Figure 4. We can see for the non-temporal causality mining model NO TEARS, the performance is worst as it contains some isolated nodes such as hotel. And NO TEARS contains some unreasonable causal relationships such as school has only cause factory. For VI-SG, it also misses some causal relationships such as school and general educated places.

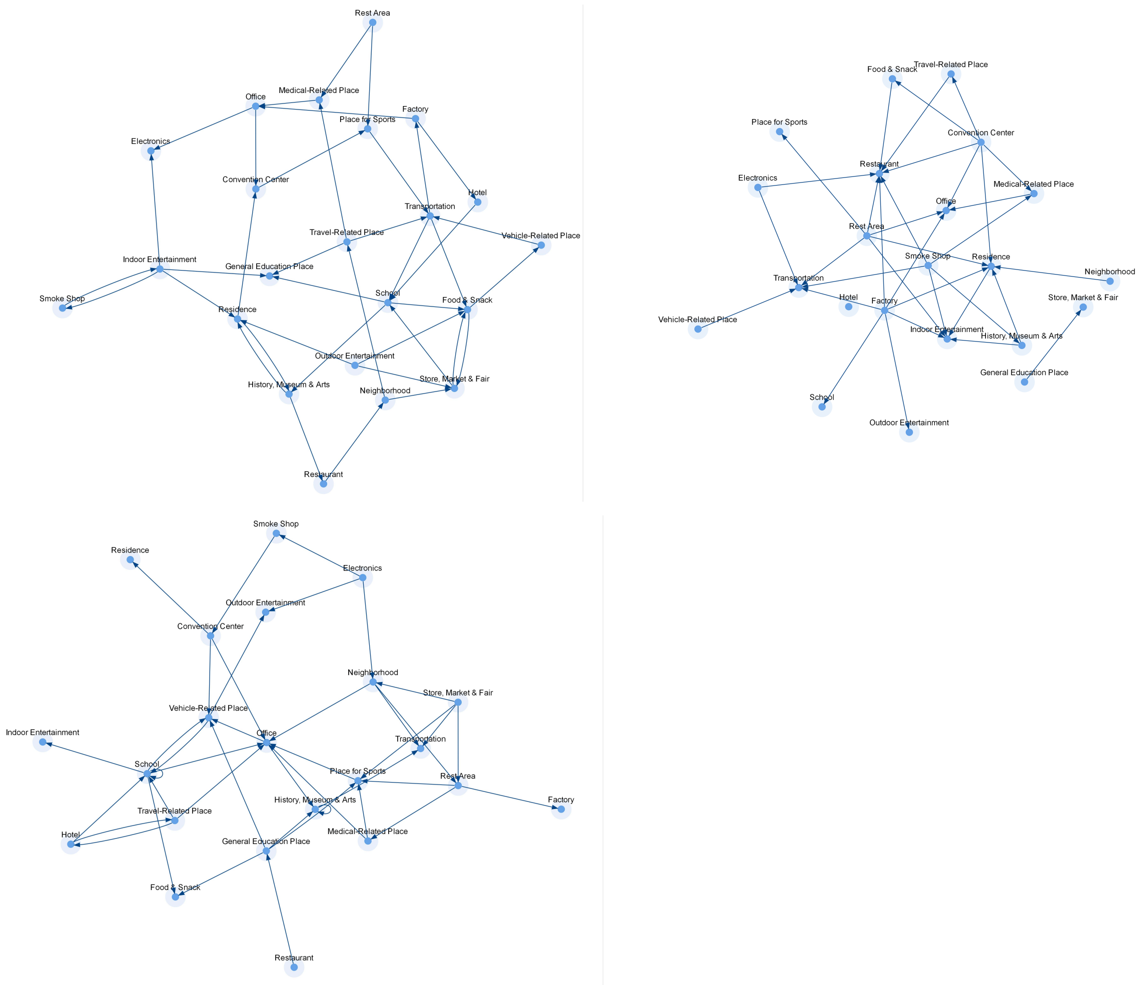

Figure 4: Experimental results on ours (top-left), VI-SG (bottom-left), and NO TEARS (top-right).

6 Individual Contributions

SU Hong mainly contributes to proposing the Granger Causality Model including the variables used and relevant hypothesis tests. LU Weiqi mainly contributes to designing the algorithms for selecting/pruning. And both of us work together to conduct the experiments including pre-processing the datasets and further analysis.

7 Future Work

1. We will find a more appropriate and reasonable approach to compare our results with others such as defining a partial ground-truths by case studies.
2. We will conduct one more experiment on a massive dataset with 4 million records over 20 cities from America, France, England and Russia. Then we will analyze and compare the behaviours and preferences of people in difference cities.

References

[1] Harry Kai-Ho Chan, Cheng Long, Da Yan, and Raymond Chi-Wing Wong. Fraction-score: a new support measure for co-location pattern mining. In *2019 IEEE 35th international conference on data engineering (ICDE)*, pages 1514–1525. IEEE, 2019.

[2] Alice Chuang. Time series analysis: univariate and multivariate methods, 1991.

[3] Xuemeng Fan, Yaolai Wang, and Xu-Qing Tang. Extracting predictors for lung adenocarcinoma based on granger causality test and stepwise character selection. *BMC bioinformatics*, 20(7):83–96, 2019.

[4] Yan Huang, Shashi Shekhar, and Hui Xiong. Discovering colocation patterns from spatial data sets: a general approach. *IEEE Transactions on Knowledge and data engineering*, 16(12):1472–1485, 2004.

[5] Junli Lu, Lizhen Wang, and Yuan Fang. Mining causal rules hidden in spatial co-locations based on dynamic spatial databases. In *2016 International Conference on Computer, Information and Telecommunication Systems (CITS)*, pages 1–6. IEEE, 2016.

[6] Anna Clara Monti. A proposal for a residual autocorrelation test in linear models. *Biometrika*, 81(4):776–780, 1994.

[7] Katarzyna Obral. *Simulation, estimation and applications of hawkes processes.* PhD thesis, Master's thesis, University of Minnesota, 2016.

[8] Farnood Salehi, William Trouleau, Matthias Grossglauser, and Patrick Thiran. Learning hawkes processes from a handful of events. *Advances in Neural Information Processing Systems*, 32, 2019.

[9] Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search.* MIT press, 2000.

[10] Hongteng Xu, Mehrdad Farajtabar, and Hongyuan Zha. Learning granger causality for hawkes processes. In *International conference on machine learning*, pages 1717–1726. PMLR, 2016.

[11] Dingqi Yang, Daqing Zhang, Vincent W Zheng, and Zhiyong Yu. Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(1):129–142, 2014.

[12] Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. Dags with no tears: Continuous optimization for structure learning. *Advances in Neural Information Processing Systems*, 31, 2018.

[13] Julie Yixuan Zhu, Chenxi Sun, and Victor OK Li. An extended spatio-temporal granger causality model for air quality estimation with heterogeneous urban big data. *IEEE Transactions on Big Data*, 3(3):307–319, 2017.