

# Knowledge transfer for learning subject-specific causal models

Verónica Rodríguez-López

*Universidad Tecnológica de la Mixteca, Huajuapan de León, Oaxaca, México*

VERORL@INAOEP.MX

Luis Enrique Sucar

*Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México*

ESUCAR@INAOEP.MX

## Abstract

Subject-specific causal models are appropriate for domains such as biology, medicine, and neuroscience, where the causal relations vary across the individuals of a population. However, its learning could be challenging, particularly under limited data sets. Although some works have addressed this issue, they are restricted to discovering up to Markov equivalence classes. In this work, we propose a method for the causal relations identification of subject-specific models. We hypothesized that transferring related data sets and locally performing interventions improves the causal direction identification of relations. The experimental results on true and imperfect Markov equivalence classes of synthetic causal Bayesian networks show that our method performing interventions over several subsets of the candidate parents and using related data according to their differences with target, recovers a higher number of correct oriented edges.

**Keywords:** Causal discovery, Probabilistic graphical models, Subject-specific causal models, Transfer learning

## 1. Introduction

Probabilistic graphical models (PGMs) could incorporate causal knowledge for making predictions under interventions. Experimental data should be used in their learning, but it may be difficult or impractical to collect enough experimental samples, particularly from specific subjects. For that reason, from observational data, causal discovery methods learn Markov equivalence classes (MECs) representing a set of equivalents PGMs with the same probability distribution and including the true causal PGM that could be identified by performing experiments.

Subject-specific causal PGMs include the particular causal relations of a member of a population. These models could be more appropriate for domains that have observed variations in causal relations across individuals. For example, studies in genetics have found that some somatic genome alterations produce expression changes in specific tumors (Cooper et al., 2018). Some works in neuroscience have revealed that causal relations might vary across patients between brain regions because of their differences in the degree of disease affection and the recovery process (Grefkes and Fink, 2014). Learning subject-specific causal PGMs from limited data sets has not been sufficiently explored. Most causal discovery methods have been designed to search for the common causal relations of a set of individuals from data sets with enough samples. Although some works have addressed this problem, they are limited to learning MECs (Jia et al., 2018; Rodríguez-López and Sucar, 2022), bipartite models (Cooper et al., 2018), or models with homogeneous causal structure and variations in causal effects across the individuals (Li et al., 2018).

In this work, we propose a method for causal direction identification of relations in subject-specific PGMs. Our main contribution is a active learning strategy based on the invariant causal prediction method (ICP) for using data sets related to the target subject and performing interven-

tions over selected variables. Our experimental results show that our method with weighted related samples recovers a higher number of correct oriented edges than from only target samples or from related sources ignoring their differences with the target.

The paper is organized as follows. Related work to our proposal is described in Section 2. Section 3 provides a description of the invariant causal prediction method. Our proposal is presented in Section 4, and the experimental results in Section 5. Finally, in Section 6, conclusions and future work are drawn.

## 2. Related work

Our proposal is a knowledge transfer method that aims to solve the lack of data for the target subject, using data from other ones closely related to the target (Pan and Yang, 2010). Despite the advantages of knowledge transfer strategies for learning from limited data sets, only a few works have explored their application in causal discovery. Most of these works have relied on the identification of causal predictors for a single response variable from multiple data sets (Gamella and Heinze-Deml, 2020; Peters et al., 2016; Rojas-Carulla et al., 2018). Although other methods have been proposed for discovering the complete causal structure, they do not analyze subject-specific causal discovery from limited target data sets. Assuming that all data sets include a representative number of samples, these methods discover the common causal relations from multiple observational (Ramsey et al., 2010; Tillman and Spirtes, 2011) or interventional data sets (Claassen and Heskes, 2010; Mooij et al., 2020; Triantafillou and Tsamardinos, 2015). Only a few works have explored the learning of subject-specific causal models. Some are limited to learning simple bipartite causal models from sufficient data of the target subject (Cooper et al., 2018), for finding the same causal structure for all individuals with only variations in their causal effects (Li et al., 2018), for discovering up to Markov equivalence classes from only observational data of the target subject (Jabbari et al., 2018) or from data sets closely related to the target subject (Jia et al., 2018; Rodríguez-López and Sucar, 2022). In contrast with previous work, we propose a method for causal direction identification of subject-specific causal models by transferring related data sets and actively performing interventions, assuming variations in causal relations across subjects and small target sample size.

## 3. Invariant causal prediction

The invariant causal prediction method (ICP) (Peters et al., 2016) identifies the direct causes of a response variable from a set of environments including samples from different experimental conditions. It focuses on Gaussian linear models and considers that there exists a causal prediction model that is invariant under different experimental conditions.

In ICP, it is assumed that given a set of variables  $\{X_0, X_1, \dots, X_p\}$  in which  $X_0$  is the response variable, and,  $\mathbf{Y} = \{X_1, \dots, X_p\}$  its set of candidate causal predictors, a set of environments  $\mathcal{E}$  is composed by samples for  $(X_0, X_1, \dots, X_p)$  from different experiments  $e \in \mathcal{E}$ , that were generated by linear Gaussian structural equation models (SEM),  $X_j^e = \sum_{k \neq j} \beta_{j,k}^e X_k^e + \epsilon_j^e$ ,  $j = 0, 1, \dots, p$ ; with  $\epsilon_j^e \sim \mathcal{N}(0, \sigma^2)$ , and  $\beta_{j,k}^e \neq 0$  if  $X_k^e = Pa(X_j^e)$ . In  $\mathcal{E}$ ,  $e = 0$  correspond to observational samples, and for  $e > 0$ , samples from SEMs in which  $\mathcal{I}^e \subset X_1, \dots, X_p$  were intervened with the do-operator in the form:

$$X_j^e = \begin{cases} a_j^e & \text{if } j \in \mathcal{I}^e; \\ \sum_{k \neq j} \beta_{j,k}^0 X_k^e + \epsilon_j^0 & \text{if } j \notin \mathcal{I}^e. \end{cases} \quad (1)$$

ICP does not need knowledge about the localization of the interventions. It only requires that interventions do not occur on the response variable and the set of candidate causal predictors includes all causal predictors. To identify the set of causal predictors of a variable, ICP tests with each  $\mathcal{S} \subset \mathbf{Y}$  the null hypothesis:  $H_{0,\mathcal{S}}(\mathcal{E}) : \exists \beta \in \mathbb{R}^p$ , such that,  $\forall e \in \mathcal{E}, \hat{\beta}^e(\mathcal{S}) \equiv \beta$ , and  $X^e = \mathbf{Y}^e \beta + \epsilon^e$ ,  $\epsilon^e \perp X_{\mathcal{S}}^e$ , where  $\hat{\beta}^e$  are the least-squares regression coefficients for  $e \in \mathcal{E}$ . Several sets could satisfy  $H_{0,\mathcal{S}}(\mathcal{E})$ , hence the final set of causal predictors of  $X$  is,  $\hat{\mathcal{S}}(\mathcal{E}) = \bigcap_{(S):H_{0,S}(\mathcal{E}) \text{ not rejected}} \mathcal{S}$ .

## 4. Method

The proposed method, called Knowledge transfer for Subject-Specific causal learning - KSS, identifies the causal direction of subject-specific models with limited observational data using related data sets and locally performing interventions over selected variables. KSS searches the set of causal parents of each variable assuming that does not exist latent variables and using the invariant causal prediction method (Peters et al., 2016).

KSS takes as input the skeleton of a causal Bayesian network (BN)  $\mathcal{G}_T = (\mathbf{X}, \mathbf{E})$ , and finds the causal direction of the edges from a target data set  $\mathbf{D}_T$ , and a set of additional source data sets  $\{\mathbf{D}_s\}_{s=1, \dots, S}$ . We assume that target and sources data sets contain observational samples of the  $\mathbf{X}$ ,  $\mathbf{D}_T = \{(x_{T1}, \dots, x_{Tp})_i\}_{i=1, \dots, N_T}$  and  $\mathbf{D}_s = \{(x_{s1}, \dots, x_{sp})_i\}_{i=1, \dots, N_s}$ , with  $\forall s : P_T(\mathbf{x}) \neq P_s(\mathbf{x})$ ,  $N_T \ll N_s$ , with small  $N_T$ . With these elements, the knowledge transfer problem for causal relation identification is defined as follows:

**Definition 1** *Given a target domain  $\mathcal{D}_T$  and a set of source domains  $\{\mathcal{D}_s\}, s = 1, \dots, S$ , the knowledge transfer problem for causal relation identification consists in how to identify causal directions of a target causal BN  $\mathcal{G}_T = (\mathbf{X}, \mathbf{E})$ , using the knowledge in  $\mathcal{D}_T$  and  $\{\mathcal{D}_s\}$ , with  $\forall s : \mathbf{V}_T = \mathbf{V}_s$ ,  $\mathcal{D}_T \neq \mathcal{D}_s$ .*

Our knowledge transfer method by transferring related sources aims to identify higher number of correct causal directions than those recovered by a method that only uses  $\mathcal{D}_T$ . We consider that source data sets contain samples from causal BNs related to the target causal BNs. That is, source causal BNs have some parts of their structures in common with the target causal BNs hence they represent structure modifications of the target causal BN.

For each  $X_i$ , KSS locally searches for its set of parents by applying ICP, transferring source data sets, and performing interventions over subsets of candidate parents. Source data sets are transferred to estimate the parameters of the linear Gaussian model for  $X_i$ . To improve the estimations of the parameters, we consider differences between source and target data sets in their probability distributions. These differences are encoded in weights for each source sample that are estimated following the proposal of Rodríguez-López and Sucar (2022):

$$W_{sj}(\mathbf{D}_T, \mathbf{D}_s) = \exp\left(-\frac{1}{K_\gamma} d_{MMD}(\mathbf{D}_T, \mathbf{D}_s)\right) * \exp\left(-\frac{1}{K_\psi} d_\psi(x_{sj}, \mathbf{D}_T)\right) \quad (2)$$

where  $K_\gamma = \frac{1}{S} \sum_{s=1}^S d_{MMD}(\mathbf{D}_T, \mathbf{D}_s)$ ,  $K_\psi = \frac{1}{N_s} \sum_{j=1}^{N_s} d_\psi(x_{sj}, \mathbf{D}_T)$ , with  $d_{MMD}$  the maximum mean discrepancy (MMD) metric (Borgwardt et al., 2006), and

$$d_\psi(x_{sj}, \mathcal{D}_T) = \frac{\frac{1}{N_T} \sum_{k=1}^{N_T} d_{MMD}(x_{Tk}, x_{sj})}{\max_{k=1, \dots, N_T} \{d_{MMD}(x_{Tk}, x_{sj})\}} \quad (3)$$

An overview of the KSS is given in Algorithm 1. KSS takes the skeleton of a causal BN as input to reduce the search of causal parents of a specific variable. First, KSS applies the *multipleICP* procedure (see Algorithm 2) to identify source data sets most related to the target. We consider that the *multipleICP* procedure can not find parents for  $X_i$  with some source data sets because their probability distributions are most similar to that of the target. Thus, weighted samples of source data sets most similar to the target are added to the target data set to estimate the parameters of the linear Gaussian model for  $X_i$ . Then, it applies the *performNExperiments* procedure that is described in Algorithm 3. This procedure obtains samples for  $X_i$  and its candidate parents  $\mathbf{Y}_i$  after performing interventions over several subsets of the candidate parents. We suppose that all candidate parents should be intervened to ensure the identifiability of the true causal parents. For that, the *performNExperiments* procedure at least generates one environment with the intervention of all candidate parents. These new interventional samples and the new target data set are used to find the parents of  $X_i$ . Finally, based on the width of the corresponding confidence interval, the direction of each undirected edge in  $\mathcal{G}_{IN}$  is determined.

Because some source data sets might include observational or interventions samples not helpful for identifying the parents of  $X_i$ , the procedure *multipleICP* (see Algorithm 2) applies ICP with each combination of target and source data. Following the strategy in Gamella and Heinze-Deml (2020), it identifies as parents of  $X_i$ , the variables that appear on at least half of all sets accepted by ICP. The union of all parents sets and its confidence interval define the direction of the edges.

---

**Algorithm 1:** KSS

---

**Procedure** *KSS*

**Input:**  $\mathbf{D}_T$ : target data set,  $\{\mathbf{D}_s\}$ : set of source data sets,  $\mathcal{G}_{IN}$ : undirected graph,  $nExp$ : the maximum number of experiments,  $nEnv$ : the number of environments for each experiment,  $\alpha$ : confidence level

**Output:**  $\mathcal{G}_{OUT}$  estimated DAG

Compute  $\mathbf{W}_{si}$  for each  $x_i \in D_s$  using 2

**foreach**  $X_i$  in  $\mathcal{G}_{IN}$  **do**

$\mathbf{Y}_i \leftarrow$  adjacent nodes to  $X_i$  in  $\mathcal{G}_{IN}$

$\mathbf{S}_{ACC} \leftarrow multipleICP(X_i, \mathbf{Y}_i, \mathbf{D}_T, \{\mathbf{D}_s\}, \alpha)$

$\mathbf{D}_{Tnew} \leftarrow \mathbf{D}_T(X_i, \mathbf{Y}_i) \cup \{\mathbf{W}_{ri} * \mathbf{D}_r(X_i, \mathbf{Y}_i)\}_{r \notin \mathbf{S}_{ACC}}$

$\mathbf{D}_{Src} \leftarrow \{\mathbf{D}_r(X_i, \mathbf{Y}_i)\}_{r \in \mathbf{S}_{ACC}}$

$\{\mathcal{E}_t\} \leftarrow performNExperiments(X_i, \mathbf{Y}_i, \mathbf{D}_{Tnew}, nExp, nEnv, |\mathbf{D}_{Tnew}|)$

$(\mathbf{Z}_i, \mathbf{S}_{ACC}, \mathbf{IC}_i) \leftarrow multipleICP(X_i, \mathbf{Y}_i, \mathbf{D}_{Tnew}, \{\mathcal{E}_t\}, \alpha)$

$\mathcal{G}_{OUT} \leftarrow \mathcal{G}_{IN}$ ,  $\mathbf{T} \leftarrow$  the set of undirected edges in  $\mathcal{G}_{IN}$

**foreach**  $(X - Y) \in \mathbf{T}$  **do**

$A_{X \rightarrow Y} \leftarrow IC_Y(X), B_{Y \rightarrow X} \leftarrow IC_X(Y)$

**if**  $A_{X \rightarrow Y} \neq B_{Y \rightarrow X}$  **and**  $A_{X \rightarrow Y} > B_{Y \rightarrow X}$  **then**

**orient**  $(X - Y)$  as  $(X \rightarrow Y)$  in  $\mathcal{G}_{OUT}$

**if**  $A_{X \rightarrow Y} \neq B_{Y \rightarrow X}$  **and**  $A_{X \rightarrow Y} < B_{Y \rightarrow X}$  **then**

**orient**  $(X - Y)$  as  $(Y \rightarrow X)$  in  $\mathcal{G}_{OUT}$

**return**  $\mathcal{G}_{OUT}$

---

---

**Algorithm 2:** multipleICP

---

**Procedure** *multipleICP*

**Input:**  $\mathbf{Y}$ : set of potential parents for  $X$ ,  $\mathcal{E}_0$ : observational environment,  $\{\mathcal{E}_s\}$ : set of experimental environments and,  $\alpha$ : confidence level

**Output:**  $\mathbf{Z} \subset \mathbf{Y}$ : set of accepted parents for  $X$ ,  $\{IC_X(W)\}$ : set with the width of the confidence interval for each  $W \in \mathbf{Z}$ ,  $\mathbf{S}_{ACC}$ ; set of source indexes that accept some  $W \in \mathbf{Z}$

```
 $\mathbf{S}_{ACC} \leftarrow \emptyset$ 
foreach  $\mathcal{E}_s$  do
   $\mathcal{E} \leftarrow \mathcal{E}_0 \cup \mathcal{E}_s$ 
   $(\mathbb{S}_{\mathcal{E}}, \mathbf{IC}_{\mathcal{E}}) \leftarrow ICP(\mathcal{E}, \mathbf{Y}, \alpha)$ 
   $\mathbf{Z}_s \leftarrow$  the set of variables that appear on at least half of all sets in  $\mathbb{S}_{\mathcal{E}}$ 
   $\mathbf{IC}_s \leftarrow \{|IC_{\mathcal{E}}(W)|\}_{W \in \mathbf{Z}_s}$ 
   $\mathbf{S}_{ACC} \leftarrow \mathbf{S}_{ACC} \cup \{s\}$ 
 $\mathbf{Z} \leftarrow \bigcup_s \mathbf{Z}_s$ 
 $\mathbf{IC} \leftarrow \max_s(\mathbf{IC}_s)$ 
return  $(\mathbf{Z}, \mathbf{IC}, \mathbf{S}_{ACC})$ 
```

---

**Algorithm 3:** performNExperiments

---

**Procedure** *performNExperiments*

**Input:**  $X$ : target variable,  $\mathbf{Y}$ : set of potential parents for  $X$ ,  $\mathbf{D}$ : samples for  $X$  and  $Y$ ,  $nExp$ : the maximum number of experiments,  $nEnv$  the number of environments for each experiment,  $nsamples$ : the number of samples for each experiment

**Output:**  $\{\mathcal{E}_t\}$  the set of environments

```
 $modelX \leftarrow$  parameters of  $X \sim \mathbf{Y}$  using  $\mathbf{D}$ 
 $\mathcal{E}_1 \leftarrow performExperiment(modelX, \mathcal{I} = \mathbf{Y}, nsamples)$ 
 $i \leftarrow 1$ 
while  $i < nExp$  and  $i < |\mathbf{Y}|$  do
  foreach  $k \leftarrow 1$  to  $nEnv$  do
     $\mathcal{I} \leftarrow \mathcal{I} \subset \mathbf{Y}, |\mathcal{I}| = i$ 
     $\mathcal{E}_t \leftarrow performExperiment(modelX, \mathcal{I}, nsamples)$ 
     $i \leftarrow i + 1$ 
return  $\{\mathcal{E}_t\}$ 
```

---

## 5. Experimental Results

In this section, we present the evaluation of *KSS* using simulated data sets from synthetic and benchmark causal BNs. The experimental evaluation aimed to analyze the performance of *KSS* with different target and source data sets configurations. We hypothesized that by transferring related sources datasets, according to their differences with target data sets, *KSS* will get its best performance. Hence, we analyzed the performance of *KSS* using only samples of target data sets (*KSS-TARGET*),

transferring source data sets ignoring (*KSS-ALL*) and considering their differences with target data sets (*KSS-WEIGHTED*). *KSS-ALL* joins all available data, that is, all source samples with weight one are used to perform experiments. The performance of *KSS* was compared with the *SLICE* method (Montero-Hernandez et al., 2018) using only target samples. Additionally, we evaluated the performance of methods when increasing the number of experimental environments and starting from the perfect and imperfect skeletons of causal BNs.

We evaluated the methods using normalized structural Hamming distance (NSHD), arrowhead precision (AHP) and recall (AHR) metrics. NSHD is the minimum number of edge insertions, deletions, and changes needed to transform a model into another. Arrowhead precision is the ratio  $TP/(TP+FP)$ , and  $TP/(TP+FN)$  is the arrowhead recall. Where  $TP$  is the number of common edges in the estimated and the true models with the same orientation;  $FP$  and  $FN$  represent false orientations, that is when an oriented edge  $X \rightarrow Y$  is present in one model, but in the other one there is  $X \leftarrow Y$ ,  $X - Y$ , or no edge between  $X$  and  $Y$  (Jabbari et al., 2018). Differences between the methods were assessed by applying the Wilcoxon signed-rank test with a significance level of 0.05. All simulations and implementations were performed on *R* 4.0.1 using the *bnlearn* (Scutari, 2010), *pcalg* (Kalisch et al., 2012), and *InvariantCausalPrediction* (Peters et al., 2016) packages.

## 5.1 Synthetic data sets

We generated data sets from the benchmark causal BN MAGIC-NIAB ( $p = 44$  nodes, 66 arcs, and  $k = 9$  maximum number of parents for a node) and ten synthetic causal BNs. Following the procedure of (Hauser and Bühlmann, 2012), we generated ten synthetic causal Bayesian networks representing the target causal BNs, with  $p = 30$  nodes,  $k = 6$ , and edge weights selected uniformly from the range  $[0.1, 1.0]$ . Edge weights were considered as parameters of each variable, together with  $\epsilon \sim N(0, 1)$ . Using the procedure of (Luis et al., 2010), source causal BNs were generated by modifying in certain percent ( $e_{mod}$ ) the edges of the target BNs, adding  $e_{mod}$  edges, followed by deleting edges in the same  $e_{mod}$  percent. Source BNs less related to the target are simulated increasing  $e_{mod}$  (see Figure 1). Five source BNs, from each target causal BN and each  $e_{mod} \in \{10, 20, 30, 40, 50\}\%$ , were generated. Each target and source data set were sampled from its corresponding BN using forward sampling and linear Gaussian models.

## 5.2 Results

We evaluated our proposal starting from the true and estimated skeletons of synthetic and MAGIC-NIAB causal BNs. Estimated skeletons were obtained by the method of (Rodríguez-López and Sucar, 2022) using the three most related synthetic source data sets (with a discrepancy level of 10%). The skeletons for synthetic causal BNs were estimated using source and target data sets with 300 and 30 samples, respectively. From source and target data sets with 3080 and 44 samples, respectively, was estimated the skeleton for MAGIC-NIAB. These estimated skeletons include false and missing edges and incorrect v-structures.

First, we analyzed the impact on the performance of *KSS* of using source data sets with increasing levels of discrepancy. In Figures 2 and 3 we present the average performance for the correct and estimated skeletons of the synthetic causal BNs and the MAGIC-NIAB, respectively. The x-axis represents the level of the discrepancy between source and target data sets. *KSS-WEIGHTED*, *KSS-ALL*, and *KSS-TARGET* were configured to generate one experimental environment with  $p$  samples. In the plots for synthetic causal BNs (Figure 2), each point represents the average performance over

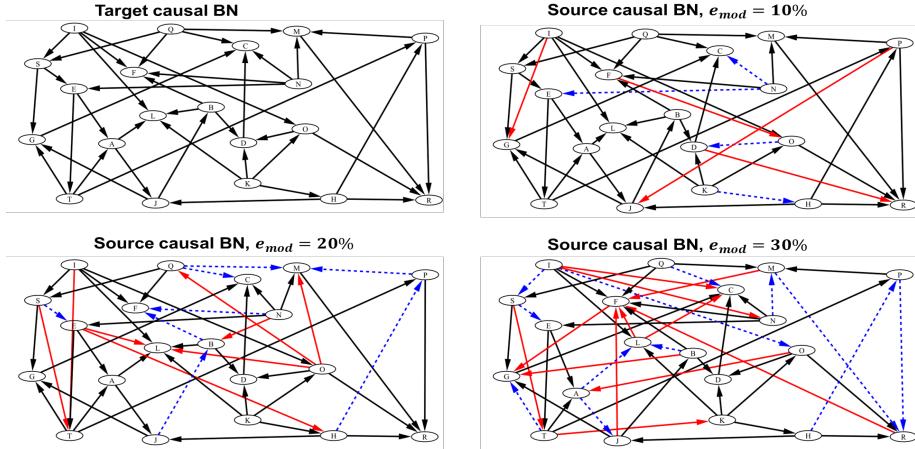


Figure 1: An example of synthetic target and source causal BNs used in the experiments. In red, edges in the source BNs that do not appear in the target causal BN, and in blue, target edges that do not appear in the source causal BNs. (Best seen in color.)

the ten synthetic causal BNs, five target data sets with 15 samples, and five source data sets with the same level of discrepancy and 900 samples. The plots suggest that the level of discrepancy of source data sets impacts the performance in the causal direction identification. They indicate that including the most related source data sets significantly outperformed the other configurations of KSS. We found a significantly superior performance of *KSS-WEIGHTED* with respect to *KSS-TARGET* and *KSS-ALL* in arrowhead recall and NSHD. This confirms that considering the differences between the source and target data sets is important. The performance of *KSS* for true and estimated skeletons is explained by the identifiability conditions of ICP and the number of experimental environments. Because *KSS-WEIGHTED*, *KSS-ALL*, *KSS-TARGET* generate one experimental environment, and ICP has problems identifying single causal parents, they fail in the causal identification of some edges. Furthermore, ICP requires that sets of candidate parents include all causal parents, but estimated skeletons have missing edges. It is important to note that our proposal identifies causal directions from imperfect skeletons without background knowledge.

In Figure 3 we present the average performance for the true and estimated skeletons of the MAGIC-NIAB. Each point in the plots represents the average performance over ten target data sets with 44 samples, and five source data sets with the same level of discrepancy and 6160 samples. The plots shows that starting from the true skeleton, *KSS-WEIGHTED* gives significant superior performance than *KSS-TARGET* and *KSS-ALL*, but not than *SLICE*. Due to the density of the MAGIC-NIAB, one experimental environment is insufficient to identify more causal directions and improve the performance of *SLICE*. This last, also explains the performance of *KSS* with estimated skeletons of MAGIC-NIAB. In this case, the number of experimental environments and the missing edges of estimated skeletons limit the causal identification.

We also analyzed the performance of our proposal when increasing the number of experiments. The maximum number of experimental environments was increased from one to eleven in the following form. For the case of one experiment, it corresponds to the environment generated by the intervention of the  $n$  candidate parents of a variable. The subsequent experiments correspond to environments that were created by the intervention of  $k$  randomly selected candidate parents with

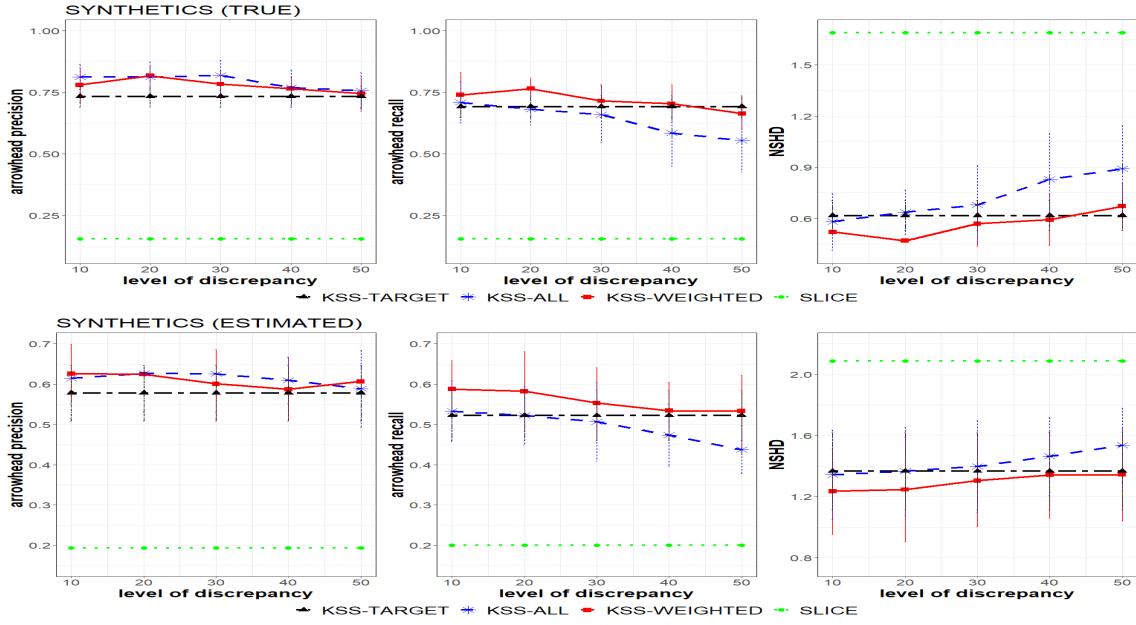


Figure 2: Plots of the averages in arrowhead precision and recall -higher is better-, and NSHD -lower is better- across the target-source discrepancy for the true (top) and estimated (bottom) skeletons of synthetic causal BNs. (Best seen in color.)

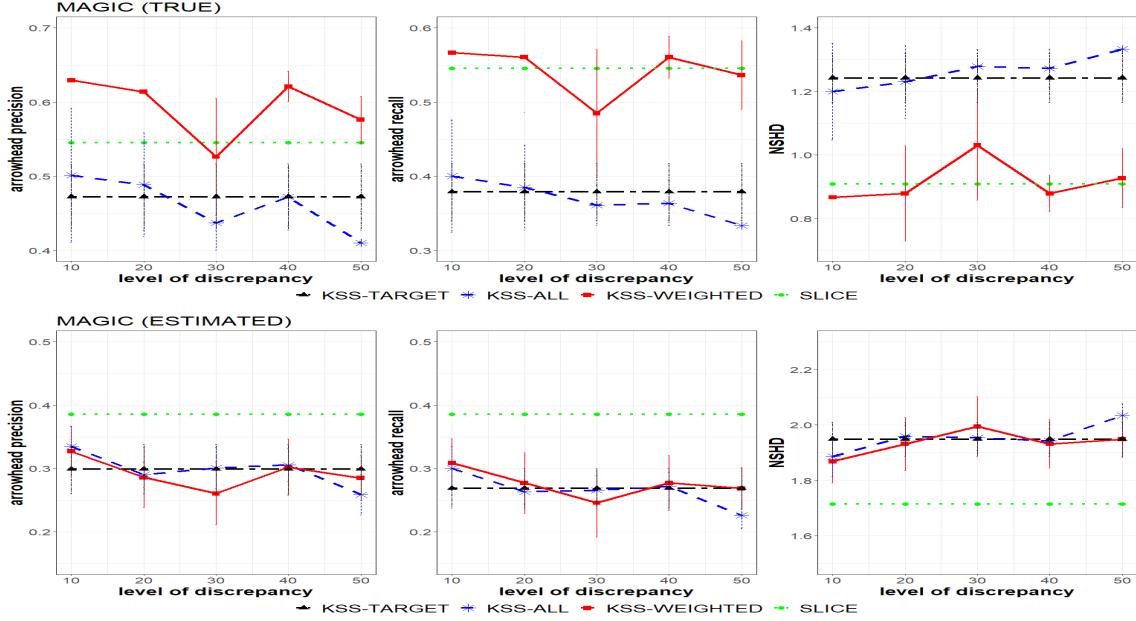


Figure 3: Plots of the averages in arrowhead precision and recall, and NSHD, across the target-source discrepancy for the true (top) and estimated (bottom) skeletons of MAGIC-NIAB. (Best seen in color.)

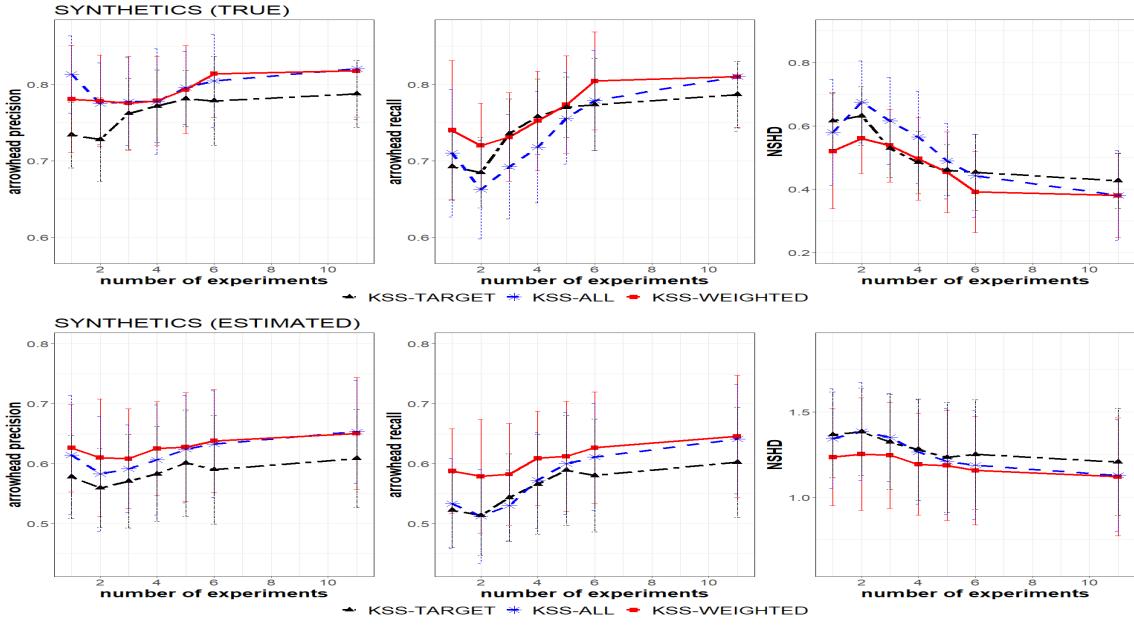


Figure 4: Plots of the averages in arrowhead precision and recall, and NSHD across the number of experiments for the true (top) and estimated (bottom) skeletons of synthetic causal BNs. (Best seen in color.)

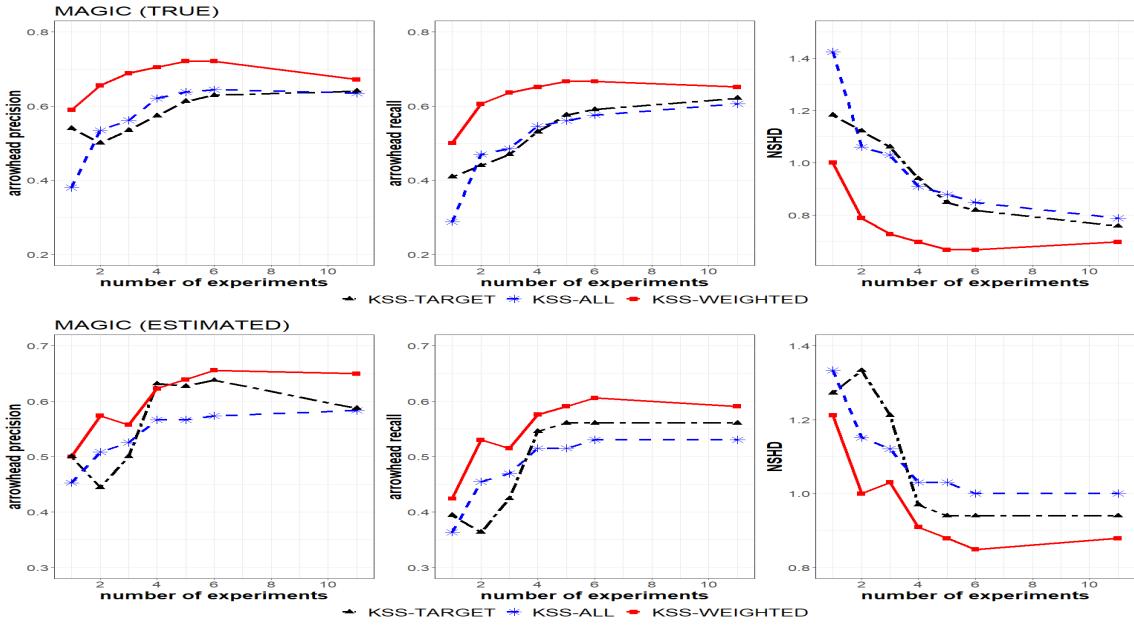


Figure 5: Plots of the averages in arrowhead precision and recall, and NSHD across the number of experiments for the true (top) and estimated (bottom) skeletons of synthetic causal BNs. (Best seen in color.)

$k = 1, 2, 3, \dots, n - 1$ . The remaining experimental data sets were generated by the intervention of each candidate parent. In Figure 4, we present the average performance across the number of experiments for the true and estimated skeletons of synthetic causal BNs. We followed a similar configuration to that of the experiment across the level of discrepancy. The plots suggest that increasing the number of experiments improves the performance of *KSS-WEIGHTED*, *KSS-ALL*, and *KSS-TARGET*. For true and estimated skeletons, we observed significant differences between the performance of *KSS-WEIGHTED* with *KSS-TARGET* and *KSS-ALL* in arrowhead recall and NSHD. The plots indicate that one experimental environment, including interventions of all candidate parents, is enough to obtain acceptable performance. They indicates that even with imperfect skeletons, our proposal improves the causal identification. We found that this performance could be significantly improved by transferring weighted source data sets and adding experimental environments with the interventions of at least half of the candidate parents. Including experimental environments with few interventions is not helpful, particularly with one random intervention that produces edges with incorrect direction. These findings are confirmed with the plots for true and estimated skeletons of MAGIC-NIAB in Figure 5. The plots suggest that *KSS-WEIGHTED* is more appropriate to identify the causal direction of edges in dense causal BNs. The plots shows the significantly superior performance of *KSS-WEIGHTED* with respect to *KSS-ALL* and *KSS-TARGET*.

## 6. Conclusions

In this paper, we have addressed the issue of learning subject-specific causal models from limited data sets through a knowledge transfer method based on ICP. Our proposal is an active method that identifies the direction of causal relations by transferring observational related data sets and performing interventions over selected variables. This is one of the first approaches to try to recover the complete causal structure of subject-specific models with limited data.

Our experimental results with simulated data sets from synthetic causal BNs indicate that our proposal identifies causal directions when starting from the skeleton of causal BNs, including weighted samples of the most related data sets and locally performing interventions over several subsets with at least half of the candidate parents. Even though our proposal requires a skeleton, it could identify causal direction starting from imperfect skeletons, without background knowledge of the possible direction for some edges. Our experimental results also suggest that it is necessary to include strategies to include extra edges in the imperfect skeletons for solving missing potential causal parents.

As future work, we plan to validate our proposal with neuro-images of neuro-rehabilitation patients, and extend it to identify the causal direction using subsets or supersets of the target variables.

## Acknowledgements

This work was supported in part by CONACYT project A1-S-43346. V. Rodríguez-López was sponsored by a CONACYT scholarship No. 719288.

## References

K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*

*ics*, 22(14):e49–e57, 07 2006. ISSN 1367-4803. doi: <https://doi.org/10.1093/bioinformatics/btl242>. URL <https://academic.oup.com/bioinformatics/article-pdf/22/14/e49/616383/bt1242.pdf>.

T. Claassen and T. Heskes. Causal discovery in multiple models from different experiments. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23, pages 415–423. Curran Associates, Inc., 2010. URL <https://proceedings.neurips.cc/paper/2010/file/5807a685d1a9ab3b599035bc566ce2b9-Paper.pdf>.

G. Cooper, C. Cai, and X. Lu. Tumor-specific causal inference (TCI): A Bayesian method for identifying causative genome alterations within individual tumors. *bioRxiv*, page 225631, 2018. doi: <https://doi.org/10.1101/225631>.

J. L. Gamella and C. Heinze-Deml. Active invariant causal prediction: Experiment selection through stability. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 15464–15475. Curran Associates, Inc., 2020. URL <https://proceedings.neurips.cc/paper/2020/file/b197ffdef2ddc3308584dce7afa3661b-Paper.pdf>.

C. Grefkes and G. R. Fink. Connectivity-based approaches in stroke and recovery of function. *The Lancet Neurology*, 13(2):206–216, 2014. doi: [https://doi.org/10.1016/S1474-4422\(13\)70264-3](https://doi.org/10.1016/S1474-4422(13)70264-3).

A. Hauser and P. Bühlmann. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. *J. Mach. Learn. Res.*, 13(1):2409–2464, aug 2012. ISSN 1532-4435.

F. Jabbari, S. Visweswaran, and G. F. Cooper. Instance-specific Bayesian network structure learning. In V. Kratochvíl and M. Studený, editors, *Proceedings of the Ninth International Conference on Probabilistic Graphical Models*, volume 72 of *Proceedings of Machine Learning Research*, pages 169–180, Prague, Czech Republic, 11–14 Sep 2018. PMLR. URL <http://proceedings.mlr.press/v72/jabbari18a.html>.

H. Jia, Z. Wu, J. Chen, B. Chen, and S. Yao. Causal discovery with Bayesian networks inductive transfer. In *International Conference on Knowledge Science, Engineering and Management*, pages 351–361. Springer, 2018. doi: [https://doi.org/10.1007/978-3-319-99365-2\\_31](https://doi.org/10.1007/978-3-319-99365-2_31).

M. Kalisch, M. Mächler, D. Colombo, M. H. Maathuis, and P. Bühlmann. Causal inference using graphical models with the R package `pcalg`. *Journal of Statistical Software, Articles*, 47(11):1–26, 2012. ISSN 1548-7660. doi: <https://doi.org/10.18637/jss.v047.i11>. URL <https://www.jstatsoft.org/v047/i11>.

X. Li, S. Xie, P. McColgan, S. J. Tabrizi, R. I. Scahill, D. Zeng, and Y. Wang. Learning subject-specific directed acyclic graphs with mixed effects structural equation models from observational data. *Frontiers in Genetics*, 9, 2018. ISSN 1664-8021. doi: 10.3389/fgene.2018.00430. URL <https://www.frontiersin.org/article/10.3389/fgene.2018.00430>.

R. Luis, L. E. Sucar, and E. F. Morales. Inductive transfer for learning Bayesian networks. *Machine Learning*, 79(1-2):227–255, 2010. doi: <https://doi.org/10.1007/s10994-009-5160-4>.

S. Montero-Hernandez, F. Orihuela-Espina, and L. E. Sucar. Intervals of causal effects for learning causal graphical models. In V. Kratochvíl and M. Studený, editors, *Proceedings of the Ninth International Conference on Probabilistic Graphical Models*, volume 72 of *Proceedings of Machine Learning Research*, pages 296–307. PMLR, 11–14 Sep 2018. URL <https://proceedings.mlr.press/v72/montero-hernandez18a.html>.

J. M. Mooij, S. Magliacane, and T. Claassen. Joint causal inference from multiple contexts. *Journal of Machine Learning Research*, 21:17–123, 2020. URL <https://jmlr.org/papers/v21/17-123.html>.

S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010. doi: <https://doi.org/10.1109/TKDE.2009.191>.

J. Peters, P. Bühlmann, and N. Meinshausen. Causal inference using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 78(5):947–1012, 2016. doi: 10.1111/rssb.12167. URL <http://arxiv.org/abs/1501.01332>. (with discussion).

J. D. Ramsey, S. J. Hanson, C. Hanson, Y. O. Halchenko, R. A. Poldrack, and C. Glymour. Six problems for causal inference from fMRI. *Neuroimage*, 49(2):1545–1558, 2010. doi: <https://doi.org/10.1016/j.neuroimage.2009.08.065>.

V. Rodríguez-López and L. E. Sucar. Knowledge transfer for causal discovery. *International Journal of Approximate Reasoning*, 143:1–25, 2022. ISSN 0888-613X. doi: <https://doi.org/10.1016/j.ijar.2021.12.010>. URL <https://www.sciencedirect.com/science/article/pii/S0888613X21002073>.

M. Rojas-Carulla, B. Schölkopf, R. Turner, and J. Peters. Invariant models for causal transfer learning. *Journal of Machine Learning Research*, 19(1):1309–1342, Jan. 2018. ISSN 1532-4435.

M. Scutari. Learning Bayesian networks with the bnlearn R package. *Journal of Statistical Software, Articles*, 35(3):1–22, 2010. ISSN 1548-7660. doi: <https://doi.org/10.18637/jss.v035.i03>. URL <https://www.jstatsoft.org/v035/i03>.

R. Tillman and P. Spirtes. Learning equivalence classes of acyclic models with latent and selection variables from multiple datasets with overlapping variables. In G. Gordon, D. Dunson, and M. Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 3–15, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR. URL <http://proceedings.mlr.press/v15/tillman11a.html>.

S. Triantafillou and I. Tsamardinos. Constraint-based causal discovery from multiple interventions over overlapping variable sets. *Journal of Machine Learning Research*, 16(66):2147–2205, 2015. URL <http://jmlr.org/papers/v16/triantafillou15a.html>.