

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ**

**ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ**

**Ρουμπελάκη Άννα**

**Μεταπτυχιακή Φοιτήτρια**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

Επόπτης Μεταπτ. Εργασίας: Αναπλ. Καθηγητής Ι. Τσαμαρδίνος

**Τρίτη, 18/10/2016, 12:00**

**Αίθουσα B108, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

**" Συνέπεια αιτιακών σχέσεων κατά την εκμάθηση περιθωρίων αιτιακών γράφων με αλγορίθμους βασισμένους σε περιορισμούς "**

#### **ΠΕΡΙΛΗΨΗ**

Οι Μέγιστοι Προγονικοί Γράφοι είναι πιθανότατα γραφικά μοντέλα που μπορούν να αντιπροσωπεύσουν αιτιακές σχέσεις ανάμεσα σε ομάδες μετρήσιμων μεταβλητών υπό την παρουσία λανθανόντων παραγόντων. Ο αλγόριθμος FCI είναι ευρέως χρησιμοποιούμενος για την ανάκτηση, μέσα από σύνολα δεδομένων, αμετάβλητων χαρακτηριστικών ανάμεσα σε ζεύγη μεταβλητών, τα οποία ανήκουν σε μια κλάση Markov ισοδύναμων Μέγιστων Προγονικών Γράφων.

Ερευνούμε τη συνοχή των αιτιακών χαρακτηριστικών που λαμβάνουμε από τον FCI και των παραλλαγών του ανάμεσα σε διάφορα οριακά σύνολα δεδομένων. Υπό τέλεια στατιστική γνώση, οι αιτιακές σχέσεις που λαμβάνουμε από τον αλγόριθμο σε οριακά σύνολα δεδομένων δεν θα έπρεπε να συγκρούονται με τις αιτιακές σχέσεις που λαμβάνουμε όταν τον εφαρμόζουμε σε ολόκληρο το σύνολο των δεδομένων. Παρ' όλα αυτά, ο FCI είναι επιρρεπής σε στατιστικά λάθη και σε διάδοση σφαλμάτων. Συνεπώς, τα αποτελέσματα του αλγορίθμου σε διαφορετικά οριακά σύνολα δεδομένων να είναι αντικρουόμενα.

Για να μετρήσουμε την αιτιακή συνοχή ανάμεσα σε οριακές κατανομές δεδομένων, χρησιμοποιήσαμε μια νέα μέθοδο η οποία βρίσκει όλες τις αμετάβλητες αιτιακές σχέσεις που υπάρχουν ανάμεσα στις μετρημένες μεταβλητές. Επίσης, παρουσιάζουμε έναν αλγόριθμο για την ταξινόμηση αιτιακών σχέσεων. Συγκρίναμε με τη μέθοδο bootstrapping και μελετήσαμε την επίδραση του συνδυασμού των δύο μεθόδων. Μελετάμε το πρόβλημα της αιτιακής συνοχής σε περιθώριες κατανομές υπό διάφορες συνθήκες, ακόμα και σε μη ιδανικές για τον αλγόριθμο.

**Roubelaki Anna**

**M.Sc. Thesis**

**Computer Science Department**

**University of Crete**

**Master's Thesis Supervisor: Associate Professor I. Tsamardinos**

**Tuesday, 18/10/2016, 12:00**

**Room B108, Computer Science dept., University of Crete**

**“Marginal Causal Consistency in Constraint-based Causal Learning”**

#### **ABSTRACT**

Maximal ancestral graphs are probabilistic graphical models that can represent the causal relationships among a set of measured variables in the presence of latent confounders. The Fast Causal Inference algorithm is broadly used to retrieve invariant pairwise features of a class of Markov Equivalent MAGs from observational data sets.

We investigate the consistency of causal features obtained by FCI and its variations in different marginals of a data set. Under perfect statistical knowledge, the causal relationships obtained by the algorithm when applied on marginal data sets should not conflict the causal relationships obtained when it is applied on all the measured variables. However, in practice FCI is prone to statistical errors and error propagation. As a result, the output of FCI in different marginals may be conflicting.

In order to measure causal consistency among marginal data sets, we employed a novel method to identify all invariant causal relationships that exist among the measured variables. We also introduce an algorithm for ranking causal relationships. We compared with bootstrapping and studied the effect of combining both scores. We studied the problem of marginal causal consistency in various settings, and even in cases of non-ideal scenarios for the algorithm.