**3.3 Causal Decision Making**

Dr Cheng Zhang (Microsoft Research)

<https://www.microsoft.com/en-us/research/people/chezha/>

<http://www.dagitty.net/>

**Casual machine learning**

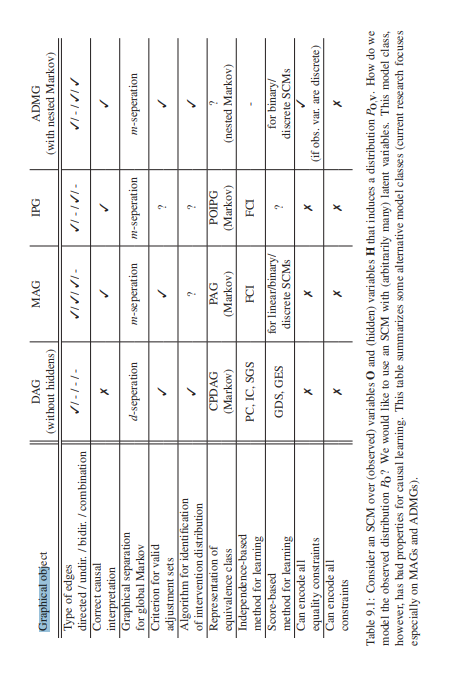
* Key of decision making: “what if?”
* Understand the casual relationship: Causal Discovery
* Understanding the impact of the actions: Casual Inference

**If there is a (Partial) Causal Graph?**

* Do-calculus (Pearl)
* Potential outcome (Rubin)

**Causal discovery**

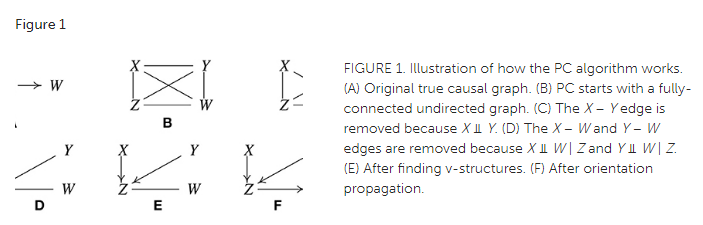
* Tasks: relationship between variables
* Directed Acyclic Graph
  + G = { V, E}
  + V = a set of variables
  + E = a set of edge indicating casual relationship
* D-blocked paths
  + Given a path *t*, and a set of nodes *S,*
    - S d-blocks the path *t* if *t* contains:
      * A non-collider which is in S
      * A collider which is not an ancestor of S
* D-seperation
  + <http://bayes.cs.ucla.edu/BOOK-2K/d-sep.html#:~:text=d%2Dseparation%20is%20a%20criterion,ness%22%20or%20%22separation%22>.
* Markov condition/ assumption
  + Causal graph -> data distribution
* Faithfulness assumption
  + Causal graph <- data distribution
  + Paths cannot cancel out
* Markov equivalence of graph
  + A set of DAGs that are Markov equivalent
  + Completed partial DAG (CPDAG)



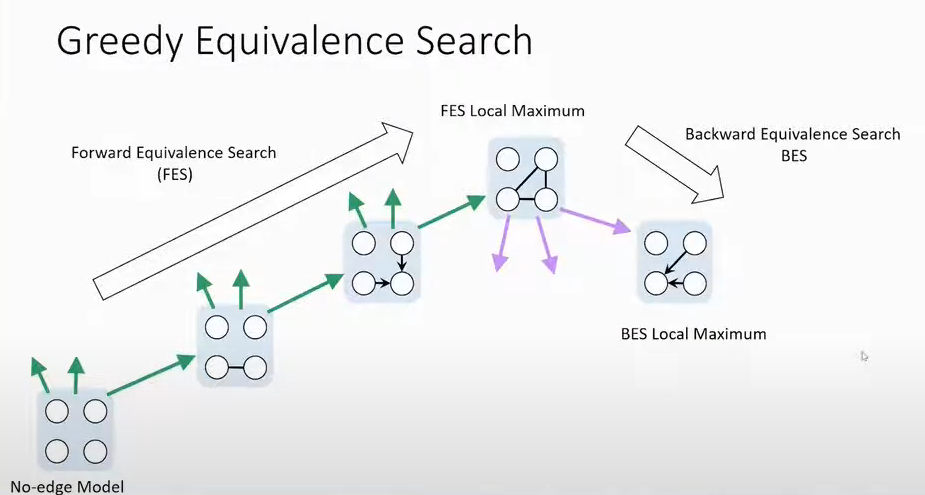
* + - <https://library.oapen.org/bitstream/handle/20.500.12657/26040/11283.pdf?sequ>

**Three types of methods**

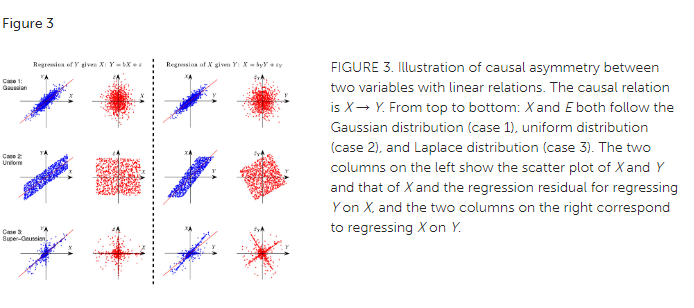
* **Review of Causal Discovery Methods Based on Graphical Models**
  + <https://www.frontiersin.org/articles/10.3389/fgene.2019.00524/full>
* Constraint-based
  + PC (Peter Spirtes; Clark Glymour)
    - Assumptions:
      * Markov assumption
      * Faithfulness assumption Acyclicity
      * Causal sufficinecy



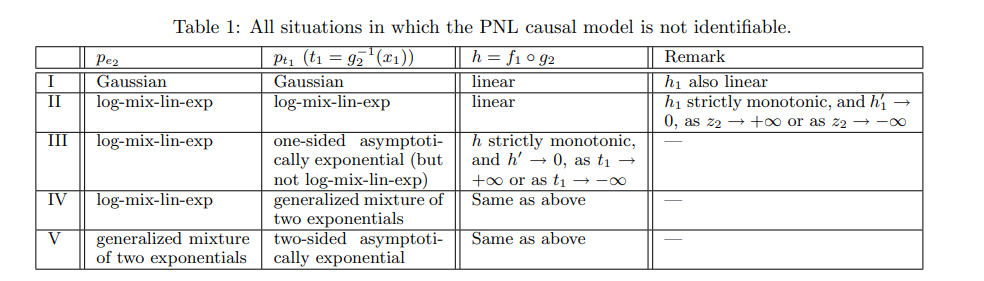
* From: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00524/full>
* PC returns CPDAG
  + Uncertainty
* FCI (Fast casual inference)
* Beyond PC:
  + Relax other assumptions: allow cycles
  + Other difficult situations: e.g. missing data (MVPC, MVFCI)
  + Use any other orientation method (function based etc.)
  + Use any other constrains (pattern-based method)
* Score-based
  + Greedy Equivalent Search (GES)
    - Find graphs best fitting the data
    - BIC
    - Naive score - based methods
    - Super exponential number of graphs with number of data
    - GES
    - **Statistically Efficient Greedy Equivalence Search**
    - <http://proceedings.mlr.press/v124/chickering20a/chickering20a.pdf>



* Functional Causal Models
  + LinGAM (Linear non-gaussian model)
    - Instead of , 



* + - Find out X causing Y rather than Y causing X
    - Linear Gaussian model is not identifiable
* Beyond LiNGAM: PNL (Post-linear)
  + **On the Identifiability of the Post-Nonlinear Causal Model**
    - <https://arxiv.org/ftp/arxiv/papers/1205/1205.2599.pdf>



* Multiple variables: ICA-LINGAM
  + The cause and noise should be independent
    - In matrix form *X = BX + E*
      * *E = (I - B)X*
      * Learn B through ICA
      * *Z = WX*
    - Use Z as E above and we just need to make W have lower triangular shape
    - Linear, Non-Gaussian, Acyclic

**Continuous optimization**

* **Dags with no tears:**
  + <https://arxiv.org/pdf/1803.01422.pdf>
* Functional Causal Models and Score-based
* Adding DAG constrain

**Causal Discovery + Causal Inference**

* Separate development
  + **Minimal Enumeration of All Possible Total Effects in a Markov Equivalence Class**
    - <http://proceedings.mlr.press/v130/guo21c/guo21c.pdf>
* Different assumptions

**Deep end-to-end casual inference**

* **Deep End-to-end Causal Inference**
  + <https://arxiv.org/pdf/2202.02195.pdf>
* Normalizing flow and GNN in deep generative model -> graph and function

**Ongoing research**

* Improved accuracy:
  + Incomplete information
  + accurate casual discovery
* Time-series data adaptation
  + Time dependence noise form
  + Multimodal data

**Scalable Causal AI**

* Casuality:
  + Lots of theory but no impact
* Deep Learning:
  + Great algorithm and impact but not correlation based
* With scalable casual AI