NASA DAG ML Validation Project Report

Jasper Doan
Donald Bren School of Information and Computer Sciences
University of California, Irvine
Irvine, CA 92697
jasperd1@uci.edu

September 14, 2023

ABSTRACT

Project consists of developing an approach to validate expert-drawn graphs by building probabilistic graph model Directed Acyclic Graphs, through populating Machine-Learning with empirical biological data from the NASA Open Science Data Repository. Tasked with researching IAMB, Fast-IAMB, Inter-IAMB, and IAMB-FDR algorithms for medical risk assessment to let NASA HSRB formalize a shared causal flow of risk model among Risk Board stakeholders. Though the 4 assigned algorithms' results were inconsistent with what NASA scientist thought the relationship should be.

1 Introduction

The goal of our project is to validate expert-drawn Directed Acyclic Graphs (DAGs) for Human Spaceflight Risks, for tracking and researching risks that astronaut crews face during spaceflight. This paper goes over the process of using Bnlearn's IAMB, Fast-IAMB, Inter-IAMB, and IAMB-FDR algorithm DAG generation to formalize a shared causal flow of risk model.

Directed Acyclic Graphs (DAG) DAGs are network maps which have unidirectional arrows (directed) and do not allow feedback loops (acyclic).

In the context of the The Human System Risk Board (HSRB), DAGs are used to represent the chain of events that lead from spaceflight exposures to negative mission-level outcomes. This enables two immediate uses as well as sets the stage for further evolution of the causal networks as tools of inference.

Why DAG? Enable mathematical analysis of the relationships between factors and can potentially assess the strength of influence if quantitative values are assigned to nodes and edges.
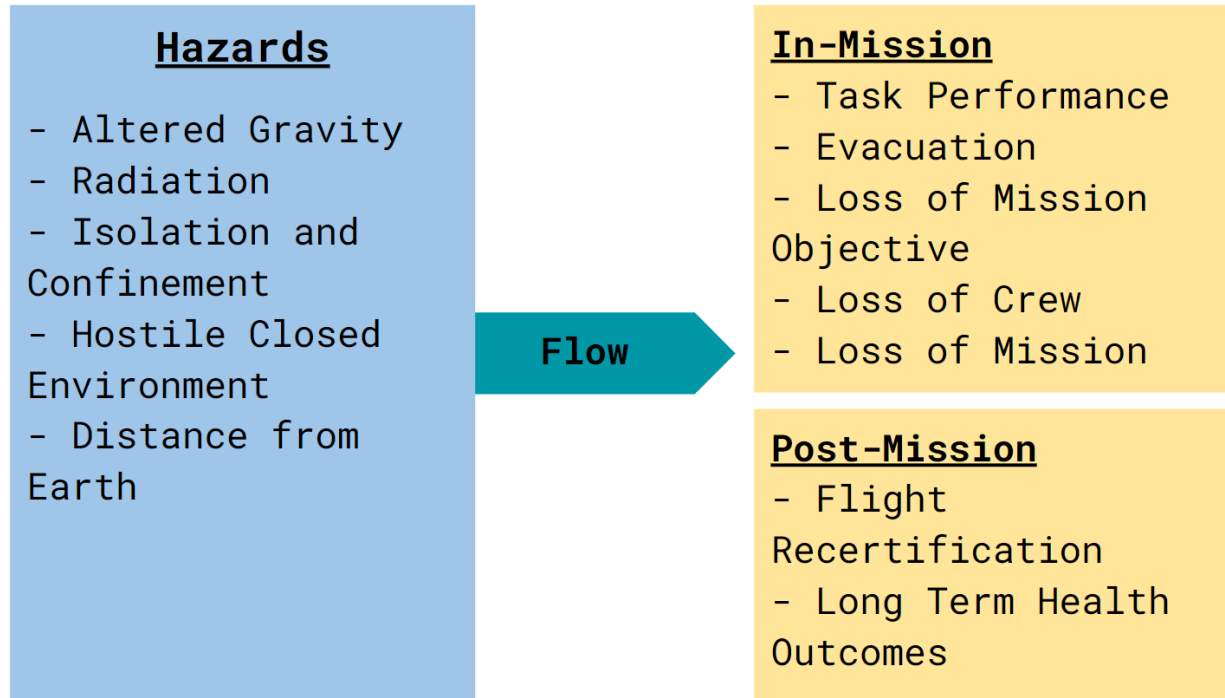
Subject to challenge and improvement based on evolving evidence. If new evidence suggests a lack of causal connection, the corresponding connection can be removed. It can aid in conveying high-level and aggregated concepts that link key components of causal flow to downstream effects in the risk domain. They facilitate communication and the development of shared mental models during board or stakeholder meetings.

DAGs for communication of complex human spaceflight risks Limits the provision of in-mission support capabilities and resources, while simultaneously increasing the need for them

Limits on these capabilities and resources stem from constraints on mass, volume, power, and data bandwidth allocations available to the vehicle's systems/habitats used by astronauts; the further a mission takes astronauts from Earth, the greater these constraints and thus the less support capability they will have. The need for capabilities and resources is increased because the further a mission goes from Earth, the longer astronauts are exposed to degradation by the spaceflight environment.

DAGs aid in prioritizing research and development Evaluation of Human System Risks is necessary to prioritize the allocation of limited research, surveillance, and technology development resources.

The previous scoring system (Red, Yellow, Green) did not consider the complex interactions and synergies between risks, which can amplify risks in other body systems or at a later time. Directed acyclic graphs (DAGs) help analyze the structure of risks and identify important factors in the causal network. Nodes in the DAG represent factors that have many effects, bridge or join risks together, or exist in the middle of the action. DAG analysis provides insights into the interdependencies and cumulative effects of risks faced by astronauts during missions.

Hazards - Altered Gravity - Radiation - Isolation and Confinement - Hostile Closed Environment - Distance from Earth	**Flow** ▶	**In-Mission** - Task Performance - Evacuation - Loss of Mission Objective - Loss of Crew - Loss of Mission **Post-Mission** - Flight Recertification - Long Term Health Outcomes

DAG Validation   In the context of NASA's Human System Risk Management Process, DAG validation can be used as Level of Evidence (LoE) assessment, which is a process used by Risk Custodial Teams at NASA. LoE assessment involves evaluating the strength of evidence for each individual arrow (causal relationship) in a DAG. This assessment is based on guidelines initially defined by epidemiologist Sir A. Bradford Hill in 1965. To assign an LoE rating, one needs to carefully review and synthesize published and unpublished data relevant to the causal relationships.

The key difference between LoE assessment and DAG validation is that LoE assessment focuses on determining how strong the evidence is for a single causal connection, while DAG validation looks at whether the overall causal system represented by the DAG is correct. Sometimes, there can be discrepancies between these two approaches if important causal relationships are missing in the DAG.

This can be further emphasized that DAG validation should occur after at least a preliminary LoE assessment, hence the goal of this project. And both assessments should be updated when new evidence becomes available. Additionally, the author notes that although testable implications may be validated, further evidence is required because the studies used were conducted on rodents and not humans. It is to note that specific publications, studies, and datasets used in the assessment are initially considered weak evidence for human system risks, but they can be assigned a higher LoE if they meet certain quality criteria.

We acknowledged that a formal evaluation of QoE and LoE has not been performed in their study due to resource constraints. However, they suggest that future research should focus on high-quality evidence to improve the assessment of DAG validity. But nonetheless we should explore the possibility of applying LoE ratings to entire DAGs or subsets of DAGs, which could help identify knowledge gaps and provide better communication of critical information related to human system risk in spaceflight missions.

This is the analysis of a DAG done by experts (figure below), it reveals certain relationships between variables based on the flow of causation depicted in the diagram. In this DAG, it's expected that all variables connected by arrows should have non-zero marginal correlations with each other. Here are some pairs of variables that should be marginally correlated, along with the expected direction of correlation (positive or negative):

- Skeletal unloading and bone formation: There should be a negative correlation between these two variables. In other words, when skeletal unloading occurs, we expect to see a decrease in bone formation.

- Skeletal unloading and bone resorption: There should be a positive correlation between these two variables. When skeletal unloading happens, it should lead to an increase in bone resorption.

- Bone formation and bone mass: There should be a positive correlation between these two variables. When bone formation increases, we expect to see an increase in bone mass.

- Bone resorption and bone mass: There should be a negative correlation between these two variables. An increase in bone resorption should result in a decrease in bone mass.

- Bone mass and trabecular microarchitecture: There should be a positive correlation between these two variables. Higher bone mass is expected to be associated with a denser trabecular microarchitecture.

- Bone mass and bone strength: There should be a positive correlation between these two variables. A greater bone mass should lead to higher bone strength.
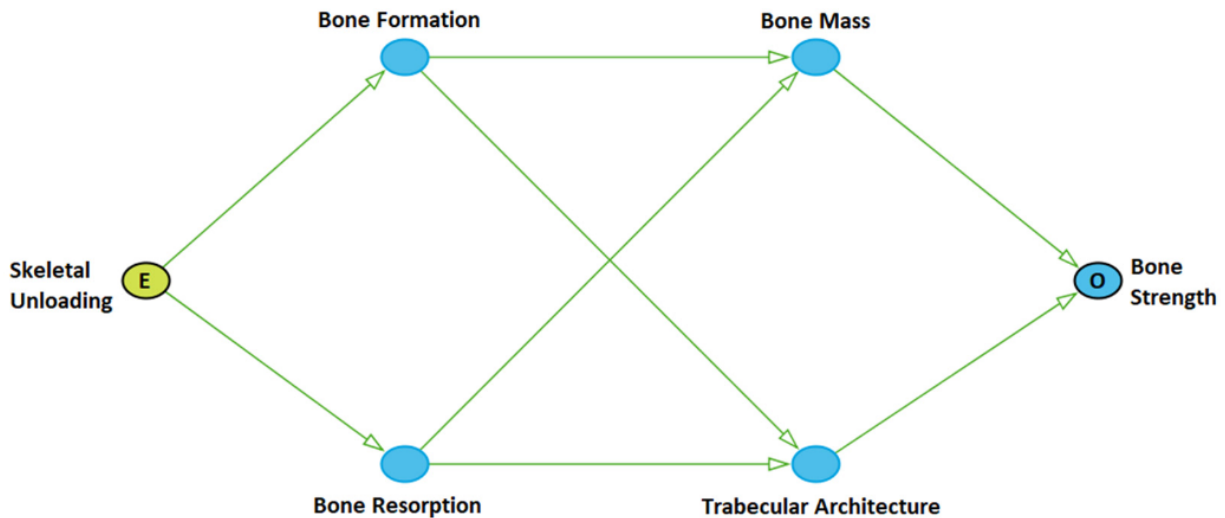
2 Theory

Bayesian networks   You can represent these relationships between variables by building a Bayesian networks. Where it shows the representation of how nodes/variables interact with each other (direct dependencies between variables) → Allows the user to affirm and make accurate predictions based on observed data.

For instance, if we observe that the weather is rainy, we can use the Bayesian network to estimate the probability of carrying an umbrella and the probability of the ground being wet. We can also do the reverse: if we know the ground is wet, we can estimate the probability that it is raining. By using probabilities and the relationships encoded in the graph, Bayesian networks allow us to reason and make inferences about the variables even when we have incomplete or uncertain information.

Constraint-based methods (What IAMB Variations do)   Discover the dependencies and relationships between nodes based on data by imposing certain constraints. Identify statistical dependencies between nodes → infer the underlying structure of the Bayesian network

- Independence Testing: Determines the statistical independence or dependence between pairs of variables in the data. Test whether the variables are conditionally independent given other variables. If two variables are found to be independent, it suggests that there is no direct edge between them in the Bayesian network.

- Skeleton Discovery: Constructs a skeleton or an initial structure for the Bayesian network. Represents the presence or absence of edges between variables. Edges are added to the graph for variables that are found to be dependent, indicating a potential causal relationship.

- Orientation of Edges: Determine the orientation of the edges in the network. Establish the direction of causality between variables. Examining conditional dependencies and using additional tests or heuristics to infer the most likely direction of causal influence.

Markov Blanket   The Markov blanket of a variable in a Bayesian network: Minimal set of variables that contains all the information necessary to predict the variable's value, given the values of other variables in the network.

Formally, the Markov blanket of a variable X in a Bayesian network consists of three sets of variables:

- Parents of X: These are the variables that directly influence the value of X. If A is influenced by B, then B would be a parent of A.

- Children of X: These are the variables that are directly influenced by the value of X. If C is influenced by A, then C would be a child of A.

- Parents of X's children: These are the variables that are parents of X's children. If C is influenced by B, then B would be a parent of C's child (C) and would be included in the Markov blanket of A.

Simpler terms: Imagine you have three variables: Cloudy, Rain, Wet Grass. From what you know (observed data), they do influence each other's decisions. Rain's Markov blanket consists of the variables that directly influence it. It will form a special group that has all the information needed to understand how rain will happen.
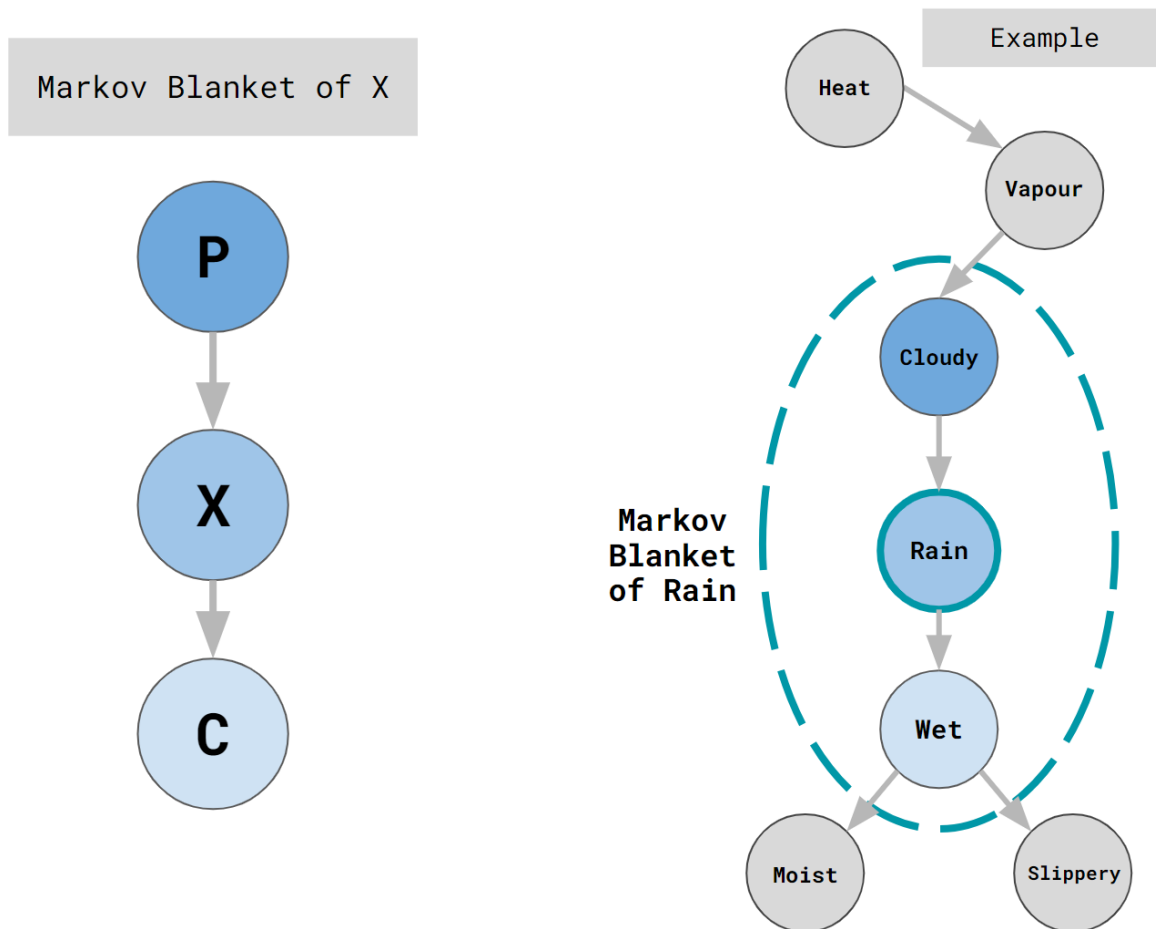
When considering Rain's Markov blanket, we should identify its core components. Firstly, it comprises the node "Cloudy" due to a consistent observation over 21 years: when it's cloudy, there's a high likelihood of rain. Secondly, it encompasses nodes directly affected by rain, such as "wet grass," "wet floor," and "cold."

Think of Rain's Markov blanket as a small group of people crucial to her decision-making. Understanding what these individuals are up to can help make reasonably accurate predictions about Rain's likely actions.

In a Bayesian network, Markov blanket (MB) of a variable is the group of variables that have a direct influence on it or are directly influenced by it. It's the minimal set of variables that you need to pay attention to in order to understand and predict the behavior of that variable, without worrying about all the other variables in the network.

The Markov blanket concept helps us simplify things by focusing on a smaller, important group rather than considering the entire network. It allows us to make predictions or perform calculations about a specific variable by looking only at the variables in its Markov blanket.

Generally the process of finding the Markov blanket of a variable is called Markov blanket discovery. It is a key step in many algorithms for learning the structure of a Bayesian network from data. And its Pseudo-code is as follows:

```
# 1. Preprocess the data: Handle data cleaning, handling missing values
data = preprocess_data(data)

# 2. Define the set of variables and their relationships
variables = get_variables_from_data(data)
causal_relationships = define_causal_relationships()

# 3. Initialize the Markov Blanket for each variable
markov_blanket = {}
for variable in variables:
    markov_blanket[variable] = find_markov_blanket(variable, data)

# 4. Learn causal relationships from the Markov Blanket
bayesian_network = create_empty_bayesian_network()
for variable in variables:
    parents = find_parents(variable, markov_blanket[variable])
    for parent in parents:
        bayesian_network.add_edge(parent, variable)

# 5. Output the Bayesian Network
bayesian_network.to_graph()

```

3 Algorithm Description

3.1 IAMB | Incremental Association Markov Blanket

Consists of two phases, a forward and a backward one. An estimate or copy of the Markov Blanket of a variable of interest T is kept in the MB. In the forward phase all variables that belong in $MB(T)$, including false positives enter CMB (Copy of MB) while in the backward phase the false positives are identified and removed so that $CMB = MB(T)$ in the end.

Phase I is the following: start with an empty candidate set for the CMB and admit into it (in the next iteration) the variable that maximizes a heuristic function $f(X;T|CMB)$. This is a measure of association between X and T given CMB. In backward conditioning (Phase II) we remove one-by-one the features that do not belong to the $MB(T)$ by testing whether a feature X from CMB is independent of T given the remaining CMB.
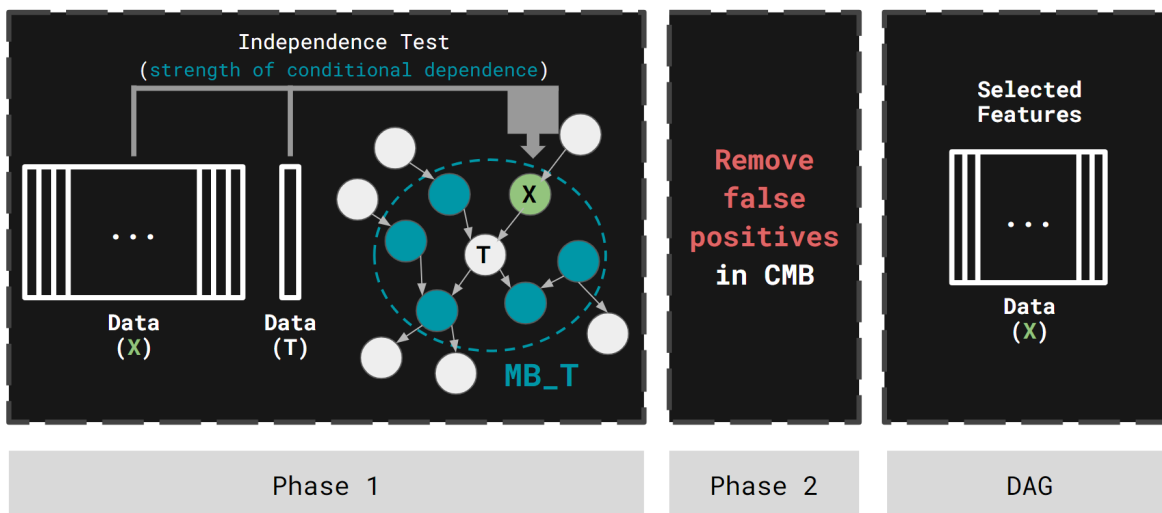
3.2 Fast-IAMB | Fast-Incremental Association Markov Blanket

Considered to be an improvement over the IAMB algorithm that aims to reduce the computational complexity. It aims to reduce the computational and time complexity of the algorithm by pruning unnecessary tests. Maintaining an order of the variables based on their likelihood of being in the Markov blanket. Uses a combination of forward and backward steps to efficiently determine the Markov blanket for each variable.

Instead of relying on traditional statistical tests like chi-squared (kai-squared). It uses a combination of forward and backward steps to determine the MB for each variable. Computationally this is just $O(n)$ while IAMB is $O(2n)$ where it needs go forward and backward, here it does it in 1 loop.
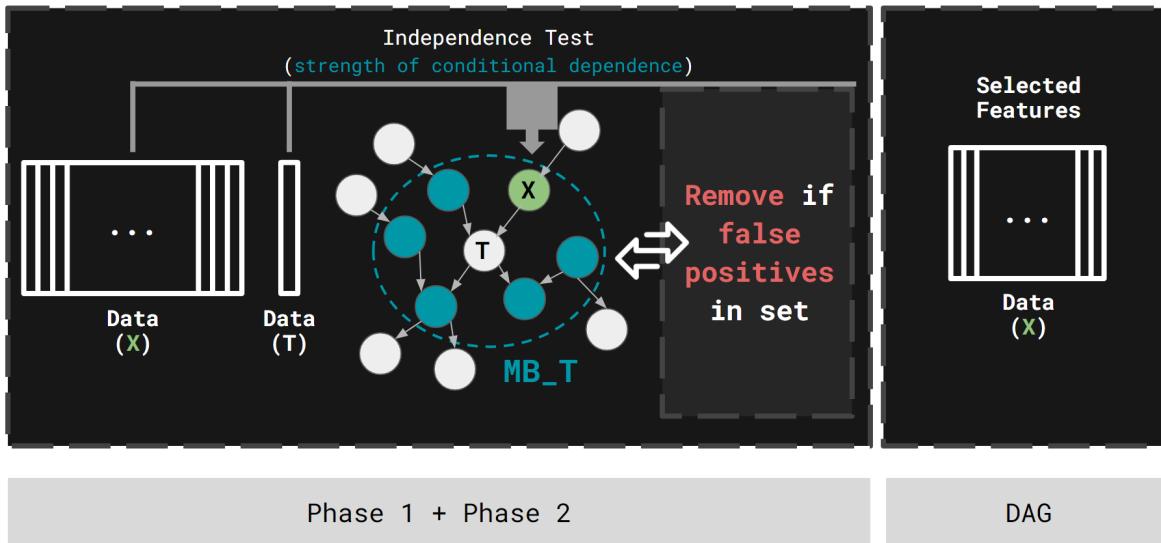
3.3 Inter-IAMB | Interleaved Incremental Association Markov Blanket

Extension of the IAMB algorithm that introduces an interleaved strategy for constructing the Markov blanket. Instead of iterating and processing one target variable at a time, it selects multiple target variables from the dataset and performs interleaved passes over the variables, gradually refining the Markov blankets. Allows for more efficient identification of variable dependencies and can lead to improved results in certain scenarios.

For each selected target variable T, it just perform the regular IAMB algorithm based on independence tests. But after a short duration of processing, it switch to another target variable and continue the process. The switching between target variables is the "interleaving" part of the algorithm. By interleaving the runs, the algorithm can potentially reduce the total number of independence tests required and, thus, improve computational efficiency.

6
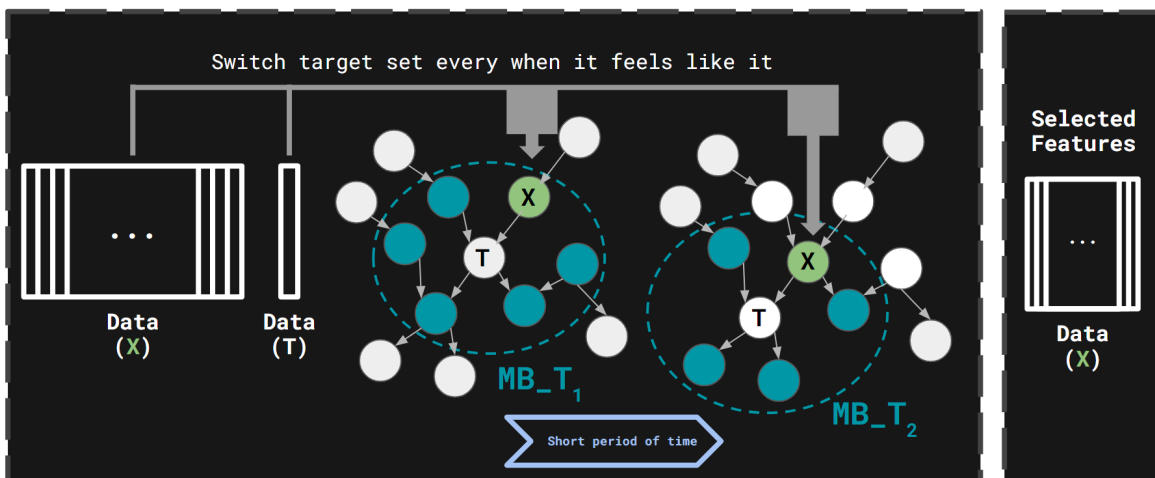
3.4 IAMB-FDR | Incremental Association Markov Blanket with False Discovery Rate Control

Variant of the IAMB algorithm that incorporates false discovery rate (FDR) control to address the issue of multiple hypothesis testing. In structure learning, multiple statistical tests are performed, and without proper control, this can lead to an increased chance of false discoveries. IAMB-FDR applies FDR control techniques to adjust the statistical significance thresholds used in independence tests, thereby mitigating the risk of false discoveries and improving the reliability of the learned structure.
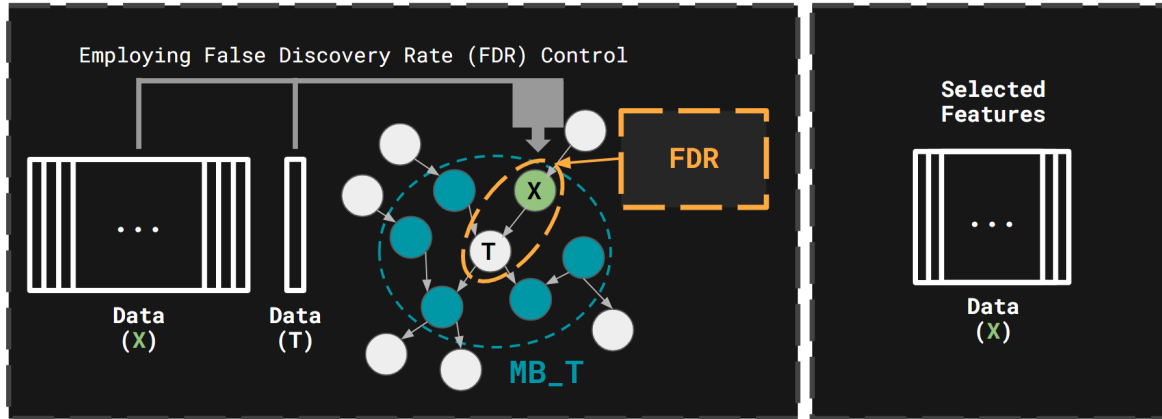
4 Datasets and Results

GLDS-366 | Coalescence of DNA double strand breaks induced by galactic cosmic radiation is modulated by genetics in 15 inbred strains of mice The study investigates how different mouse strains respond to deep space radiation and explores the role of DNA damage markers in assessing radiation toxicity. It also explores the genetic factors influencing these responses using genome-wide association studies.
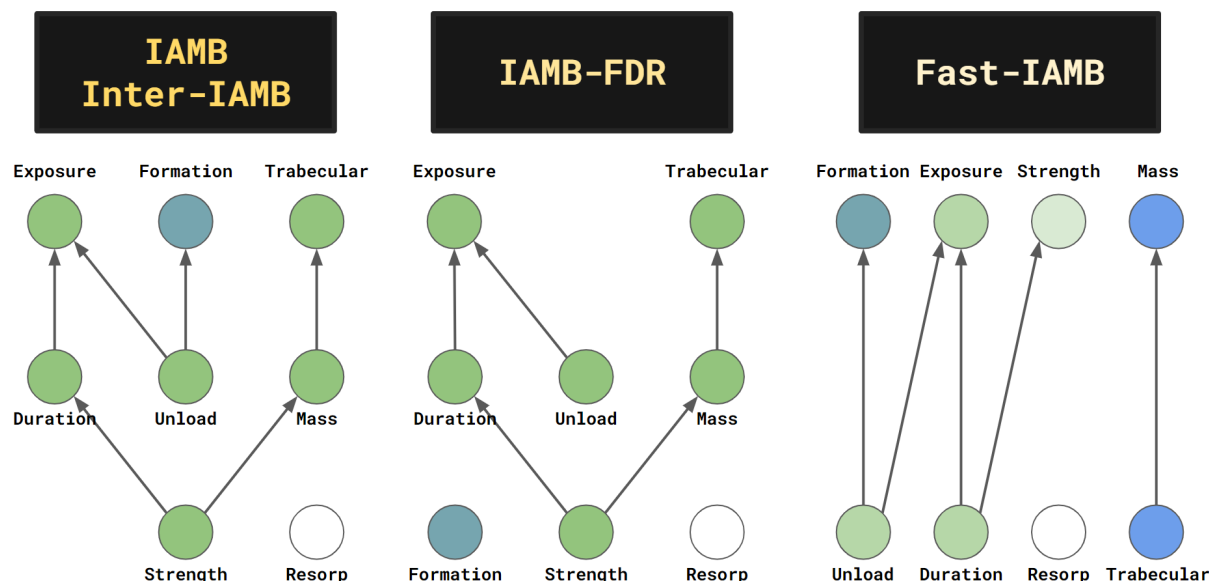
Examination: Investigate the variations in DNA double strand break coalescence induced by galactic cosmic radiation across autosome, mitochondrial, and sex chromosome regions in 15 inbred mouse strains, determining whether there are any genetic factors influencing radiation-induced DSBs on these different genetic elements:
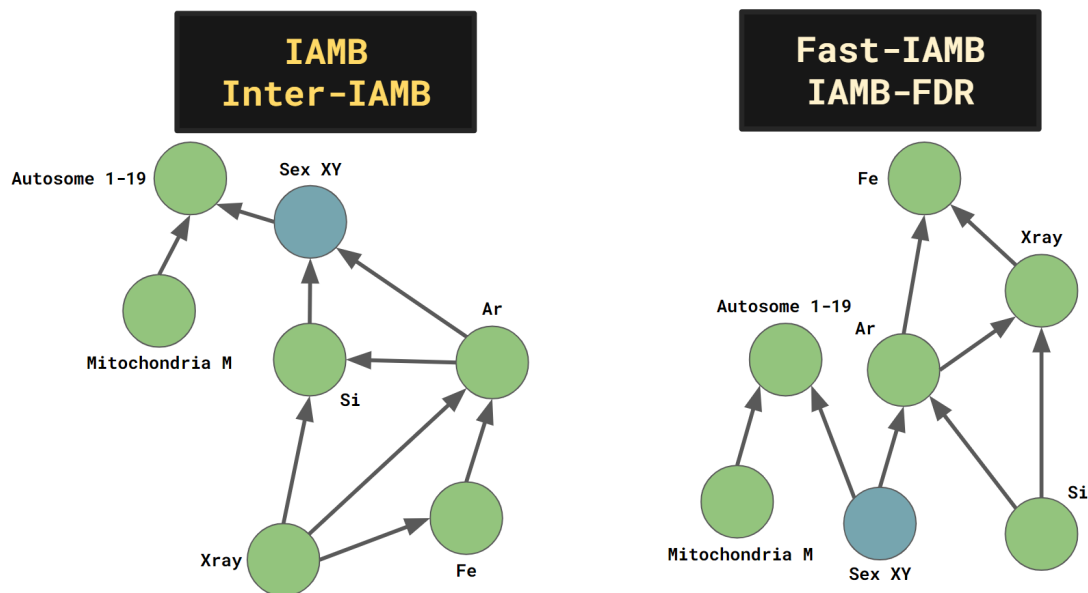
OSD-477 and OSD-608 | Dose-dependent skeletal deficits due to varied reductions in mechanical loading in rats (Femur - microCT, three-point bending, histomorphometry) (Tibia - pQCT) OSD-477 demonstrates that this rat model of partial weight bearing leads to progressive deterioration of trabecular bone density, which is proportional to the

degree of partial unloading. The study mainly focuses on results obtained from mechanical testing, bone microstructure analysis using microCT, and bone histomorphometry assays using femur tissue. OSD-608 demonstrates that partial weight bearing in rats leads to a progressive deterioration of trabecular bone (the spongy, inner part of bone) that is proportionate to the degree of weight reduction. Importantly, this model had limited effects on cortical bone. The researchers used pQCT to gather these findings, primarily focusing on the tibia as their primary area of interest.

Examination: Investigate the relationship between mechanical unloading, bone mass, trabecular architecture, and strength in rats to gain insights into bone resorption and formation processes under different loading conditions:

IAMB Inter-IAMB

Exposure · Formation · Trabecular · Duration · Unload · Mass · Strength · Resorp

IAMB-FDR

Exposure · Trabecular · Duration · Unload · Mass · Formation · Strength · Resorp

Fast-IAMB

Formation · Exposure · Strength · Mass · Unload · Duration · Resorp · Trabecular

OSD-310 | Spaceflight-induced (STS-62) vertebral bone loss in ovariectomized rats is associated with increased bone marrow adiposity and no change in bone formation This study sheds light on the complex relationship between bone marrow adipocytes and bone loss during spaceflight, particularly in a context of hormonal deficiency. It suggests that increased MAT may be associated with bone loss primarily through enhanced bone resorption, rather than direct inhibition of osteoblast activity.

Examination: Look at the impact of microgravity exposure on bone mass, resorption, and formation in ovariectomized rats, analyzing the mechanisms driving vertebral bone loss in the context of hormonal deficiency:

IAMB Fast-IAMB Inter-IAMB

Exposure · Formation · Resorp · Mass

IAMB-FDR

Exposure · Formation · Resorp · Mass
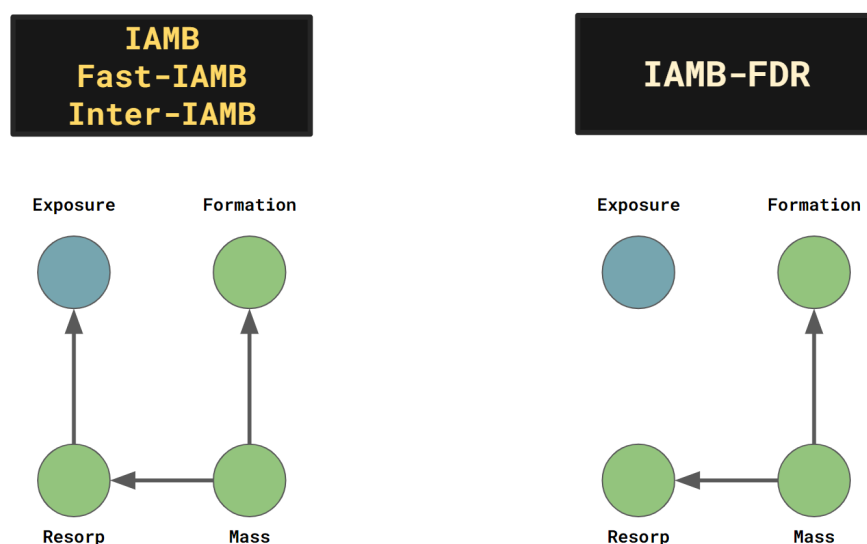
OSD-489 | Quantifying Cancellous Bone Structural Changes in Microgravity: Axial Skeleton Results from the RR-1 Mission   The study aimed to confirm that bone degradation in microgravity predominantly affects specific types of bones, especially those not subjected to weight-bearing forces. It also sought to deepen our understanding of how mechanical factors influence the dynamic process of bone remodeling. The data for this study were collected from rodent lumbar 4 bones during a space mission and were analyzed using micro-computed tomography techniques.

   Examination: See if exposure to microgravity affects the microarchitecture and trabecular architecture, and whether mass of subject plays a vital role:
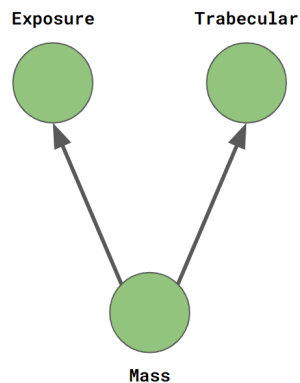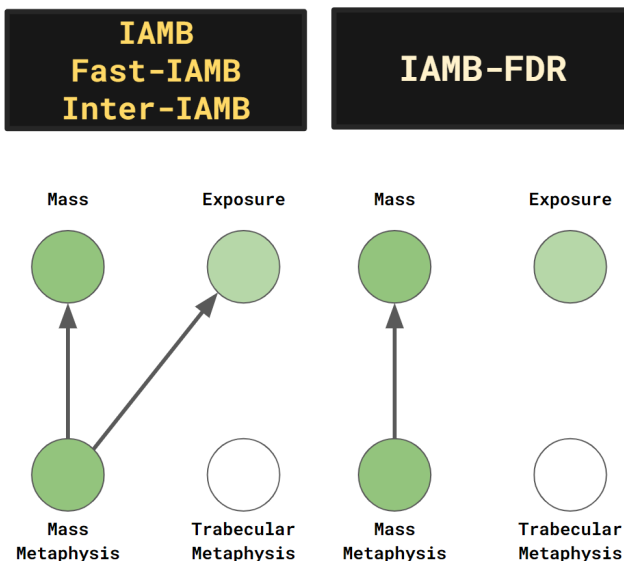

OSD-351 | Effects of Spaceflight on Bone Microarchitecture in the Axial and Appendicular Skeleton in Growing Ovariectomized Rats from STS-62   This study revealed that the impact of spaceflight on bone microarchitecture in ovariectomized rats was bone- and bone compartment-specific. Interestingly, the effects were not solely related to whether the bones were weight-bearing or non-weight-bearing. The study used micro-computed tomography (mCT) assays to obtain detailed insights into bone structure and health.

   Examination: Look at the impact of spaceflight on bone mass and trabecular architecture in various bone compartments, analyzing how weight-bearing and non-weight-bearing bones respond to microgravity conditions:
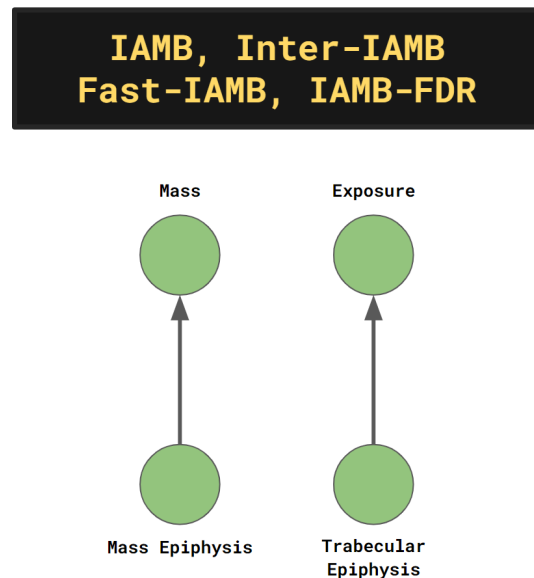
5 Analysis and Discussion

From first glance, seems like IAMB and its variations are not good candidates for validating expert's DAG. Results shows that IAMB failed to capture the right relationships between variables, and is inconsistent with the expert's DAG. Perhaps another algorithm should be used instead, or maybe the data is not large enough to correctly infer the relationships between variables.

Nonetheless, it is to note that building causal models is hard, as causal discovery algorithms are used to infer causal relationships from observational and simulated data. Meaning:

- There's no ground truth.
- No domain knowledge or prior information was taken into consideration, relied solely on observed data.
- Simulations only reflect common preconceptions.
- Built on assumptions such as faithfulness or independence principles.

As such, in the first place we needed a method for falsifying the output of a causal discovery algorithm in the absence of ground truth. Even if these algorithms produces reliable and compatiable DAG results to the expert-drawn graph. Passing such compatibility test makes way for a necessary criterion for good performance, and provides strong evidence for how how variables co-occur, increasing confidence of the causal models.

Whatever the premise of the algorithm's design or operation is based on, it might work under certain assumptions, these assumptions might not always hold true in others. Reliability and effectiveness are subjects of uncertainty and debate. I suggest we find a way to quantify the level of incompatibility of the outputs of causal discovery. As it will serve as a proxy for measures like Structural Hamming distance (SHD) which require access to ground truth.

Since the true causal relationships are often unknown, this distance can't be directly calculated. The compatibility score could provide a more accessible and meaningful metric to assess the performance of the algorithm. You could achieve this by looking at the algorithm application to different subsets of variables are.

For example, if you have a dataset with variables A, B, C, and D, you might run the algorithm on subsets like $\{A, B\}$, $\{A, C\}$, $\{A, D\}$, $\{B, C\}$, and so on. Trying to find causal relationships between different pairs of variables or sets of variables. Instead of just saying it is completely compatible or not (yes or no, 1 or 0, binary).

This score could provide a more nuanced measure of compatibility by considering how much the algorithm's outputs align across different subsets. If the algorithm's outputs are inconsistent when applied to different subsets, it suggests a lack of compatibility. This inconsistency could indicate that the algorithm's results might not be reliable in which we can discredit the algorithm. But then again, we wouldn't be able to guarantee that a causal discovery algorithm with a high self-compatibility score will accurately predict system behavior.

6 Conclusion

I still have no idea what I'm doing and what this project is about

References

[1] Erik A; Avalon M; Jacqueline C; Erin C; Robert R; Ahmed A; Kristina M; Charlotte B. Directed Acyclic Graphs: A Tool for Understanding the NASA Human Spaceflight System Risks - Human System Risk Board. In *NTRS - NASA Technical Reports Server, 1 Oct. 2022*, ntrs.nasa.gov/citations/20220015709.

[2] Erik A. Human System Risk Management Plan. In *NTRS - NASA Technical Reports Server, 1 Oct. 2022*, https://ntrs.nasa.gov/citations/20205008887.

[3] Erik A; Robert R; Avalon M; Jacqueline C; Devan P; Erin C; Mark S; Mary B; Ahmed A; Kristina M; Daniel B; Wilma A; Charlotte Brown. Directed Acyclic Graph Guidance Documentation. In *NTRS - NASA Technical Reports Server, 1 Jun. 2022*, https://ntrs.nasa.gov/citations/20220006812.

[4] Robert R, Ryan S, Russell T, Urszula I, Mary B, Lauren S, Erik A. Validating Causal Diagrams of Human Health Risks for Spaceflight: An Example Using Bone Data from Rodents. In *National Library of Medicine, 5 Sep. 2022*, https://pubmed.ncbi.nlm.nih.gov/36140288/.

[5] Costes S; Karpen G; Cekanaviciute E; Pariset E Coalescence of DNA double strand breaks induced by galactic cosmic radiation is modulated by genetics in 15 inbred strains of mice Pub *NASA Open Science Data Repository*, 2021, doi: 10.26030/v8w4-rg83

[6] Rutkove S; Bouxsein M; Ko F; Mortreux M; Riveros D; Nagy J Dose-dependent skeletal deficits due to varied reductions in mechanical loading in rats (Femur - microCT, three-point bending, histomorphometry) Pub *NASA Open Science Data Repository*, 2023, doi: 10.26030/b09t-mw60

[7] Rutkove S; Bouxsein M; Ko F; Mortreux M; Riveros D; Nagy J Dose-dependent skeletal deficits due to varied reductions in mechanical loading in rats (Tibia - pQCT) Pub *NASA Open Science Data Repository*, 2023, doi: 10.26030/emsm-0648

[8] Turner R; Keune J; Philbrick K; Branscum A; Iwaniec U Spaceflight-induced (STS-62) vertebral bone loss in ovariectomized rats is associated with increased bone marrow adiposity and no change in bone formation Pub *NASA Open Science Data Repository*, 2023, doi: 10.26030/kb2k-2150

[9] Alwood J; Dubeé P; Scott R; Thomas N; Pendleton M Quantifying Cancellous Bone Structural Changes in Microgravity: Axial Skeleton Results from the RR-1 Mission Pub *NASA Open Science Data Repository*, 2023, doi: 10.26030/8wja-w380

[10] Turner R; Keune J; Branscum A; Iwaniec U Effects of Spaceflight on Bone Microarchitecture in the Axial and Appendicular Skeleton in Growing Ovariectomized Rats from STS-62 Pub *NASA Open Science Data Repository*, 2023, doi: 10.26030/cztm-cx29