RAWLSNET:

Altering Bayesian Networks to Encode Rawlsian Fair Equality of Opportunity†

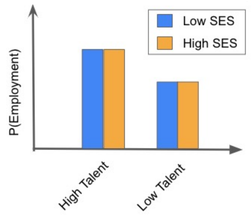
David Liu*, Zohair Shafi, William Fleisher, Tina Eliassi-Rad, Scott Alfeld
Northeastern University, Amherst College

Problem Definition

- We draw on John Rawls' notion of fairness known as the principle of Fair Equality of Opportunity (FEO).
- Question: *Given an unfair outcome and the capability to alter some (but not all) decision making processes, how can one satisfy FEO?*

Fair Equality of Opportunity

- FEO requires that any two people with similar talent and ambition receive the same chance of achieving an advantageous social position.
- Under the FEO framework there are four variables of interest:
  - **Justified** (*J*): variable to consider when allocating position (e.g. talent)
  - **Sensitive** (*S*): a protected attribute that should not affect allocation (e.g. race)
  - **Control** (*C*): an intermediate decision that the RAWLSNET user can affect (e.g. college admissions)
  - **Outcome** (*O*): a variable indicating whether the individual received the position (e.g. employment)
- FEO can be formalized as the conditional independence: $O \perp S \mid J$



Use Cases

There are two main use cases for RAWLSNET:

1. **Generating fair data**
   A Bayesian Network that has been altered with RAWLSNET can be sampled for fair data that satisfy the FEO conditional independence. These data can then be used to train downstream applications. Such data address growing concerns of bias in the training data. In contrast to de-biasing methods, data from an altered Bayesian Network can be sampled infinitely.
2. **Guiding policy**
   The optimized parameter values for the control variable can be used to guide policy decisions. These policies will ensure that the system as a whole satisfies FEO. Users of RAWLSNET can collaborate with domain experts to best identify the variables of relevance.

Our Contribution

RAWLSNET is a system that alters Bayesian Networks to satisfy FEO in three steps, outlined below:

1. **Input**
   The user provides a Bayesian Network (structure and parameter values) or a dataset from which a network is learned.
2. **Identification**
   The user identifies the justified, sensitive, control, and outcome variables.
3. **Optimization**
   RAWLSNET determines the optimal parameter values for the control variable to satisfy FEO. Specifically, the system solves a system of linear equations in which the equations are constraints to satisfy FEO and the variables are parameter values for the control variable.

Example: College Admissions

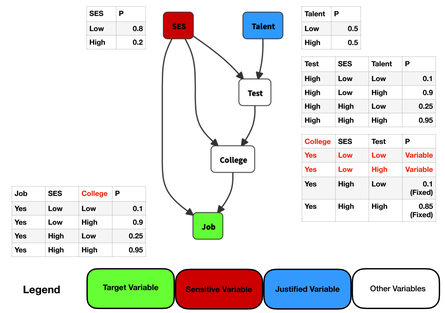
SES	P
Low	0.8
High	0.2

Talent	P
Low	0.5
High	0.5

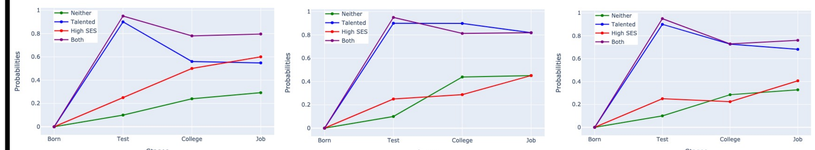
Test	SES	Talent	P
High	Low	Low	0.1
High	Low	High	0.9
High	High	Low	0.25
High	High	High	0.95

College	SES	Test	P
Yes	Low	Low	Variable
Yes	Low	High	Variable
Yes	High	Low	0.1 (Fixed)
Yes	High	High	0.85 (Fixed)

Job	SES	College	P
Yes	Low	Low	0.1
Yes	Low	High	0.9
Yes	High	Low	0.25
Yes	High	High	0.95



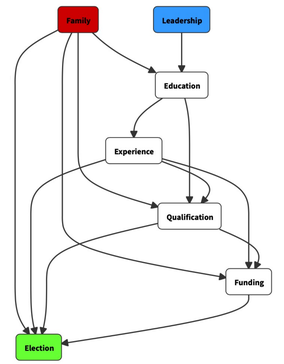
The above Bayesian Network is an example of an FEO application. Individuals are born with a certain level of Talent and Socioeconomic status (SES). Then, decisions are made about whether an individual scores well on a test, is admitted into college, and receives a job. Here, RAWLSNET will optimize the college acceptance rates for low SES individuals so that Job allocations are fair.



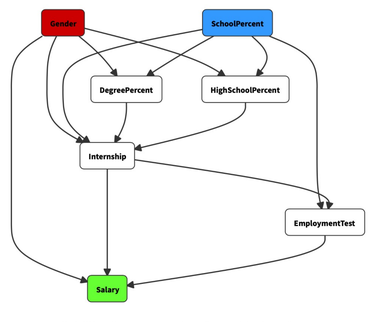
(A) (B) (C)

Each graph shows the probability of obtaining an outcome of interest separated by demographic. In (A), the parameters have not been optimized and individuals of high SES receive jobs with higher probability. In (B), the intersections of the blue and purple lines as well as the green and red lines show that the probability of receiving a job is determined by only talent. In (C), we determine the optimal college admissions policy under feasibility constraints (the college can only admit a maximum number of students). RAWLSNET determines the policy that most closely satisfies FEO.

Experiments



Lead.	Family	Not Elected	Nominee	Elected
Poor	Not Political	16.34%	24.10%	59.50%
Poor	Political	22.34%	11.10%	66.60%
Good	Not Political	15.96%	21.40%	62.60%
Good	Political	21.14%	10.70%	68.20%



SchoolPercent	Gender	Salary
Low Score	Male	43.06%
Low Score	Female	43.06%
High Score	Male	47.76%
High Score	Female	47.76%

On the left, we present a more involved synthetic example in which elections are won by individuals based on their leadership abilities. In this case, there was not an exact solution and RAWLSNET determined the closest solution. On the right, we trained a Bayesian Network on real campus recruitment data from Kaggle. Here RAWLSNET determines an Internship allocation policy such that gender and salary are independent given talent (as measured by performance on the earliest standardized test "SchoolPercent").

Identifying FEO Applications

To determine whether a decision domain is a valid application for RAWLSNET, three criteria must be satisfied:

1. The decision must be centered on the allocation of advantageous social positions, such as jobs.
2. The control variable that RAWLSNET modifies must represent a decision that occurs prior to the social position allocation decision. For example, in the college admissions example, we determine the optimal college admissions policy so that job allocation satisfies FEO.
3. For each individual we have access to their justified and sensitive attributes.

We note that talent is inherently immeasurable and proxies must be used. As such, RAWLSNET should only be used when proxies that capture innate talent for a job exist – independent of sensitive attributes.

Ultimately, RAWLSNET is not designed to unilaterally determine decision making policy. Instead, the system should be used in collaboration with domain experts who are then informed by RAWLNET's recommendations.

* The point of contact is liu.davi@northeastern.edu

† David Liu, Zohair Shafi, William Fleisher, Tina Eliassi-Rad, and Scott Alfeld. 2021. RAWLSNET: Altering Bayesian Networks to Encode Rawlsian Fair Equality of Opportunity. In AIES'21. https://doi.org/10.1145/3461702.3462618