Exposee Bayesian Networks

Problem domain

Drug use is something that has been around for a very long time and is a field which gets a lot of attention and is well represented in research. Drugs is also a topic a lot of people tend to have an opinion on. One thing people tend to have prejudices on in this topic is about drug users themselves. It is common for people to think that educational level and drugs see a strong relationship, or that certain personality traits are "typical" for drug users. A Lot of these prejudices however are not founded on facts and are heavily biased, so they might not hold to be true in the real world.
In this project we will further study the relations that might exist between certain social aspects, personality traits and drug use. We will build a Bayesian network which is based upon a public dataset that covers these factors.

Data

In our project we will be using data from the Drug consumption (quantified) Data Set. This data set is available at http://archive.ics.uci.edu. The database contains 1885 responses. There are 12 attributes known for every response: Personality measurements which include NEO-FFI-R (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness), BIS-11 (impulsivity), and ImpSS (sensation seeking), level of education, age, gender, country of residence and ethnicity. These 12 attributes represent a variety of personality traits and a variety of social information.

Next to those 12 attributes the use of 18 legal and illegal drugs were recorded (alcohol, amphetamines, amyl nitrite, benzodiazepine, cannabis, chocolate, cocaine, caffeine, crack, ecstasy, heroin, ketamine, legal highs, LSD, methadone, mushrooms, nicotine and volatile substance abuse and one fictitious drug (Semeron) which was introduced to identify over-claimers. Our Bayesian network will focus only on a selected few of these drug types.

Implementation Plan

We will implement our Bayesian Network in R, using the R packages 'dagitty' for building causal diagrams and the R package 'bnlearn' for constructing our Bayesian Network. Next to that we will use the standard R packages for statistical analysis and Exploratory Data Analysis (EDA) to further familiarize ourselves with the data. Furthermore, we will be using the package 'lavaan' to fit our parameters.

Application

The topic we are most interested in is measuring the influence social/personality traits have on certain drug use levels. For example, we would like to further investigate if it is possible to sketch a profile of a person that uses a specific kind of drug based on the relations we discover in our network. We would also be interested if this information can be used to make certain predictions for example: does a person who is very "sensation seeking" have a bigger risk of using certain drugs? Another topic of interest could be the relation between alcohol and drug use.

The last topic we are interested into test is the fake drug they introduced in the dataset "Semeron". This drug does not exist but was purely added to spot over claimers in the dataset. We are interested in finding relations in our network with social/personality traits and this over claiming of their drug usage.

Lieuwe Meijdam, Pratik Kulkarini and Krutika Dhavale