

Leveraging Online Social Network Data and External Data Sources to Predict Personality

Daniel Chapsky*

*Contact author: danchapsky@gmail.com

Keywords: Bayesian Networks, Personality, Facebook

People express their personalities through online social networks in a variety of ways, such as their relationships with their friends and their listed interests. In this work I present a method for automatically predicting an individual's personality by combining his Facebook profile information with external data sources using a machine learning method known as a Bayesian Network. The developed models use representations of people's connections to other people, places, cultures, and ideas, as expressed through Facebook. Due to the nature of Bayesian Networks, the semantics underlying the models are clear enough to not only predict personality, but also use knowledge of one's personality to predict his behavioral attributes and actions. I will present some of the more interesting models of personality that my systems have produced thus far. These models demonstrate the potential of my methodology in two ways: First, they are able to explain up to 70% of all variation in a personality trait from a sample of 615 individuals. Second, they are able to clearly describe underlying relationships in the model through findings such as how to predict a man's agreeableness based on his age, hometown, number of Facebook wall posts, and his willingness to disclose his preference for music made by Lady Gaga.

I will also present the necessary background in Bayesian Networks and Personality theory to understand the above results, and present how all data collection and modeling was automated using an academic edition of R Evolution R Enterprise and various R packages.

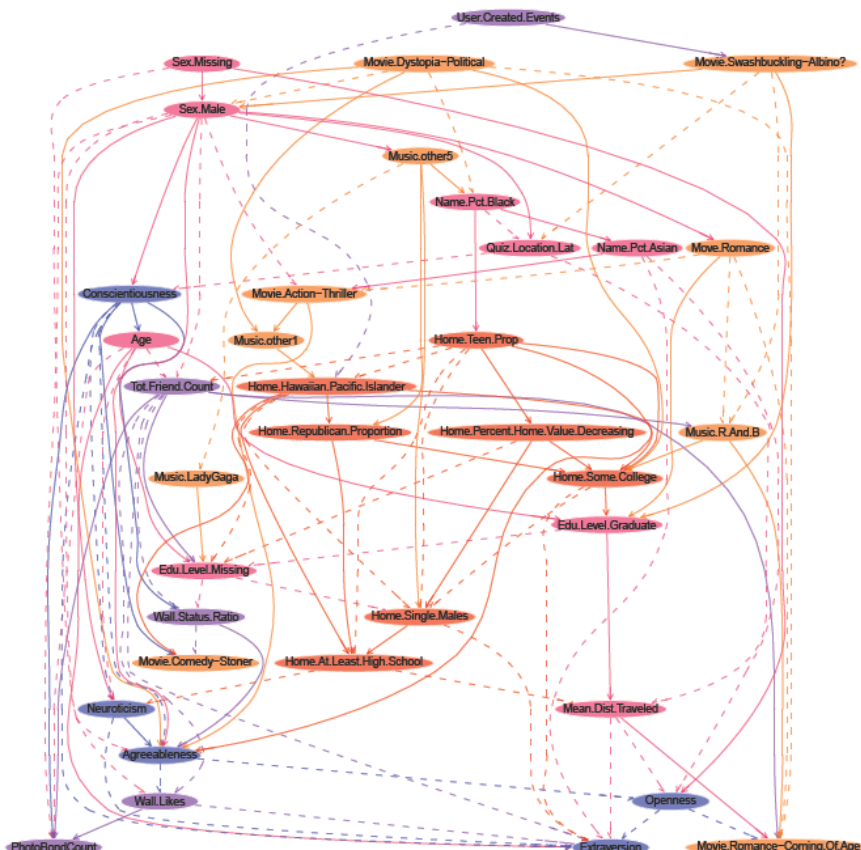


Figure 1: An Example Bayesian Network modeling personality. Blue nodes are personality traits. Dotted lines denote a positive relationship while straight lines denote a negative one.