Privacy in Multiple On-line Social Networks – Re-identification and Predictability

This software demonstrates how overlapping Online Social Networks can affect user privacy. The full paper will be published soon and will be made available:

> David F. Nettleton, Vladimir Estivill-Castro, Julian Salas, "Privacy in Multiple On-line Social Networks – Re-identification and Predictability", Transactions on Data Privacy (in press).

First the program calculates the SAN metrics which are based on the graph structure and assigned attributes. Then it uses these to calculate information gain including/excluding different metrics and with different predictors. Then it outputs the data file which is then imported to Weka where J48 and SVM-SMO are used to predict the different combinations of metrics/predictors.

The main input files are "SEP4.csv" and "ATTRIBDATA.csv".

All sources files are copyright David F. Nettleton (GNU General Public License V3.0) with the exception of the following which are copyright Vladimir Estivill-Castro (GNU General Public License V3.0):

AttributeInformationPair.java
DecisionTreePath.java
DecisionTreeQuestion.java
Exhaustive.java
Heuristic.java
SensitivityRule.java