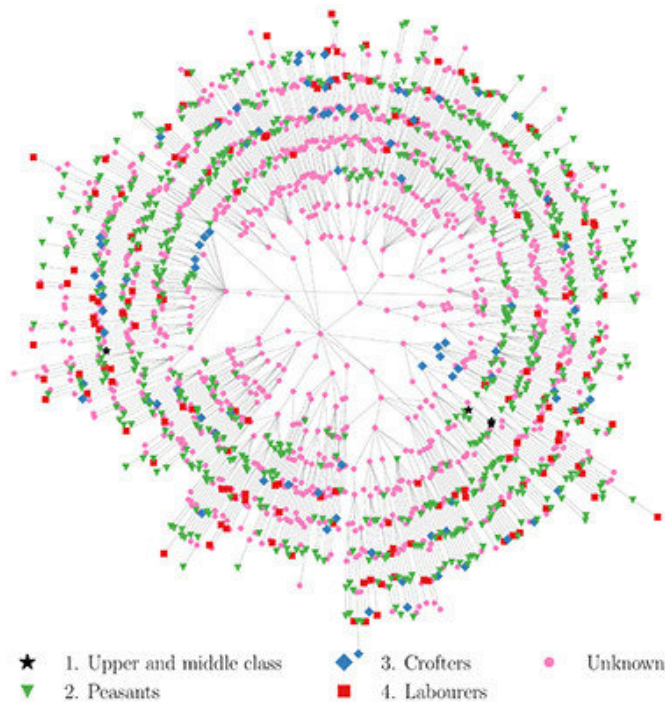# AI-created family trees confirm 18th and 19th century class divisions in Finland

May 4 2018

A small section of a family tree covering 13 generations that was derived by the algorithm. The colours show the socio-economic status of the individual. Credit: Eric Malmi

It would take 100 person-years for a genealogist to map and find all the parents for five million people – with a rate of one person per minute. The AncestryAI algorithm can do the same work in an hour using 50 parallel computers and with a success rate of 65 per cent. The algorithm

can also measure the level of uncertainty for each connection so that unreliable results can be ignored. Genealogists and demographers can use the algorithm to shed light on societal change and history.

'The [algorithm](#) does not replace the work of genealogists; it is simply a tool for helping them in their work. The genealogy algorithm can suggest connections which are probably correct, but on its own it is not as precise as a careful genealogist. The algorithm can also search for parents from nation-wide data, while a genealogist may need to limit their search to just one parish,' explains Eric Malmi, doctoral student at Aalto University who currently works for Google in Zürich.

Using AncestryAI, launched in 2017, genealogists have indeed succeeded in finding new ancestors, such as familial ties between with individuals of whom some have relocated to different regions in Finland. Currently, AncestryAI is being used to derive the genealogical relationships for people who died in the Finnish Civil War in 1918 to give, for instance, a more precise estimate of the number of war orphans.

Class division in Finland remained unchanged for 150 years

The genealogy algorithm helps examine huge amounts of data to analyse social change over long periods of time rather than at only particular and narrow timeframes. Malmi's work has confirmed, for example, that class division in Finland remained virtually unchanged between 1735 and 1885.

'We studied the effect of [socio-economic status](#) on the choice of spouse and found that they are clearly connected. Against our expectations, however, the strength of the connection did not decrease over time, but

rather stayed the same,' explains Malmi.

Socio-economic status was deduced based on the profession of a spouse's father. Farmhands and other landless peasants represented the lowest class, and the rest were then divided into tenant farmers, farmers, middle-class and upper-class.

AncestryAI makes use of statistical deduction and machine learning procedures developed for genealogical use. The basic algorithm seeks to separately deduce the mother and father for each individual based on their name, locality and date of birth. A supplementary algorithm then improves the accuracy of the basic algorithm by taking into account other factors, such as that people usually have children with the same spouse.

AncestryAI makes use of data in the HisKi database maintained by the Genealogical Society of Finland. The data consist of a total of 5 million births and 3.3 million deaths during 1648–1918. The algorithm has made a total of 7.3 million connections between children and their parents.

More information: Computationally Inferred Genealogical Networks Uncover Long-Term Trends in Assortative Mating. arxiv.org/abs/1802.06055

Provided by Aalto University