Households and their dwellings as communities in administrative data

Helle Visk
helle.visk@stat.ee

Households in register-based census

In register-based census, households and families are derived from administrative data. A household in register-based setting consists of all people living in the same dwelling. Estonia used its Population Register (PR) data on place of residence to divide people into households, but the resulting statistics was biased towards lone parent families (Figure 1).
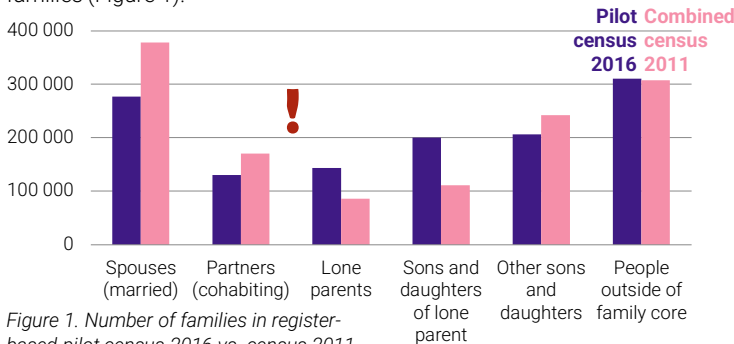

Figure 1. Number of families in register-based pilot census 2016 vs. census 2011.

This happens because family members are often registered on different addresses. About 20% people don't have correct place of residence in PR.

Let us consider a family of four: mother, father and two kids.

If one parent with a child registers themselves elsewhere,

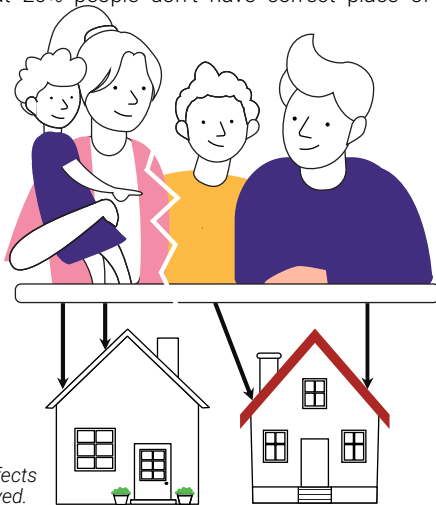this family appears as two lone parent families in the register.

Figure 2. False registering affects how family structure is perceived.

Fixing families for statistics

We would like to rejoin families, using administrative data.
In PR, children are linked to their parents and married people are linked to their spouses. Cohabiting couples can also be detected from registers. They may co-own property, have mutual children, etc.

Reuniting families means that place of residence needs also revisiting. For example, family on Figure 1 may live in either spouse's registered place of residence, their property, by their parents, etc.

Dividing people into households and assigning dwellings for them are closely related. If forming households is done first, then couples that can be linked solely by their relationship with some place are missed. PR based household formation started with dwellings and failed to detect even married couples. To make the most of the available data, we approach both tasks simultaneously.
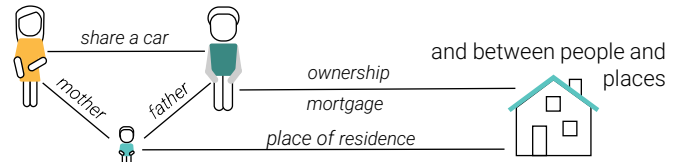
Definitions

Household – all people living in a housing unit

Family – couple with or without children or lone parent with children
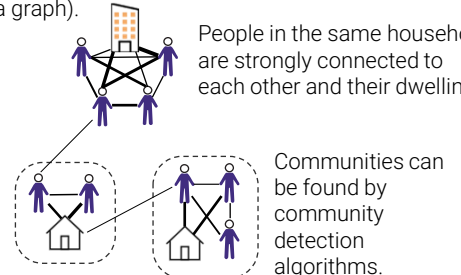A household may contain one, multiple, or no families.

Households as communities

Links between people

and between people and places

form a large network
(or in other words, a graph).

People in the same household are strongly connected to each other and their dwelling.

Household with its dwelling is a densely connected subgraph, that is: a community.

Communities can be found by community detection algorithms.

Figure 3. Households as communities in graph of people and places

Estonian Labor Force Survey and Estonian Social Survey (part of EU-SILC) were used to model edge weights: probabilities that people live in the same household or people live on certain address. Covariates were extracted from 17 registers. Edge weights between people were calculated using logistic regression on 24 types of connections. For edges linking people to places, random forest on 23 types of connections was used.

The resulting graph had 5.2M nodes and 7.8M edges. First, Louvain method [1] was used to break the graph to subgraphs of up to 5000 nodes. Then, Infomap [2] algorithm was applied iteratively until the communities were small enough. After that, some too large households were broken further according to heuristics. Also, it was ensured that each household would have exactly one place of residence.

With graph-based method, the household and family statistics matches closely with surveys (Figure 4). Place of residence from graph is used for geographical breakdown of yearly population statistics.
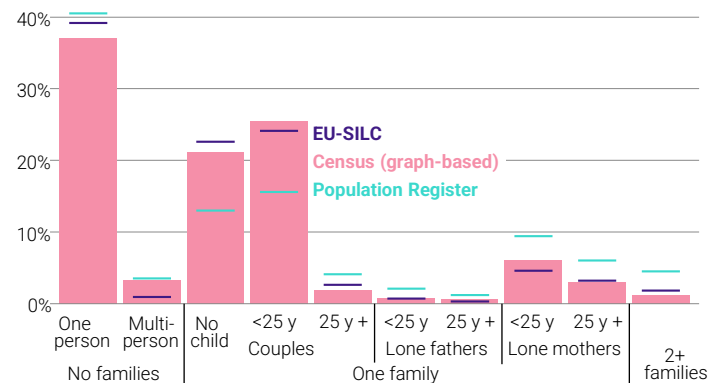

Figure 4. Distribution by type of household in 2022 with graph-based method (official census results), in EU-SILC, and in Population Register.

References

[1] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech., vol. 2008, no. 10 (2008), P10008, doi: 10.1088/1742-5468/2008/10/P10008.

[2] M. Rosvall and C. T. Bergstrom, Maps of random walks on complex networks reveal community structure, PNAS, vol. 105, no. 4 (2008), 1118–1123, doi: 10.1073/pnas.0706851105.