



StatHybrid

Combining statistics to inform you with big data to guide you

Portugal

Aims

- Develop the concept of labour market (LM) attractiveness
- Relate LM attractiveness with skills mismatch using data mining tools

Data

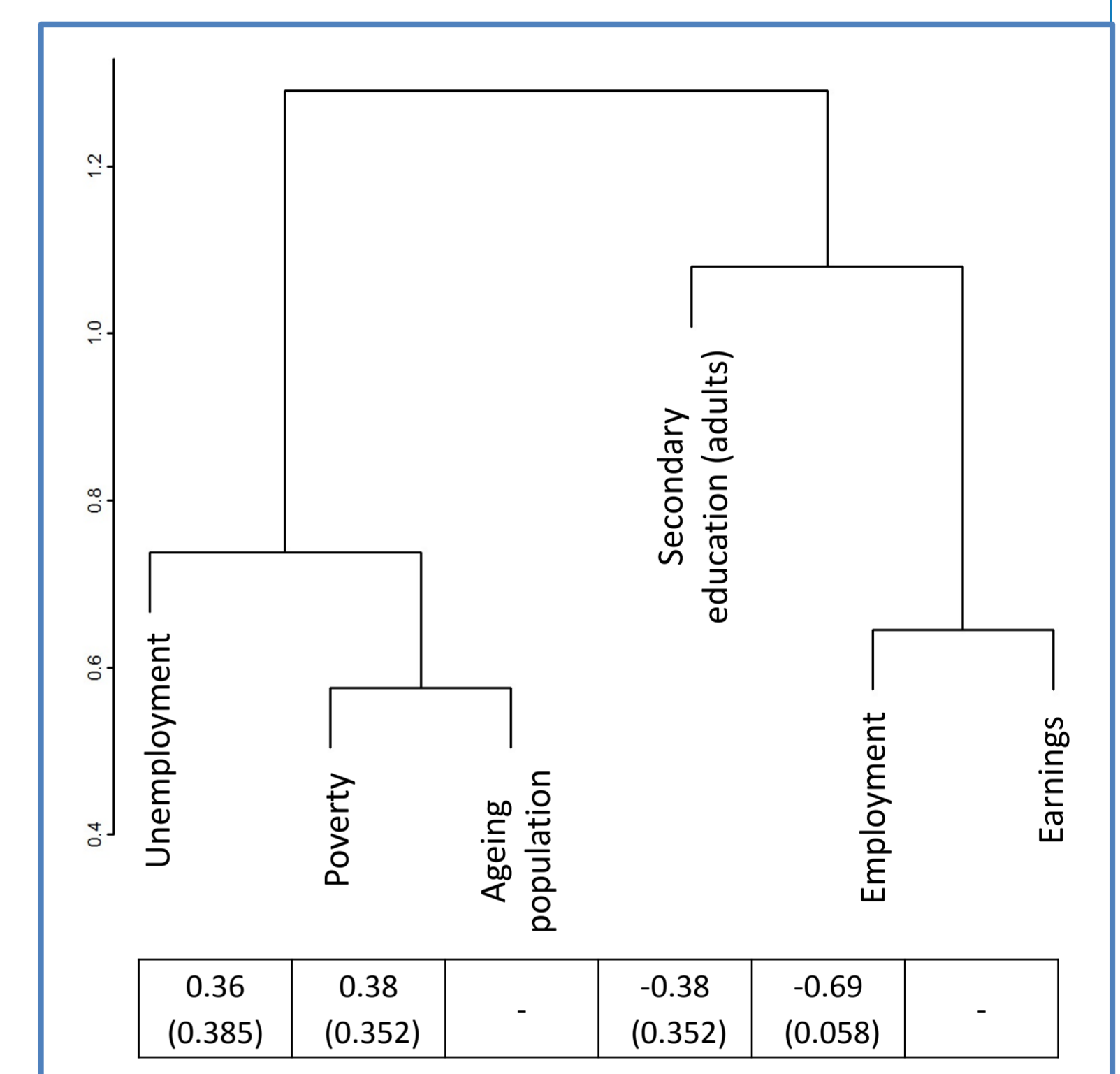
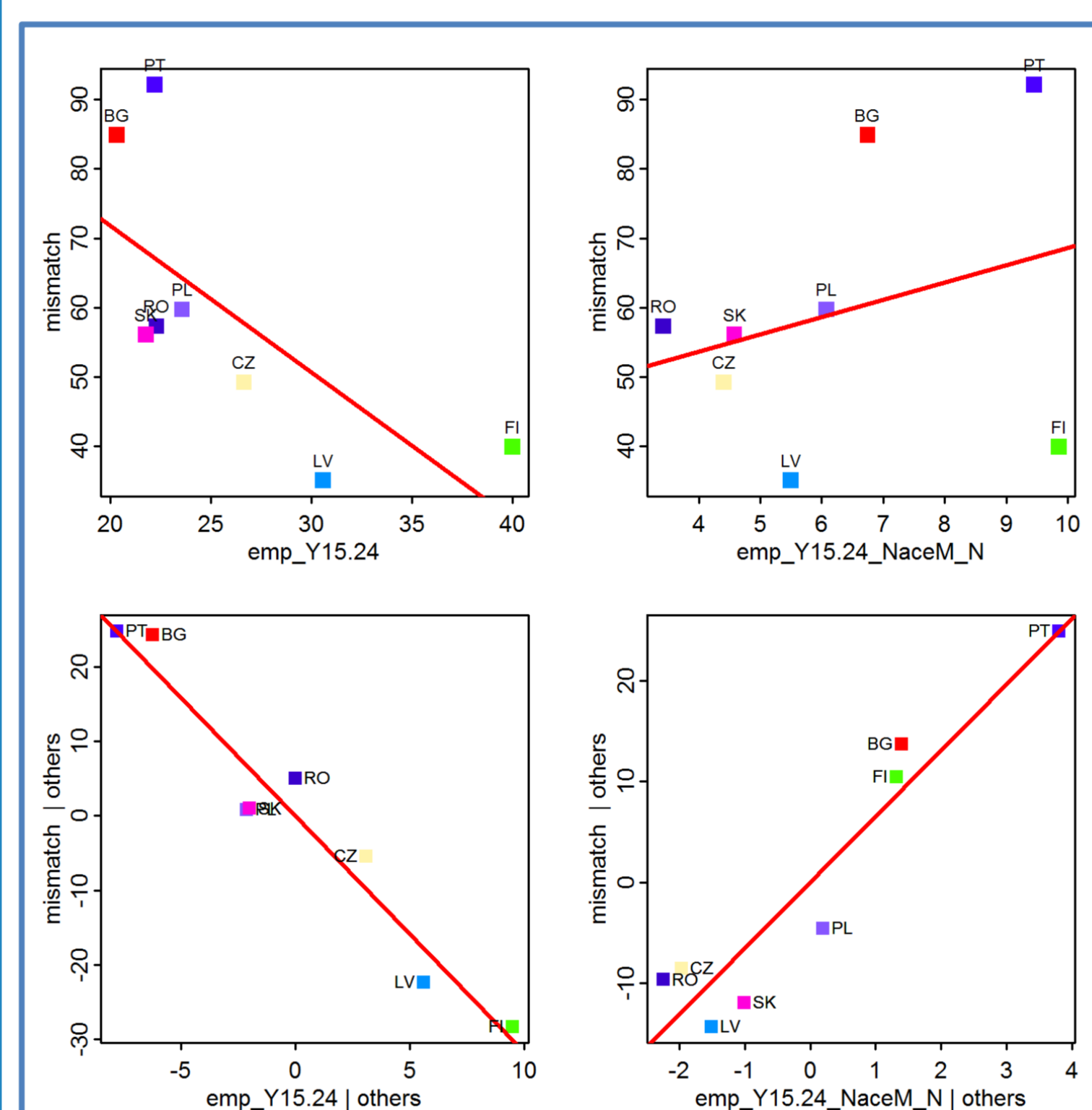
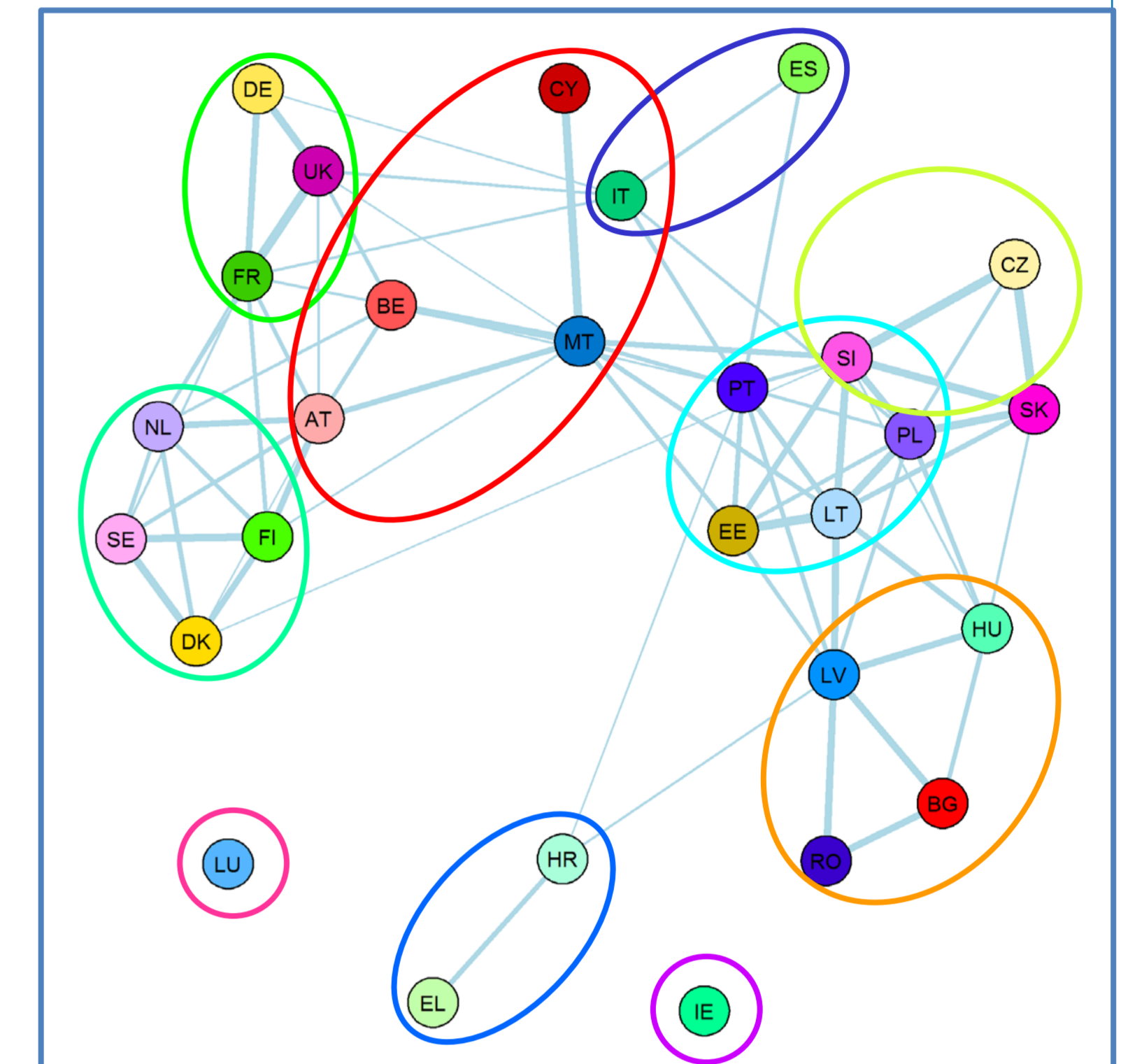
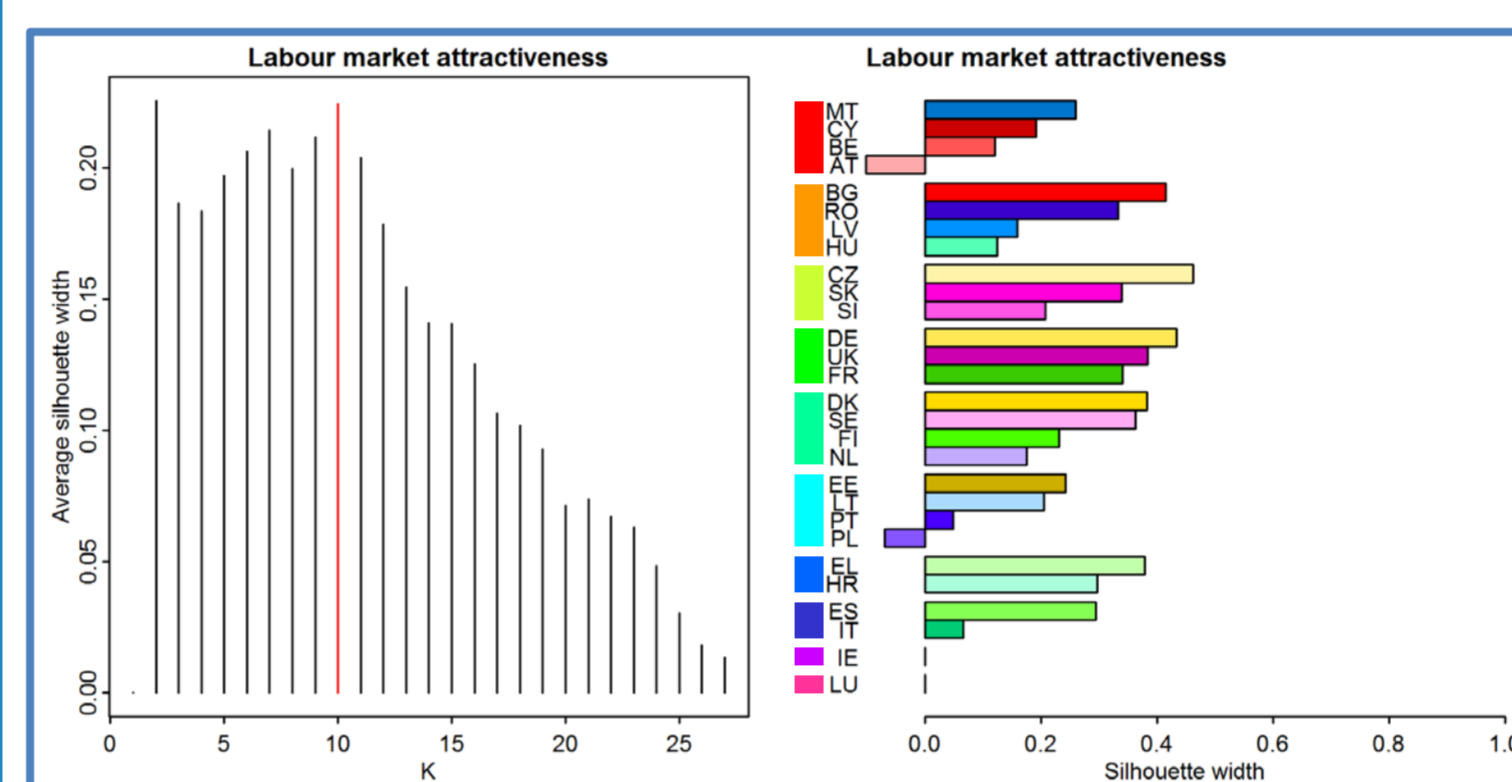
- LM attractiveness: > 70 variables (i.e. 17 variables broken down by age group, level of education, economic sector, occupation title, type of contract) from 6 Eurostat data sets: *reg_demo*; *earn*; *edtr*; *ilc*; *employ*; and *na10*
- Skills mismatch: graduated data by education fields (*educ_uoe_grad02*); job vacancies data by economic sector and occupation title (*jvs_a_nace2*)

Methods, technologies and tools

- Social network analysis (SNA)
- Partitioning around medoids (PAM) analysis
- Model selection using multinomial logistic regression via ANN
- Model selection using multivariate linear regression via general LM
- Weighted correlation network analysis
- Analysis at levels NUTS0 (country-level), NUTS1 and NUTS2
- Performed in R using *car*, *cluster*, *glmulti*, *Hmisc*, *MASS*, *nnet*, *sna* and *WGCNA*

Results

- PAM on LM attractiveness
- SNA on LM attractiveness
- GLM on skills mismatch
- WCNA on skills mismatch



Conclusions

Key achievements:

- LM attractiveness data set forms consistent clusters at all NUTS-levels;
- LM attractiveness data set can be reduced to only a few eigenvariables;
- skills mismatch is negatively associated with employment and secondary education and positively associated with unemployment and poverty.

Main challenges:

- provided data sets needed considerable expertise for cleaning and structuring data, and required detailed demographic data for normalisation;
- mapping between qualifications (ISCED-F 13) and cross between occupations (ISCO-08) and economic sector (NACE Rev. 2) is not available.