Understanding EU's Regional Potential in Low-Carbon technologies

Enrico Bergamini, Tanja Linta, Donovan Tokuyama, Georg Zachmann

Context

- Markets tend to underinvest in green-tech innovation
- Climate goals will push decarbonisation → race to green tech on a diverse technological portfolio
- Builds on COP21 Ripples Report (particularly, Roth and Zachmann 2018)

Presentation structure

1. Data

2. Green innovation landscape in EU's regions
   1. Estimation of regional potential in green technologies
   2. Network analysis (RTA, pRTA, co-patenting)

3. Which labour market, economic and demographic conditions are associated, together with potential specialization, to a stronger advantage?
   1. Building the dataset (second stage)
   2. Data-driven selection of variables

4. Limitations
5. Results

Data

- Patents: OECD's REGPAT
  - IPC technological codes and CPC Y-class codes (JRC)

    - Solar PV
    - Solar thermal
    - Electric vehicles
    - Batteries
    - Energy management
    - Efficient heating and cooling
    - Biofuels
    - Efficient lightning
    - Efficient rail
    - Efficient combustion
    - Hydro energy systems
    - Wind energy systems
    - Nuclear
    - Insulation

- All the rest:
  - Eurostat (bulk)
  - Urban Data Platform (EC + JRC)

Potential advantage: conceptual framework

- Complexity and technological specialization
  - Hidalgo (2007)
  - Hausmann (2014)
  - Yan and Luo (2017)
  - Stellner (2014)
  - Roth and Zachmann (2018)

Estimating potential advantage 1/4

- Structure of the dataset:
  - Stacks of non-overlapping 3 years period by technology IPC/CPC Y-classes to compose the panel
  - Fractional counting of patents

- From crosstables to RTAs:

$$RTA = \frac{\frac{x_{il}}{\sum_i x_{il}}}{\frac{\sum_l x_{il}}{\sum_{il} x_{il}}} \qquad sRTA = \frac{\frac{RTA-1}{RTA+1} + 1}{2}$$

where:

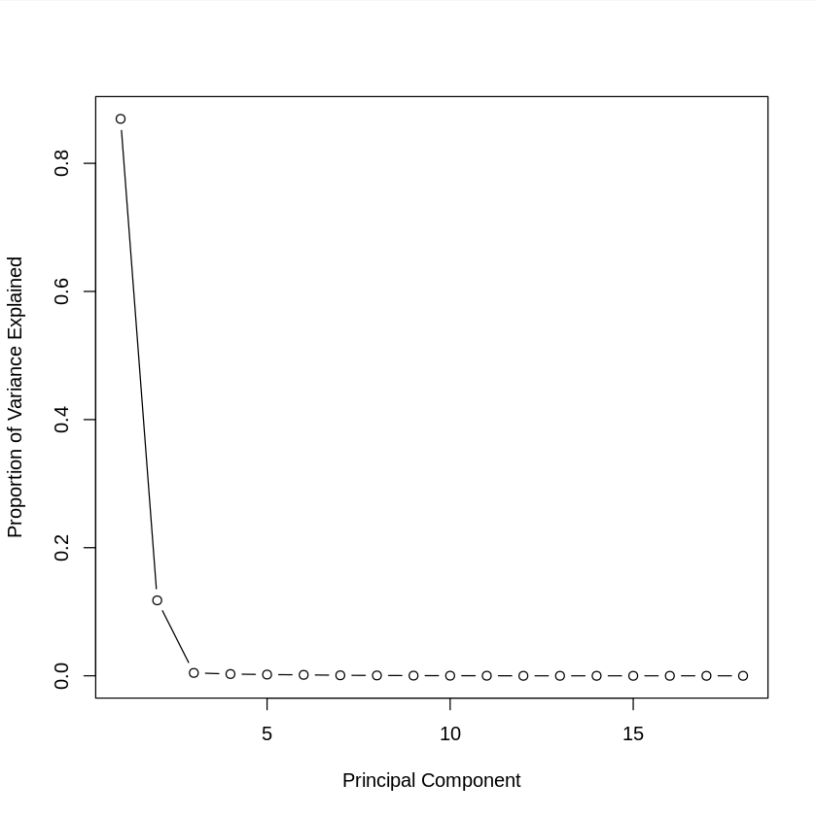
x_{il} is the number of patents of technology i in region l

$\sum_l x_{il}$ is the sum of patents of technology i across all regions

$\sum_{il} x_{il}$ is the sum of all patents across all regions

Estimating potential advantage 2/4

- Networks:
  - We calculate different types of technological proximity networks based on 18 combinations of level-method
- Levels:
  - NUTS2, country, person, applicant
- Methods:
  - Correlation, Minimum Pairwise Correlation Probability, Co-occurrence…
- From the 18 networks of NUTS-tech (highly collinear) we obtain the first two principal components



Estimating potential advantage – 3/4

Different reasons for correlation

- Latent factors
  - Factor cost
  - Infrastructure
  - Geography
  - Domestic market size
  - …
- Technological links
  - Similar value chain
  - Tech spill-over
  - Education
  - …

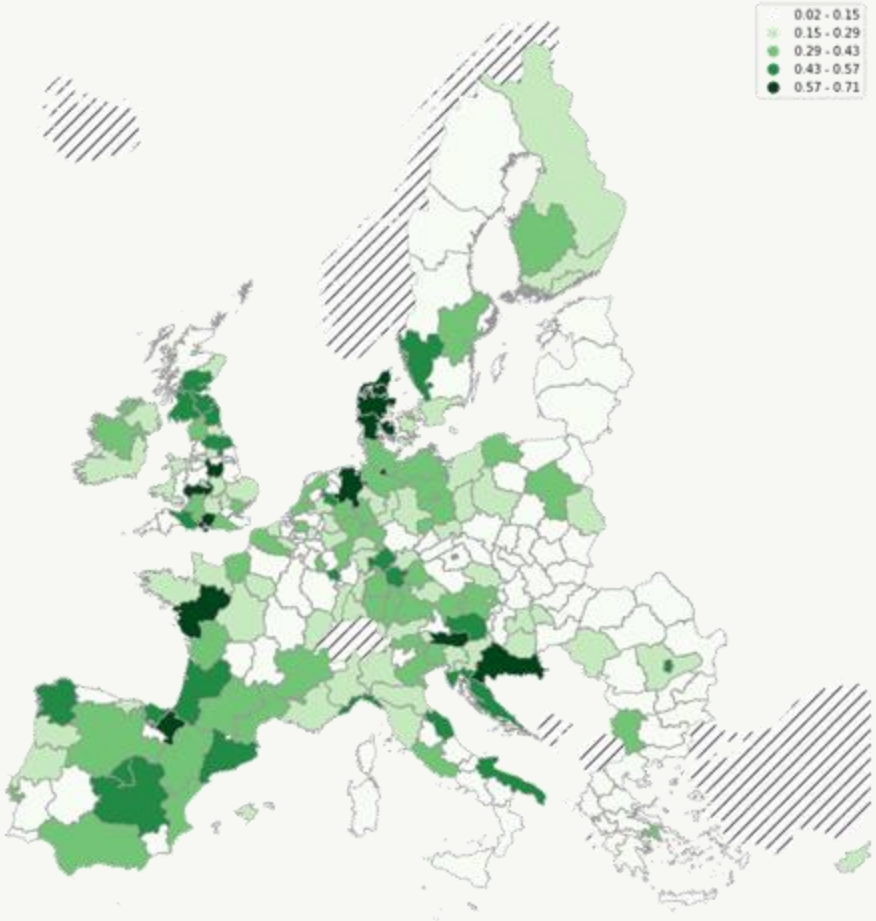
Estimating potential advantage 4/4

- Regression based estimations:
  - $RTA_{t+1} = \beta\, PC1_t + \beta\, PC2_t$
  - Zero Inflated Beta regression on the several cross sections
  - 4 stacks (2006, 2009, 2012, 2015)
  - We regress, minimize MSE, and obtain the pRTA values for *t+2* by fitting the model on the next period:
    - Fit betas obtained in RTA (t2) → predicted RTA *(t+2)*
    - R^2 is not higher than 0.35 (low predictive power?)

Potential advantage – Wind (2015 stack)

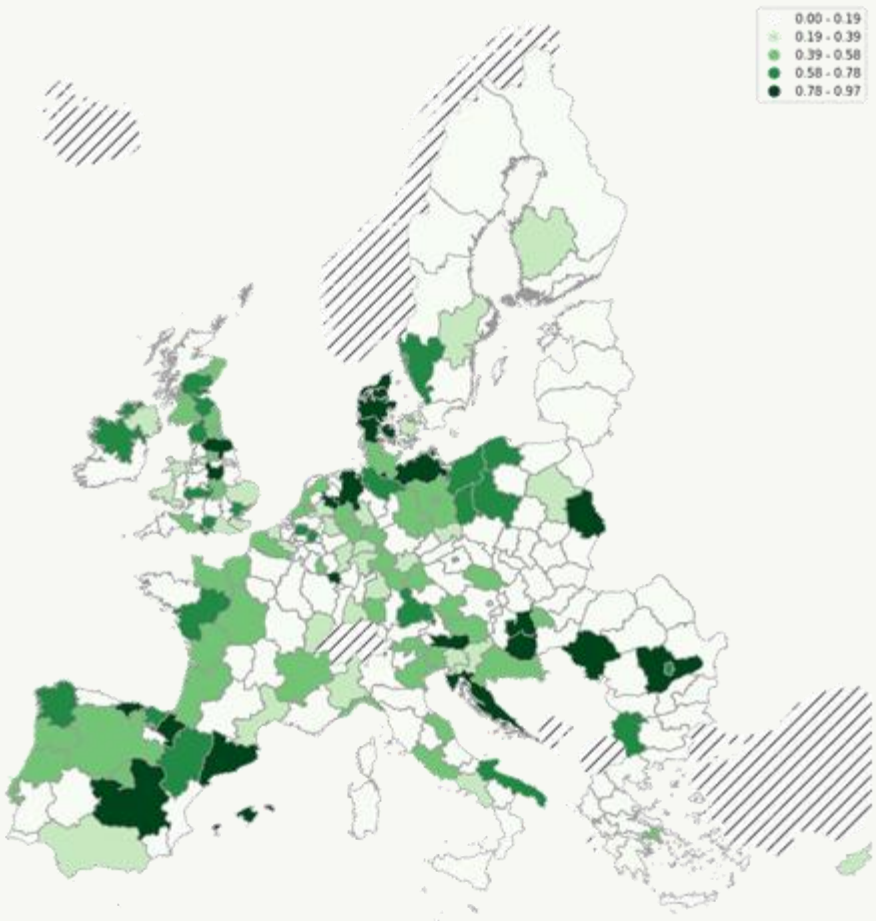
RTA

wind



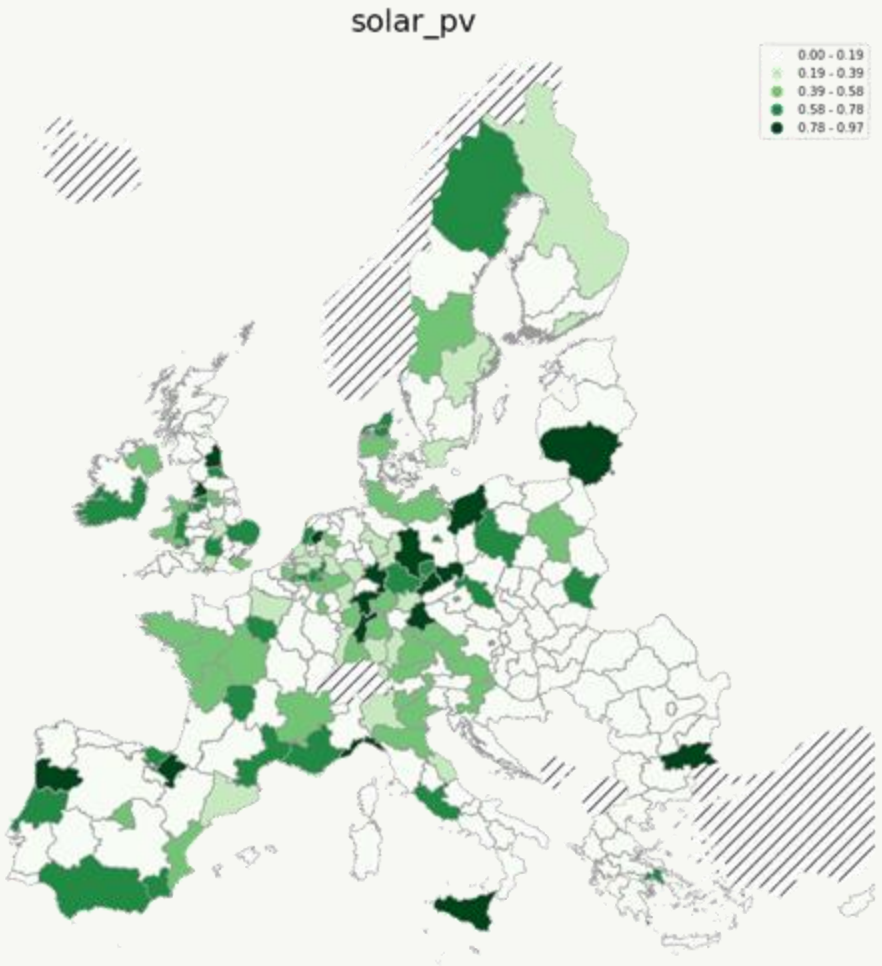
pRTA

wind

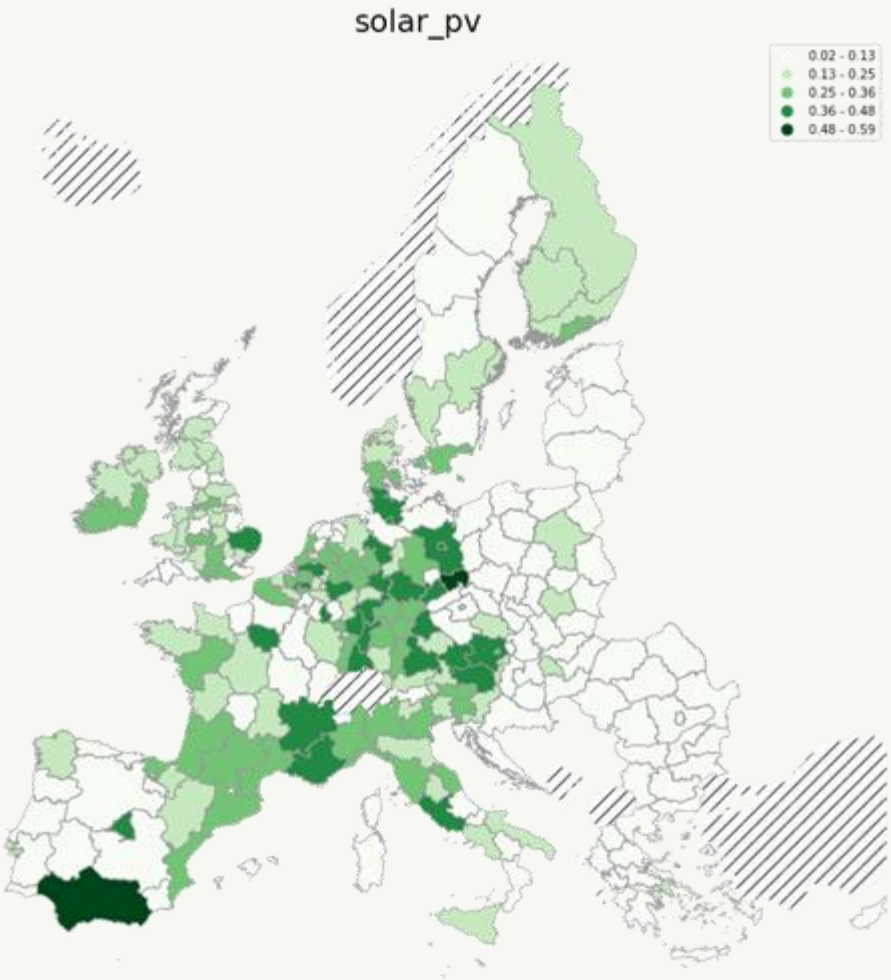


Potential advantage – Solar PV (2015 stack)

RTA

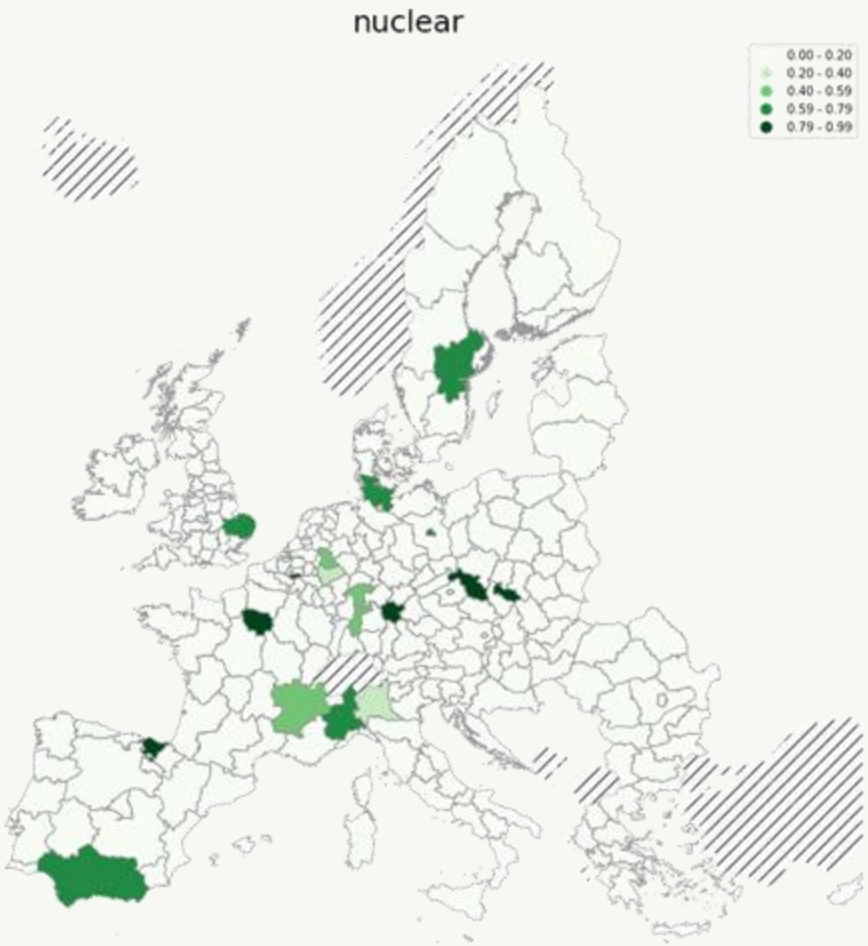


pRTA

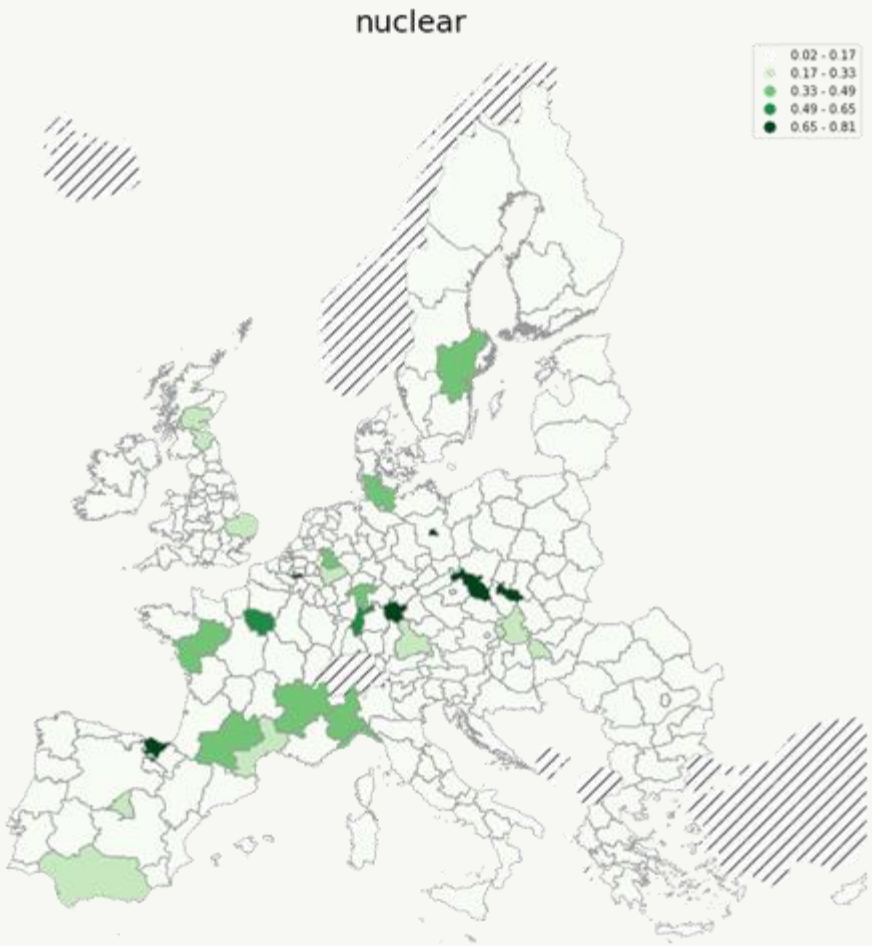


Potential advantage – Nuclear (2015 stack)

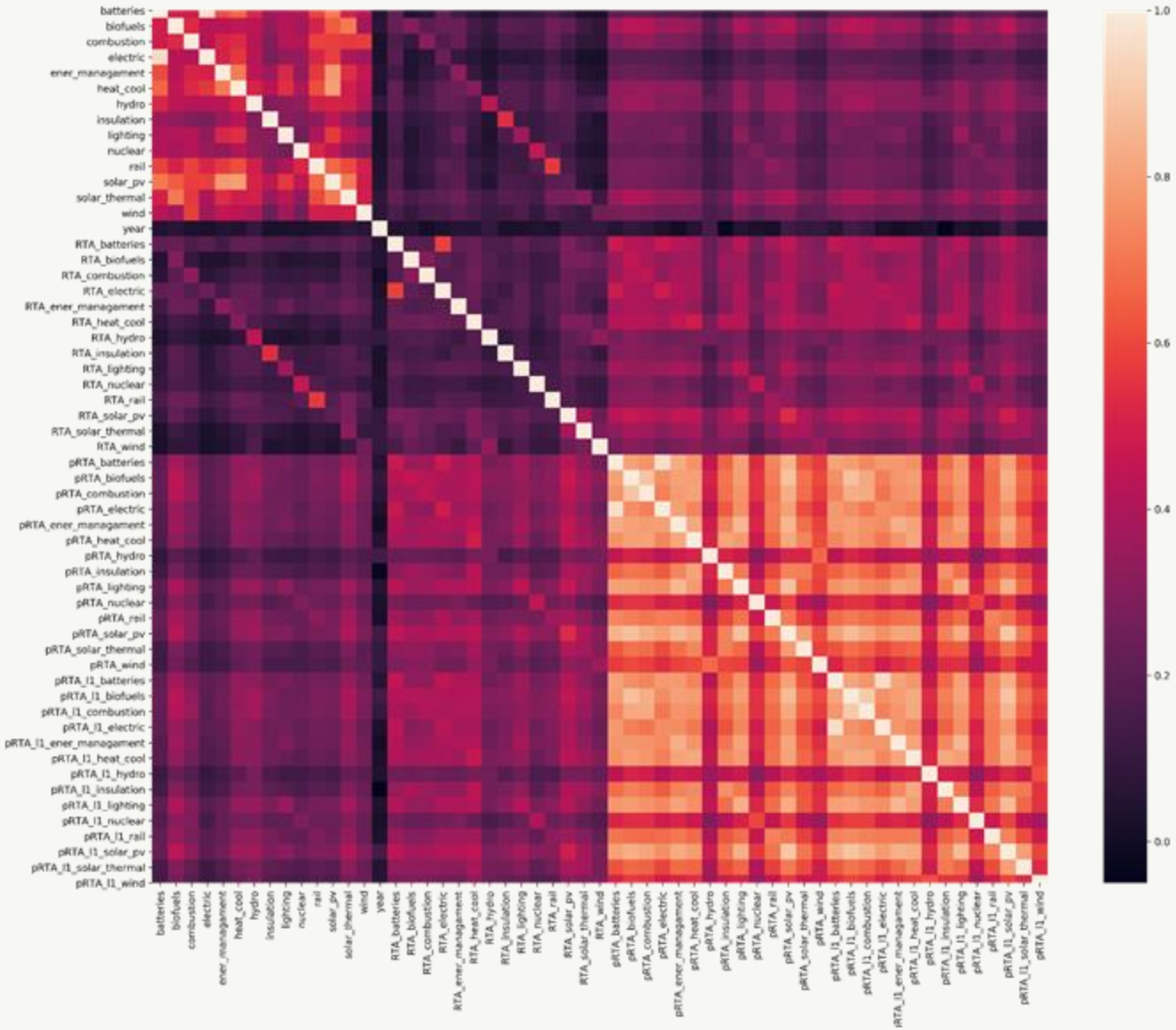
RTA



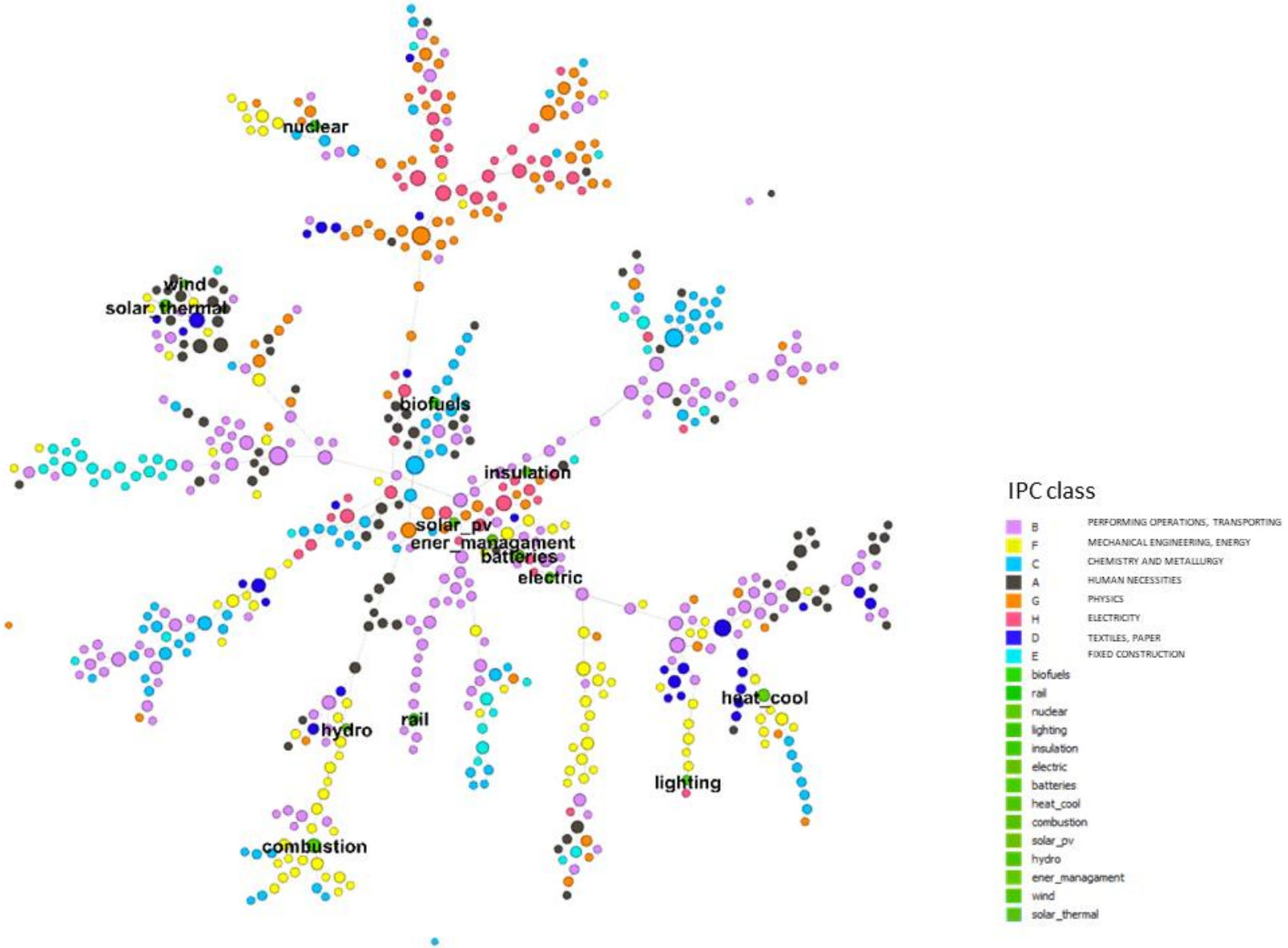
pRTA



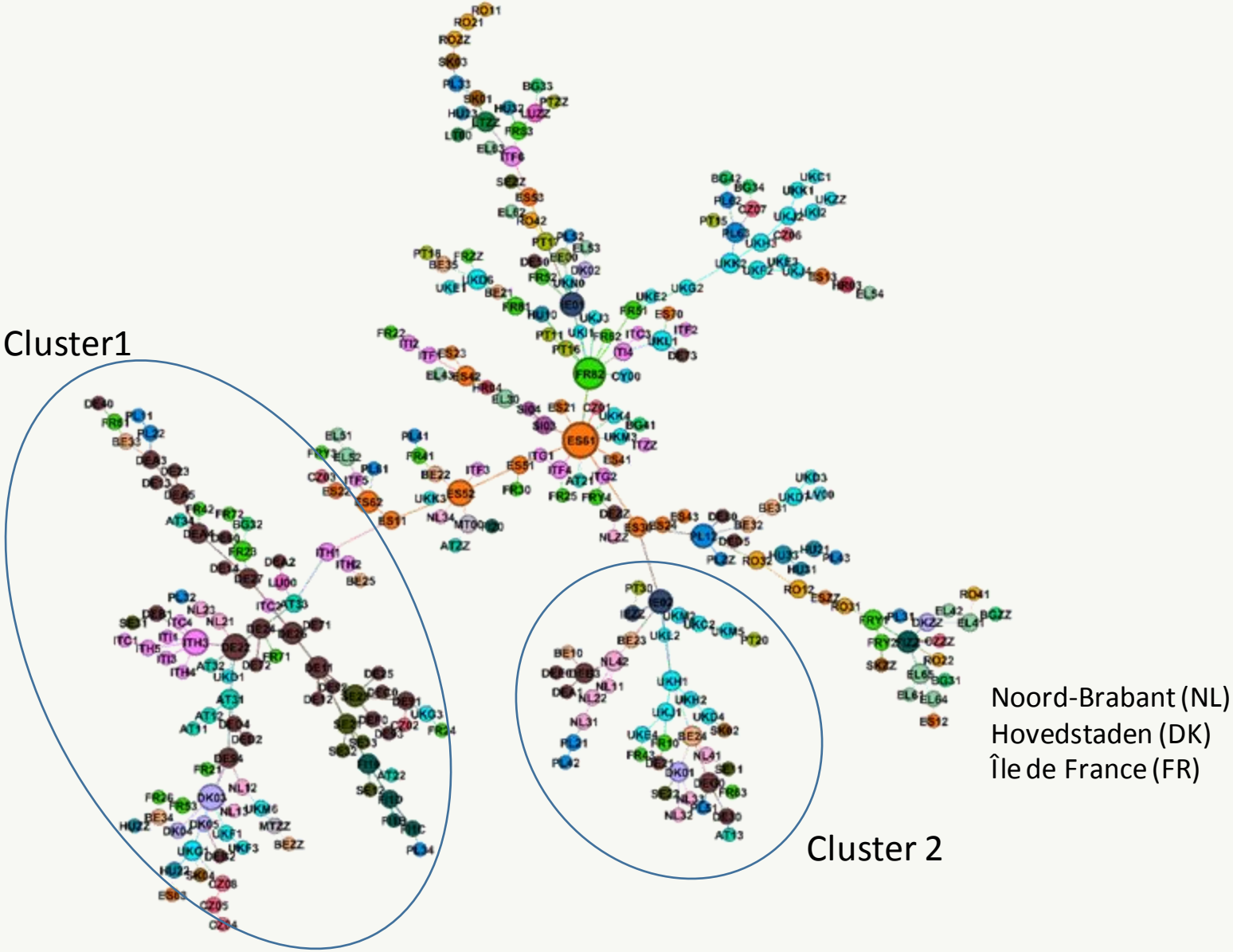
Potential advantage – Nuclear (2015 stack)



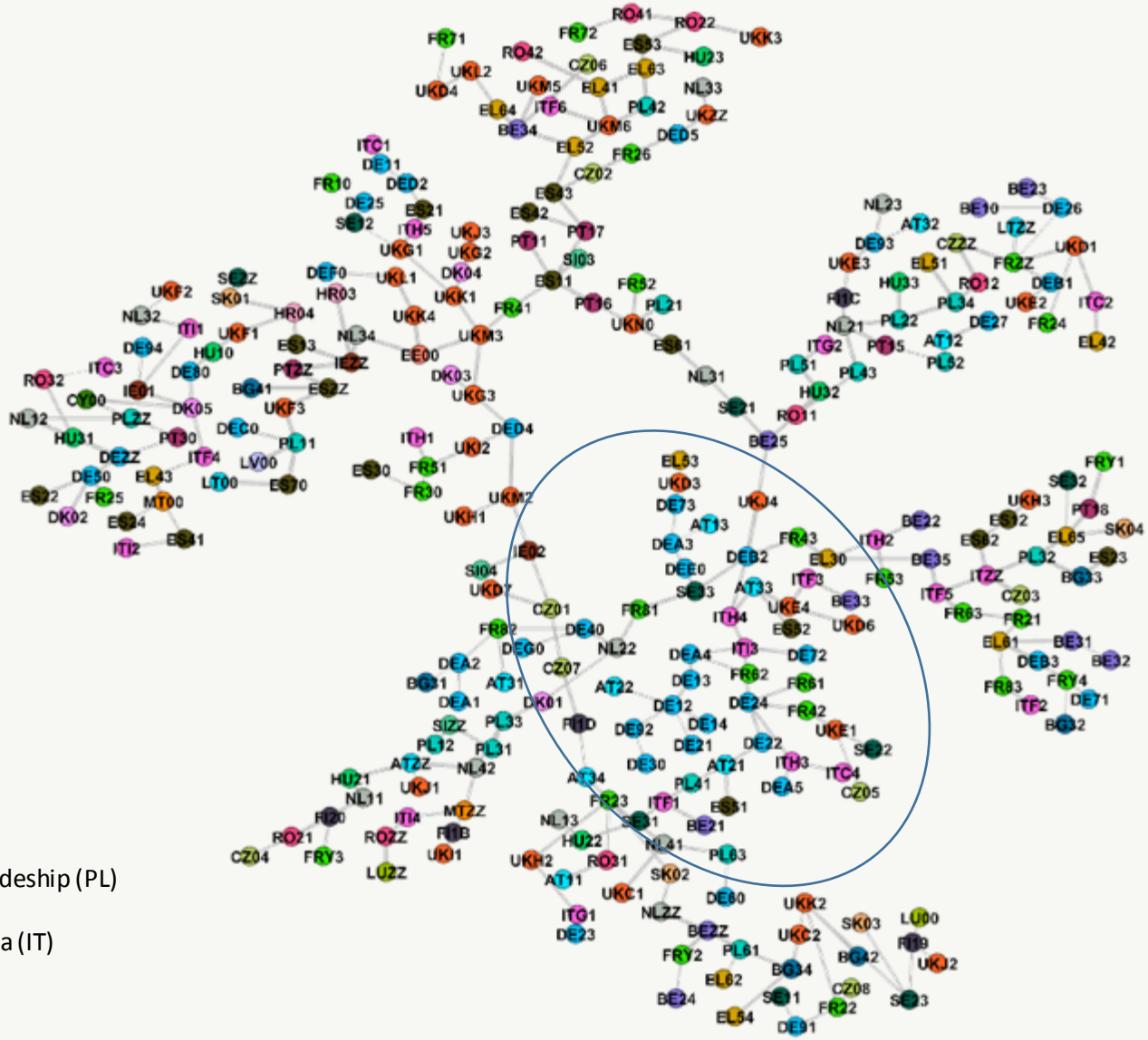
Technological proximity



Regional correlations (RTA 2012)



Regional correlations (pRTA - 2012)

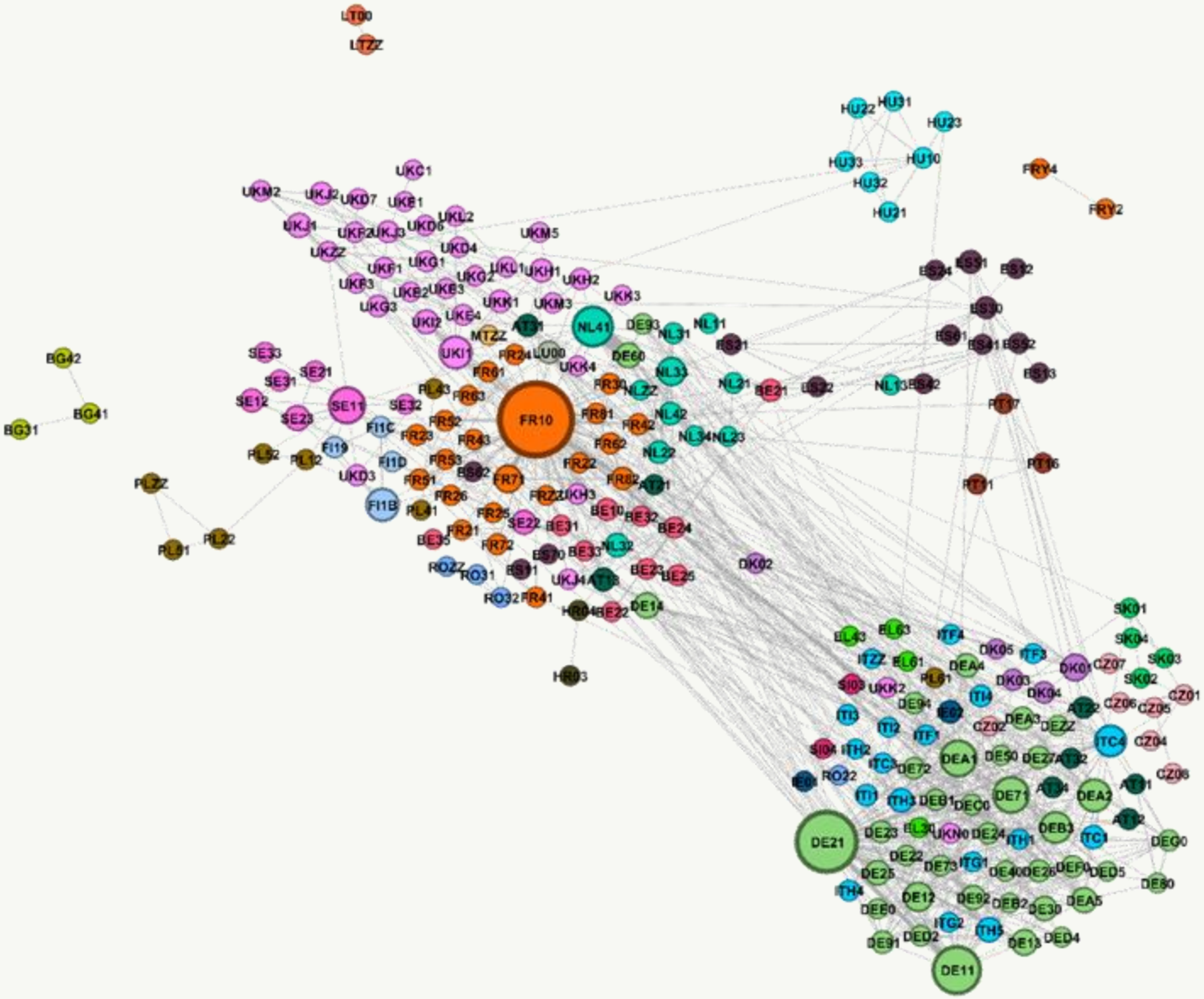


Alsace (FR)
Aquitanie (FR)
Antwerpen (BE)
Cataluña (ES)
Dussledorf (DE)
Detmold (DE)
Dytiki Ellada (EL)
Kent (UK)
Manchester (UK)
Pomeranian Voivodeship (PL)
Veneto (IT)
Friuli Venezia Giulia (IT)

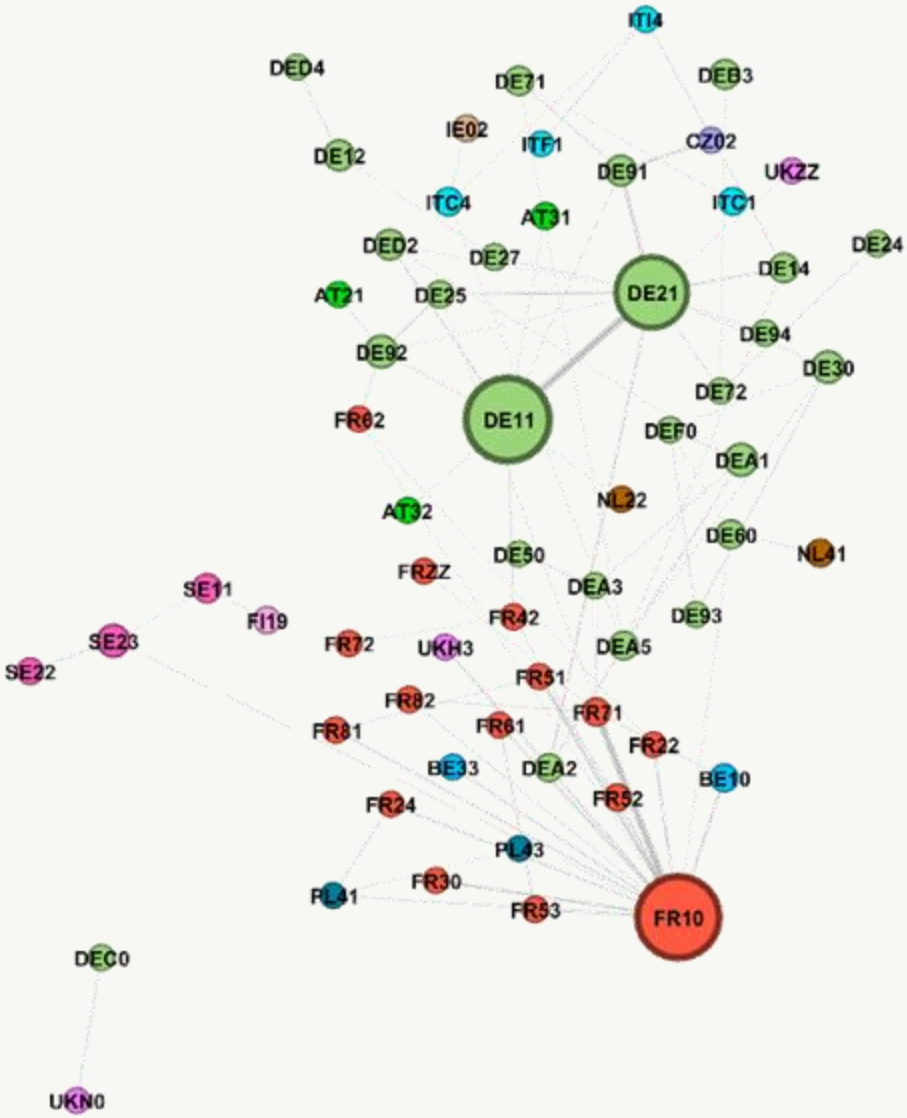
Co-patenting

- OECD’S REGPAT database, simple patent counts
- NUTS2 co-patenting figures are built on fractional counting (applicants residency)
- Based on green technologies definitions from the JRC definition based on CPC Y-codes

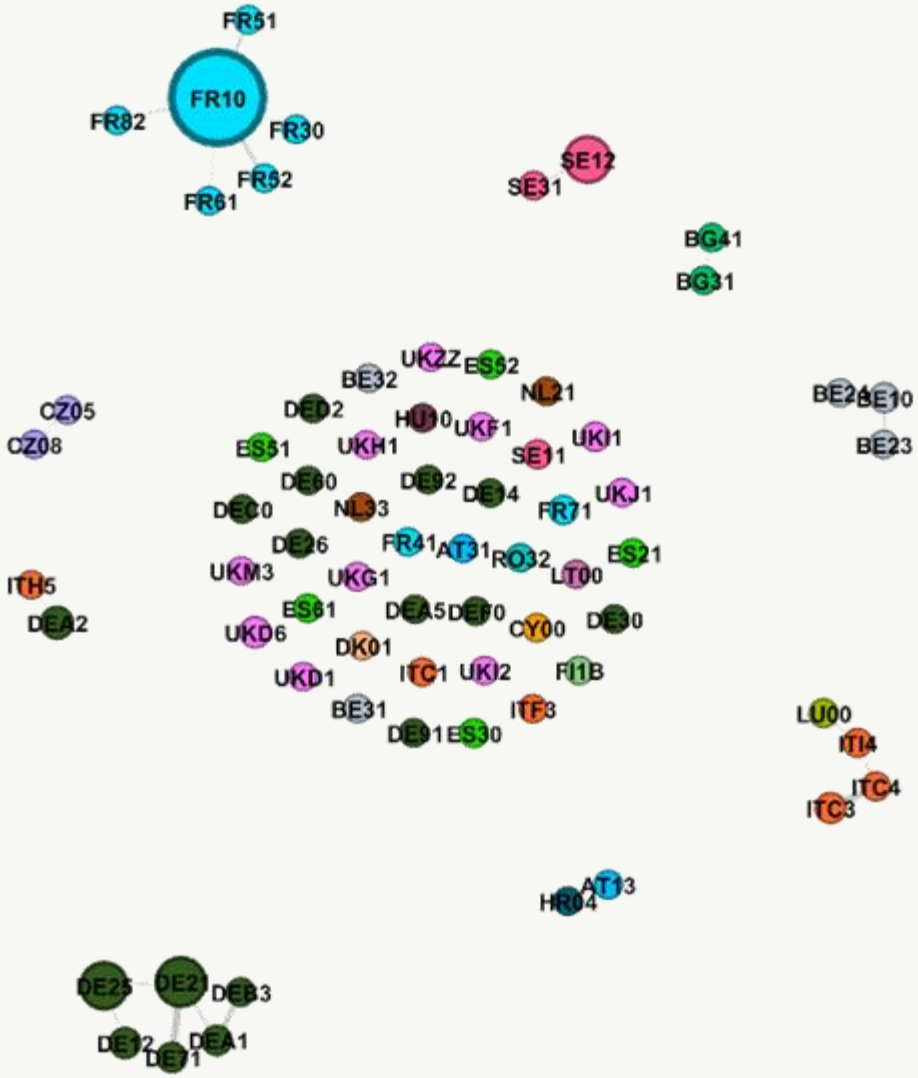
Co-patenting in green technologies



Co-patenting batteries



Co-patenting in nuclear



Second stage

- What is associated with a higher RTA? (main effects)
- What affects RTA if combined with a potential? (interaction factors)

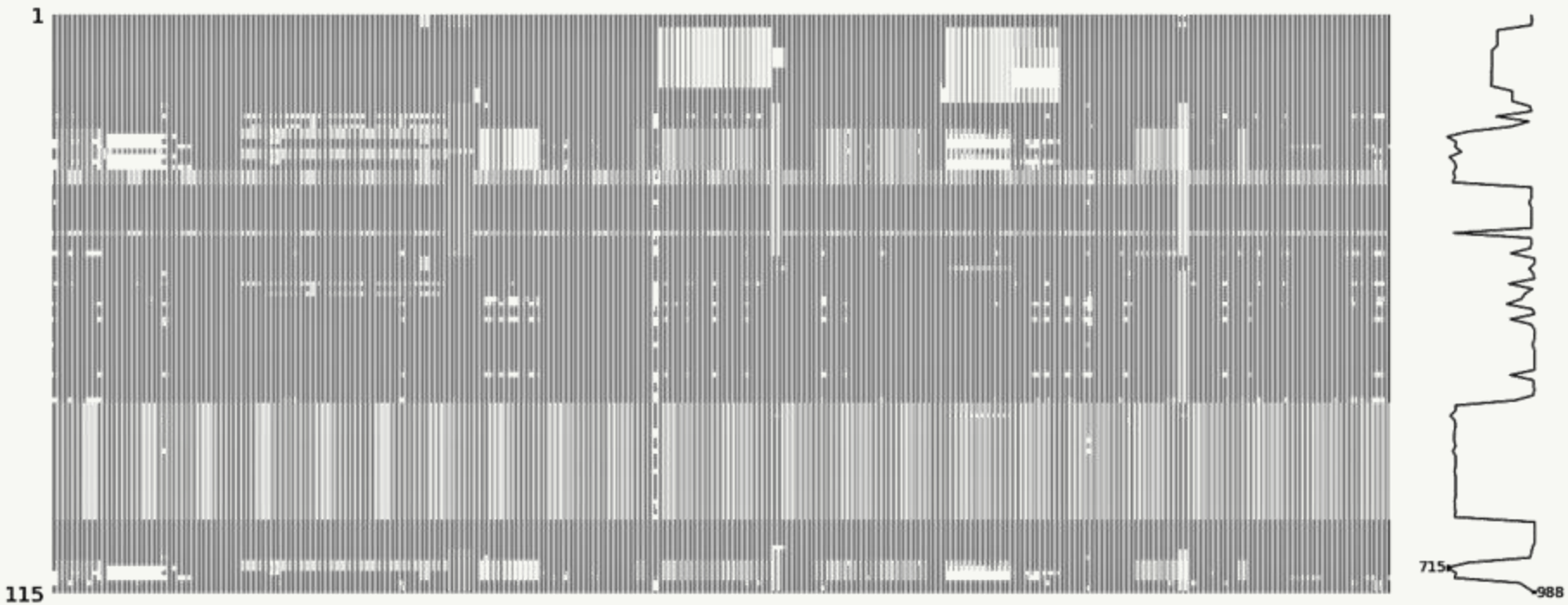
1. Building a large dataset (Eurostat+JRC)

2. Exploratory approach:
   We are not looking for causation but mere associations

3. Data-driven selection of variables (via Regularization)

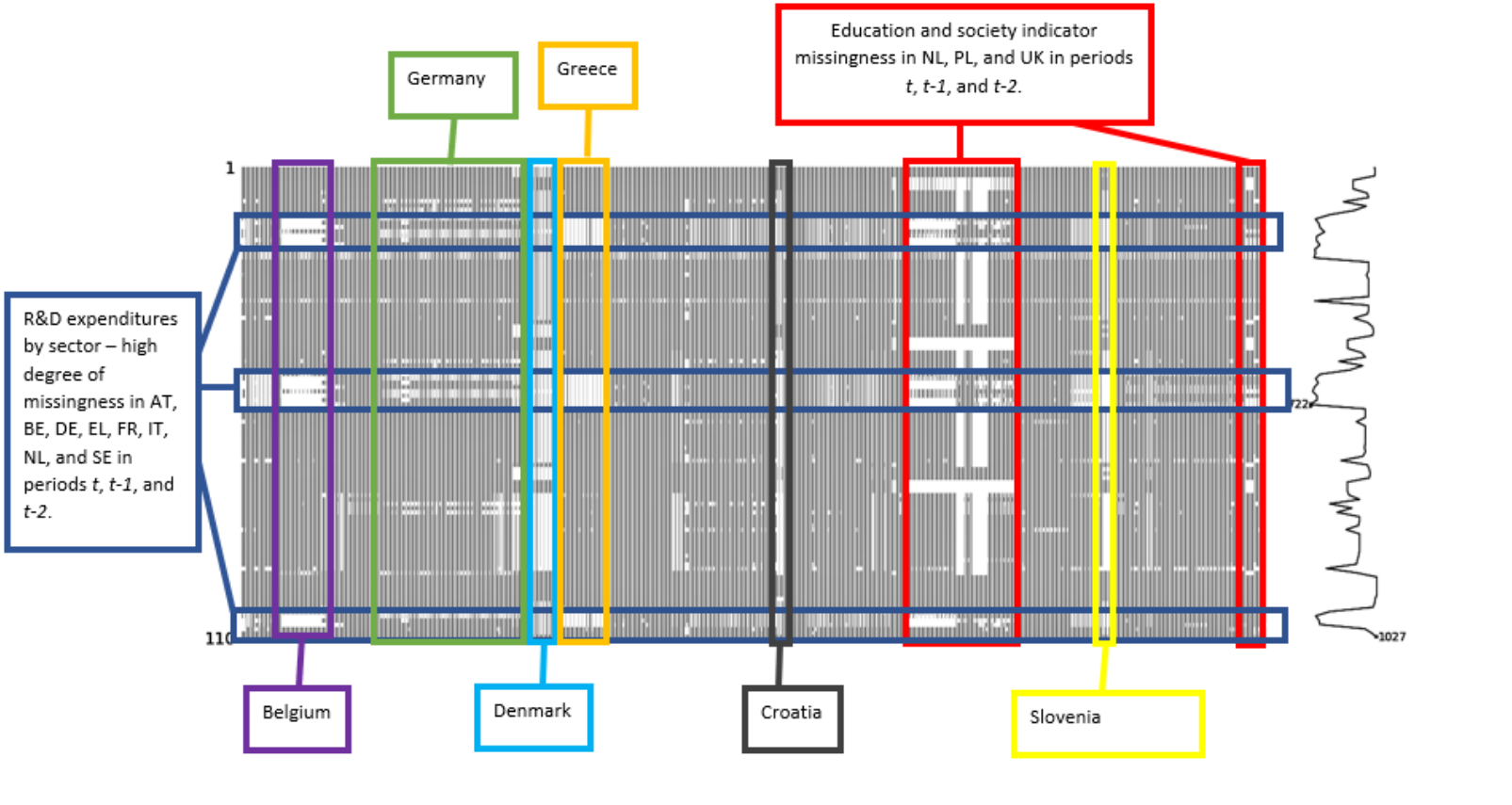
Building the dataset: need for imputation 1/3

- Regularized regression techniques are often used to reduce dimensionality through variable selection → fail in case of missing data

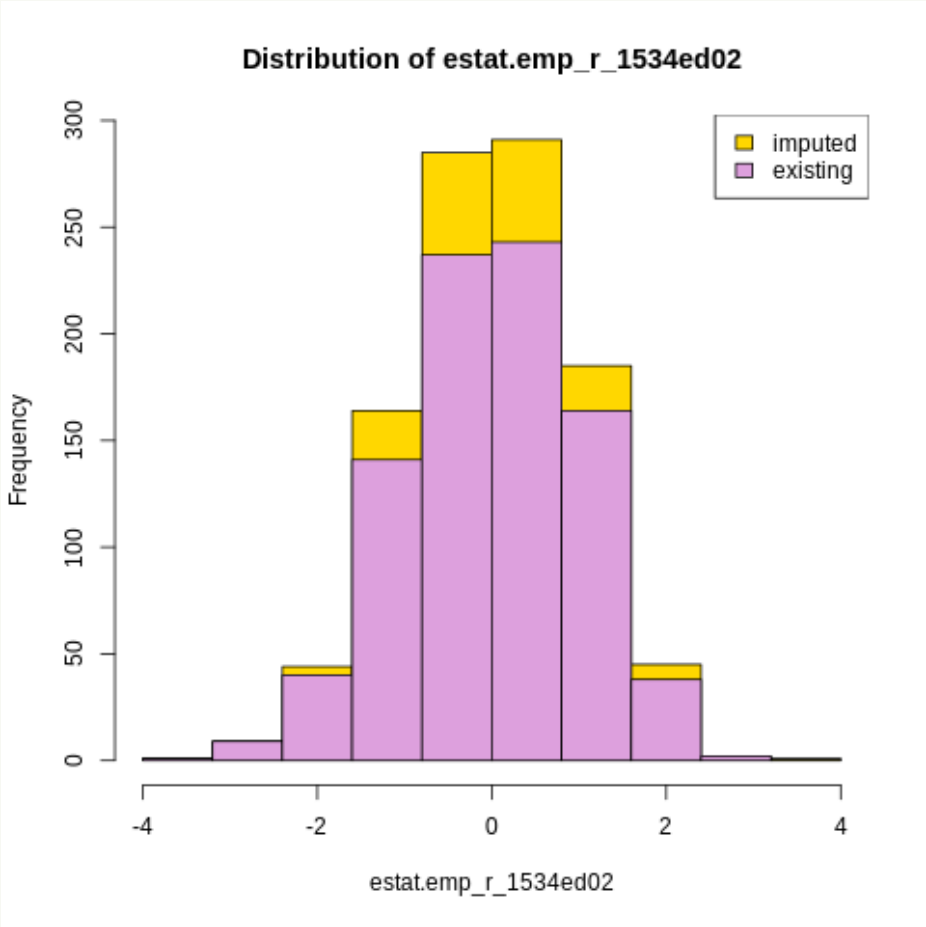


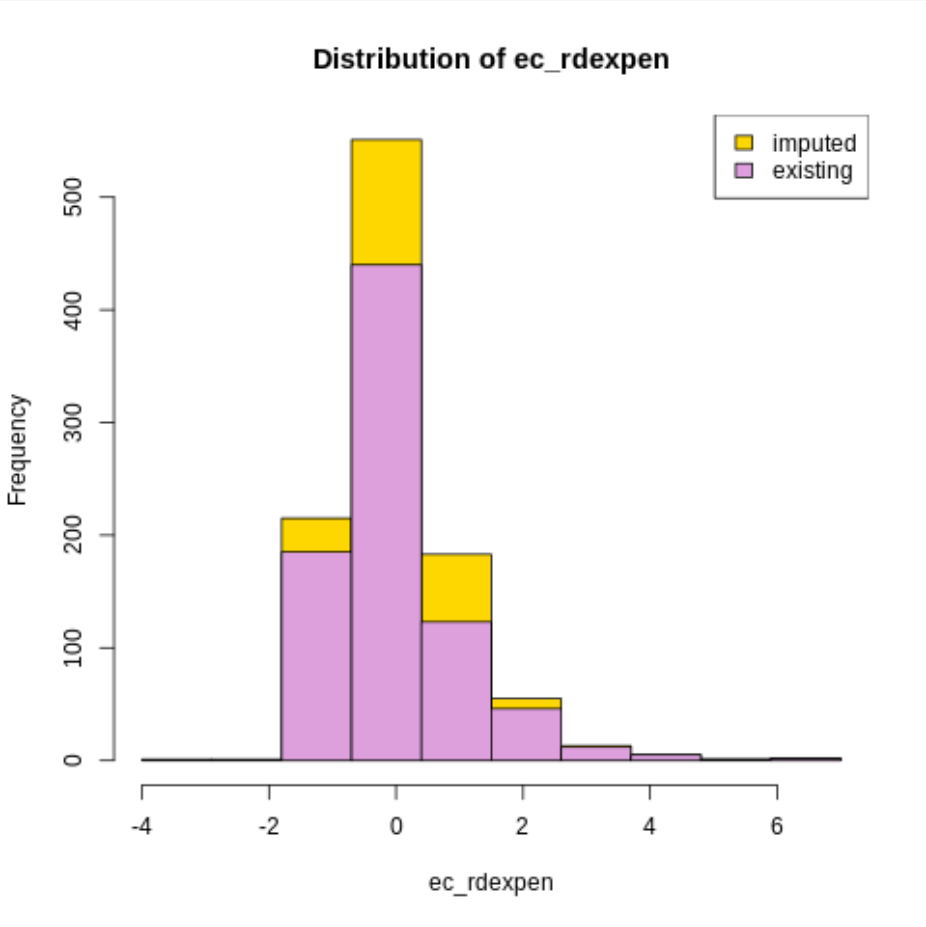
Building the dataset: need for imputation 2/3

- Multiple Imputation by Chained Equations (MICE)
- Key assumption:
  - Missing values are randomly distributed



Building the dataset: need for imputation 3/3





Shrinkage algorithms: regularization and variables selection

- 4 time stacks (2006, 2009, 2012, 2015)
- 258 NUTS2 regions
- 130 variables for 1032 observations
  - 258x130 when using cross-sections

Shrinkage algorithms: regularization and variables selection

- Our specification:

$$RTA_{r,t,i} = \beta_0 + \sum_{k=1}^{K} \omega_k x_{r,\,t,\,i,\,k} + \sum_{k=1}^{K} \gamma_k x_{r,\,t\text{-}1,\,i,\,k} + \sum_{k=1}^{K} \theta_k x_{r,\,t\text{-}2,\,i,\,k} + \theta pRTA_{r,\,t\text{-}1,\,i}$$

- OLS

$$\sum_{i=1}^{M} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{M} \left(y_i - \sum_{j=0}^{p} w_j \times x_{ij} \right)^2$$

- LASSO

$$\sum_{i=1}^{M} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{M} \left(y_i - \sum_{j=0}^{p} w_j \times x_{ij} \right)^2 + \lambda \sum_{j=0}^{p} |w_j|$$

- Ridge

$$\sum_{i=1}^{M} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{M} \left(y_i - \sum_{j=0}^{p} w_j \times x_{ij} \right)^2 + \lambda \sum_{j=0}^{p} w_j^2$$

- Elastic Net

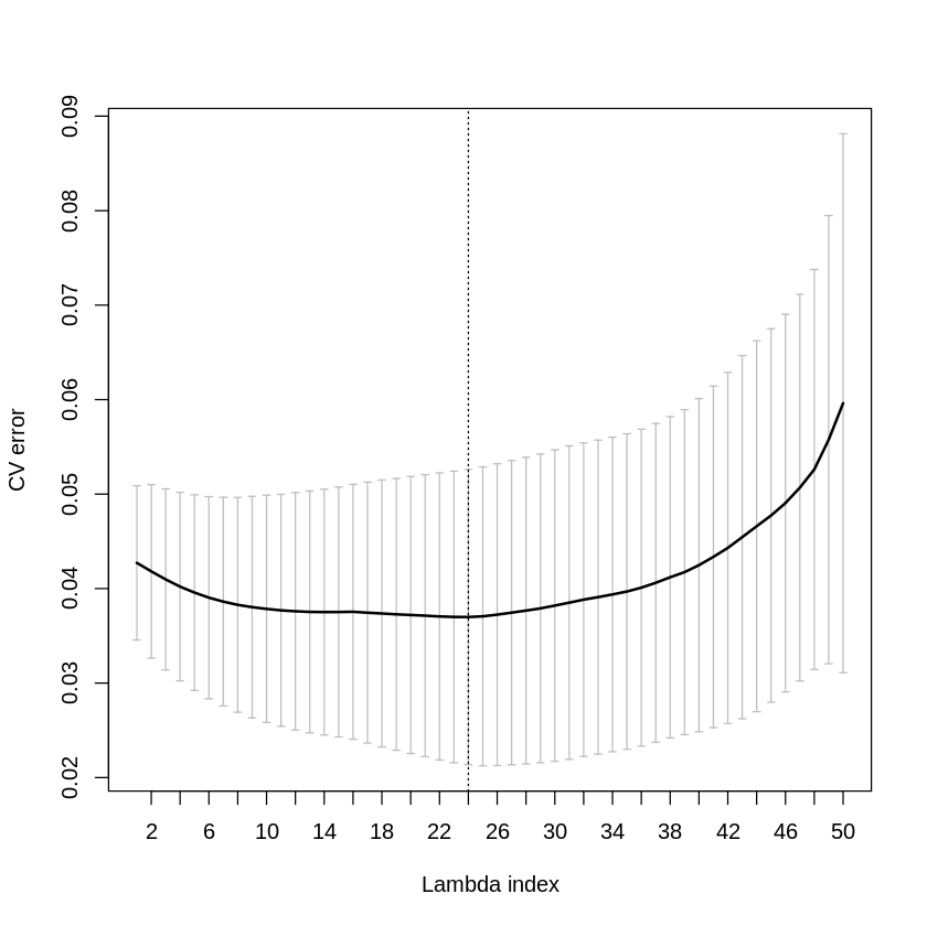
$$\hat{\beta} = \arg\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1$$

Shrinkage algorithms: regularization and variables selection

Implementation → Elastic Net (no interactions) and LASSO (hierarchical interactions) with optimized choice of MSE for 10 fold cross-validation:

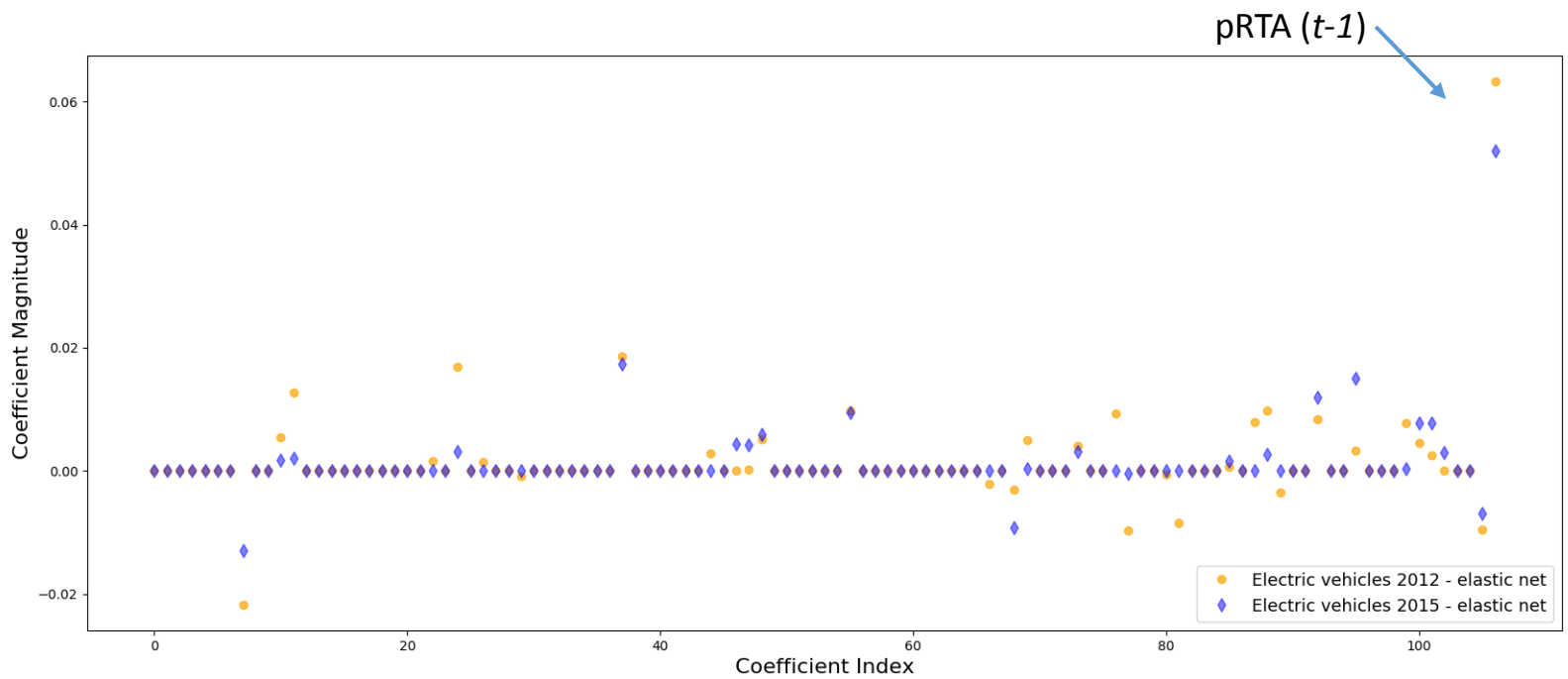
Parameters tuning:

- Lambdas (magnitude of penalty)
- L1/L2 ratio
- Accuracy scores of around 0.3



Regularization results

- Cross-sectional selections:
  - Cross-sections from the panel
  - No interaction factors
  - No time effects



Regularization results (2012 cross section)

	t	lag1	lag2	all
Total duration	12	12	17	41
Activity rates ISCED>3	11	12	13	36
Age dependency ratio (0-19 and over 60 to pop. aged 20-59)	4	5	5	14
Scientists and engineers	6	4	4	14
Long term uneployment (12 months or longer) in thousands	6	2	6	14
Students (ISCED 5-6) at regional level - as % of total country level students (ISCED 5-6)	0	6	7	13
Persons employed in science and technology	8	3	2	13
Gross domestic expenditure on R&D million EUR bussines enterprise sector	6	4	3	13
HH Paid current taxes on income wealth etc. mil EUR	2	5	4	11
SBS Manufacturing Share of employment in manufacturing total - percentage	11	0	0	11
Motorways networks in KM	6	2	3	11
Unemployment rate by age	3	4	4	11
Average number of usual weekly hours in main job by age in hours	4	3	4	11
Gross domestic expenditure on R&D million EUR all sectors	4	4	1	9
Participation rate in education and training (last 4 weeks) total age 25-64	0	5	4	9
Gross domestic expenditure on R&D million EUR government	5	2	1	8
Total R&D personnel higher education sector full time equivalent (FTE)	4	2	2	8
Proportion of population aged 20-39	4	2	2	8
Intramural R&D expenditure - Government	8	0	0	8

Second stage: regularization and variables selection

- De-meaning (time-effects):
  - Dummies could be, econometrically, a problem for LASSO / Elastic Net (Choi, Park and Seo, 2012)
  - Whole panel were the observations by regions are one-way demeaned over the years (fixed-effects estimator)

Regularization results (one-way demeaned panel)

	t	lag1	lag2	all
Total duration	12	12	17	41
Activity rates ISCED>3	11	12	13	36
Age dependency ratio (0-19 and over 60 to pop. aged 20-59)	4	5	5	14
Scientists and engineers	6	4	4	14
Long term uneployment (12 months or longer) in thousands	6	2	6	14
Students (ISCED 5-6) at regional level - as % of total country level students (ISCED 5-6)	0	6	7	13
Persons employed in science and technology	8	3	2	13
Gross domestic expenditure on R&D million EUR bussines enterprise sector	6	4	3	13
HH Paid current taxes on income wealth etc. mil EUR	2	5	4	11
SBS Manufacturing Share of employment in manufacturing total - percentage	11	0	0	11
Motorways networks in KM	6	2	3	11
Unemployment rate by age	3	4	4	11
Average number of usual weekly hours in main job by age in hours	4	3	4	11
Gross domestic expenditure on R&D million EUR all sectors	4	4	1	9
Participation rate in education and training (last 4 weeks) total age 25-64	0	5	4	9
Gross domestic expenditure on R&D million EUR government	5	2	1	8
Total R&D personnel higher education sector full time equivalent (FTE)	4	2	2	8
Proportion of population aged 20-39	4	2	2	8
Intramural R&D expenditure - Government	8	0	0	8

Second stage: regularization and variables selection

- Interactions:
  - Hierarchical interactions (Lim and Hastie, 2014)
  - *Glinternet* R-package
  - Interactions forced with pRTA
    - 1032 observations and 260 variables

Hierarchical interactions (Panel FE) – Interactions

	t	l1	l2	tot
Total duration	0	2	3	5
Activity rates ISCED >3	2	1	1	4
HH Net social contributions mil EUR	0	1	1	2
Self-employed persons	0	2	0	2
Total R&D personnel bussiness enterprise sector full time equivalent (FTE)	1	1	0	2
Unemployment Rate (Females)	2	0	0	2
Adult Participation in Learning	1	0	0	1
Economic activity rates age over 15 %	1	0	0	1
Gross domestic expenditure on R&D million EUR government	0	1	0	1
Other knowledge-intensive services	1	0	0	1
Participation rate in education and training (last 4 weeks) total age 25-64	0	1	0	1
Tertiary Educational Attainment (females)	1	0	0	1

Limitations and further research

1. pRTA estimations (might have too low predictive power)
   1. RTA smoothing over the stacks
   2. Different source of geolocated data
2. Imputation methodology
3. Exploit the pipeline to add more variables
   1. Relationship with RCA and regional trade figures
4. Better understanding of differences among green technologies
5. Third-stage approach:
   - Causal approach (Balland et al 2019)
   - Different path for green innovation vs other technological developments?

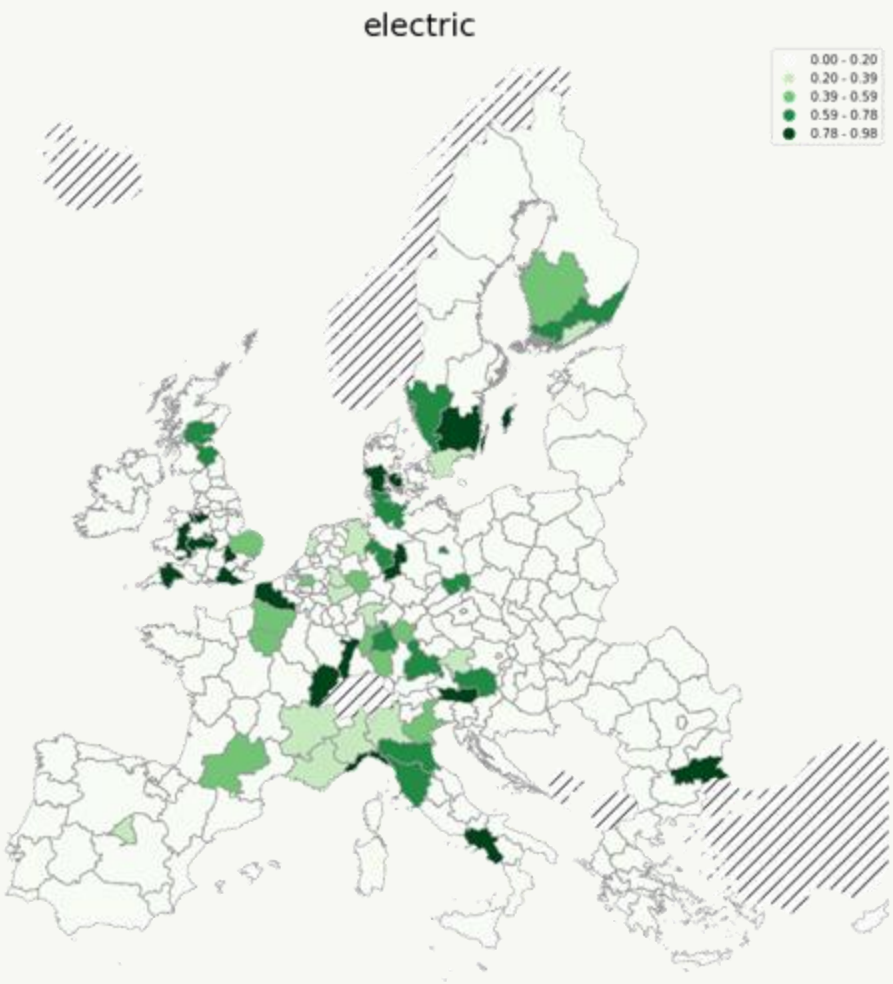
Conclusions

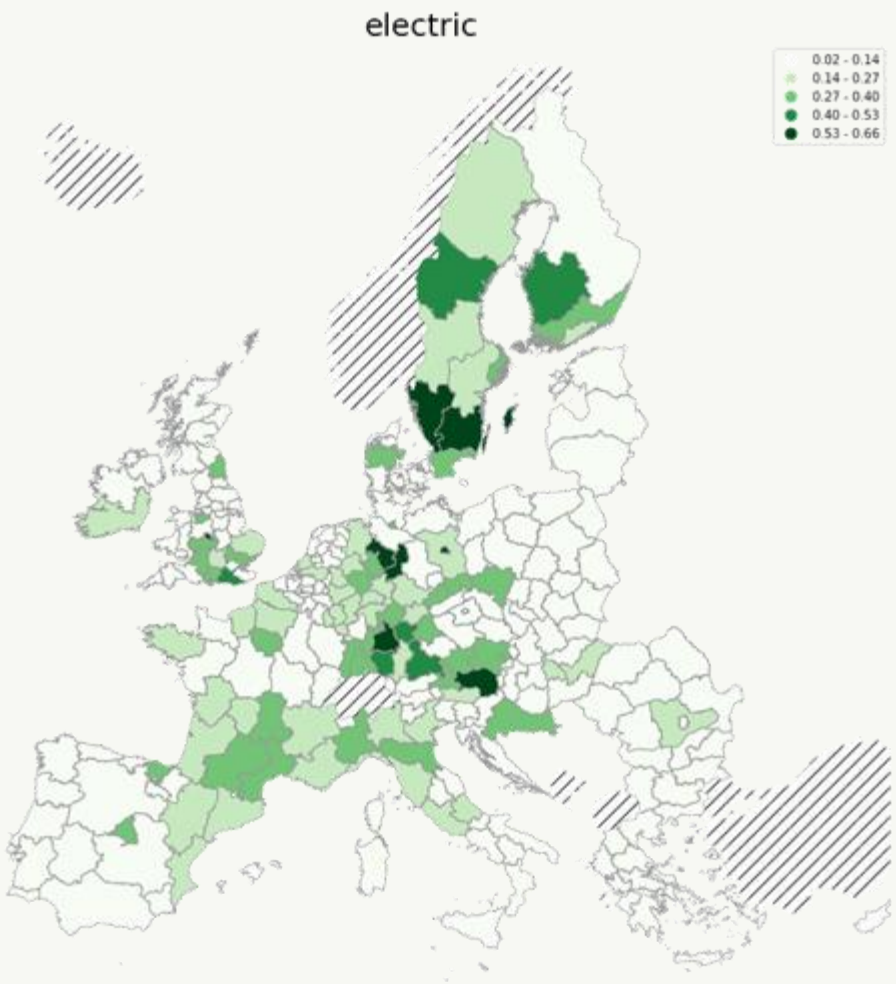
- Specialization and technological potential are largely differentiated within European countries

- Two hubs in France and Germany are dominating the scene for green innovation in the EU

- Our exploratory approach identifies groups of variables that are associated, via a potential advantage, with technological advantage:
  - the activity rate and employment of highly educated workers
  - total duration of employment
  - public and private R&D spending
  - Self-employment, adult participation in learning and female participation to the labour market

Thank you!


electric
0.00 - 0.20
0.20 - 0.39
0.39 - 0.59
0.59 - 0.78
0.78 - 0.98


electric
0.02 - 0.14
0.14 - 0.27
0.27 - 0.40
0.40 - 0.53
0.53 - 0.66