Examples of FP network analyses: SNA measures and selected regression methods

Thomas Scherngell

AIT Innovation Systems Department

Course

SOCIAL NETWORK ANALYSIS. INTRODUCTION TO METHODS AND APPLICATIONS TO THE EUPRO DATABASE

LUGANO, FEBRUARY 16-18 2016

Outline

Part 1: Examples of SNA approaches to describe FP networks

Part 2: Examples of regression approaches to analyse FP network

- Determinants of university participation in FPs
- Determinants of European FP networks from a spatial interaction modelling perspective

Examples of SNA approaches to describe FP networks

Definition of FP networks using EUPRO

Three main choices in EUPRO

I. organisation to organisation (one-mode)

II. project to project (one-mode)

III. organisation to project (two-mode, bipartite)

Bipartite network has more information than the others; the two projection networks throw away information (do you actually need it?)

Part 1: FP analysis from a SNA perspective

- Visualising and describing networks is an important exploratory step in our analyses of R&D networks
  - How many organisations and interactions can be observed in the network under consideration?
  - Where can we find intensive collaboration, dense network areas, etc.?
  - What is the thematic and/or spatial distribution of the network?

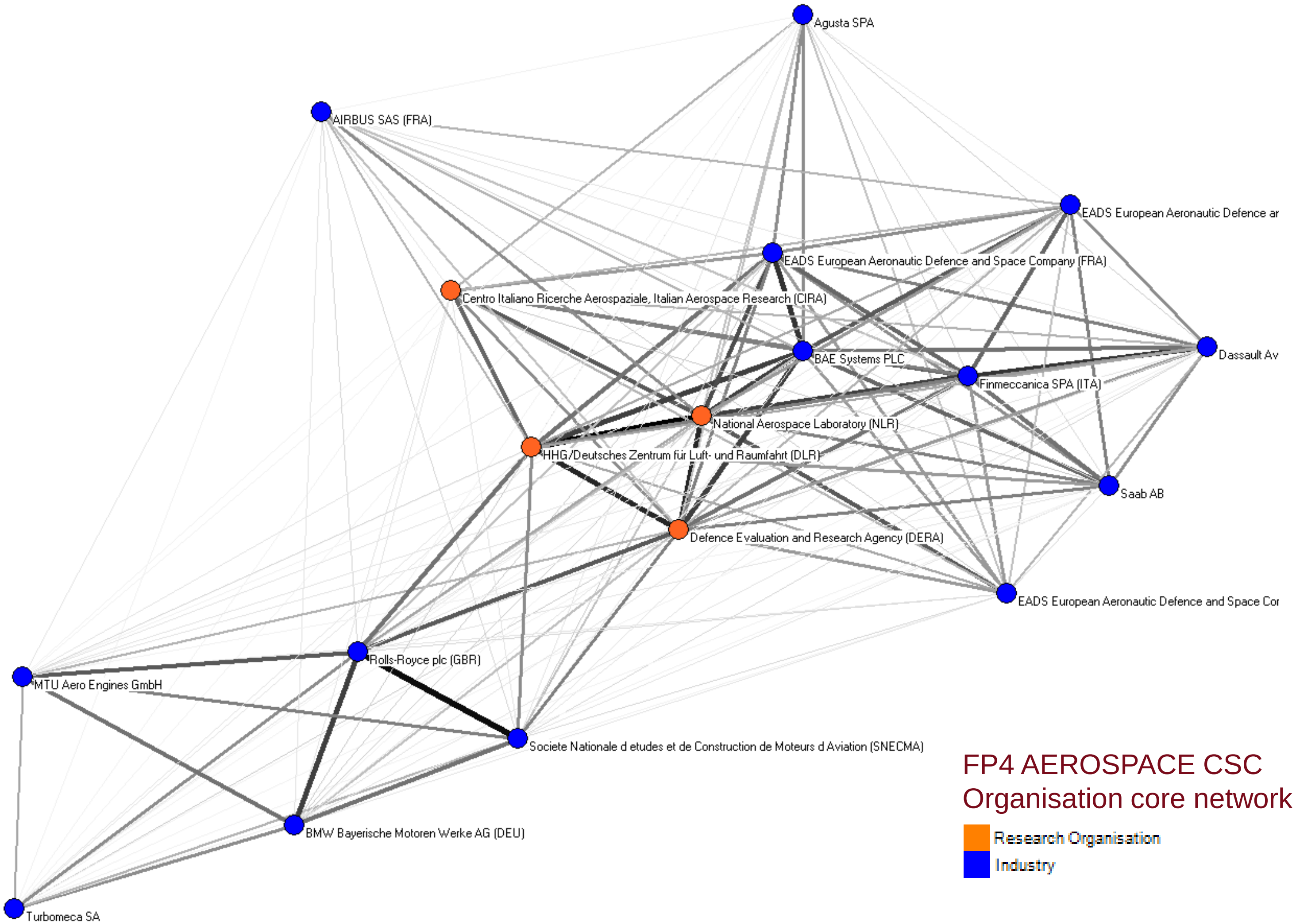
- Investigating network structures of R&D networks
  - concerning their connectedness, centralisation, cliquishness, and heterogeneity
  - Do network structures differ across themes?
  - Do network structures differ across political instruments?
  - What role do different organisation types play?
  - What role do different countries play?
  - What is the backbone of the network?
  - Is the network characterized by relevant, thematically distinct community groups?

Example from EUPRO: Structure of FP networks

- Background and Objectives
  - Analysis of collaborative networks promoted by the European Framework Programmes
  - Exploit the richness of FP data through social network analysis to contribute to the progress of monitoring the move towards the European Research Area (ERA)

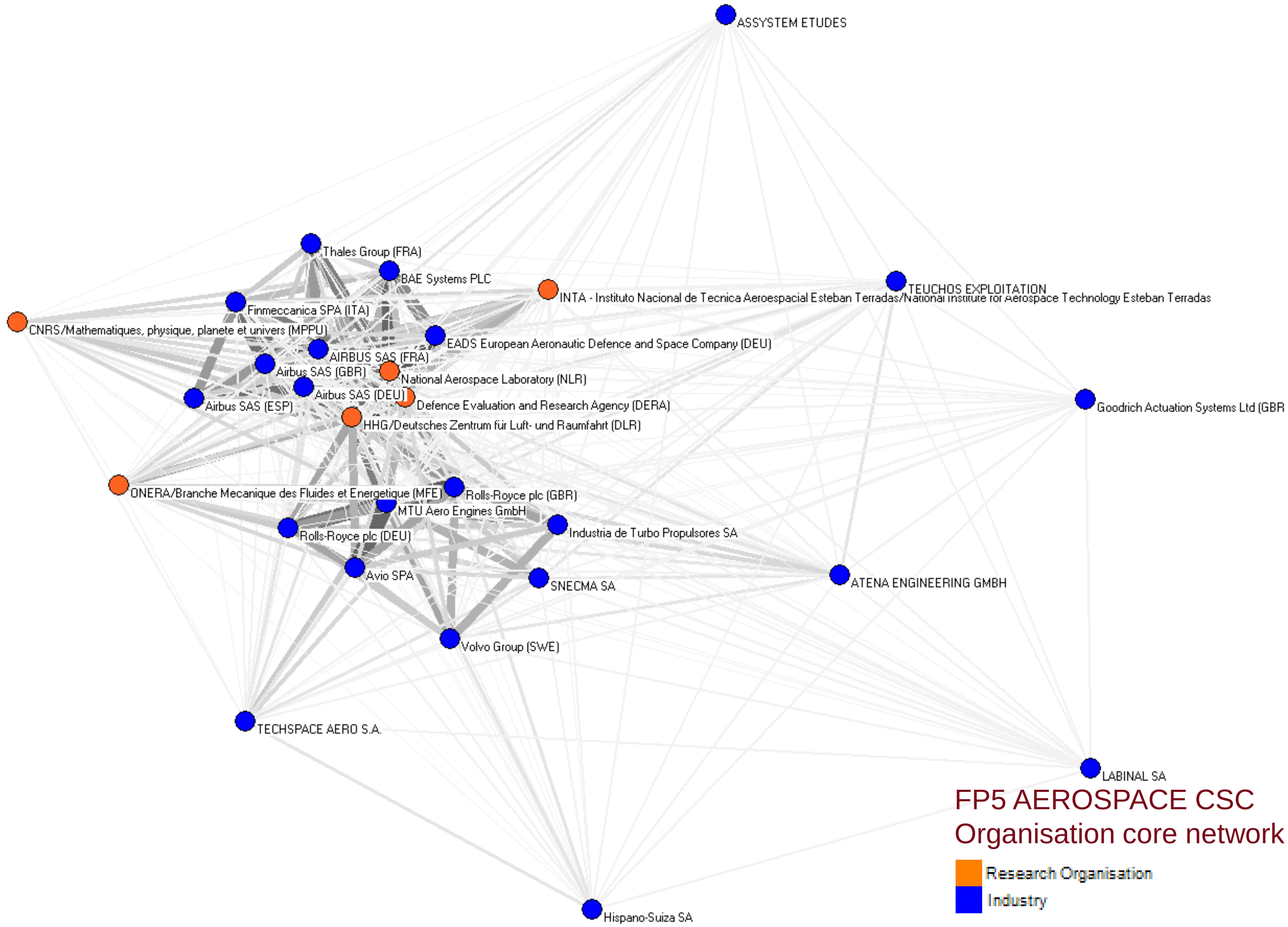
- Focal points
  - Global characteristics of (thematically or spatially distinct) FP networks
  - Local characteristics, i.e. who are the central players, and how are they inter-connected?
  - Dynamics of global and local characteristics

Network visualisations (I) (spring model approach)



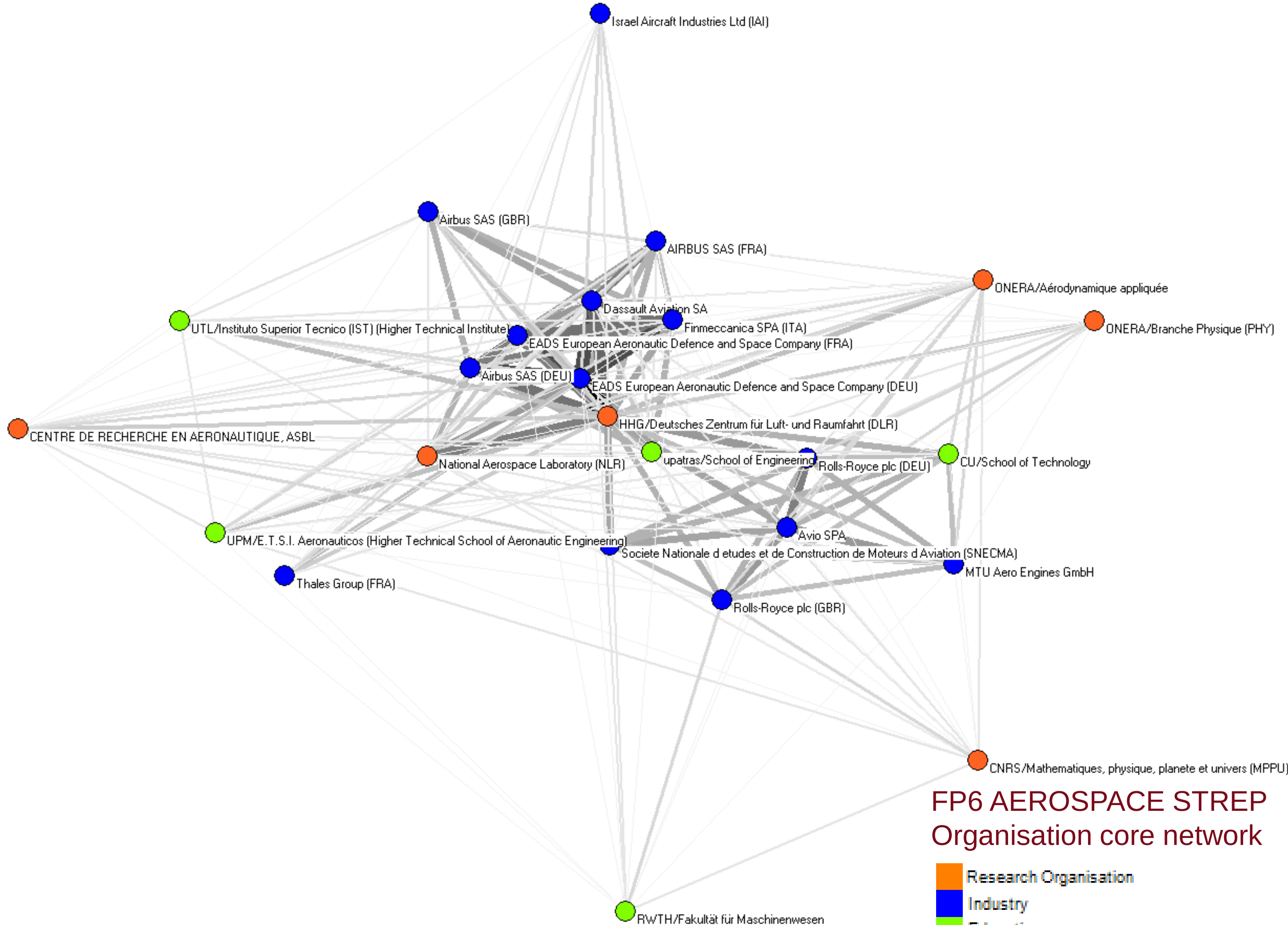
FP4 AEROSPACE CSC
Organisation core network

Network visualisations (II) (spring model approach)



FP5 AEROSPACE CSC
Organisation core network

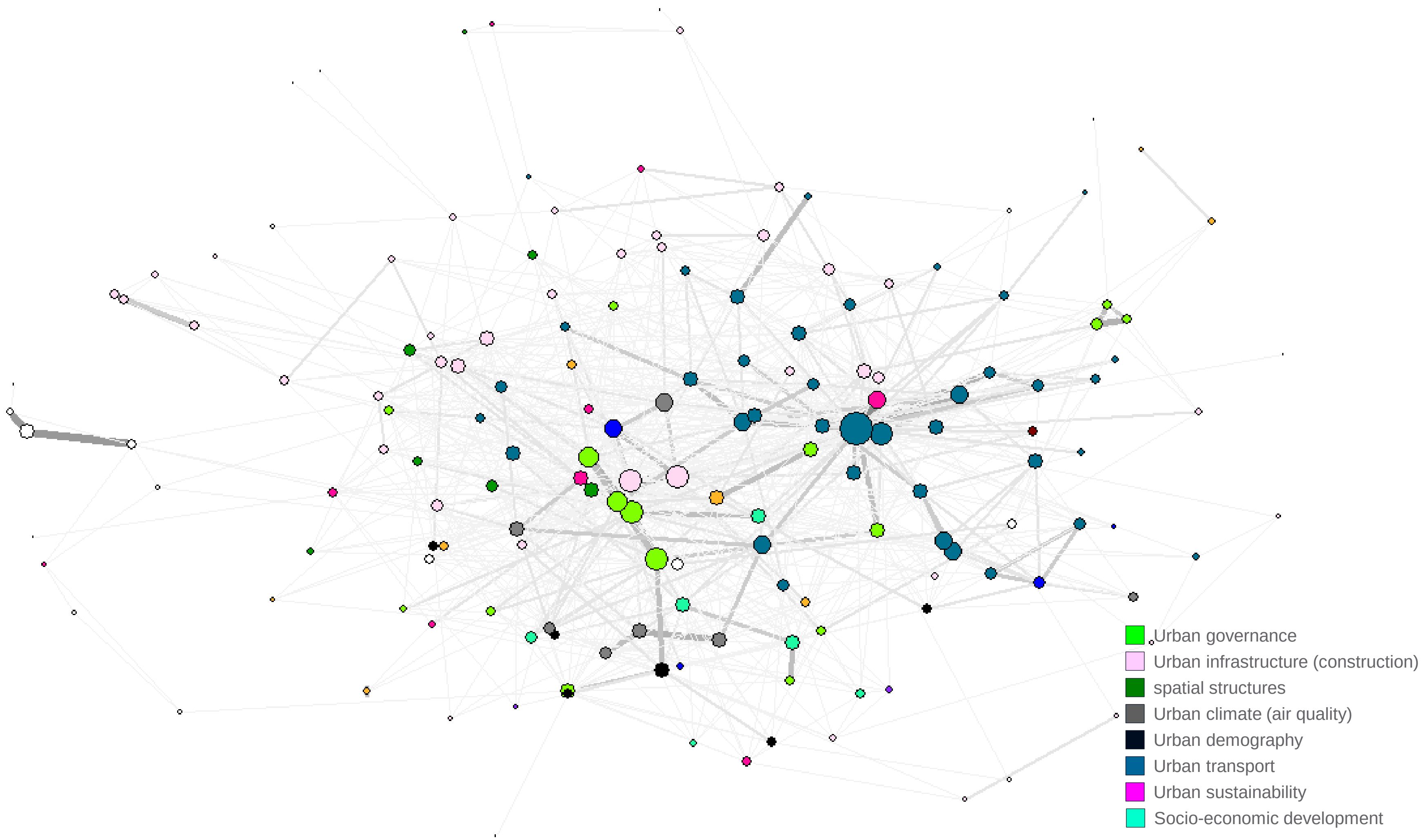
Network visualisations (III) (spring model approach)



FP6 AEROSPACE STREP
Organisation core network

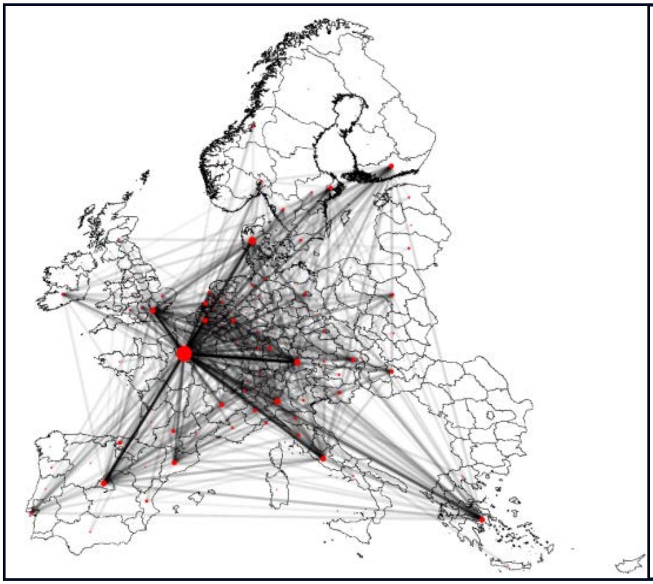
Research Organisation
Industry
Education

Network visualisations (IV) (spring model approach)

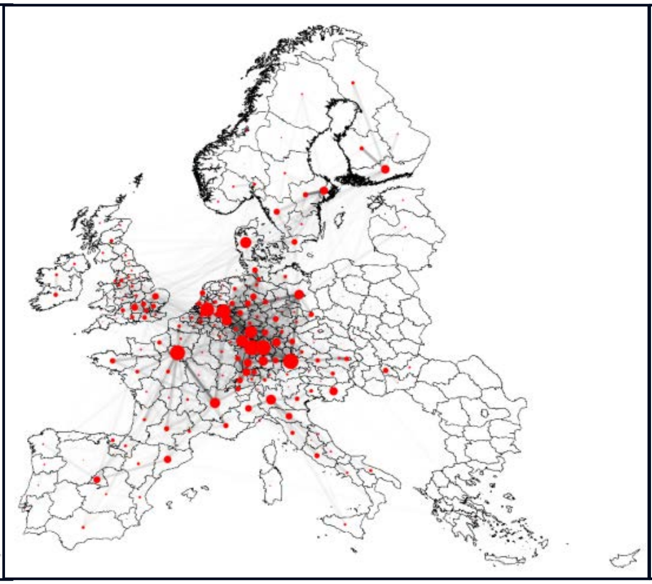


Network visualisations (V) (spatial approach)

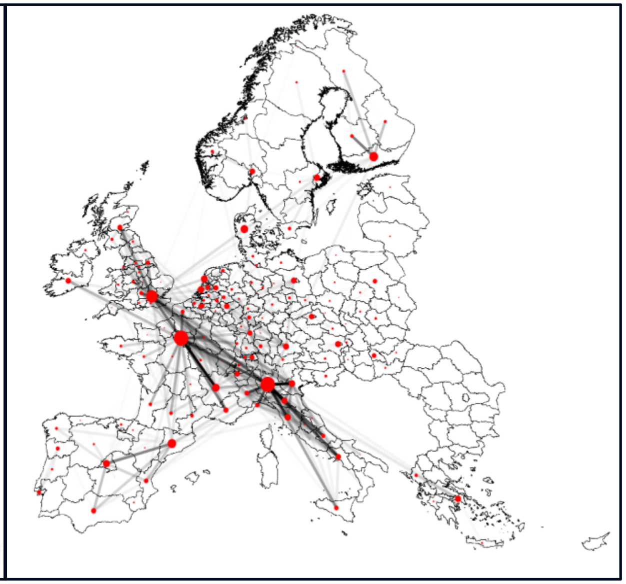
FP-network



co-patent network



co-publication network



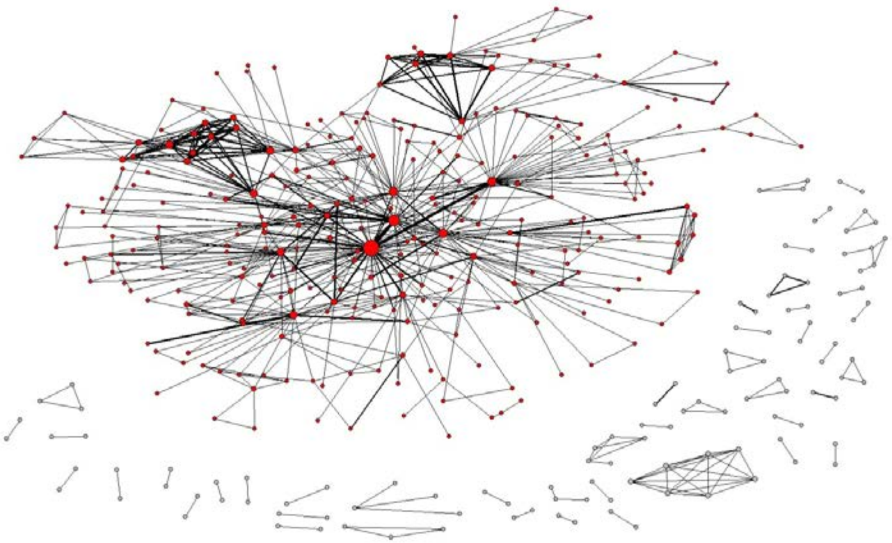
Positive links	
Matrix elements	43,693
Sum	1015,336

Positive links	
Matrix elements	8,896
Sum	268,498

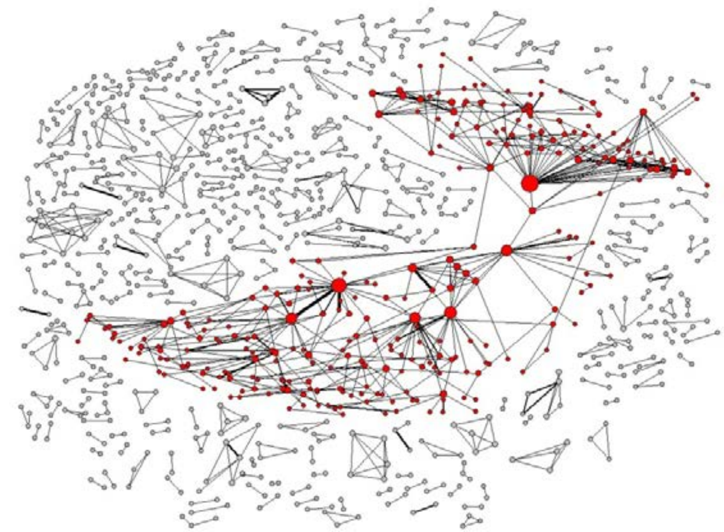
Positive links	
Matrix elements	36,391
Sum	1868,686

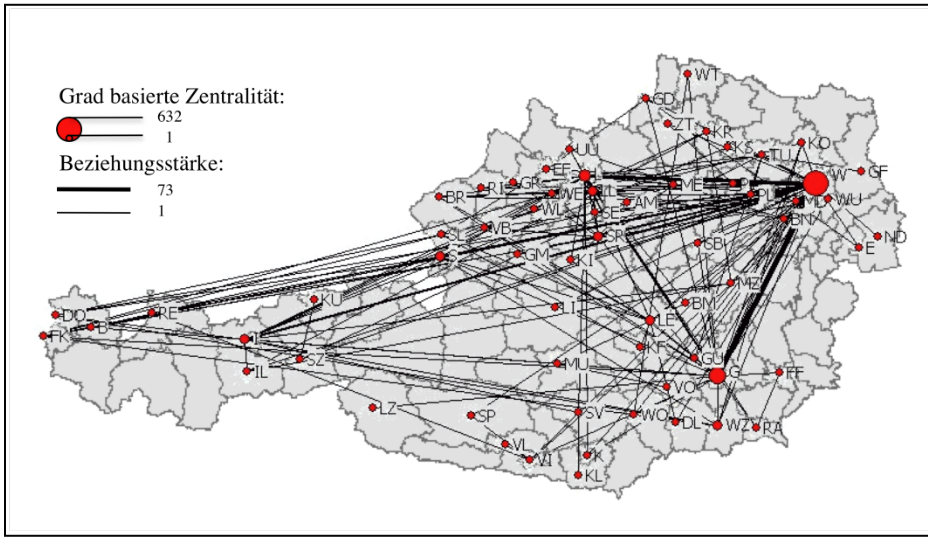
Network visualisations (VI)
(spring model vs. spatial approach)

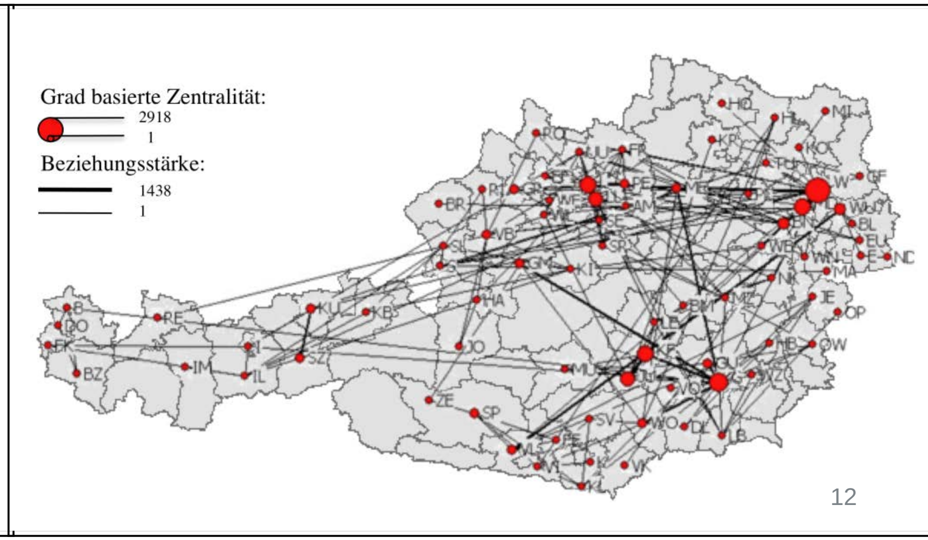
A: FP network of Austria



B: Co-patent network of Austria







Basic SNA measures across different FPs

Graph Characteristic	FP1	FP2	FP3	FP4	FP5	FP6
No. of vertices N	2,116	5,758	9,035	21,599	25,840	17,632
No. of edges M	9,489	62,194	108,868	238,585	385,740	392,879
No. of components	53	45	123	364	630	26
N for largest component	1,969	5,631	8,669	20,753	24,364	17,542
Share of total (%)	93.05	97.79	95.95	96.08	94.29	99.49
M for largest component	9,327	62,044	108,388	237,632	384,316	392,705
Share of total (%)	98.29	99.76	99.56	99.60	99.63	99.96
N for 2nd largest component	8	6	9	10	12	9
M for 2nd largest component	44	30	72	90	132	72
Clustering coefficient	0.65	0.74	0.74	0.78	0.76	0.80
Diameter of largest component	9	7	8	11	10	7
l largest component	3.62	3.21	3.27	3.45	3.30	3.03
Mean degree	9.0	21.6	24.1	22.1	29.9	44.6
Fraction of N above the mean (%)	29.4	28.0	23.6	22.4	23.5	26.1
Mean vertex size	3.0	3.1	3.3	3.0	2.8	2.7
Standard deviation	5.0	6.1	7.7	7.9	6.8	5.4

Basic SNA measures across different FP communities

Table 2 Properties of the FP5 Communities

	Aerospace	Aquatic resources	Electronics	Environment	Ground transport	Information processing	Life sciences	Sea transport
Vertices *n*	1,146	81	2,307	1,855	686	40	2,366	218
Edges *m*	13,870	451	30,456	23,155	5,251	226	33,178	2,978
Average path length	2.669	2.199	2.732	2.797	2.549	1.731	2.713	2.030
Density	0.021	0.139	0.010	0.013	0.022	0.290	0.012	0.126
Skewness	4.263	1.169	5.132	4.512	6.739	1.097	4.749	1.718
Mean degree	24.206	11.136	26.403	24.965	15.309	11.300	28.046	27.321

Basis centrality measures on the position of universities

Overview on top ranked universities in FP7

(three highest degree ranks marked in bold and different colours)

Univ.	Degree	Univ.	Eigen-vector	Univ.	Between-ness	Univ.	Close-ness
www.dtu.dk	0.095286	www.cam.ac.uk	0.128687	**www.kuleuven.be**	0.018065	**www.dtu.dk**	0.513087
www.kuleuven.be	0.093381	www.ox.ac.uk	0.126532	**www.dtu.dk**	0.017988	**www.kuleuven.be**	0.510476
www.tudelft.nl	0.080219	www.ethz.ch	0.125548	**www.tudelft.nl**	0.013404	**www.tudelft.nl**	0.507084
www.imperial.ac.uk	0.077413	**www.dtu.dk**	0.122414	www.imperial.ac.uk	0.010850	www.imperial.ac.uk	0.505279
www.manchester.ac.uk	0.073707	www.imperial.ac.uk	0.120929	www.manchester.ac.uk	0.010150	www.manchester.ac.uk	0.504193
www.ethz.ch	0.073014	www.manchester.ac.uk	0.113072	www.ethz.ch	0.010126	www.ethz.ch	0.503183
www.cam.ac.uk	0.072079	**www.kuleuven.be**	0.110657	www.unibo.it	0.009620	www.cam.ac.uk	0.502995
www.ox.ac.uk	0.071283	www.ucl.ac.uk	0.109182	www.ucl.ac.uk	0.009562	www.ox.ac.uk	0.502301
www.ucl.ac.uk	0.068685	www.epfl.ch	0.102404	www.epfl.ch	0.009555	www.unibo.it	0.502035
www.epfl.ch	0.065810	**www.tudelft.nl**	0.098362	www.rwth-aachen.de	0.009486	www.ucl.ac.uk	0.500980

Note: Full ranking and url assignment to university names given in attached xls file

Examples of regression approaches to analyse FP network

- Determinants of university participation in FPs
- Determinants of European FP networks from a spatial interaction modelling perspective

Regression approaches: Typical questions

- Investigation of mechanisms of network constitution and dynamics
  - What are the determinants of network participation? (e.g., Lepori et al. 2015)
  - What are the determinants of partner choice in a network? (e.g., Paier and Scherngell 2011)
  - Which economic, technological, social and geographic conditions affect the constitution of observed collaboration patterns and dynamics? (e.g., Scherngell and Barber 2009 and 2011, Scherngell and Lata 2013)
- Most basic form: regression models focusing on node characteristics (such as number of links, e.g. Lepori et al. 2015)
- Advanced form: regression models explaining links and their magnitude (often spatial interaction models, see, e.g., Scherngell and Lata 2013)

A short introduction on the basics of regression analysis

A statistical instrument to investigate the relationship between on dependent and one or more independent variables

Regression analysis tests pre-assumed structures, i.e. it is based on a strong theoretical foundation

Terminology

y	x_1, x_2, x_3, x_4, …
Dependent variable	independent variable
endogenous	exogenous
explained	explaining
response variable	predictor variable

The classical linear regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_K x_K + \varepsilon$$

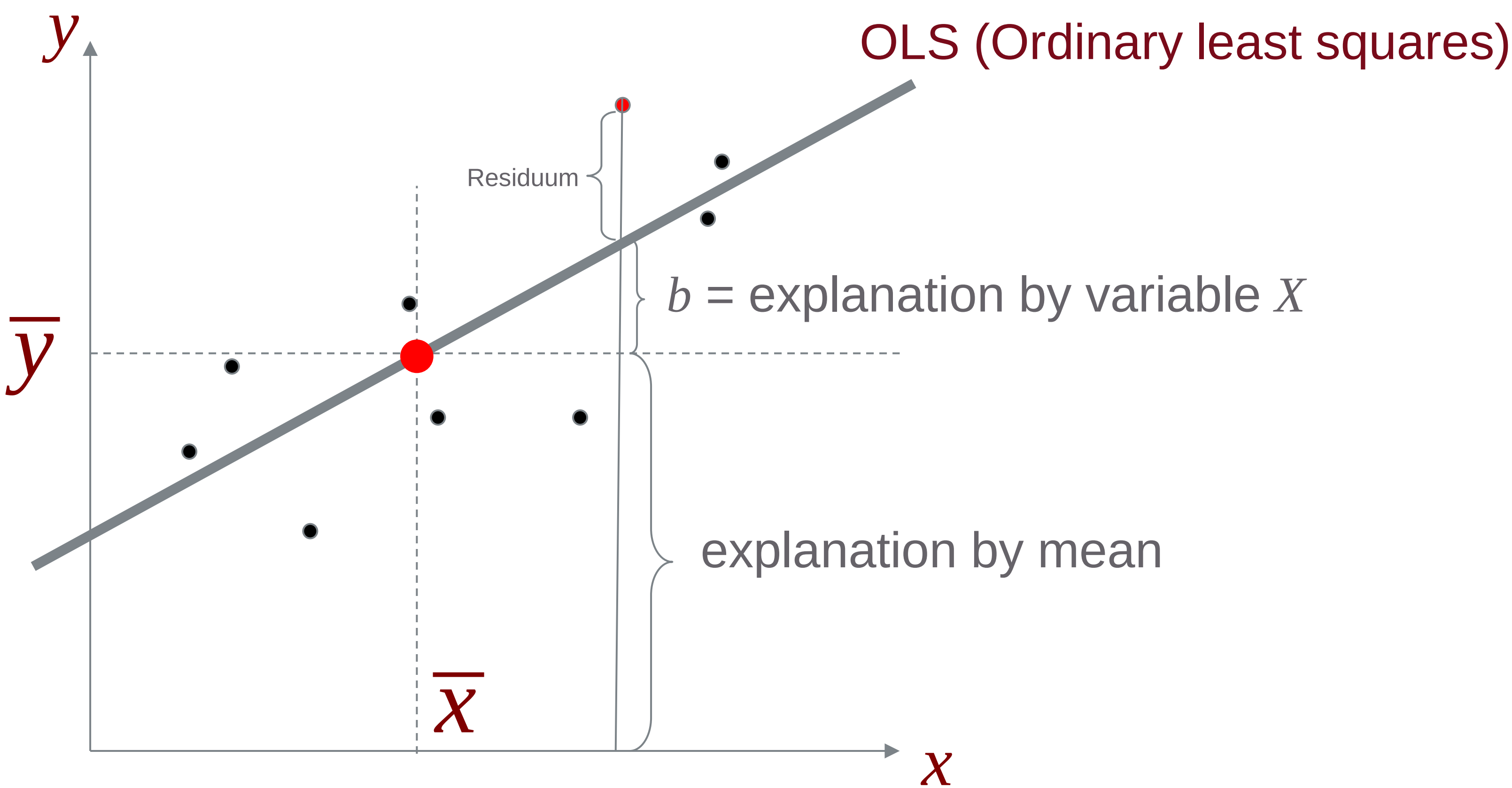
y	dependent variable
$x_1, \ldots, x_K$	independent variable
$\beta_1, \ldots, \beta_K$	regression coefficients
β_0	constant/ intercept
ε	error term

In matrix notation

$$\mathbf{y} = \boldsymbol{\beta}\mathbf{X}\ \boldsymbol{\varepsilon} +$$

$\mathbf{y}$	$(n, 1)$ vector of observations on the dependent variable
$\mathbf{X}$	(n, K) matrix of observations on the independent variables
$\boldsymbol{\beta}$	$(K, 1)$ vector of regression coefficients (inlc. constant β_0)
$\boldsymbol{\varepsilon}$	$(n, 1)$ vector of error term

The OLS estimator



$$\boldsymbol{b} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{\beta} + \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}$$

Main assumption of OLS regression

- Linear functional form
- Independent and identically distributed (iid, i.e. same probability distribution among all variables)
- The regressors must be linearly independent from each other (no multicollinearity, full rank)
- Spherical errors (homoskedisticity, no autocorrelation)
- Normality of residuals

Regression evaluation

- Global evaluation
  - Coefficient of determination (R-squared)
  - F-Test
  - Log-Likelihood and information criteria

- Evaluation of regression coefficients (t-test)

- Model diagnostics for assumptions
  - Test statistics for heteroskedasticity (e.g. Breuch-Pagan test)
  - Test of normality (e.g., Jarque-Bera test)
  - Multicollinearity condition number
  - Tests for non-spherical disturbances (e.g. unit root tests, Moran´s for spatial autocorrelation)
  - Residual plots (linear functional form assumption)

Limited dependent (non-metric) variables …

… as one major source for violating basic assumptions

→ alternative model specifications, most importantly (Long and Freese 2001)

- Binary dependent variables: Logit or probit regression models
- Ordinal or multinomial categorical outcomes: Ordinal or multinomial logit regression models
- Count nature of dependent variables: Poisson and Negative Binomial regression models (inlc. zero inflated models)
- Censored dependent variables: Tobit regression models

An example of basic regressions analysing FP networks

- Such basic regressions explain node characteristics by regression approaches

- The represent the most basic approach
  - Understanding how much network structure is associated with organizational covariates
  - Disregarding network effects.

- Example: Explaining the participation (degree) of universities in FP networks by university characteristics
  - Testing to which extent we can associate the number of participations with the university characteristics, like size, international reputation, disciplinary characteristics
  - Based on the matching of EUPRO with ETER.

Benedetto Lepori, Valerio Veglio, Barbara Heller-Schuh, Thomas Scherngell, Michael Barber (2015), *Participations to European Framework Programs of Higher Education Institutions and their association with organizational characteristics*, Scientometrics, 105(3), 2149-2178.

Regression approach

- Dependent variable: count of the number of participations per HEI
- Independent variables
  - Size (FTE staff)
  - International reputation = publications * NI / staff
  - Subject composition

- Dependent is count and highly skewed with a large of “0” (more than half of the sample): OLS will not work
  - Poisson/NB: suitable, but the estimator is less powerful
  - Two-stage approach:
    - Logistic for the probability of participating
    - Truncated regression on log(participations)
  - Dealing with country effects
    - Clustering standard errors
    - Introducing country dummies (FE)

Summary results and lessons

- The two models predict with high precision the probability of participating
- Logistic model: 88% of correctly classified cases (against 61% of the null model)
- Truncated regression
  - 77% of the variance explained in the number of participations
  - Most important factor: size. Scale factor = 1
  - Reputation is very important.
  - Country effects are not relevant
- What we learn
  - Organizational characteristics do matter in network structure
  - University networks have a simple reputational structure (core-periphery)
  - Simple models provide important preliminary insights

Advanced regression techniques …

… focus is not on node characteristics, but on the explanation of links and their magnitude

Spatial interaction model has come into fairly wide use in this context:

- used to model network links between discrete units in geographical space (see, e.g., Sen and Smith 1995),
- such as migration, commuting, telecommunication flows, or, more recently, knowledge flows and R&D collaborations
- Relating magnitude of links to different kinds of variables

The general spatial interaction model

$$Y(i,j) = \mu(i,j) + \varepsilon(i,j) \qquad i, j = 1, \ldots, n \qquad (1)$$

- $Y(i,j)$: Random variable, corresponding to observed interactions y_{ij} between i and j
- $\mu(i,j)$: expected mean interaction frequency
- $\varepsilon(i,j)$: Disturbance term about the mean
- i, j: spatial units (regions)

with

$$\mu(i,j) = C \quad A(i) \quad B(j) \quad S(i,j) \qquad (2)$$

- C: constant
- $A(i)$: origin function
- $B(j)$: destination function
- $S(i,j)$: separation function

Functional specification: the multivariate exponential spatial interaction model (see, e.g., Fortheringham and O'Kelly 1989, Fischer and Wang 2011)

$$\mu(i,j) = C \prod_{q=1}^{Q} \left(A_{iq}\right)^{\beta_q} \prod_{r=1}^{R} \left(B_{jr}\right)^{\gamma_r} \exp\left\{-\sum_{k=1}^{K} \theta_k S^{(k)}(i,j)\right\} \qquad (3)$$

- A_{iq}: Q origin measures
- B_{jr}: R destination measures
- $S^{(k)}$: K separation measures

The Poisson spatial interaction model

- Least squares suffers from major drawbacks (Flowerdew and Aitkin 1982) due to the discrete nature of spatial interactions leading to biased OLS estimates (Cameron and Trivedi 1998)

- ML estimation of the parameters under more realistic distributional assumptions: $(Y_{ij}) \sim Poisson$ is one common assumption, so that

$$\Pr[Y(i,j) = y(i,j) \mid \mu(i,j)] = \frac{\exp[-\mu(i,j)][-\mu(i,j)]^{y(i,j)}}{y(i,j)!} \qquad i,j = 1, \ldots, n \qquad (4)$$

with

$$\begin{aligned} \mu(i,j) &= E[y(i,j) \mid A(i), B(j,\gamma), S(i,j)] \\ &= \exp[A(i,\beta)\, B(j,\gamma)\, S[(i,j),\theta]] \end{aligned} \qquad i,j = 1, \ldots, n \qquad (5)$$

- Equidispersion assumption: Equality of conditional mean and variance

- For spatial interaction data, overdispersion is common due to spatial heterogeneity → Negative Binomial specification as promising solution

An example (Scherngell and Lata 2013)

- objective is to estimate progress towards ERA,
- by identifying the evolution of separation effects affecting the probability of cross-region R&D collaborations in the European network of R&D cooperation,
    - Separation effects involve geographical, technological, cultural and institutional barriers
    - The European network of R&D cooperation is captured by joint participation of organisations in R&D projects funded by the FPs
- over the time period 1999-2006,
- within a Negative Binomial Spatial Filter Interaction Model Framework

Data

The core data set used for the study is extracted from EUPRO

→ tracing of the pan-European network of actors performing joint R&D

- based on 20,123 collaborative R&D projects, producing about 2.5 million collaboration links in total,
- disaggregated to the years 1999-2006, and to
- 255 NUTS-2 regions of the 25 pre-2007 EU member states, as well as Norway and Switzerland

Capturing space-time patterns of cross-region networks

$$\boldsymbol{y}_t(i,j) = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{pmatrix} \longrightarrow$$

The cross-region collaboration matrix $\boldsymbol{Y}$ for time period t

$$\vdots \qquad \vdots$$

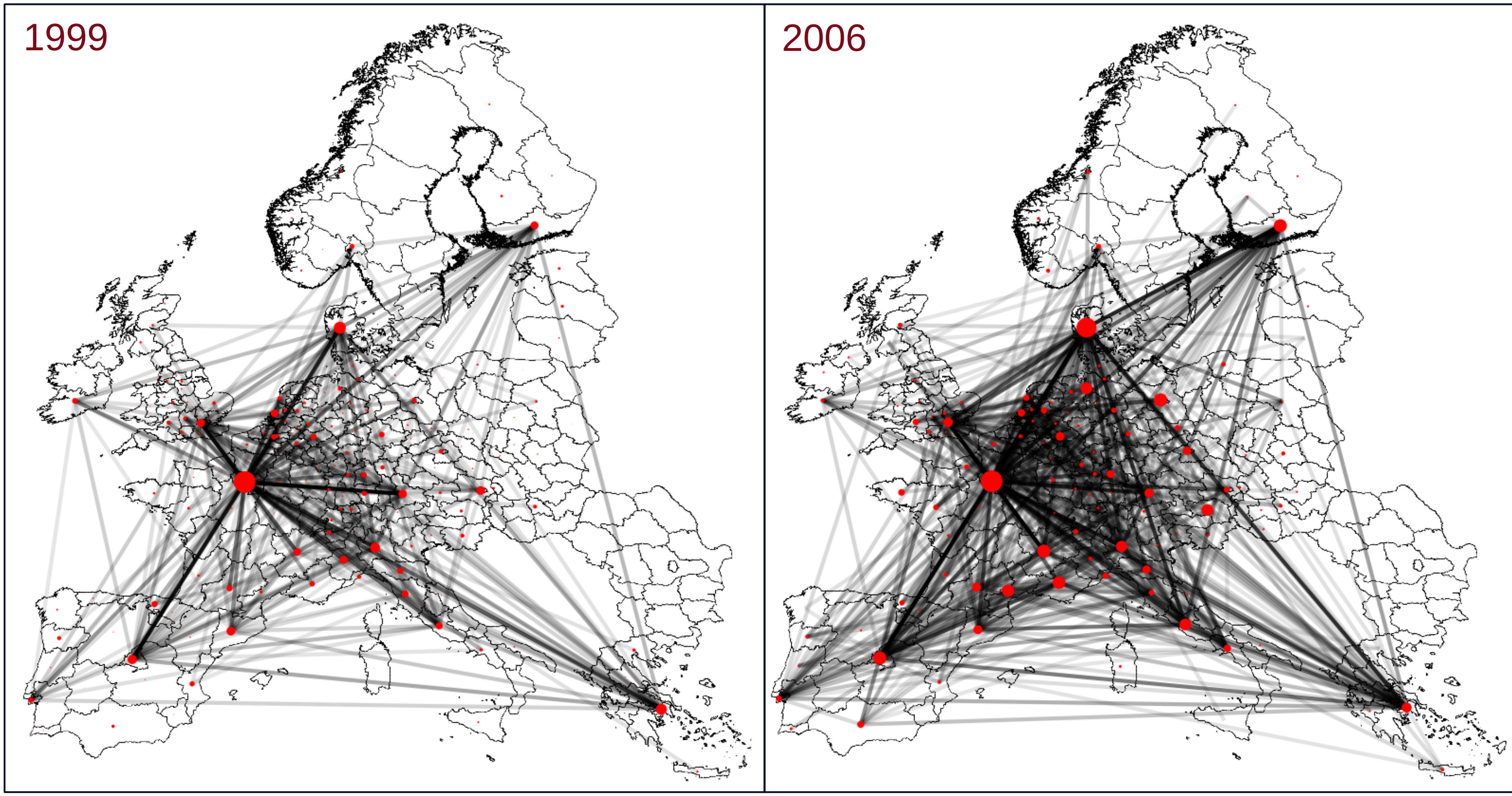
$$\boldsymbol{y}_T(i,j) = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{pmatrix} \longrightarrow$$

The cross-region collaboration matrix $\boldsymbol{Y}$ for time period T

for $i, j = 1, ..., n; \quad t = 1, ..., T$

The regional collaboration matrix $\boldsymbol{Y}$ for a given year t contains the collaboration intensities between all (i, j)-region pairs, given the $i = 1, ..., n = 255$ regions in the rows and the $j = 1, ..., n = 255$ regions in the columns

Spatial patterns of the R&D networks



Notes: node size corresponds to regional degree centrality, line transparency with the number of joint projects between two regions

The spatial interaction model

$$y_{ijt} = X_{ijt} + \varepsilon_{ijt} \qquad i, j = 1, \ldots, n; \quad t = 1, \ldots, T \tag{1}$$

Disturbance term

Function of covariates

Number of R&D collaborations between regions i and j at time t

255 Nuts2 regions

Time Period (1999-2006)

with

$$X_{ijt} = O_{it} \; D_{jt} \; S_{ijt} \tag{2}$$

$$O_{it} = o_{it}^{\alpha_{1t}} \tag{3}$$

Origin function: Number of organizations in region i in time period t

$$D_{jt} = d_{jt}^{\alpha_{2t}} \tag{4}$$

Destination function: Number of organizations in region j in time period t

$$S_{ijt} = \exp\left[\sum_{k=1}^{K} \beta_{kt} \; s_{ijt}^{(k)}\right] \tag{5}$$

Separation measures
$k = 1$ geographical distance between region i and j
$k = 2$ neighbouring region effects
$k = 3$ neighbouring country effects
$k = 4$ country border effects
$k = 5$ language border effects
$k = 6$ technological distance between region i and j

Incorporating (2) and (3)-(5) in (1) yields

$$y_{ijt} = o_{it}^{\alpha_{1t}} \; d_{jt}^{\alpha_{2t}} \; \exp\left[\sum_{k=1}^{K} \beta_{kt} \; s_{ijt}^{(k)}\right] + \varepsilon_{ijt} \tag{6}$$

The Negative Binomial spatial interaction model

- Least squares suffers from major drawbacks (Flowerdew and Aitkin 1982) due to the discrete nature of spatial interactions (Cameron and Trivedi 1998)

- ML estimation under more realistic distributional assumptions: $(Y_{ij}) \sim \mathit{Poisson}$ with heterogeneity ($\mathit{Negative\ Binomial}$)

$$\Pr\left(y_{ijt} \mid \underbrace{X^{*}_{ijt} = \mu_{ijt}}_{\text{Conditional mean}}\right) = \frac{\Gamma\left(y_{ijt} + \gamma^{-1}\right)}{\left[y_{ijt}!\,\Gamma(\underbrace{\gamma^{-1}}_{\text{Dispersion parameter}})\right]} \theta_{ijt}^{1/\gamma} \left(1 - \theta_{ijt}\right)^{y_{ijt}} \qquad \text{with } \theta = \gamma^{-1} / (\gamma^{-1} + \mu_{ijt}) \tag{6}$$

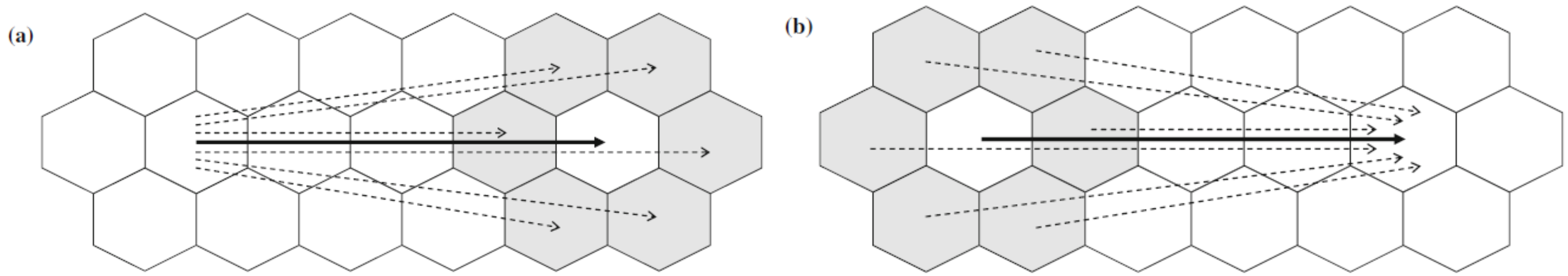
$$X^{*}_{ijt} = \mu_{ijt} = \exp\left[\alpha_{0t} + \alpha_{1t} \log\left(o_{it}\right) + \alpha_{2t} \log\left(d_{jt}\right) + \sum_{k=1}^{K} ß_{kt} s_{ijt}^{(k)} + \underbrace{\xi_{ijt}}_{\text{Stochastic heterogeneity term}}\right] \tag{7}$$

with

$$\exp\,(\xi_{ijt}) \sim \underbrace{\Gamma}_{\text{Gamma function}}(\gamma) \tag{8}$$

Recent extensions: Accounting for spatial autocorrelation

- The problem of spatial autocorrelation of flows violating ML estimates has been highlighted recently (see Chun 2008)



- Theoretical and statistical motivation for taking into account spatial autocorrelation of flows (Scherngell and Lata 2013)

- Eigenvector spatial filtering approach (Fischer and Griffith 2008)
  → applicable to different distributional assumptions but spatial information is filtered out

- Spatial autoregressive form (LeSage and Fischer 2010)
  → not yet efficiently applicable to Poisson specifications

Spatial filtering approach

- Spatial filtering approach, following Griffith (2007), extracts $\boldsymbol{E}_n$ eigenvectors from a modified spatial weights matrix $\boldsymbol{W}^*$

$$\boldsymbol{W}^* = (\underbrace{\boldsymbol{I}}_{\substack{n\text{-by-}n \\ \text{identity} \\ \text{matrix}}} - \boldsymbol{\iota}\underbrace{\boldsymbol{\iota}^T}_{\substack{n\text{-by-}1 \\ \text{vector} \\ \text{of ones}}} \tfrac{1}{n}) \underbrace{\boldsymbol{W}}_{n\text{-by-}n \text{ spatial weights matrix, with elements } w_{ij} = \begin{cases} 1 & \text{if } s_{ij}^{(1)} \leq s_{ig_{(i)}}^{(1)} \\ 0 & \text{otherwise.} \end{cases}} (\boldsymbol{I} - \boldsymbol{\iota}\boldsymbol{\iota}^T \tfrac{1}{n}) \qquad (9)$$

- Eigenvectors serve as surrogates for spatially autocorrelated missing origin and destination variables (Tiefelsdorf and Boots 1995, Griffith 1996)

Spatial filtering specification

- Select a subset of eigenvectors $\boldsymbol{E}_m$ with a Moran's I value higher than 0.25;
- Adaption of the selected eigenvectors to the n^2-by-n^2 dimension of our spatial interaction modelling framework is done by Kronecker products

Application of the extracted origin and destination filters leads to the Negative Binomial Spatial Filter Interaction Model:

$$X_{ijt}^{*\prime} = \exp\left[\alpha_{0t} + \underbrace{\sum_{q=1}^{Q} E_{qt}\,\psi_{qt}}_{\textit{Origin spatial filters}} + \alpha_{1t}\log\left(o_{it}\right) + \underbrace{\sum_{r=1}^{R} E_{rt}\,\varphi_{rt}}_{\textit{Destination spatial filters}} + \alpha_{2t}\log\left(d_{jt}\right) + \sum_{k=1}^{K} ß_{kt} s_{ijt}^{(k)} + \xi_{ijt} \right] \tag{10}$$

Estimation results: Model fit

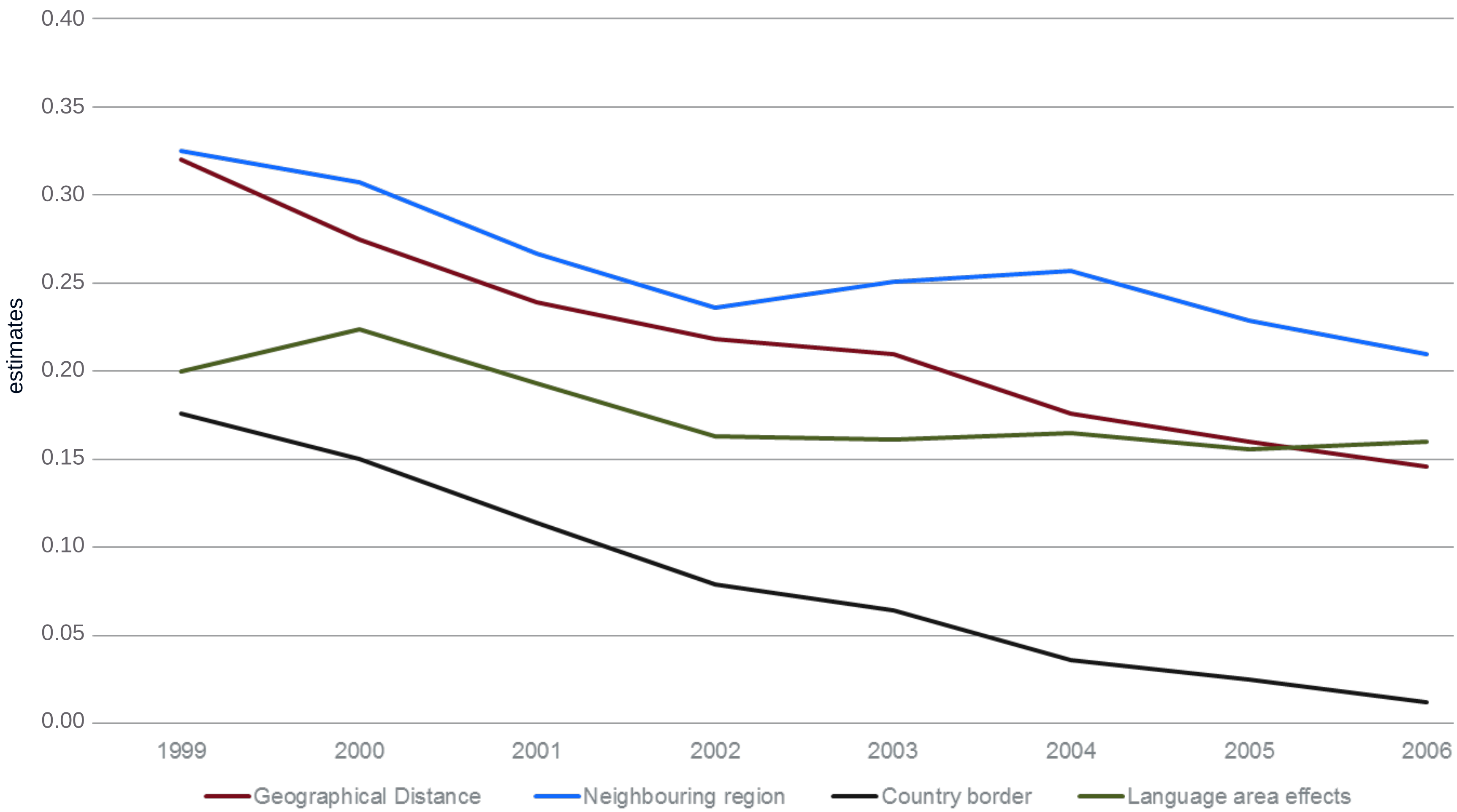
	1999	2000	2001	2002	2003	2004	2005	2006
Negative Binomial Spatial Interaction Models								
Log Likelihood	-217,282.23	-231,895.05	-248,731.40	-261,892.51	-264,208.46	-286,688.18	-282,816.69	-269,961.74
Residual deviance	63,994	64,927	68,104	68,564	69,023	72,396	71,854	72,184
Moran's *I* (error)	**0.043*****	**0.039*****	**0.032*****	**0.043*****	**0.029*****	**0.031*****	**0.029*****	**0.026*****
Negative Binomial Spatial Filter Interaction Models								
Log Likelihood	-215,933.28	-230,571.75	-247,491.04	-260,764.37	-263,125.21	-285,703.82	-282,021.01	-269,352.88
Residual deviance	63,738	64,646	67,918	68,458	68,941	72,309	71,749	72,117
Moran's *I* (error)	**0.003**	**0.002**	**0.000**	**0.004**	**0.001**	**0.003**	**0.005**	**0.004**
LR-Test	256.241***	281.032***	185.548***	106.303***	82.454***	87.159***	105.668***	67.575***

*** significant at the 0.001 level, ** significant at the 0.01 level, * significant at the 0.05 level

	Negative Binomial Spatial Filter Interaction Models							
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
Origin and Destination $[\alpha_1] = [\alpha_2]$	0.970*** (0.003)	0.977*** (0.003)	0.977*** (0.002)	0.976*** (0.002)	0.979*** (0.002)	0.980*** (0.002)	0.980*** (0.002)	0.980*** (0.002)
Geographical distance **[$ß_1$]**	**-0.320*** (0.006)**	**-0.275*** (0.006)**	**-0.239*** (0.005)**	**-0.218*** (0.005)**	**-0.210*** (0.005)**	**-0.176*** (0.008)**	**-0.160*** (0.005)**	**-0.146*** (0.005)**
Country border effects **[$ß_2$]**	**-0.176*** (0.018)**	**-0.150*** (0.017)**	**-0.114*** (0.016)**	**-0.079*** (0.015)**	**-0.064*** (0.015)**	**-0.036* (0.014)**	**-0.025 (0.014)**	**-0.012 (0.015)**
Language area effects **[$ß_3$]**	**-0.200*** (0.034)**	**-0.224*** (0.015)**	**-0.193*** (0.014)**	**-0.163*** (0.013)**	**-0.161*** (0.013)**	**-0.165*** (0.012)**	**-0.156*** (0.013)**	**-0.160*** (0.013)**
Neighbouring region **[$ß_4$]**	**0.325*** (0.023)**	**0.307*** (0.022)**	**0.267*** (0.020)**	**0.236*** (0.019)**	**0.251*** (0.019)**	**0.257*** (0.018)**	**0.229*** (0.018)**	**0.210*** (0.019)**
Neighbouring country **[$ß_5$]**	**0.004 (0.010)**	**0.005 (0.009)**	**0.019* (0.008)**	**0.010 (0.008)**	**0.013 (0.008)**	**0.010 (0.007)**	**0.019* (0.007)**	**0.011 (0.008)**
Technological Distance **[$ß_6$]**	**-0.683*** (0.083)**	**-0.622*** (0.077)**	**-0.452*** (0.068)**	**-0.400*** (0.065)**	**-0.295*** (0.063)**	**-0.303*** (0.061)**	**-0.342*** (0.062)**	**-0.341*** (0.062)**
# of origin filters *Q*	19	19	19	20	19	20	18	15
# of destination filters *R*	18	19	21	20	19	19	15	14
Constant $[\alpha_0]$	-9.596*** (0.093)	-10.213*** (0.088)	-10.820*** (0.078)	-11.171*** (0.075)	-11.387*** (0.073)	-11.819*** (0.071)	-11.844*** (0.072)	-11.819*** (0.073)
(γ)	4.149*** (0.065)	4.628*** (0.068)	5.814*** (0.087)	6.193*** (0.088)	6.519*** (0.093)	6.436*** (0.088)	6.494*** (0,089)	6.529*** (0,097)
AIC	216,027	230,668	247,591	260,864	263,221	285,802	282,107	269,431

*** significant at the 0.001 level, ** significant at the 0.01 level, * significant at the 0.05 level; standard errors in brackets

Magnitudes of the estimates



Conclusions

Evidence that FPs contribute to the integration of European research

- Geographical distance effect gradually declines over time, i.e. results suggest increasing probability for large distance collaborations
- FPs are conducive to abolishing barriers for research collaborations within the FPs constituted by country borders
- Negative language area effects seem to be reduced in general, but relatively slowly
- Technological distance is the most important determinant of cross-region R&D collaborations, but also decreases over time

Methodologically the results provide evidence for the importance of taking into account spatial autocorrelation in an interaction context

References

Scherngell, T. and Lata, R. (2013): Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data, *Papers in Regional Science* 92(3), 555-577

Scherngell, T. and Barber, M. (2009): Spatial interaction modelling of cross-region R&D collaborations. Empirical evidence from the 5th EU Framework Programme, *Papers in Regional Science* 88(3), 531-546 (awarded with the Martin Beckmann price for the best article published in PIRS 2009)

Barber, M.J. and Scherngell, T. (2013): Is the European R&D network homogenous? Distinguishing relevant network communities using graph theoretic and spatial interaction modeling approaches, *Regional Studies* 47(6), 1283-1298

Lata, R., Scherngell, T. and Brenner, T. (2015): Integration processes in European R&D: A comparative spatial interaction approach using project based R&D networks, co-patent networks and co-publication networks, *Geographical Analysis* 47, 349-375