



WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS



Topics in regional science: An exploratory text mining approach to ERSA conference contributions from 2001 to 2013

Tanja Sinozic¹, Gunther Maier², Miklós Hornyák⁴, Ferenc Kruzslicz³,
^{1,2} WU Vienna University of Economics and Business,^{3,4} University of Pecs

1. Review of bibliometric mapping of regional science literature
2. Gaps and research aims
3. Method and rationale
4. Creation of the dataset
5. Results 1: Identification of meta-constructs and their change over time
6. Results 2: Exploration of key construct and attributes
7. Summary and next steps in analysis

Review

Papers mapping Regional Science literature

- Topics and authors in individual Regional Science journals (Allen and Kau, 1991; Stratham, 1992; Anselin, 1995; O'Kelly, 1999; Florax 1999)
- Change in publication patterns, analysis techniques and data in multiple Regional Science journals (Rey and Anselin 2000; Surinach et al. 2003)
- Interaction between Regional Science journals (Taylor and Jones, 1992)
- Interdisciplinarity (Isserman 1993)
- Central scholars (Isserman 2004)
- Contributions by author, institution and country (Florax and Plane, 2004)
- Co-authorship (Maier and van Dijk, 2006)
- Important books in Regional Science (Waldorf et al 2004)
- Historical account of changes in Regional Science (Boyce 2004)

- Exploring shifts in the discourse on regional science
- Contention that linguistic constructions are part of an order and supports (Perren and Dannreuther, 2013)

Aims:

- Change in meta-constructions of core terms
- Stable and changing themes
- Convergence and diffusion
- Symbolism
- Shared beliefs

Method

Corpus-based linguistics/Text mining

- We try to follow Perren and Dannreuther (2013)
- Linguistic turn (e.g. manual content analysis, narrative analysis, semantics)
 - Difficult when dealing with large amounts of text over extended time periods
 - Danger of „cherry picking“ and difficult to replicate
 - Ideal approach: systematic dataset condensation while making it easy to see changes in meaning through qualitative analysis (hybrid approach)
 - Corpus-based linguistics makes this possible
 - Based on the idea that meaning is constructed through context (Callon, 1983)
 - Context is defined variably as collocated words, co-occurrence of words
 - Uses processes from computer science, mathematics and statistics approaches to treat text as numbers

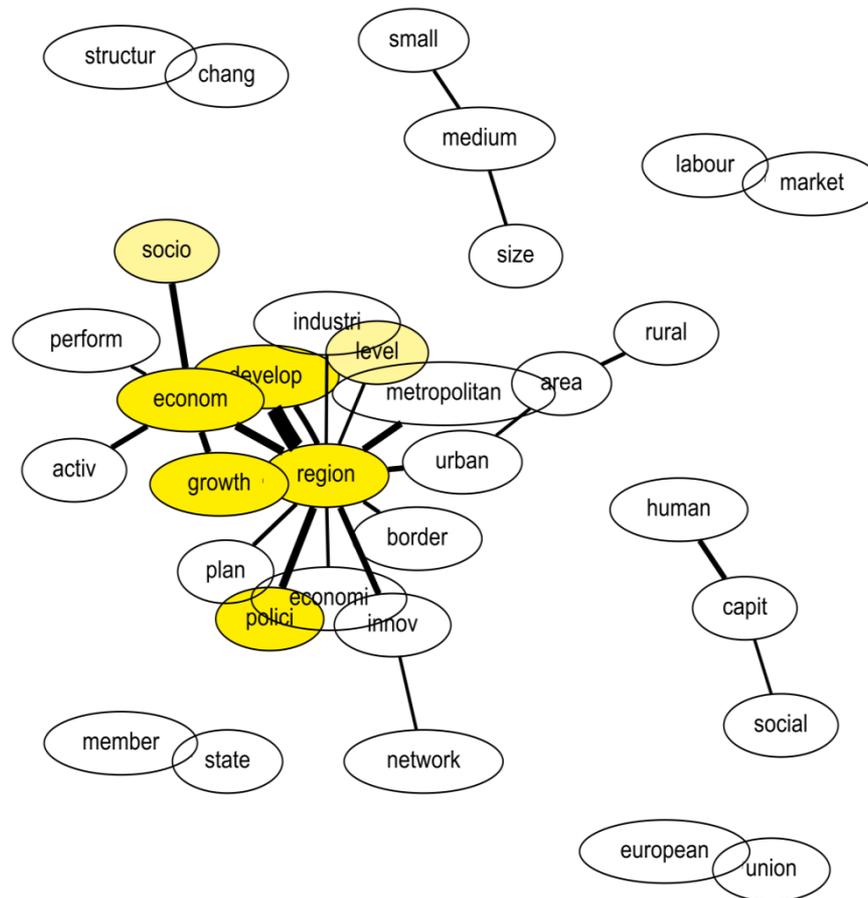
Creation of the dataset

- Establishing the dataset
 - Download and store ERSA conference abstracts from 2001 to 2013
 - 3,789 documents
 - Sum of words: 527,024
 - Different words: 20,280
- Pre-process text data
 - Carry out tokenization process to remove punctuation marks
 - Filter using stopwords, stripping and stemming

Creation of the dataset

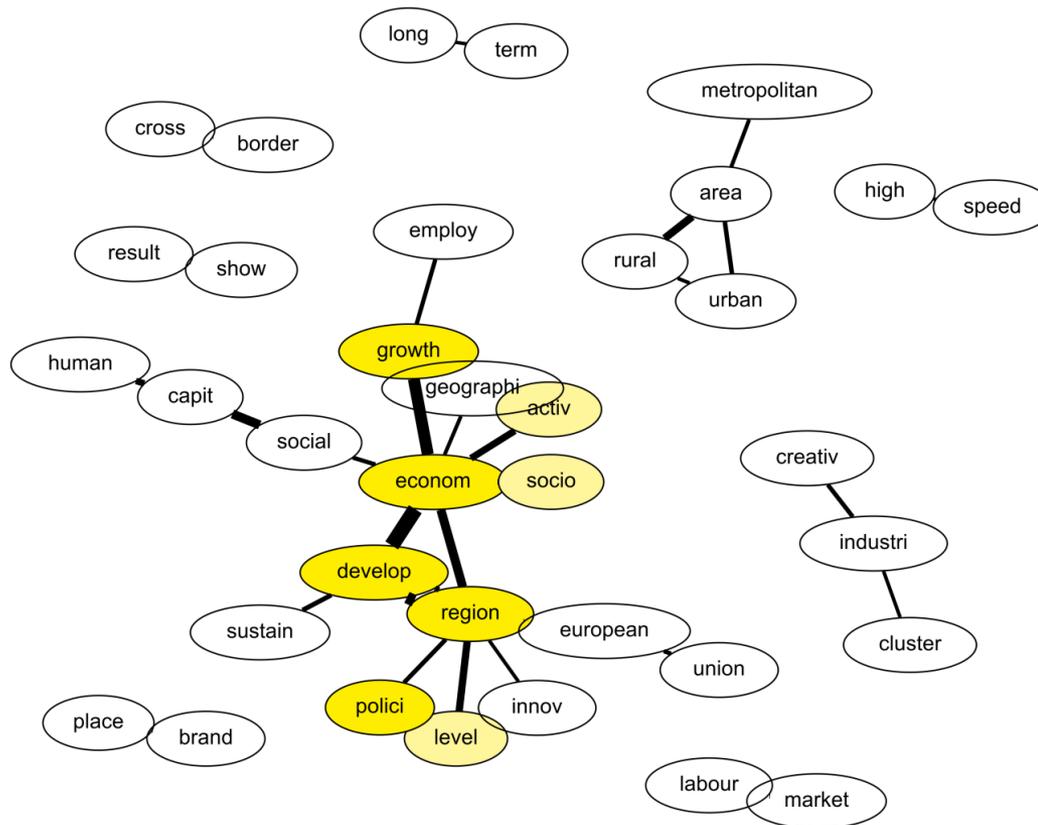
- Analysing meaning through word frequencies and word co-occurrence
 - Network diagrams organised by year
 - Extraction of most frequent terms and collocates
 - Identify stable terms for comparison („region +econ+develop“)
- Explore patterns through qualitative analysis
 - Data display

Stable terms and collocates (2002)



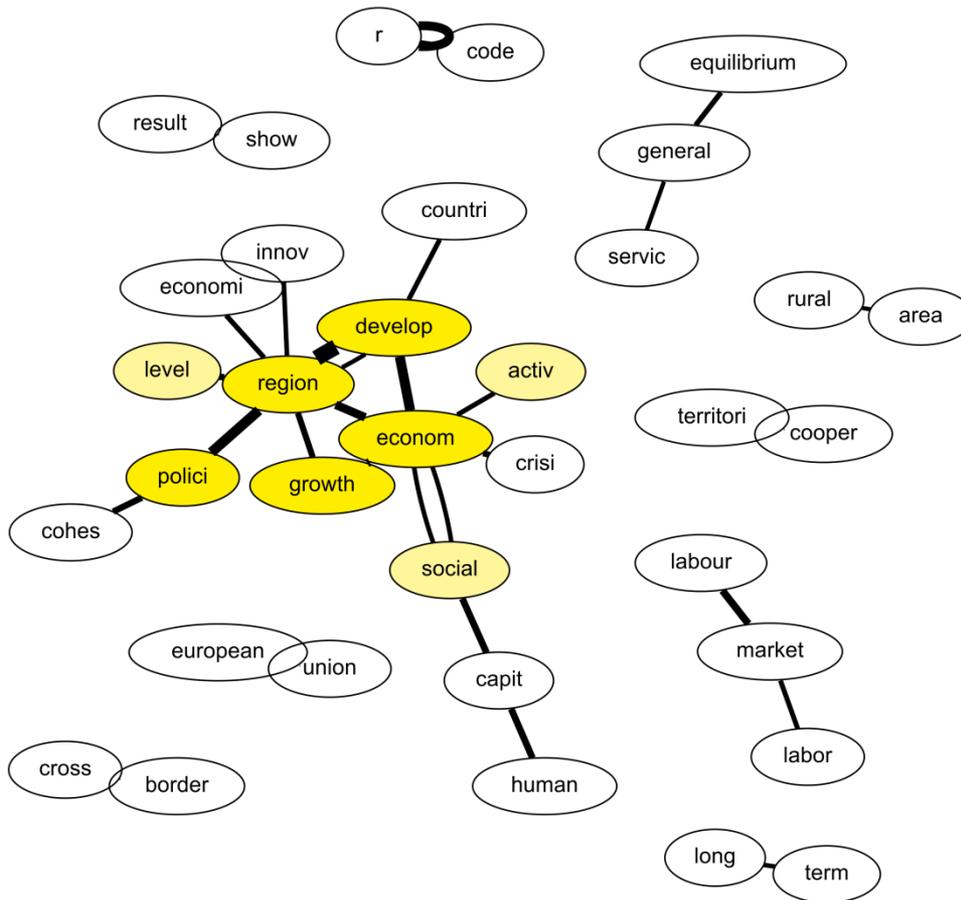
- Stable terms more frequently associated
- New topics: structural change and human and social capital

Stable terms and collocates (2010)



- Creative industries cluster becomes a prominent topic>but it does not sustain itself in the later years

Change in stable terms in response to events (2012)



- Core terms remain stable, as to terms (activ*, level*, social and human capital) associated with it
- Crisi* emerges and is immediately part of core

Results 2

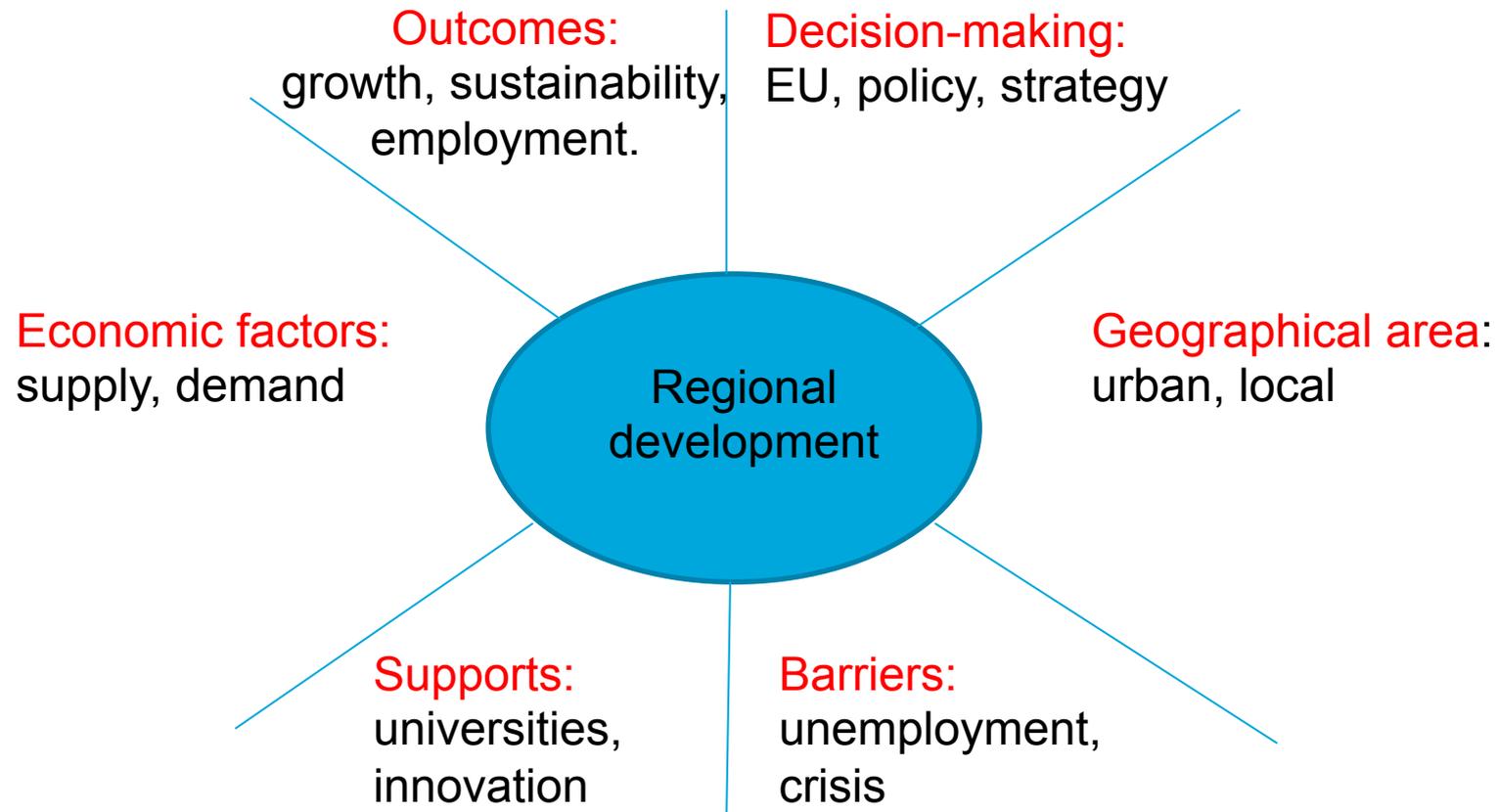
Exploration of key construct attributes

Collocates

Nr.	2001-2006, Radius = 1				2010-2013, Radius = 1			
	Count	region	Count	develop	Count	region	Count	develop
	469	develop	469	region	318	develop	318	region
1	289	polici	275	econom	202	econom	301	econom
2	200	econom	113	sustain	166	level	94	countri
3	153	develop	106	countri	146	polici	93	sustain
4	135	level	92	local	106	european	63	polici
5	131	innov	89	polici	93	growth	59	rural
6	129	growth	81	spatial	86	innov	53	urban
7	120	urban	65	tourism	76	dispar	53	local
8	117	dispar	58	model	59	local	40	process
9	108	economi	55	urban	56	differ	36	growth
10	105	european	54	strategi	55	nation	35	tourism
11	100	local	52	plan	54	competit	34	strategi
12	87	differ	52	rural	54	citi	28	plan
13	86	border	47	process	53	metropolitan	27	territori
14	81	inequ	34	level	50	inequ	22	land
15	78	metropolitan	32	innov	47	countri	22	model
16	76	citi	30	model	47	rural	20	innov
17	74	labour	30	research	45	effect	18	industri
18	71	spanish	29	industri	44	nut	18	develop
19	68	nation	28	platform	42	scienc	17	level
20	67	competit	27	project	39	industri	17	empir
21	62	countri	27	growth	37	urban	17	index
22	60	unemploy	27	recent	35	italian	16	futur
23	60	inter	25	land	33	product	16	area
24	59	peripher	22	agenc	32	converg	16	citi
25	58	analysi	21	econom	32	spanish	16	program
26	55	product	21	futur	31	context	15	spatial
27	55	plan	20	activ	31	employ	14	research
28	55	govern	20	area	31	govern	14	recent
29	50	industri	20	system	31	peripher	14	path
30	49	nation	19	european	31	determin	14	social

Qualitative results

Topics coded from collocates



Summary of preliminary results

- Stable terms over the years (2001-2012) region*, econom*, develop*, polici* and growth
- Stable in term environment (recurring in majority of years): socio* and level*
- Novel topics emerging over time, some stay in periphery and disappear (e.g. creative industries), some become stable (e.g. human and social capital)
- Some immediate collocates with stable terms (e.g. crisis) but temporary
- Common overarching themes: barriers (unemployment, crisis), supports (universities, innovation), outcomes (growth, sustainability, employment), geographical level (urban, local) and decision-making aspects (EU, strategy, policy)

Next steps

- Identify collocates that are statistically significant in providing context of meaning to the term *region* by comparing with general (disordered) natural language in British National Corpus
- Assumption is that the larger the difference the more meaning the term brings (Barnbrook, 1996)
- Analyse significant collocates using Key Word in Context (KWIC) lines (Miles and Huberman, 1994)

Thank you for listening!