UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Content and networks of relationships in the real and virtual domains.

Promoting the system of “innovation poles”

Matteo Di Cristofaro, Valentina Fiordelmondo,
Stefano Ghinoi, **Margherita Russo**

«Poli.in project» www.poliinnovazione.unimore.it
Università di Modena e Reggio Emilia, Italy

Di Cristofaro et al. Promozione del sistema dei poli di innovazione nello spazio web: analisi dei contenuti e delle reti di relazioni virtuali, pp. 36 | http://merlino.unimo.it/campusone/web_dep/wpdemb/0074.pdf

EUSN 2016 Second European Conference on Social Networks | June 14-17, 2016, Paris

Research questions

When using contents and relational information published online to characterise agents' behaviours:

To which extent does the virtual domain mirror the real one?

Which information can be detected from those domains?

Empirical analysis

Focus on

intermediaries of technological transfer
«innovation poles» in Tuscany, Italy:

policy to support SMEs and to foster the regional system of innovation

Research strategy

- Linguistic analysis
  qualitative and quantitative analysis of the content of the poles' communication
- Network analysis to identify and analyse:
  - the relationships with other agents

    such as the organization leading the consortium or the other organizations managing the pole, their laboratoris or incubators
  - the shared contents, i.e. their «information space»

Data

- **«virtual space»**: innovation poles' websites

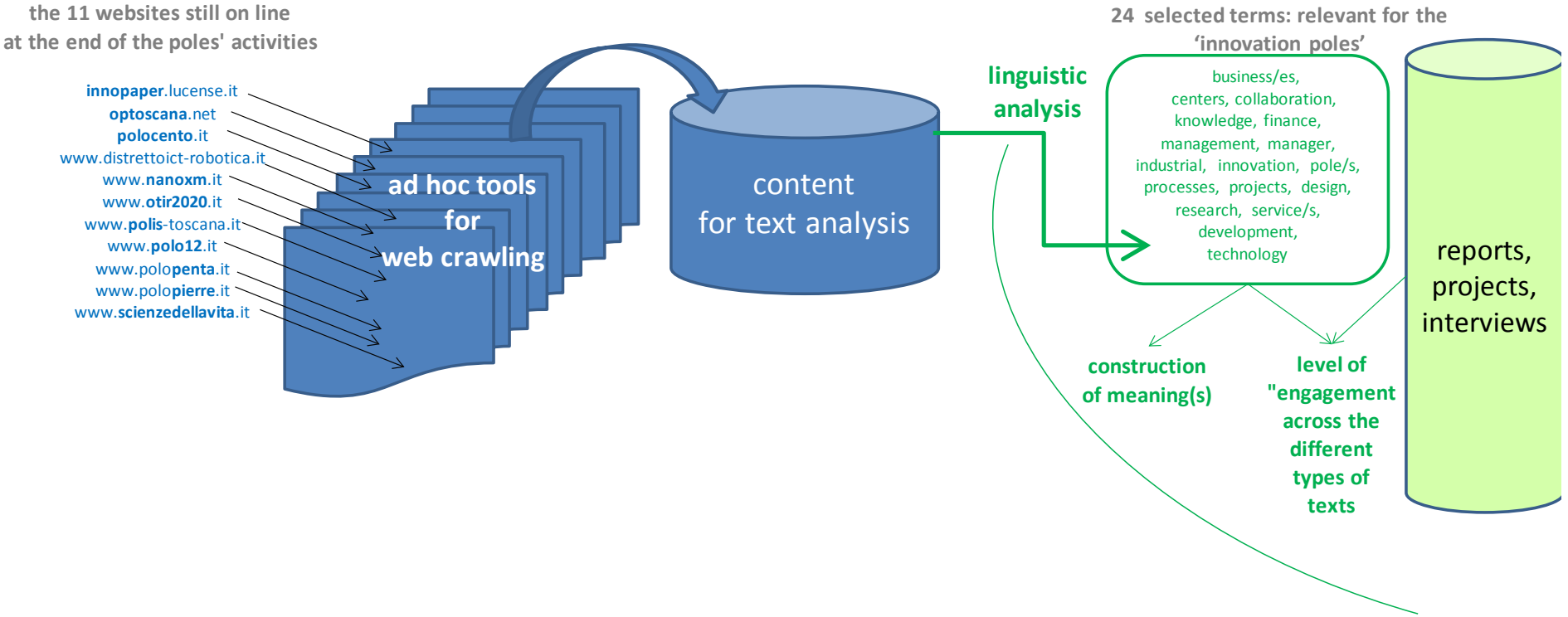
vs

- **«real» domain»**: data from the *«poli.in research project»*

Official documents and data produced
in three years by the 12 poles
to report on their activities
(as demanded by the accounting of
public funds received by the Tuscany region)

Methodology_Content analysis

What is being <u>said</u> in the virtual space?
And what in the real space?



Methodology_Network analysis

What is being done in the virtual space?
And what in the real space?

the 11 websites still on line
at the end of the poles' activities

innopaper.lucense.it
optoscana.net
polocento.it
www.distrettoict-robotica.it
www.nanoxm.it
www.otir2020.it
www.polis-toscana.it
www.polo12.it
www.polopenta.it
www.polopierre.it
www.scienzedellavita.it

ad hoc tools
for
web crawling

content
for text analysis

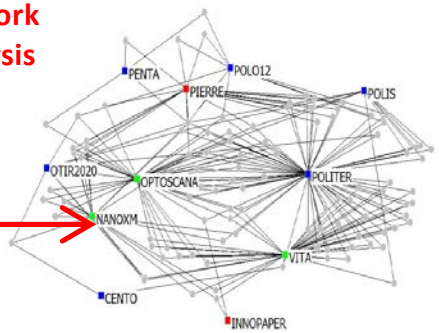
linguistic
analysis

24 selected terms: relevant for the
'innovation poles'

business/es,
centers, collaboration,
knowledge, finance,
management, manager,
industrial, innovation, pole/s,
processes, projects, design,
research, service/s,
development,
technology

list of url

network
analysis



<

Citations of organizations
Citations of contents
(events…)

Content analysis: research questions

1. **How does each pole 'present itself' to the public?**

2. **To which extent their communication is original vs. reported?**

→

Quantitative analysis of contents

tokens, types, TTR

	Tokens	Types	TTR
innopaper.lucense.it	9,921	2,128	21.45%
optoscana.net	26,776	2,470	9.22%
polocento.it	64,399	4,375	6.79%
www.distrettoict-robotica.it	94,497	7,749	8.20%
www.nanoxm.it	46,680	5,027	10.77%
www.otir2020.it	9,329	1,130	12.11%
www.polis-toscana.it	1,588,739	1,380	0.09%
www.polo12.it	119,416	2,930	2.45%
www.polopenta.it	96,270	2,516	2.61%
www.polopierre.it	16,736	2,376	14.20%
www.scienzedellavita.it	627,217	7,308	1.17%

Tokens: Words, including repetitions.

Types: Unique words.

TTR: *Type-Token Ratio*, «used to assess how varied the vocabulary is» (Baker, McEnery and Hardie 2006:81).

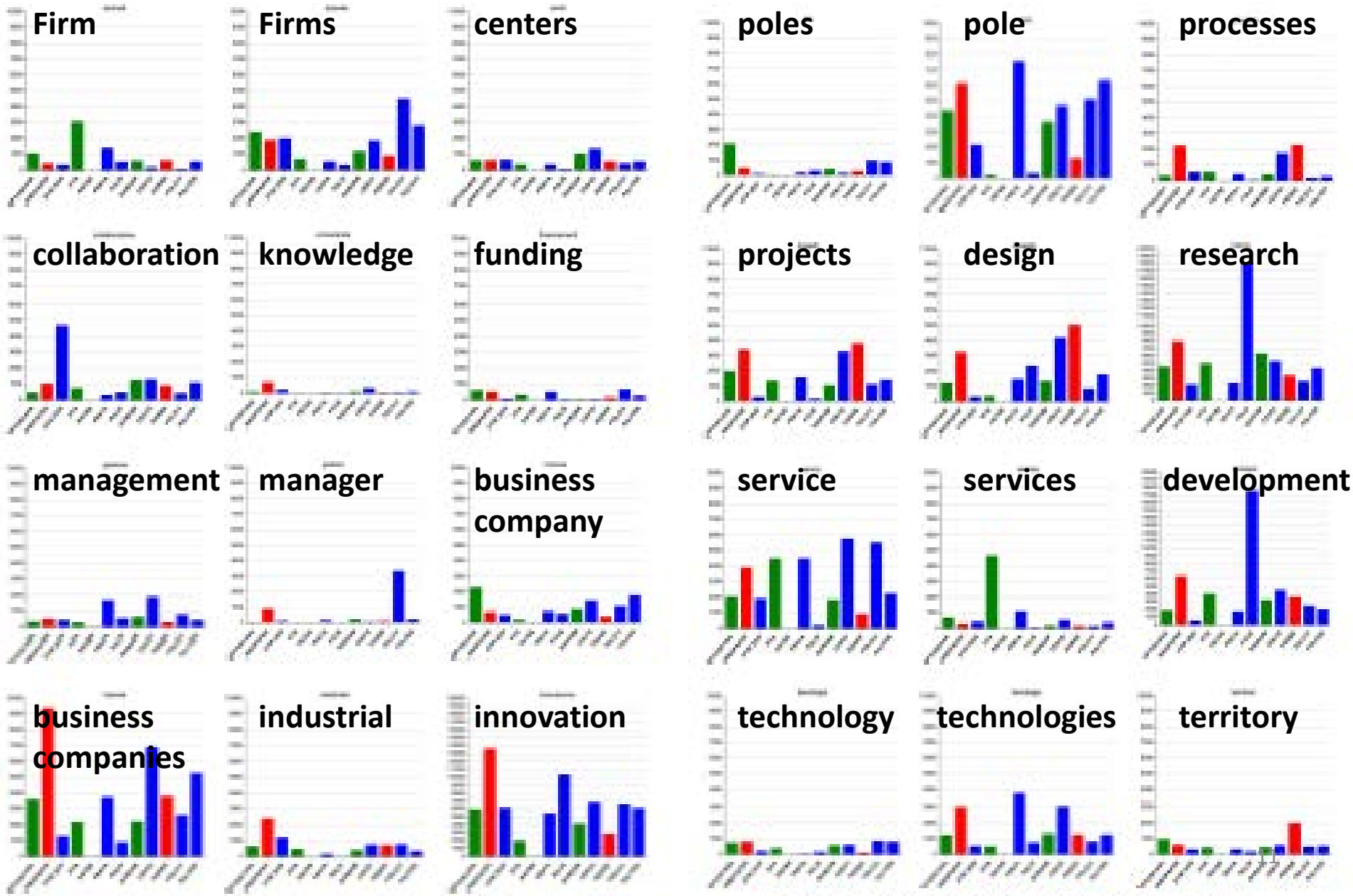
Qualitative analysis of contents
Communication

From the use of the 24 words:

- insights on how each pole ‘presents itself’ to the public.
- over/under use of a specific terms signals how each innovation pole focusses on one (or more) particular aspect rather than another when promoting itself.

Quantitative analysis of contents

Terms' frequency by pole (colors: band=pole's size)



Qualitative analysis of contents

Communication_an example: «*research*»

- It is used across the 11 corpora with four different meanings:
  - **‘to look for someone/something’,** as in «ricerca di aziende» (‘searching for a company’; polo-penta.it)
  - ‘**data analysis**’, as in «ricerca di settore» (‘statistical study for the estimation of taxes due by firms in specific industries; otir2020.it)
  - ‘**academic research**’ as in «finanziamento alla ricerca europea» (‘funding to the european research’; distrettoict-robotica.it)
  - ‘**research and development**’ as in «ricerca e sviluppo» (**all innovation poles use this construction**).

Qualitative analysis of contents

Communication_original vs. reported

RESULTS

- Innovation poles with ‘**smaller’ websites**
  (i.e. with the smaller numbers of contents)
  - are relatively more directly engaged in promoting themselves and their activities

- The ones with **‘bigger’ websites**
  - are mostly reporting texts from third-party sources, with relatively lower promotion of their activities
  - They are mostly acting as ‘aggregators’ of news and announcements related to the network of poles

Virtual vs real network: research questions

1. **To which extent have the web sites reported the poles connections with each other and with those involved in technology transfer?**

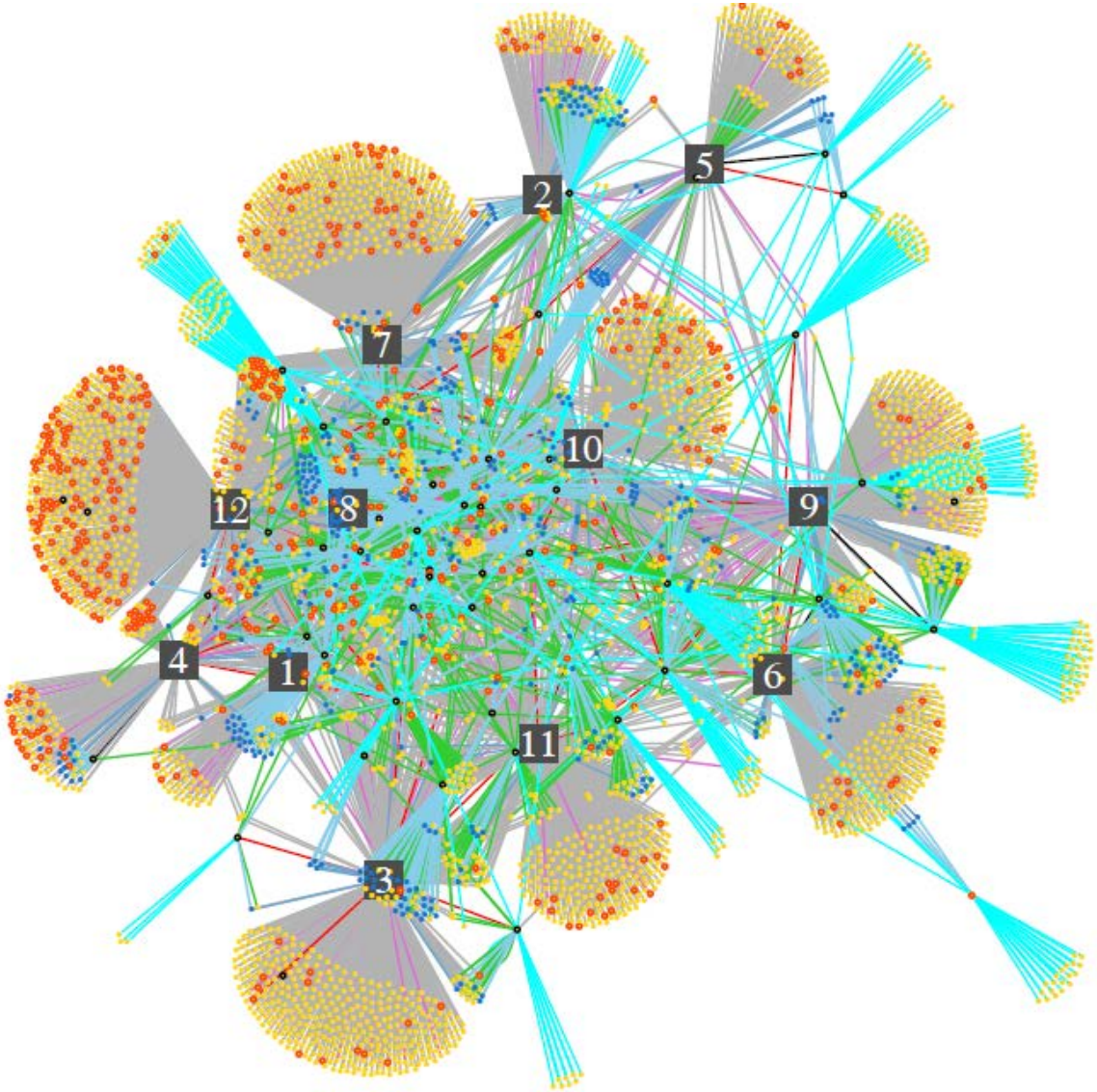
   →enhancing a systemic perspective on the regional innovation system

2. **Do the innovation poles refer to the same information space**?

   →hints on common/distinct technological domains

«Real» linkages

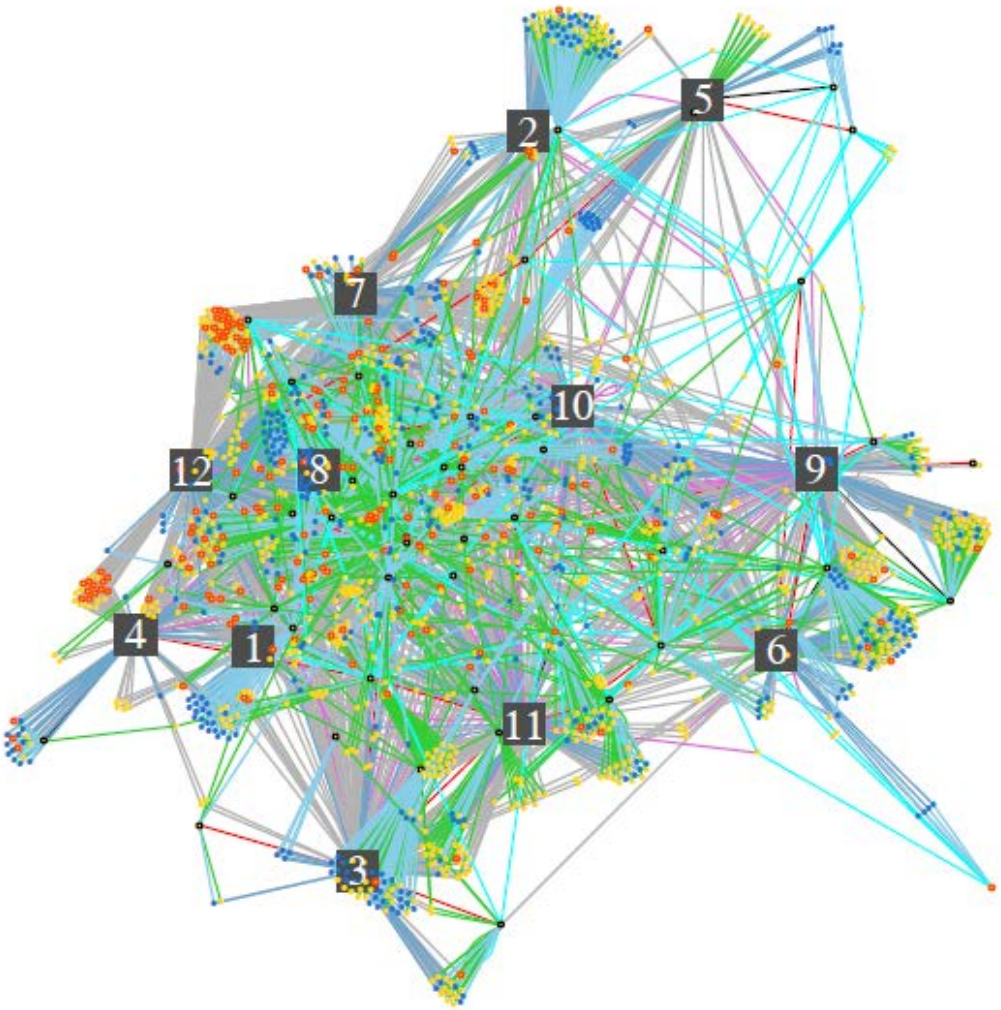
Innovation poles [1-12]
Leading organizations
Manging organizations
Shareholders
Members
Personnel



«Real» linkages

only: more than one link

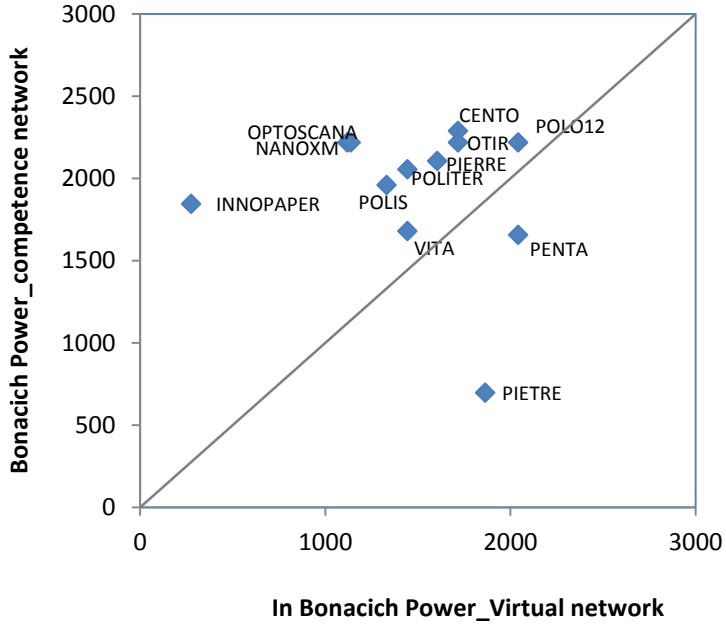
Innovation poles [1-12]
Leading organizations
Manging organizations
Shareholders
Members
Personnel



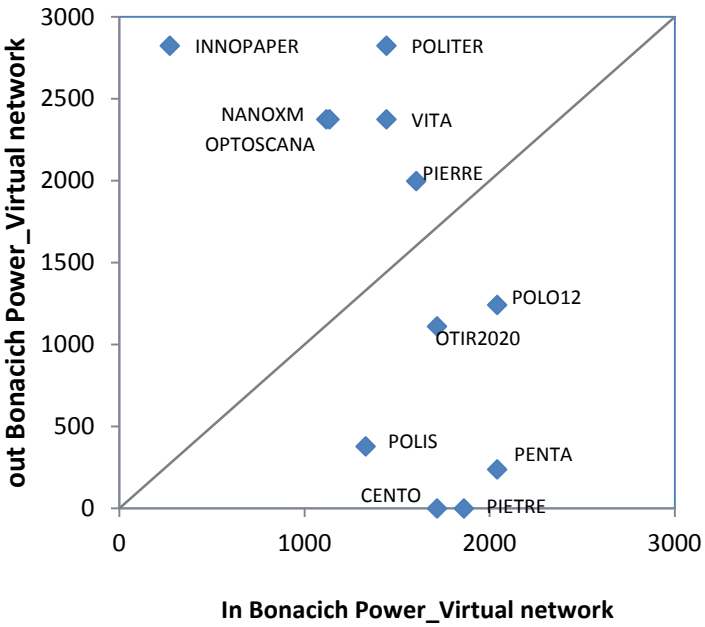
1_Poles connections

centrality positions in the real and virtual domains

Virtual (in) domain
vs.
real (competence network) domain



Virtual domain:
in vs. out



1_Poles connections

- The analysis of citations has highlighted a rather sparse network of connections among the innovation poles:
  - Only two websites (optoscana. and politer) contain links to at least one of the other innovation poles' websites.
  - There are very little connections between the poles in terms of reciprocal citations
  - Poles do not cite systematically the organizations managing them
  - Often, there is asymmetry in citation of the organizations/institutions with whom the poles have signed cooperation agreements.
- **It does not seem that the virtual space has been chosen as the tool to strengthen the systemic dimension of the poles activities**

2_Information space

- **A few poles are relatively more cited by all the others**
- **One pole** (Innopaper) is **never cited** by the others: that pole, its operators, the participating companies and the related infrastructures (such as laboratories) are not mentioned by any other pole, **but that pole is very active in quoting websites** of agents linked to other poles.
- **The poles of the third band** (Life, Optoscana and Nanoxm) **present a higher number of references to the same information** (complete links to web pages).
  - Complementarities between technological fields could explain this result, but it is also possible that these poles have mostly used the web communication precisely to cope with their small size

Perspectives on the innovation system
organizations managing the poles vs. innovation poles

The leading and managing organizations of the poles
have conceptualized these entities
(the poles as intermediaries in technology transfer)

in different ways

- **as an entity having its own identity** (specific purpose and role) **within the technology transfer system in Tuscany**

- **as a "project" of the leading organization**

Conclusion: two perspectives on the innovation system
real vs virtual dimensions

In general, both domains provide information in complementary ways, but in the case examined :

- The network of relationships across the innovation system would be largely **underestimated** by the virtual network perspective

- Analysis of web communication helps in highlighting the **diverse points of view** of the organizations managing the poles on these intermediaries in the regional innovation systems

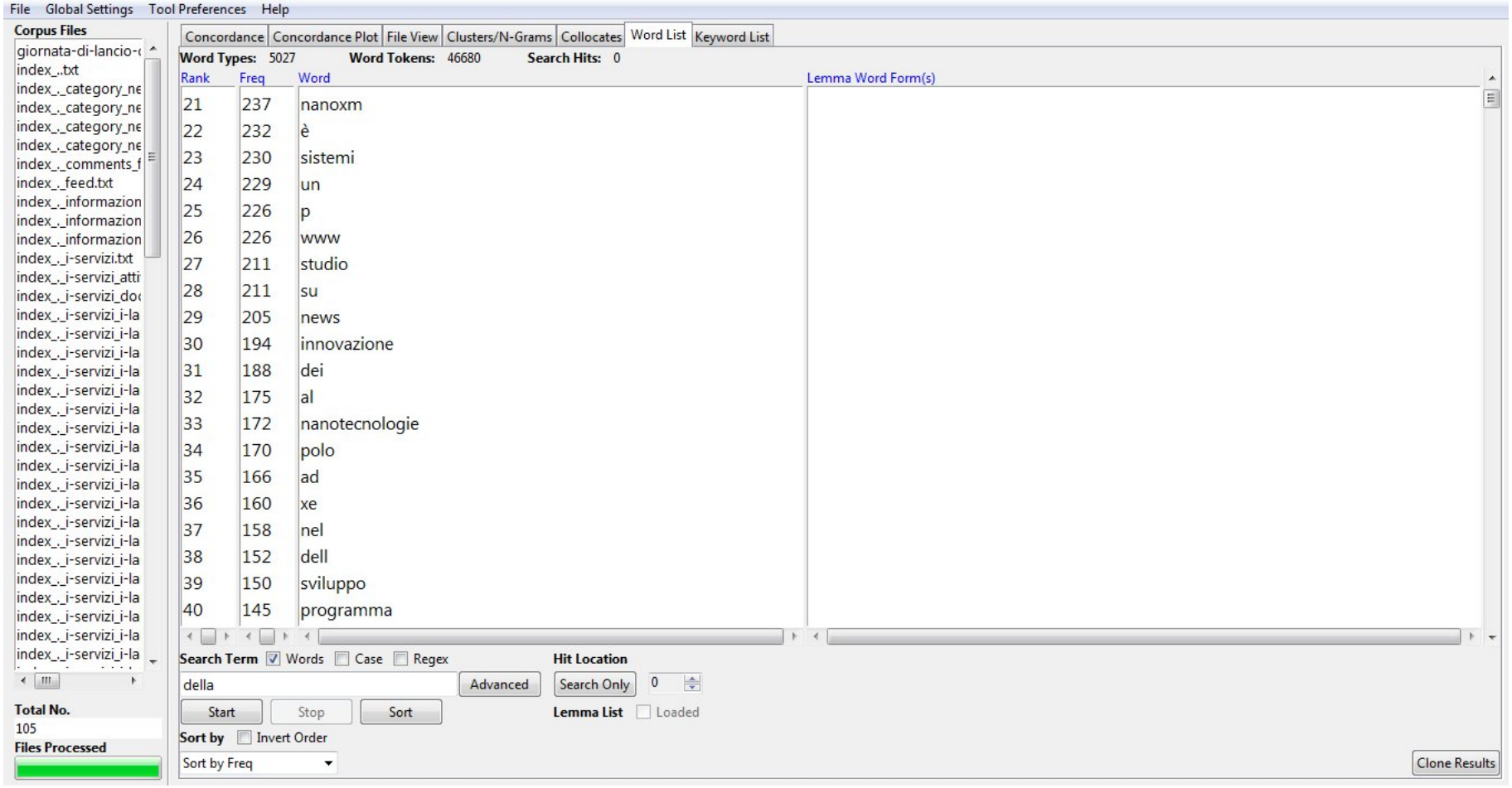
additional documentation

- Corpus Linguistics
- Tools used

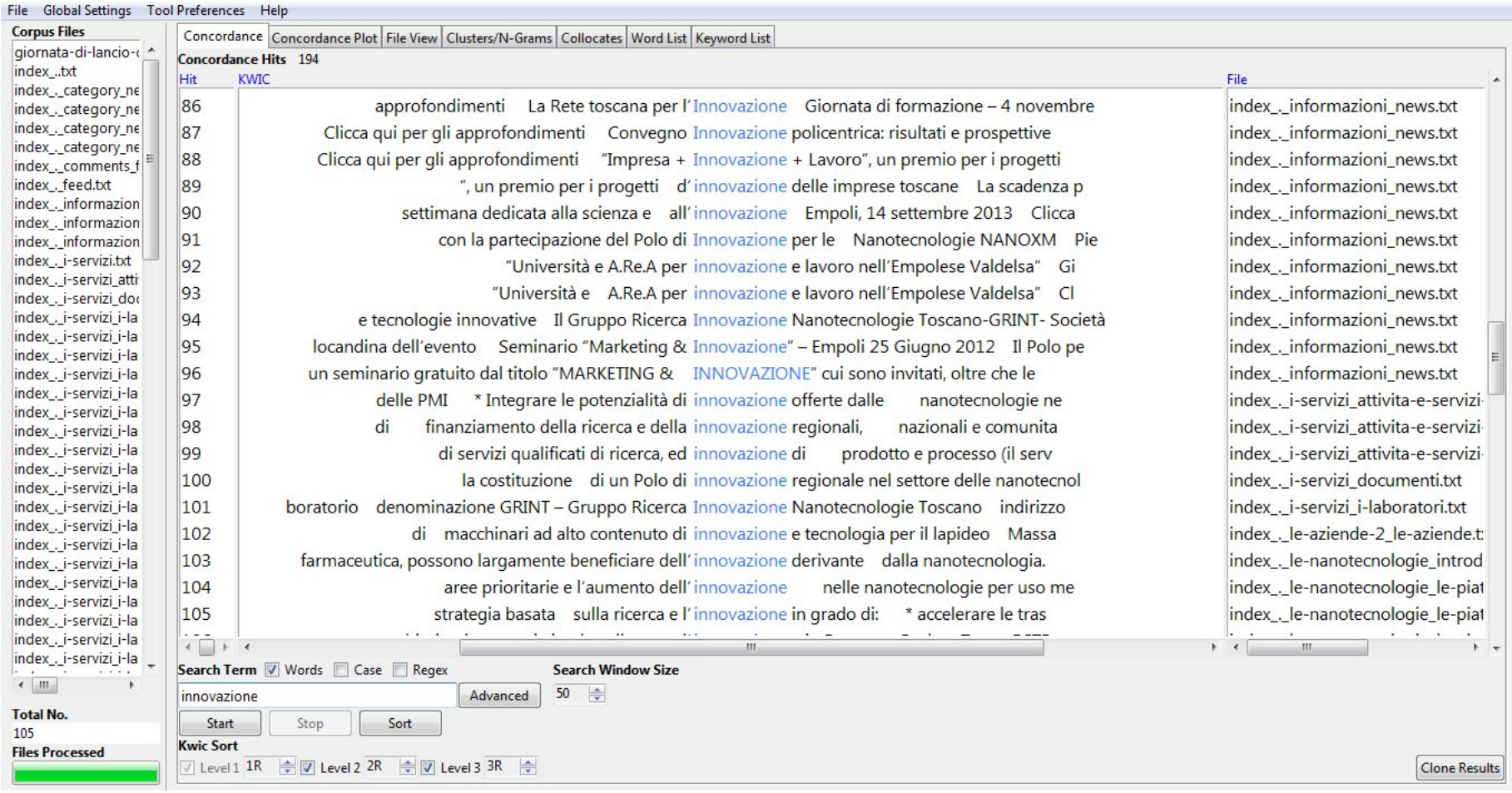
Corpus Linguistics

- Corpus Linguistics «is an area which focuses upon a set of procedures, or methods, for studying language» (McEnery and Hardie 2012:1). These methods can be extremely heterogeneous; to generalise, Corpus Linguistics deals «with some set of machine-readable texts [i.e. corpus] which is deemed an appropriate basis on which to study a specific set of research questions»
- Corpora (pl. of corpus) are then «exploited using tools which allow users to search through them rapidly and reliably», allowing the researcher to produce e.g. concordances and word frequency lists in order to analyse both quantitatively and qualitatively analyse the data.

Word frequency list: an example



Concordances: an example



Tools used for content analysis & hyperlink extraction

- The linguistic analysis and the hyperlinks extraction, were conducted using free and libre/open-source software – with the exception of AntConc, which is free and closed-source.

Tools used for the content analysis

Webpages collection and cleaning

Httrack (https://www.httrack.com/): tool to create local copies of websites

lynx (http://lynx.browser.org/): command-line browser, used to dump the text without html code from webpages

find (http://www.gnu.org/software/findutils/manual/html_mono/find.html: command for filename searches.

grep (http://www.gnu.org/software/grep/manual/grep.html) : command to search inside files' contents

perl (https://www.perl.org/) : command-line version, used for batch text-substitutions using regexes

rm (https://www.gnu.org/software/coreutils/manual/html_node/rm-invocation.html#rm-invocation) : command to delete files

wc (http://www.linfo.org/wc.html): command to count the number of elements (words, characters) inside files

shuf (http://www.gnu.org/software/coreutils/manual/html_node/shuf-invocation.html#shuf-invocation): command to extract random elements from a set

Linguistic analysis

AntConc (http://www.laurenceanthony.net/software/antconc/) : Software for corpus linguistics analysis

Link extraction from webpages

urifind (http://www.unix.com/man-page/debian/1p/urifind/) : tool to extract links from webpages

sort (https://www.gnu.org/software/coreutils/manual/html_node/sort-invocation.html) : command to sort the content of files according to a user-defined criteria

uniq (http://unixhelp.ed.ac.uk/CGIman-cgi?uniq) : command to identify duplicate elements inside multiple files