



Georgia Technical S3 Workshop:

Identifying scientific potential (level 2)

9 December 2020

Dr-Ing Yannis Toliás

Managing Partner, innovatia systems

toliás@innovatiasystems.eu



Contents – Level 2 Training

1

Mapping relationships among entities

2

Using Network Maps to get insights on how your science production system is working

3

Mapping Innovation Ecosystems



Some interesting questions that can be answered by scientometrics

1. Where does your country already excel or has the potential to put itself on the map as a recognized world-class place of competence?
2. Which are the specific scientific strengths and research specializations in your country?
3. Which emerging new scientific competences can be spotted in your country?
4. Who are the **key actors**? **How are they linked** with the national economy?
5. How fit is your national knowledge base to address conjointly the grand challenges of society?
6. **How do lead institutions position themselves in global chains of knowledge?**
7. How favorable are working conditions for researchers in your country? How much mobility between the public science and the private sector does exist? Do universities train scholars and graduates to become entrepreneurs?
8. Does current academic education fit to the needs of the national economy – do employers absorb graduates or are graduates forced to look elsewhere?
9. What about the **internationalization** of researchers and research collaborations?



What can you expect from Scientometrics?

Analysis of scientific production, specialization, and identification of key actors



Level1

Who has done what? Who's good in what? Who collaborates with whom?



Level1

Specialisation analysis at various levels

Analysis of knowledge trajectories and of the degree of participation in global knowledge chains



Level2

Knowledge trajectories



Level2

Access to global knowledge chains

Research evaluation

Micro-, meso- and macro-levels

Research monitoring



Level1

Who is researching what right now?

Mapping of Innovation Ecosystems



Level2

To what degree higher education and public research organisations are parts of wider innovation ecosystems?



Part I

MAPPING RELATIONSHIPS



An example – Mapping institutional relationships in projects

A simple project list:
8 institutions participating in 3 projects

OrganisationID	Project	Budget
Company1	Project1	25000
Company2	Project1	30000
ResearchCentre1	Project1	15000
University1	Project1	50000
Company3	Project2	5000
ResearchCentre2	Project2	15000
University1	Project2	20000
Company2	Project3	32000
Company4	Project3	14000
University1	Project3	22000
University2	Project3	20000

A pivot table counting
project participations

	Project1	Project2	Project3
Company1	1		
Company2	1		1
Company3		1	
Company4			1
ResearchCentre1	1		
ResearchCentre2		1	
University1	1	1	1
University2			1

$$\mathbf{A} = \{\alpha_{i,j}\} = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix}, \quad i \in [1, \dots, 8], \quad j \in [1, \dots, 3]$$

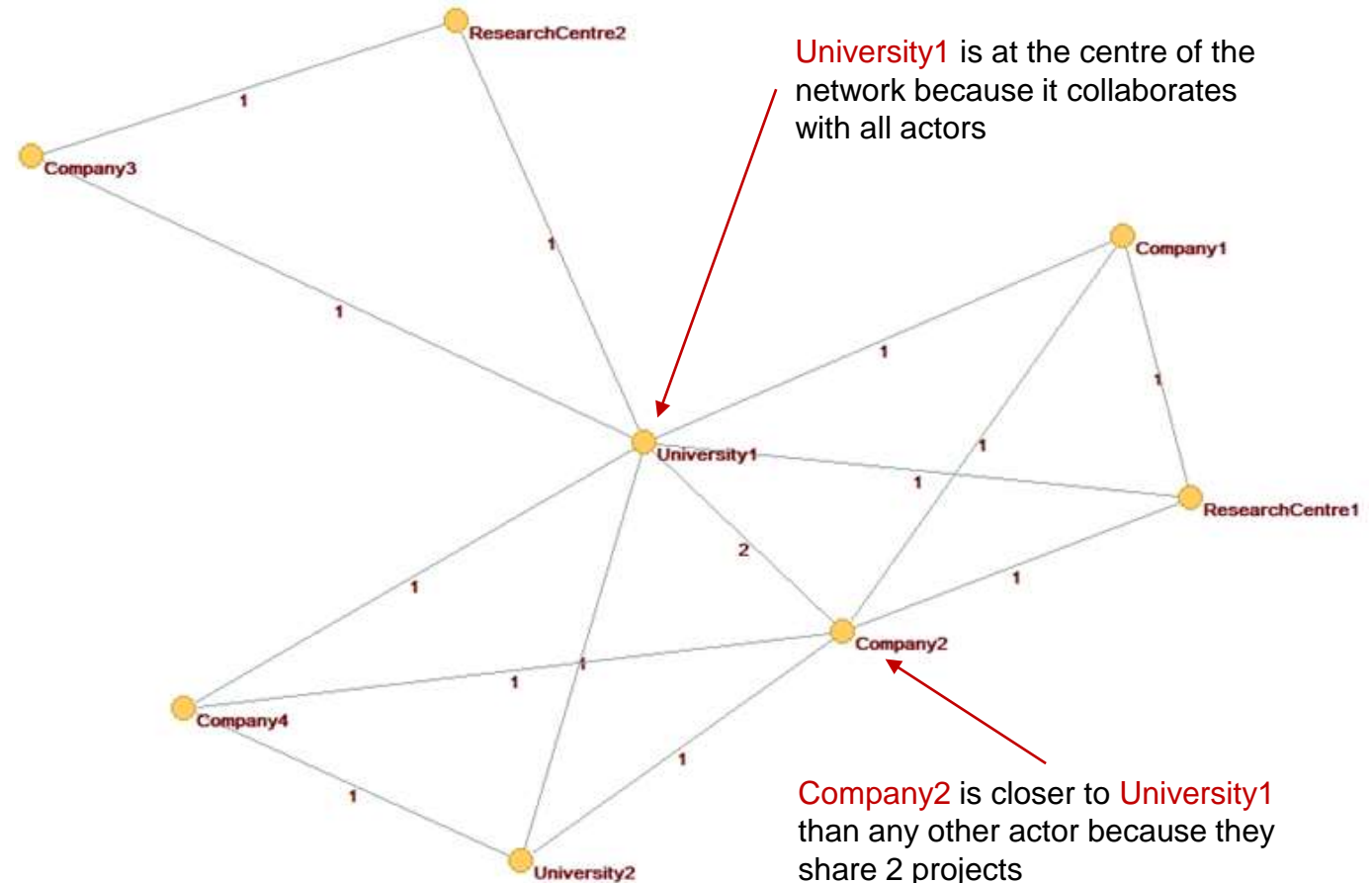


Let's do some matrix algebra

$$C = A \cdot A^T = \{c_{i,j}\}, i, j \in [1, \dots, 8]$$

	Company1	Company2	Company3	Company4	ResearchCentre1	ResearchCentre2	University1	University2
Company1	1	1			1		1	
Company2	1	2		1	1		2	1
Company3			1			1	1	
Company4		1		1			1	1
ResearchCentre1	1	1			1		1	
ResearchCentre2			1			1	1	
University1	1	2	1	1	1	1	3	1
University2		1		1			1	1

Matrix C maps project collaborations



Doing it yourself using open-source software from Slovenia & Austria

1. Download [Pajek](#)
2. Download Jürgen Pfeffer's [text to Pajek utility](#)
3. Create your dataset in a spreadsheet and export it as a tab-delimited text file (i.e., `inst-proj.txt`).
4. Run `txt2pajek.exe`:
 - a. Select the Input File(s) button and locate `inst-proj.txt` with the file selection dialog. Once you have located it, the Output File(s) text box will be auto-filled with a file with the same name and extension `.net`.
 - b. Set, if not by default set, the Line Values drop box to 1 to discard the budget information and just check organisations to projects relationships.
 - c. Press the Network Type select box below and select "2-mode undirected (*Edges)".
 - d. Set the value of Header Lines to 1 (if there are column names in the first line of your `inst-proj.txt`, otherwise leave it to 0).
 - e. Finally press the Run button in the top left of the screen.
 - f. Now you have a `inst-proj.net` file located in the same folder as your original `.txt` with the same name.
 - g. Press the Exit button to close `txt2pajek.exe`.
5. Run `pajek.exe` (filename may vary according to the version you have downloaded).
 - a. From the menu, select File | Network | Read and locate the `inst-proj.net` file using the file selection dialog. If successful, you will see in the main screen the full path to this file and [2-Mode] (xx) where xx is the number of both organisations and projects in your data set, i.e. all the vertices of the network.
 - b. From the main menu, select Network | 2-mode Network | 2-Mode to 1-Mode | Rows. If successful, you will see in the main screen something like 2. Network from ROWS in affiliation network N1 (x), where x is the number of unique entities in the first column of your data.
 - c. From the main menu, select Draw | Network.
 - d. From the main menu of the window that will appear, select Layout | Energy | Kamada-Kawai | Free
 - e. From the same menu select Options | Lines | Mark Lines | With Values.
 - f. You're done!



Doing this at a larger scale: Open-source tools

Leiden University's **VOSviewer**

<https://www.vosviewer.com/>

Univ of Naples' **Bibliometrix** package for R

<https://www.bibliometrix.org/>



Part II

USE NETWORK MAPS TO GET INSIGHTS



Step 1: Get the data from Web of Science (or Scopus)

All **5721 publications** from Georgia between 2015 and 2019

The screenshot shows the Web of Science interface with search results for 'CU-GEORGIA'. The results are sorted by Date. The 'Export' button is highlighted with a red box. The results list includes:

- 1. Electrophysiological correlates of visual backward masking in patients with major depressive disorder. By: Farooq, Ophelia; da Cruz, Jeeb R.; Reinhold, Mays; et al. PSYCHIATRY RESEARCH-NEUROIMAGING Volume: 284 Article Number: 111804 Published: DEC 30 2019
- 2. Rationale and design of the AFFIRM-AHF trial: a randomised, double-blind, placebo-controlled trial comparing the effect of intravenous ferric carboxymaltose on hospitalisations and mortality in iron-deficient patients admitted for acute heart failure. By: Peslakowski, Piotr; Kirwan, Bridget Anne; Anker, Stefan D.; et al. EUROPEAN JOURNAL OF HEART FAILURE Volume: 21 Issue: 12 Pages: 1651-1658 Published: DEC 2019
- 3. Structure of Mixed Reverse Microemulsions Based on Sodium Bis (2-Ethylhexyl) Sulfosuccinate and Sodium Cholate. By: Tikkanen, Irma; Kartaniden, Marjo; Rutilahti, Marjo; et al. JOURNAL OF SURFACTANTS AND DETERGENTS Volume: 22 Issue: 2 Pages: 339-346 Published: MAR 2020
- 4. Development of the method of production of the ultrafine macrohomogeneous composite powder. By: Mestviridze, Zviad; Bantubadze, Vakhtang; Bakradashvili, Irakli; et al. MATERIALS SCIENCE AND TECHNOLOGY Volume: 36 Issue: 3 Pages: 327-333 Published: FEB 11 2020

All **4168 journal papers in English** from Georgia between 2015 and 2019

The screenshot shows the Web of Science interface with search results for 'CU-GEORGIA'. The results are sorted by Date. An 'Export Records to File' dialog box is open, showing options for 'All records on page', 'Records from 1 to 500', 'Record Content: Full Record and Cited References', and 'File Format: Plain Text'. The dialog box also includes a 'Cancel' button and a 'Search' button.

Export **Full Record and Cited References** in **Plain Text**

(You will have to do it in batches of 500 records, one at a time, so it will take some time. Be careful with your filenames)



This is what you get

Really big text files

GE-0001-0500.txt	26/11/2020 15:44	Έγγραφο κειμένου	19 933 KB
GE-0501-1000.txt	26/11/2020 15:44	Έγγραφο κειμένου	24 737 KB
GE-1001-1500.txt	26/11/2020 15:44	Έγγραφο κειμένου	28 315 KB
GE-1501-2000.txt	26/11/2020 15:44	Έγγραφο κειμένου	20 241 KB
GE-2001-2500.txt	26/11/2020 15:45	Έγγραφο κειμένου	26 052 KB
GE-2501-3000.txt	26/11/2020 15:45	Έγγραφο κειμένου	23 300 KB
GE-3001-3500.txt	26/11/2020 15:45	Έγγραφο κειμένου	21 288 KB
GE-3501-4000.txt	26/11/2020 15:45	Έγγραφο κειμένου	30 194 KB
GE-4001-4168.txt	26/11/2020 15:45	Έγγραφο κειμένου	1 383 KB

This is how a record is structured

```
PT J
AU Favrod, O
   da Cruz, JR
   Roinishvili, M
   [...]
AF Favrod, Ophelie
   da Cruz, Janir R.
   Roinishvili, Maya
   [...]
TI Electrophysiological correlates of visual backward masking in patients with major depressive disorder
SO PSYCHIATRY RESEARCH-NEUROIMAGING
LA English
DT Article
DE Global field power; N1 component; Event-related potential; Endophenotype of schizophrenia
ID SCHIZOPHRENIA; SYMPTOMS
AB Depression and schizophrenia are two psychiatric diseases with high co-morbidity. For this reason, it is important to find sensitive endophenotypes, which may disentangle the two disorders. The Shine-Through paradigm, a visual backward masking task, is a potential endophenotype for schizophrenia. Masking is strongly deteriorated in schizophrenia patients, which is reflected in reduced EEG amplitudes. Here, we tested whether masking deficits and associated EEG changes are also found in patients with major depressive disorder. First, we replicated previous findings showing that depressive patients exhibit, at most, only weak masking deficits. Second, we found that the EEG amplitudes of depressive patients were reduced compared to controls and slightly increased compared to schizophrenia patients. As a secondary analysis, we compared the performance in the masking paradigm with three cognitive tasks, namely: the Wisconsin card sorting test, a verbal fluency WA and a degraded continuous performance test. Performance in all but the verbal fluency test could discriminate schizophrenia from depression.
C1 [Favrod, Ophelie; da Cruz, Janir R.; Brand, Andreas; Herzog, Michael H.] Ecole Polytech Fed Lausanne, Lab Psychophys, Brain Mind Inst, Lausanne, Switzerland.
   [da Cruz, Janir R.; Figueiredo, Patricia] Univ Lisbon, Inst Syst & Robot Lisboa, Inst Super Tecn, Dept Bioengn, Lisbon, Portugal.
   [Roinishvili, Maya] Beritashvili Ctr Expt Biomed, Lab Vis Physiol, Tbilisi, Georgia.
   [Roinishvili, Maya; Chkonia, Eka] Agr Univ Georgia, Inst Cognit Neurosci, Tbilisi, Georgia.
   [Berdzenishvili, Ekaterine; Chkonia, Eka] Tbilisi State Med Univ, Dept Psychiat, Tbilisi, Georgia.
RP Favrod, O (corresponding author), Ecole Polytech Fed Lausanne, SV, BMI, LPSY, Stn 19, CH-1015 Lausanne, Switzerland.
EM ophelie.favrod@epfl.ch
RI Figueiredo, Patricia/A-1194-2013OI Figueiredo, Patricia/0000-0002-0743-0869; Ramos da Cruz, Janir/0000-0001-7029-5092FU NCCR Synapsy grant from the Swiss National Science Foundation [51NF40-185897]; Portuguese Fundacao para a Ciencia e a TecnologiaPortuguese Foundation for Science and Technology [FCT PD/BD/105785/2014]; "Knowledge Foundation" of GeorgiaFX This work was supported by an NCCR Synapsy grant from the Swiss National Science Foundation (51NF40-185897), by the Portuguese Fundacao para a Ciencia e a Tecnologia, grant #FCT PD/BD/105785/2014 and by the "Knowledge Foundation" of Georgia. The authors have declared that there are no conflicts of interest in relation to the subject of this study.
CR Andreasen N. C., 1984, SCALE ASSESSMENT POS
   Andreasen N.C., 1984, RIN SEIS IGA, V13, P999
   ANDREASEN NC, 1982, ARCH GEN PSYCHIAT, V39, P789
   Bach M, 1996, OPTOMETRY VISION SCI, V73, P49, DOI 10.1097/00006324-199601000-00008
   [...]
IRELANDSN 0925-4927EI 1872-7506J9 PSYCHIAT RES-NEUROIMJI Psychiatry Res. Neuroimaging
PD DEC 30
PY 2019
VL 294
AR 111004
DI 10.1016/j.psychresns.2019.111004
PG 8
WC Clinical Neurology; Neuroimaging; Psychiatry
SC Neurosciences & Neurology; Psychiatry
GA JV0EI
UT WOS:000502039300005
PM 31704371
DA 2020-11-26
ER
```



Create...

Open...

Save...

Screenshot... ▾

Manual

About VOSviewer

Create Map



Choose type of analysis and counting method

Type of analysis: ?

- Co-authorship
 Co-occurrence
 Citation
 Bibliographic coupling
 Co-citation

Unit of analysis:

- Authors
 Organizations
 Countries

Counting method: ?

- Full counting
 Fractional counting

VOSviewer thesaurus file (optional): ?

 ▾ ... Ignore documents co-authored by a large number of countries

Maximum number of countries per document: 25 ▾

< Back

Next >

Finish

Cancel

Visualization

Scale:



Labels

Size variation:

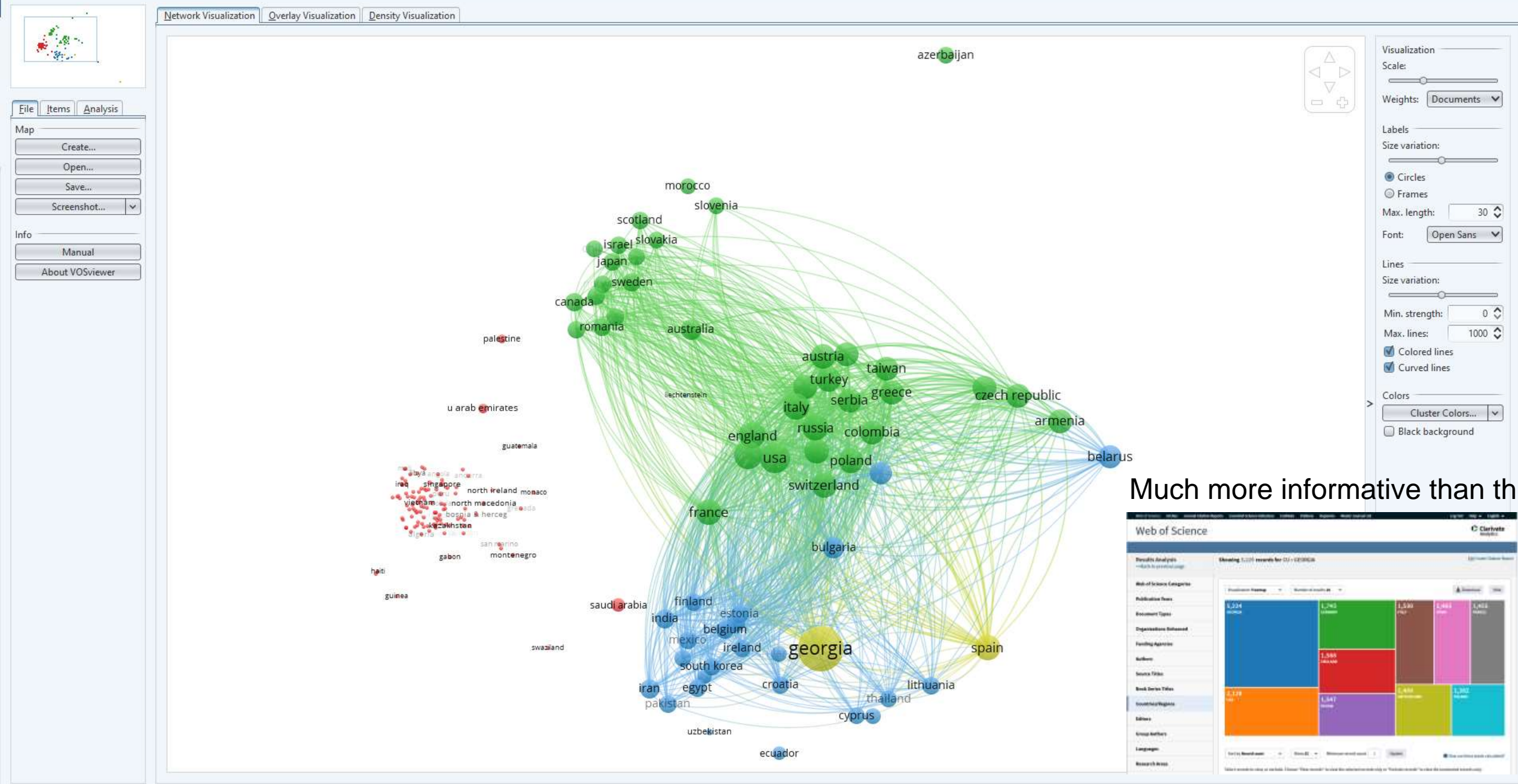
 Circles Frames

Max. length: 30 ▾

Font: Open Sans ▾

Colors

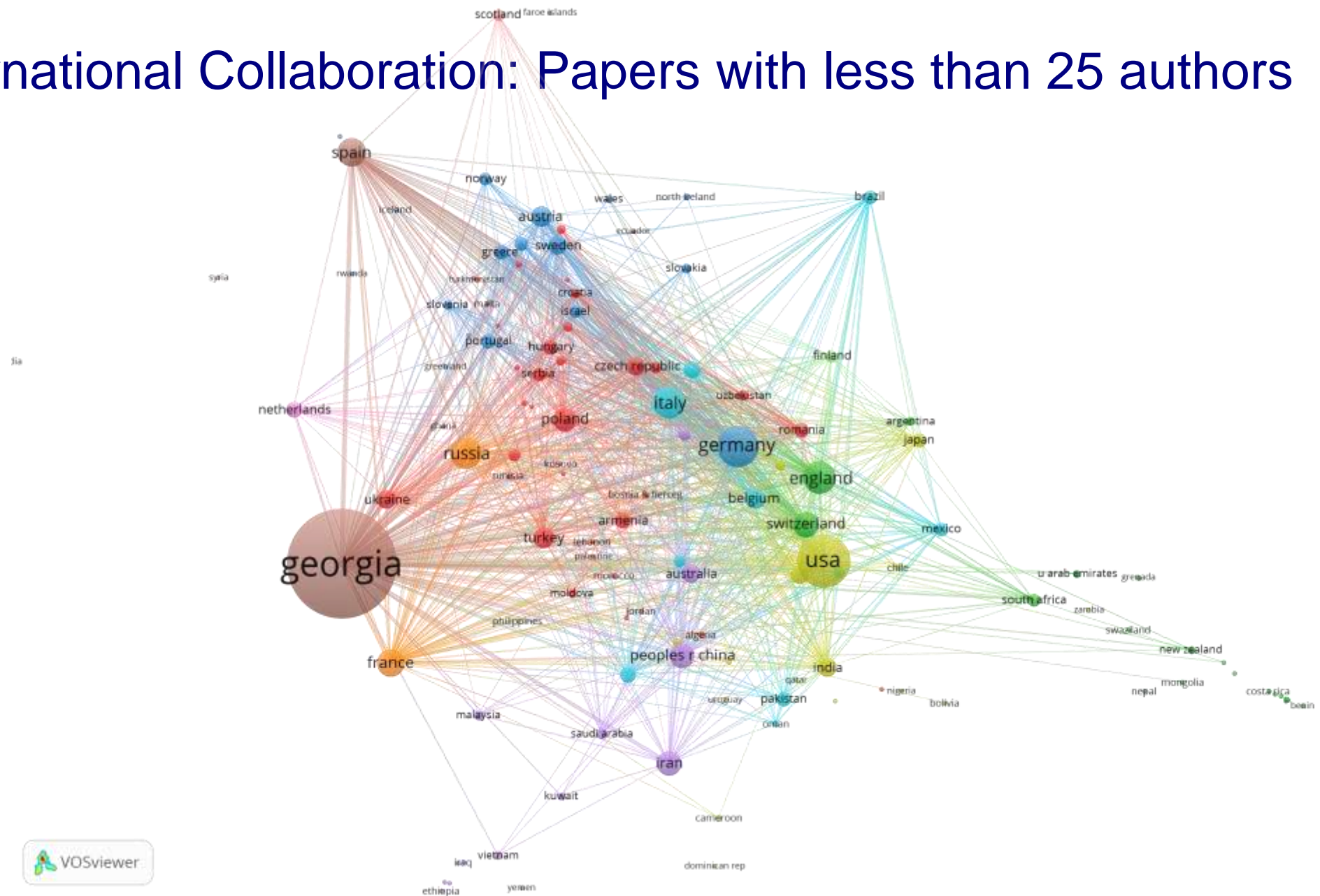
 Black background



Much more informative than this

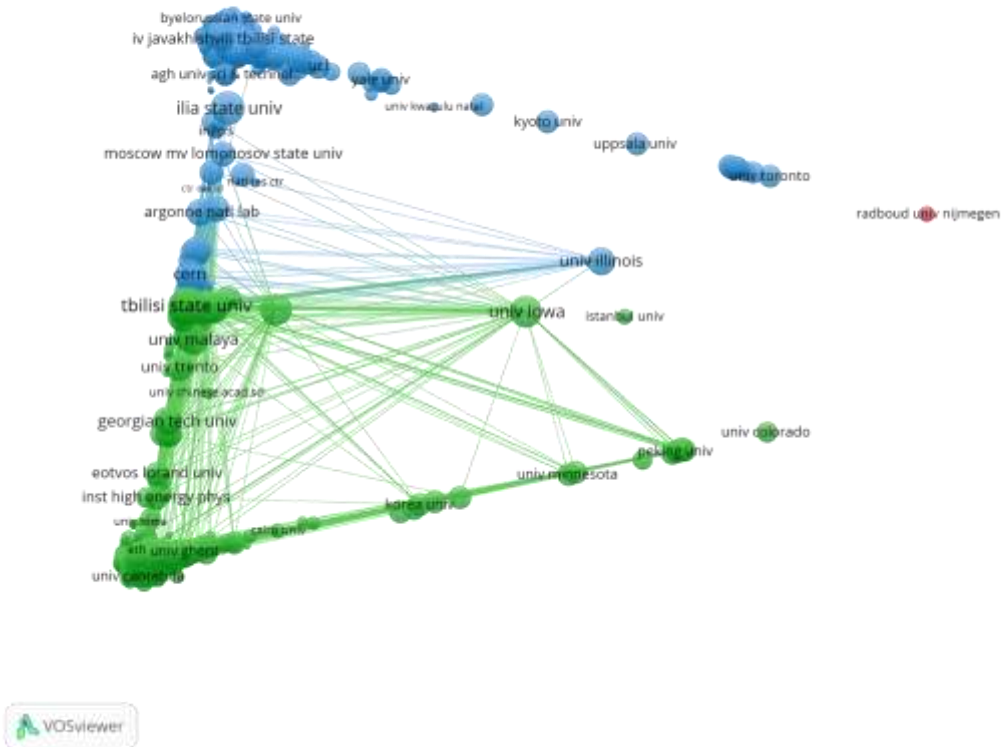


International Collaboration: Papers with less than 25 authors

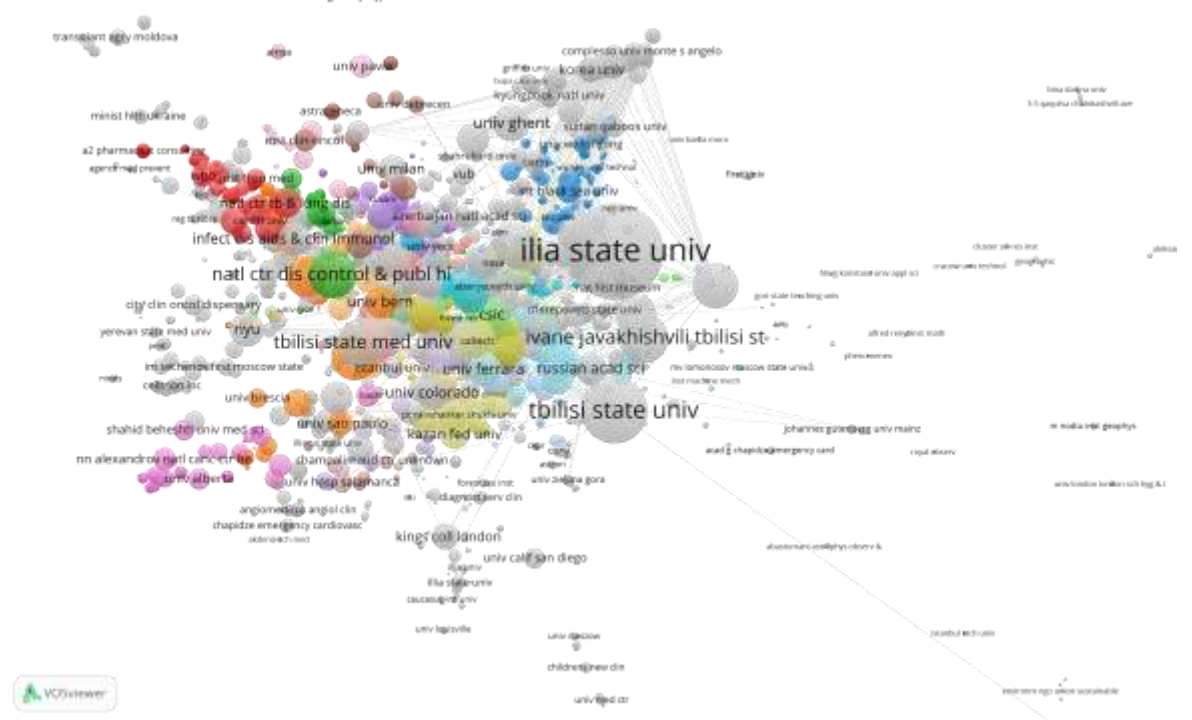


Institutional Collaborations

All publications

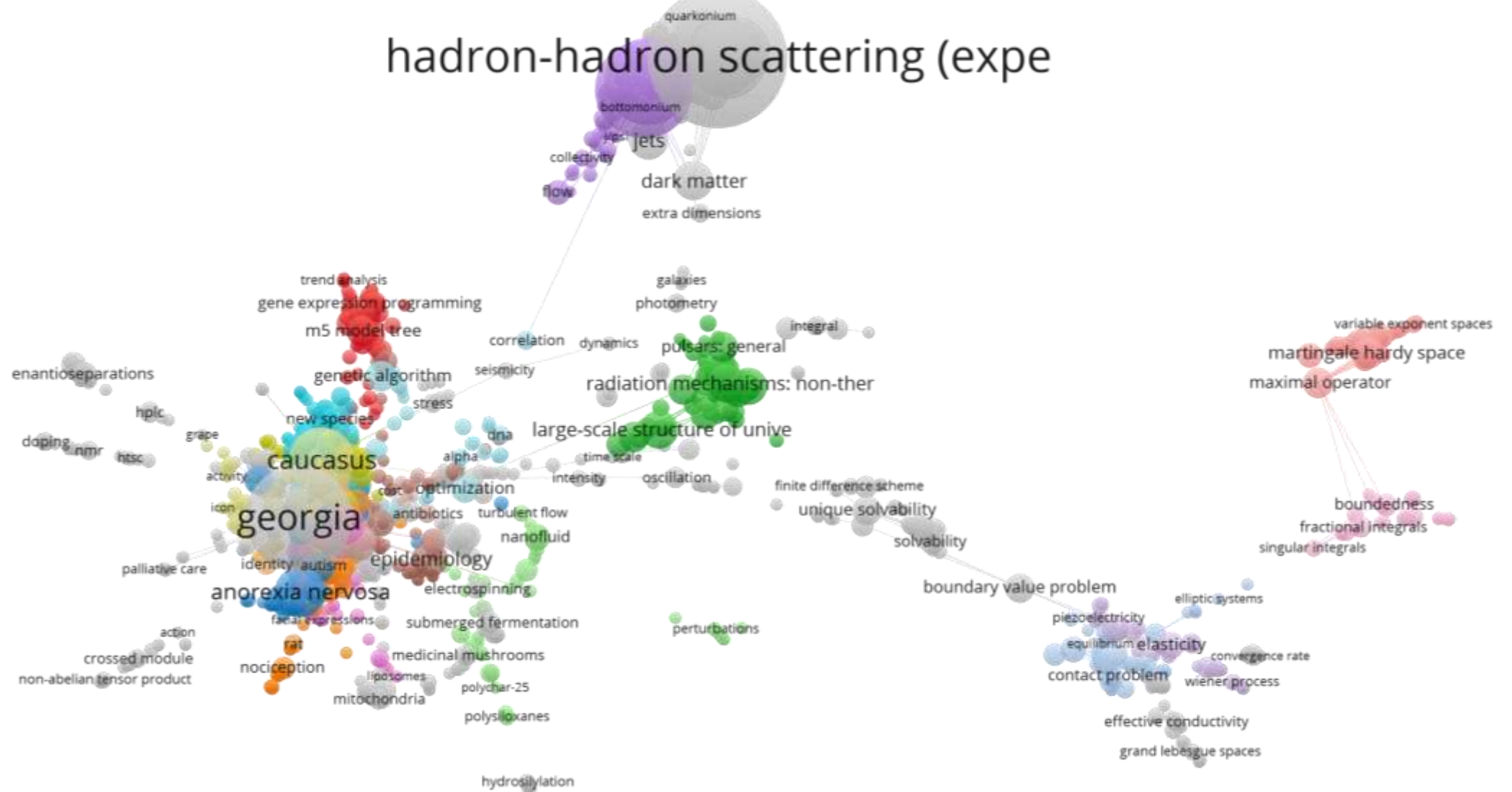


All publications with less than 25 co-authors



Keyword analysis

(all author-provided keywords appearing 3 or more times)

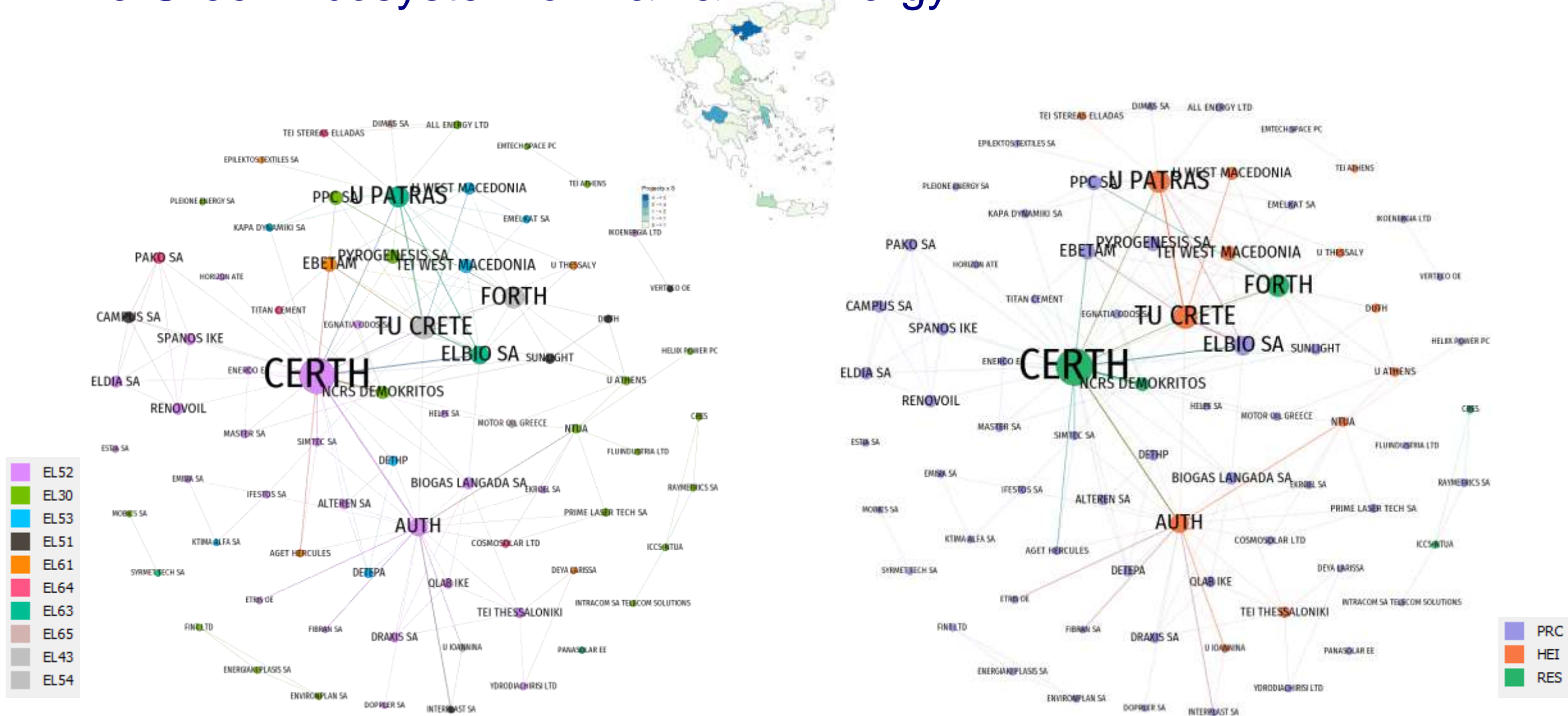


Part III

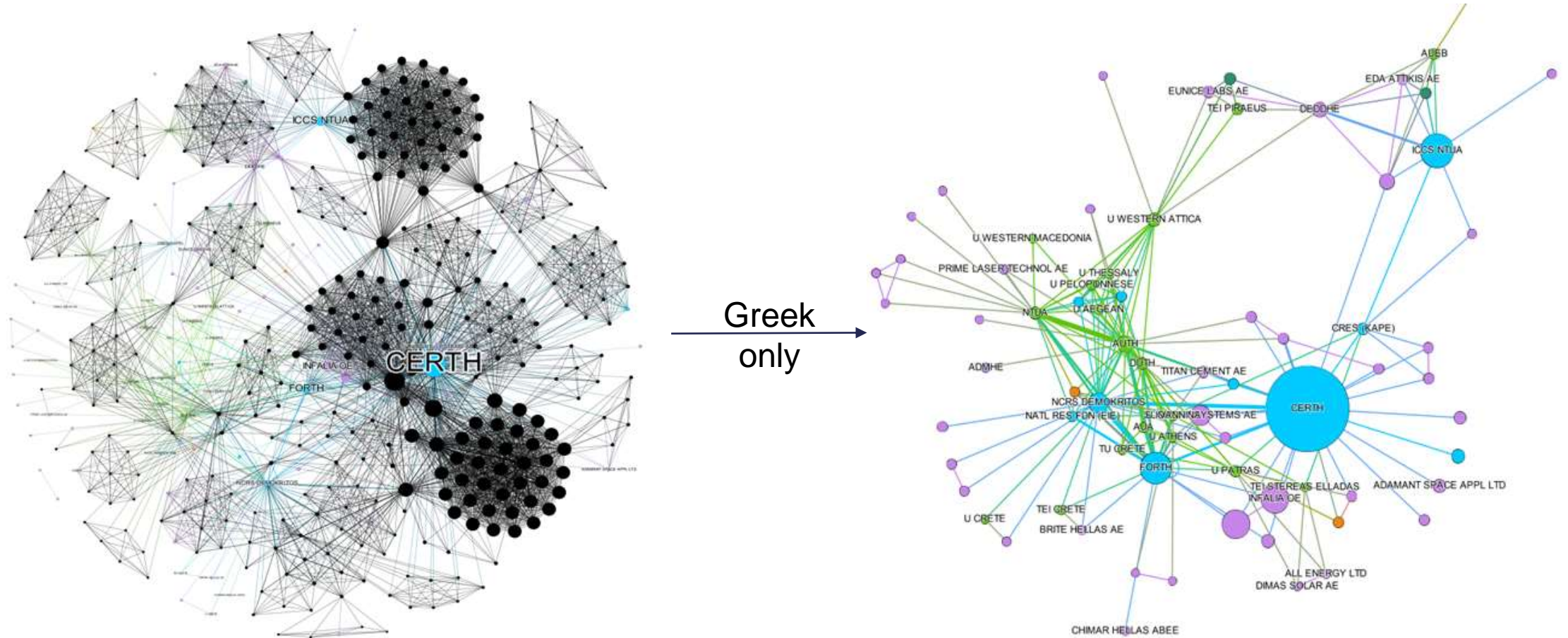
MAPPING INNOVATION ECOSYSTEMS



The Greek Ecosystem of R&D&I in Energy



Greek Participations in Horizon2020 Energy-related projects



Part IV

GETTING FULL CONTROL OF THE PROCESS (NEEDS PROGRAMMING SKILLS)



Computer-aided harvesting of Web of Science records

PT J
AU Wang, JF
Fabbiano, G
Elvis, M
Risaliti, G
Mundell, CG
Karovska, M
Zezas, A
AF Wang, Junfeng
Fabbiano, Giuseppina
Elvis, Martin
Risaliti, Guido
Mundell, Carole G.
Karovska, Margarita
Zezas, Andreas
TI A DEEP CHANDRA ACIS STUDY OF NGC 4151. II. THE INNERMOST EMISSION LINE REGION AND STRONG EVIDENCE FOR RADIO JET-NLR CLOUD COLLISION
SO ASTROPHYSICAL JOURNAL
LA English
DT Article
DE galaxies: individual (NGC 4151); galaxies: jets; galaxies: Seyfert; X-rays: galaxies
ID ACTIVE GALACTIC NUCLEI; SEYFERT-GALAXY NGC-4151; SOFT-X-RAY; HUBBLE-SPACE-TELESCOPE; H-I; CIRCUMNUCLEAR REGION; NUMERICAL-SIMULATION; EXTENDED EMISSION; VLA OBSERVATIONS; GEMINI NIFS
AB We have studied the X-ray emission within the inner similar to 150 pc radius. These maps show extended structures that are spatially correlated with the morphological correspondences with regions of X-ray enhancement, peaks of Fe VII ratios; the X-ray emission of these regions also exceeds that expected from thermal energy of the hot gas suggests that greater than or similar to 0.1% jet and the dense medium of the circumnuclear region. We find possible presence. We obtain constraints on the extended iron and silicon fluorescent emission. The alpha emission is less than or similar to 5% of the total, in disagreement with previous results.
C1 [Wang, Junfeng; Fabbiano, Giuseppina; Elvis, Martin; Risaliti, Guido; Karovska, Margarita; Mundell, Carole G.] INAF Arcetri Observ, I-50125 Florence, Italy.
[Risaliti, Guido] INAF Arcetri Observ, I-50125 Florence, Italy.
[Mundell, Carole G.] Liverpool John Moores Univ, Astrophys Res Inst, Birbeck Coll, Leazes Rd, Liverpool L69 3GB, Merseyside, England.
[Zezas, Andreas] Univ Crete, Dept Phys, GR-71003 Iraklion, Crete, Greece.

C1 [Wang, Junfeng; Fabbiano, Giuseppina; Elvis, Martin; Risaliti, Guido; Karovska, Margarita; Mundell, Carole G.] INAF Arcetri Observ, I-50125 Florence, Italy.
[Risaliti, Guido] INAF Arcetri Observ, I-50125 Florence, Italy.
[Mundell, Carole G.] Liverpool John Moores Univ, Astrophys Res Inst, Birbeck Coll, Leazes Rd, Liverpool L69 3GB, Merseyside, England.
[Zezas, Andreas] Univ Crete, Dept Phys, GR-71003 Iraklion, Crete, Greece.
RP Wang, JF (reprint author), Harvard Smithsonian Ctr Astrophys, 60 Garden St, Cambridge MA 02138, USA.
EM juwang@cfa.harvard.edu
FU NASA [G08-9101X, G01-12009X, NAS8-03060]; CXC; Royal Society; Research Councils, UK
FX We thank the anonymous referee for helpful suggestions. This work is supported by NASA grant G08-9101X (PI: Fabbiano) and grant G01-12009X (PI: Wang). We acknowledge support from the CXC, which is operated by the Smithsonian Astrophysical Observatory (SAO) for and on behalf of NASA under Contract NAS8-03060. C.G.M. acknowledges financial support from the Royal Society and Research Councils, UK. J.W. thanks G. Ferland, T. Kallman, S. Bianchi, A. Marinucci, and S. Chakravorty for extensive advices on photoionization modeling, P. Nulsen for jet power discussion, and T. Storchi-Bergmann and R. Riffel for providing the Gemini NIFS maps. This research has made use of data obtained from the Chandra Data Archive, and software provided by the CXC in the application packages CIAO and Sherpa.
NR 72
TC 8
Z9 8
PU IOP PUBLISHING LTD
PI BRISTOL
PA TEMPLE CIRCUS, TEMPLE WAY, BRISTOL BS1 6BE, ENGLAND
SN 0004-637X
J9 ASTROPHYS J
JI Astrophys. J.
PD JUL 20
PY 2011
VL 736
IS 1
AR 62
DI 10.1088/0004-637X/736/1/62
PG 10



Preparation of flat files for analysis

WOS:000242935400095 2006 J MACROMOLECULES In-plane elastic excitations in 1D polymeric photonic

WOS:000242935400095 1 CHENG, W
 WOS:000242935400095 2 GORISHNYI, T
 WOS:000242935400095 3 KRIKORIAN, V
 WOS:000242935400095 4 FYTAS, G
 WOS:000242935400095 5 THOMAS, EL

Papers x Authors (& rank of authorship):
Who has written what; Average rank;
Co-authorship networks (persons)

WOS:000242935400095 MAX PLANCK INST POLYMER RES GERMANY MAX PLANCK INST POLYMER RES, D-55128 MAINZ, GERMANY.
 WOS:000242935400095 UNIV CRETE GREECE UNIV CRETE, DEPT MAT SCI & TECHNOL, IRAKLION 71110, GREECE.
 WOS:000242935400095 FORTH GREECE FORTH, IRAKLION 71110, GREECE.
 WOS:000242935400095 MIT MA 02139 USA MIT, CTR MAT SCI & ENGN, CAMBRIDGE, MA 02139 USA.
 WOS:000242935400095 MIT MA 02139 USA MIT, INST SOLDIER NANOTECHNOL, CAMBRIDGE, MA 02139 USA.

Papers x Affiliations (Organisation, Dept., City, Country):
Co-authorship networks between Countries, Organisations,

WOS:000242935400095 Polymer Science

Departments
Papers x Subject Codes:
Filtering results, drill-down per subject codes

Papers x Times Cited:
Impact analysis

WOS:000242935400095 thin-films; bri
 poly(ethylene-terephthalate); polypropy

Papers x Author Keywords:
Approximate content analysis

Papers x Abstracts:
Better content analysis



Advanced scientometrics

Identifying and Mapping Emergent Science and Technology Domains through **Topic Modelling** → semantic characterization of a document corpus

Mapping STI Domains with a **Controlled Vocabulary** → identification of textual records related with some theme of interest

Mapping STI Domains from a **Seed Text Corpus** → Targeted analyses of STI within very specific domains

See Fuster, E., Massucci, F. A., and Matusiak, M. (2020). *Identifying specialisation domains beyond taxonomies: mapping scientific and technological domains of specialisation via semantic analyses*. In *Quantitative Methods for Place-Based Innovation Policy*, Cheltenham, UK: Edward Elgar Publishing.

<https://doi.org/10.4337/9781789905519.00014>



Summary of what you learned today


1. How to encode, process and visualize relationships
2. How to use relationships to gain insights on how your scientific production system is working
 - Internationalisation of science
 - Institutional collaborations
 - First level contextual analysis of the scientific output
 - Map Research, Development and Innovation Ecosystems
3. How to use custom tools to get deeper insights
4. Some advanced scientometric techniques





 Dr-Ing Yannis Toliás

 Innovatia Systems

 22 Dodekanissou st., GR-546 26 Thessaloniki

 +30 231 056 7442

 +30 693 605 6035

 toliás@innovatiasystems.eu

