Insights from bibliometric network properties into technology evolution - wind energy example

E. Boelman, T. Telsnig, G. Joanny

European Commission, Joint Research Centre e-mail: elisa.boelman@ec.europa.eu; thomas.telsnig@ec.europa.eu

Keywords: *wind energy, emerging technologies, bibliometrics, network analysis, metrics*

I. Introduction

The aim of this paper is to explore prospects for analysing the evolution in time of basic properties of bibliometric [1] networks to detect and better understand patterns of technology emergence and advancement. More specifically, this work explores possibilities to put some numbers to different network topologies observed in keyword co-occurrence maps, following earlier work on network densification [2] and based on insights from e.g. Bettencourt [3], Leskovec [4], Liu [5] and others.

Innovations in the area of wind energy technologies enable breakthroughs in low carbon technologies and energy system transition [6].

One of the basic physical laws governing wind energy conversion systems states that the power from a wind turbine is proportional to the square of the rotor radius and to the cube of the wind velocity. Manufacture and development of larger rotors are thus important industry and R&D strategic areas. Wind turbine rotor blades evolve towards larger designs to maximize annual electricity production and extend wind energy development to medium and low wind speed areas. To get longer and stiffer blades, manufacturers are improving wind turbine aerodynamic efficiency and reliability, and reducing weight.

II. Methodology

This study used the JRC's Tools for Innovation monitoring (TIM) software [7] to retrieve bibliometric data on wind energy from the SCOPUS database. We designed and tested targeted Boolean search strings (Annex 1) for wind turbine blades and more generally for wind energy; to put these in broader perspective, we also designed and ran a search for 'electric power'. TIM used these strings to generate bibliometric network maps of co-publication patterns by organisations as well as co-occurrences of semantically 'clean' author-keywords (as detailed in [8], TIM can group similar words into concepts and list them as semantically 'clean' keywords: clumping). TIM draws its network graphs based on the publication counts (size of the nodes) and co-occurrence of documents by two entities (edges). We exported the underlying data of these network graphs into Gephi to obtain basic network metrics for distance and connectivity. While TIM utilises the Gephi library to draw its graphs, Gephi has more extensive capabilities for providing the data in tabular format.

III. Results and Discussion

A. Co-Authorship Networks based on TIM data

Figure 1 shows densification plots (cumulative counts of organisations publishing and of co-publications) and power-law regression fits for Electric Power (◆), Wind Energy (, hereafter 'wind') and Wind Turbine Blades (●, hereafter 'blades'). While the search strings (Annex 1) are not rigorously identical, the plots clearly show that already in 1996, the search retrieved over 3000 organisations publishing on electric power (the biggest sector), ca. 250 on wind (then emerging) and only 36 for blades (clearly incipient then). The line slopes show that wind densifies at a visibly faster pace than the more established electric power. Somewhat unexpectedly, wind also densifies faster than its sub-technology 'blades'. Closer examination of the bibliometric data showed that wind densified the most between 2002 and 2008, when the counted nodes (organisations publishing) increased fivefold, and that of edges (co-publications) about eightfold. Further analysis is needed to verify e.g. whether this can be correlated to oil price hikes, and/or which sub-sectors of wind contributed the most to this remarkable intensification of scientific publication activity, as detected by our analysis.

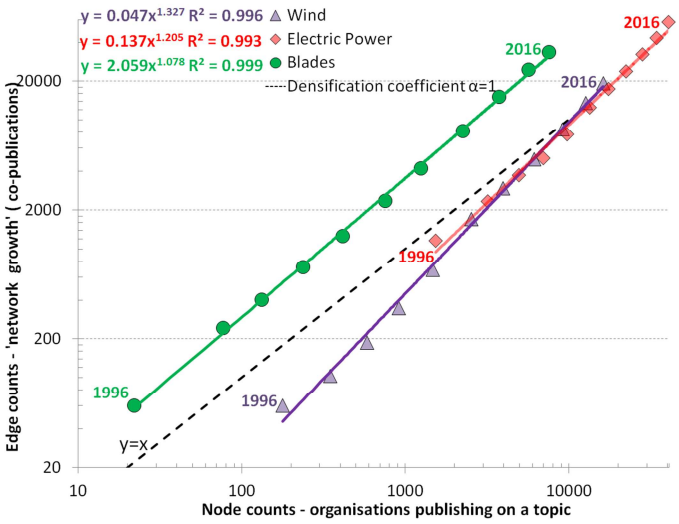

Figure 1 Size and densification of co-authorship networks, retrieved disambiguated and mapped by TIM for the search strings in Annex 1

B. *Keyword co-occurrence Networks drawn by TIM*

Figure 2 and Figure 3 show author keyword co-occurrence maps (hereafter 'keyword maps) for the 'Wind Energy' search string, for 2008 and 2009[1].

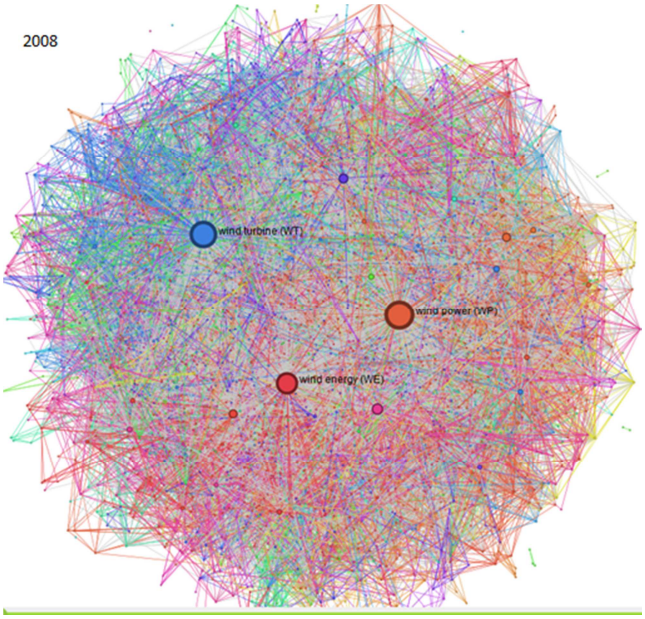

Figure 2 Co-occurrence network map of Author Keywords for the 'Wind Energy' string, year 2008 to 2008, retrieved, clumped and mapped by TIM

The keyword map clearly evolved from a relatively sparse (Figure 2) to a centre-periphery structure with a visible giant component [3], [4], [5] in Figure 3.

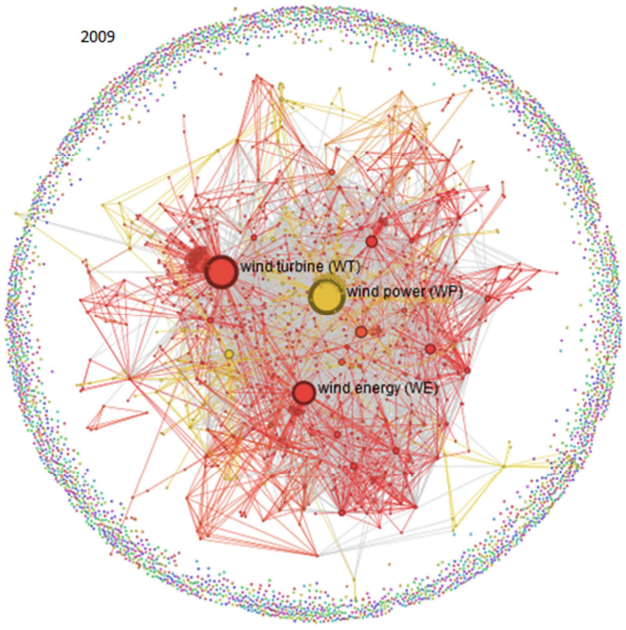

Figure 3 Co-occurrence network map of Author Keywords for the 'Wind Energy' string, year 2009 to 2009, retrieved, clumped and mapped by TIM

[1] The network maps are yearly, non cumulative

Annex 2 shows the evolution (1996-2016) of author-keywords containing 'blade', from the 'Blades' publications.

C. *Keyword network metrics based on TIM and Gephi*

Annex 3 gives a tabular yearly[1] overview of author-keyword network metrics for the 'Wind Energy' search string (Annex 1), obtained from TIM data exported to Gephi. Both node and edge counts increased tenfold from 2000 to 2008. In 2009, although node counts continued to rise, edge counts dropped by a factor of ca. 4.5, degree by a factor 6 and weighted degree by a factor 2.5. The number of connected components, on the other hand, increased by a factor of ca. 85, which might be related to the emergence of a core-periphery structure. Modularity[2], a network property often used to quantify the density of links within communities as compared to links between communities [7], steadily decreased from 1996 to 2009 and seemingly started to rise from 2014. These metrics, which need closer examination, seem to confirm the visually observed patterns in the network maps of Figure 2 and Figure 3.

IV. Conclusions

The results presented in this paper seem to confirm the potential relevance of bibliometric-network metrics for mapping technologies (at similar and complementary granularity levels) and comparing their developmental stages. Further examination of the results and their potential implications is still needed, e.g. at the levels of network properties and metrics, search string design and historic evolution of the technologies analysed.

References

[1] NOAA, “Bibliometrics.”
[2] E. Boelman, T. Telsnig, R. Shortall, G. Bardizza, and A. Villalba Pradas, “Bibliometric Network Densification Patterns for Three Renewable Energy Technologies,” in *GTM2017: 7th Global TechMining Conference*, 2017.
[3] L. M. A. Bettencourt, D. I. Kaiser, and J. Kaur, “Scientific discovery and topological transitions in collaboration networks,” *Journal of Informetrics*, vol. 3, no. 3, pp. 210–221, 2009.
[4] J. Leskovec, J. Kleinberg, and C. Faloutsos, “Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations,” *In KDD*, pp. 177–187, 2005.
[5] P. Liu and H. Xia, “Structure and evolution of co-authorship network in an interdisciplinary research field,” *Scientometrics*, vol. 103, no. 1, pp. 101–134, 2015.
[6] European Commission, *Energy Union Package - A Framework Strategy for a Resilient Energy Union with a Forward-Looking Climate Change Policy* 2015.
[7] European Commission - Joint Research Centre, “TIM technology Editor - How to make a search.” .
[8] A. Moro, E. Boelman, G. Joanny, and J. Lopez-Garcia, “A bibliometric-based technique to identify emerging photovoltaic technologies in a comparative assessment with expert review,” *Renewable Energy*, vol. 123, pp. 407–416, 2018.

[2] TIM computes modularity using the Louvain Modularity algorithm. This algorithm is a commonly accepted clustering method of nodes in network graphs and only relies on characteristics of the network, without taking into account semantic measures of similarity

ANNEX 1 – BOOLEAN SEARCH STRINGS input into TIM

Subject	Search queries
Electric Power	ti_abs_key: ("electric power" OR "electric energy" OR "power plant" OR "power station") NOT class: patent NOT class: euproject AND emm_year:[1996 TO 2016]
Wind Energy	ti_abs_key: ("wind power" OR "wind energy" OR "wind turbine" OR "wind energy converter" OR "wind farm" OR "wind park") NOT ti_abs_key: ("wind-powered sporting" OR "natural ventilation" OR "ionospheric effect" OR "ionospheric flow" OR "ocean circulation" OR "lake circulation" OR "wind forcing in the ocean boundary layer" OR "breaking surface waves" OR "stratified lake" OR "stratified water" OR "ocean mixed layer" OR "sediment transport" OR "soil erosion" OR "wind erosion" OR "sand mass flux" OR "space weather" OR "planetary atmospheric waves" OR supernova OR "molecular cloud" OR "magnetosphere" OR magnetopause OR "solar wind" OR "star formation" OR planetary OR galaxy OR interstellar OR mars OR "cable-supported bridge" OR "canopy conductance" OR "cod recruitment" OR "nest ventilation" OR phytoplankton OR "artificial aeration" OR eutrophic OR "infiltration and ventilation") NOT class: patent NOT class: euproject AND emm_year:[1996 TO 2016[3]]
Blades	ti_abs_key: ("wind power" OR "wind energy" OR "wind turbine" OR "wind energy converter" OR "wind farm" OR "wind park") AND ti_abs_key: (blades) NOT class: patent NOT class: euproject AND emm_year:[1996 TO 2016]

ANNEX 2 – BLADE-RELATED AUTHOR KEYWORD OCCURRENCES counted from co-occurrence networks produced by TIM[4]
14 most frequent occurrences (out of 75) clumped keywords, rank-ordered by decreasing frequency of occurrence

	1996	1998	2000	2002	2004	2006	2008	2010	2012	2014	2016	*Sum*
blades *(search term)*	0	**3**	1	4	6	18	34	70	155	189	169	*649*
blade element momentum	0	0	0	0	0	**2**	1	13	16	22	18	*72*
rotor blade	0	0	0	0	**1**	4	1	2	7	14	10	*39*
blade design	0	0	**1**	1	0	1	2	1	8	12	7	*33*
wind blade	0	0	0	0	0	0	**2**	1	2	10	5	*20*
straight bladed	0	0	0	0	0	**1**	1	6	3	2	2	*15*
turbine blade	0	**2**	0	0	0	1	0	2	5	2	3	*15*
blade pitch	0	0	0	0	0	1	1	3	2	4	1	*12*
flexible blade	0	0	0	0	0	0	1	4	1	4	2	*12*
blade pitch angle	0	0	0	**1**	0	0	1	3	1	3	3	*12*
blade loading	0	0	0	0	0	**1**	1	2	4	3	0	*11*
composite blades	0	0	0	0	0	0	0	0	3	4	3	*10*
blade element	0	0	0	0	0	0	0	**2**	1	3	4	*10*
straight bladed vertical axis w-t	0	0	0	0	0	0	0	0	**3**	3	3	*9*

ANNEX 3 – BASIC NETWORK METRICS FOR WIND ENERGY PUBLICATIONS, AUTHOR KEYWORD CO-OCCURRENCES from Gephi

	1996	1998	2000	2002	2004	2006	2008	2009	2010	2012	2014	2016
Giant component node counts	*158*	*124*	*280*	*612*	*1056*	*2016*	*2964*	*1064*	*1214*	*2034*	*2405*	*2538*
Giant component edge counts	*507*	*404*	*1675*	*2832*	*4612*	*10012*	*16944*	*3716*	*4151*	*7771*	*9527*	*9714*
Network nodes (author keywords)	190	154	337	688	1157	2100	3091	4304	5370	8150	9786	10205
Network Edges (co-occurrences)	579	458	1832	3004	4848	10199	**17138**	**3729**	4170	7804	9582	9773
Network average Degree	6.095	5.948	10.872	8.733	8.38	9.713	**11.089**	**1.733**	1.553	1.915	1.958	1.915
Network Avg Weighted Degree	6.695	6.065	11.318	10.183	10.054	11.624	**14.078**	**5.666**	5.261	6.756	6.684	6.624
Network Diameter, normalized	4	7	7	6	6	8	8	7	6	7	8	7
Graph Density	0.032	0.039	0.032	0.013	0.007	0.005	0.004	.0004	.0003	.00023	.0002	.00019
Modularity (resolution = 1.0)	0.68	0.714	0.663	0.597	0.566	0.507	0.446	**0.373**	**0.39**	**0.363**	0.416	0.419
Connected Components	7	8	12	18	22	19	**38**	**3231**	4140	6090	7340	7623
Avg. Clustering Coefficient	0.935	0.932	0.942	0.889	0.871	0.855	0.843	0.788	0.764	0.769	0.756	0.753
Avg. Path Length	2.681	2.888	2.952	2.869	2.909	2.979	2.828	2.788	2.755	2.704	2.808	2.801

[3] This string still retrieves a few irrelevant records, especially in the earlier years.
[4] Based on the 'Blades' search string