Unveiling Latent Relations in the Photonics Techno-Economic Complex System

Sofia Samoilis^{1,2} , Riccardo Righi^{1,2} , Montserrat Lopez-Cobo^{1,2} ,
Melisande Cardona^{1,2} , and Giuditta De Prato^{1,2}

¹ European Commission, Joint Research Centre (JRC), Seville, Spain
{sofia.samoili,riccardo.righi,montserrat.lopez-cobo,
melisande.cardona,giuditta.de-prato}@ec.europa.eu

² Edificio EXPO, Calle Inca Garcilaso, 3, 41092 Seville, Spain

Abstract. The aim of this study is to capture a technology's pathway by identifying emerging subdomains in a complex system of economic processes. The objective is to uncover indirect latent relations among agents interacting in a specific techno-economic segment (TES). A methodology, including an "Extract-Transform-Load" (ETL) process preceding the two steps aimed for analysis, is developed to analyse a TES regarding R&D economic processes of the photonics technology. In the first step, economic relevant R&D activities (EU funded projects and patents) are analysed through a multilayer network (MLN) of agents, considering their interactions in three dimensions, which represent occurred and latent relationships: co-participations in economic activities, common geographical location provenance, common use of technological terms. Then communities are detected (Infomap Algorithm for MLN), and their ongoing within and between connections are studied, as potential factors that affect the entire structured technological ecosystem. In the second step, technological subdomains associated with method-oriented and application-oriented activities are identified through topic modelling. Using the MLN structure, the textual information of the corpus of documents describing the aforementioned economic R&D activities is associated to agents, and the topic model (Latent Dirichlet Allocation) uncovers additional potential semantic connections among them. Subsequently, the results of the MLN community detection and of the topic modelling based on the descriptions of economic activities are considered. Hence, the latent relations of agents are mapped.

Keywords: Agent-artefact space · Complex networks · Bottom-up meso-structures · Community detection · Topic modelling · Semantic analysis · Natural language processing

Disclaimer: The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

© The Author(s) 2019
S. Cagnoni et al. (Eds.): WIVACE 2018, CCIS 900, pp. 72–90, 2019.
https://doi.org/10.1007/978-3-030-21733-4_6

1 Introduction

In order to follow the evolution of economic productive activities, R&D processes of economic production were investigated in the beginning of 90's [13, 15, 23], particularly by focusing in information and interactions circulation among agents (people, enterprises and institutions). While the main objective of this literature was to study the ability of expansion of national economic systems, the introduction of such approach fostered the analysis of the interactive dynamics that characterise the bottom-up development of these processes at a micro-scale [15, 28] and at a meso-scale of analysis [6, 24, 30]. Network dynamics and technology development research invoked institutional policy makers' consideration to also assess these relations.

In the prospect of comprehending interactive dynamics able to unleash the emergence of bottom-up meso-structures, the development of the techno-economic segment (hereinafter TES) analytical approach aims to analyse specific techno-economic interconnected systems and describing their dynamics [29]. Targeting also non-official heterogeneous data sources, and studying the complex underlying worldwide networks of agents, locations and technologies, TES refers to largely integrated technological (and non-technological) domains in the ICT industry that usually do not correspond to industrial or products classification. To this scope, an integrated method is developed and presented here for the R&D only part of the interconnected system, with the following stages. The flowchart of the methodological approach is presented in Fig. 1.

- A. ETL process:** "Extract-Transform-Load" (ETL) for data collection related to a specific technology, data preprocess, and keyword extraction to reduce text dimensionality
- B. Network structure:** multilayer three-level network development for unveiling latent relationships based on apparent relationships
- C. Community detection:** for analysis of strongly interacting groups of agents,
- D. Topic modelling:** for technological areas' identification, evolution, and semantic analysis.

The present first TES analysis covers the photonics industrial segment. Photonics is identified as one of the six important Key Enabling Technologies (KETs) for Europe's competitiveness [9, 34]. The technological domain of Photonics covers computational methods and applications regarding the light generation, detection and management. The methods include conversion of sunlight to electricity for renewable energy production, and the applications vary among photodiodes, LEDs, lasers and other electronic components and equipment used from a plethora of fields as ICT, life sciences, industrial manufacturing, etc. [34].

In the following sections only the stages aimed for analysis will be presented, namely stages B and C as the first analytical step, and stage D for the second step. In *Step 1* of the analysis (B and C stages), economic relevant R&D activities (EU funded projects and patents) are analysed. As the acknowledged innovation theory addresses interactions as a critical factor in the exaptation process, hence

Fig. 1. Schematic representation of methodological approach to capture a technology's pathway. The dashed line illustrates the perspective semantic analysis within the resulted communities.

in the evolution of the agent-artefact space [20–22], a network of agents is created among them. The first dimension of the network represents (a) the interactions occurred among agents through co-participations in specific type of economic activities. Then, in order to address and represent latent relationships, a second dimension regarding (b) the geographical proximity and another one regarding (c) the use of similar technological elements, are added to the network. The three dimensions are selected so as to include the three aspects that in certain works on socio-economic complex systems [21, 26] and on bio-chemical complex systems [3] are discussed as relevant in the process of the bottom-up meso-structures' formation. Correspondingly, these aspects are: (a) *activity/metabolism*, (b) *proximity/localization* and (c) *functional behaviour/information*¹. In order to inquire the aforementioned dimensions and uncover key agents and activities, a multi-layer network (MLN) is formed, in which each layer represents one of the aforementioned dimensions². Then, to identify groups of agents that, due to proximity in three dimensions, are more probable to interact intensively, the Infomap Algorithm for MLN is implemented [11]. Communities are so mapped, therefore indicating ongoing interactions within and, by means of the resulting overlaps, also between the detected groups.

In *Step 2* of the analysis (stage D), the captured dynamics are further studied, as they represent potential factors that affect the entire interconnected system. In order to unveil technological subdomains with method-oriented and application-oriented activities, a semantic analysis of the techno-economic research field of photonics is pursued. The textual information of the corpus of documents describing the aforementioned economic R&D activities that are identified and used in the previous step, are considered to structure the MLN. Topic modelling is implemented, so as to unveil technological subdomains associated with methods (e.g. chemistry) and implementations (e.g. photonic devices), that would subsequently reveal potential semantic connections among the agents. Topics are essentially summaries of the most pertinent themes of the large collection of documents. So they are used to communicate rapidly the subjects on which the agents are working. Agents grouped into the same topic do not necessarily collaborate. However, the fact that they are found active in the same topic indicates a potential aspirant collaborative relationship, or a competitive relationship. Both kinds of relationships foster innovation and suggest vibrant innovation processes, and hence prospective compelling technological sub-areas. A generative three-level hierarchical Bayesian model is employed to identify the topics that are generated based on the collected documents, namely the Latent Dirichlet Allocation (LDA). Therefore, the knowledge of a technological domain, which was only a property of experts, is approximated through this statistical, hence unbiased, method that extracts the documents' thematic content as top-

¹ In this context, the concept of “functional behaviour/information” is referred to agents' individual technological orientation.

² The layers of network can be considered one at a time to compute statistical indices and rankings of agents and activities/locations/technological terms. This part, at the moment, is out of the scope of the work.

ics [32]. A topic is a probability distribution of terms, which occurs from the underlying semantic structure of terms found in documents [4, 25, 32]. Each document is composed by one or several topics, with the latter being a combination of terms from several documents. This bag-of-words assumption is common to natural language statistical models [1, 4, 32]. Latent connections between agents are thence uncovered, as part of the natural language processing analysis. The implementation of the model is tested on the techno-economic research field of photonics with worldwide activities.

The methodological background of the network structure and community detection, and of the topic modelling will be further analysed in Subsects. 2.1 and 2.2 respectively. Subsequently, the results of the two-step analysis are presented in Sect. 3. With this last part it is possible to further investigate the co-appearance of agents and activities in topological groups, and the presence similar content in topic groups. These patterns can subsequently provide indications for the evolution of the technology. In conclusion, the presented multidimensional methodology enables the interpretation of the occurred latent relationships of the agents and their interactions.

2 Methodological Background

2.1 *Step 1: Multilayer Network Structure and Community Detection*

The set of considered economic agents is structured based on two of the world-wide most important activities of developments in productive processes. The first type of considered activities are participations in European Projects (FP7 and H2020), while the second is patent applications. The first embodies the concept of R&D processes, i.e. qualitatively new economic processes of production. The second type addresses differently R&D processes, since the included actions fix technological combinations that may not be yet formally established. After considering these two activities, the focus is set to those involved in the field of photonics. In this way, the set of agents that develop activities regarding the evolution of photonic-related technologies, is identified. Following the aim of the present study, the introduced approach focuses on interactions and information exchanges that form a network populated by identified agents. Economic agents are represented as nodes of the network, and a multidimensional space of relationships is modeled by placing different types of connections in different layers of the MLN. From a conceptual point of view, the following conjectures are established in order to justify the consideration of (i) co-participations in economic activities (first layer), (ii) common geographical location provenance (second layer), and (iii) common use of technological terms (third layer).

The first conjecture is made in order to capture relationships really occurred among agents. The hypothesis is that agents, by co-participating in a same activity (i.e. one project in FP7, or H2020, or Patent), exchange relevant information among them. Therefore, when co-participations occur, involved agents are considered as being connected. Since the analysis aims to capture the structure of

interactions for technological development, the dimension regarding information exchange necessarily represents the starting point of this work. This is the first network layer that is defined.

The second essential considered dimension addresses geographical information, as agents can establish deep connections with the place where they locate their activities. More precisely, common socio-economic habits/traditions and economic processes' integration can emerge among agents rooted in a same geographic area. This phenomenon, which favours the flourishing of fruitful and innovative interactions, is identified in economic literature with the concept of 'industrial district' [2, 7]. Therefore, connections are established when agents are located in the same region (one granularity level under the national level). The second dimension of the analysis is so justified, and the second layer of network is then structured.

The third dimension is conceptualised based on the conjecture that agents, moving towards the development of specific technological field, endorse the use of specific technology-related terms. As R&D activities are here considered, the relevant terms extracted from these activities' description³ are used to investigate the presence of agents' convergent paths/semantic proximity in terms of their technological developments [10, 12, 17]. Therefore, a connection is established among all the agents using a same term (regardless the activity). This last dimension, aimed at investigating agents' qualitative orientations, is represented by the third layer of the MLN.

Agents' interactions in three different dimensions are so represented by means of a MLN with three different layers in which the agents endorse the described types of interactions. Three additional points regarding the structure of the MLN. First of all, while initially each layer is set as a bipartite network, with agents (*first-type* node) connected to other nodes (*second-type*) whose nature depends on the layer⁴, then each layer is transformed in its one-mode projection based on the *first-type* of nodes, i.e. agents. Therefore, in the final MLN only 'agents-to-agents' connections are present. Secondly, all the connections are undirected and their weight is computed on the basis of the number of shared *second-type* nodes⁵. Finally, inter-layers connections are added to the connections considered until now (i.e. intra-layer connections). Therefore, the three 'state-nodes' of agent X , i.e. X_a and X_b and X_c where a and b and c indicate the layers, are always connected [11].

After having structured the MLN, the Infomap community detection algorithm for MLN is implemented [11]. This algorithm uses a two-level binary description based on Huffman coding and on MapEquation [14, 18, 27] and its

³ Terms are extracted from EU funded projects' abstracts and from patent applications' descriptions.

⁴ In the first layer, the types of nodes are "agents" and "activities"; in the second layer, "agents" and "locations"; in the third layer, 'agents' and 'technological terms'.

⁵ For example, if agents X and Y (*first-type* nodes) co-participate in 3 activities (*second-type* nodes), the weight of the edge $\overline{XY}$ equals 3. In the second layer the weights are all equal as agents can have one and only one location.

implementation in socio-economic complex-networks is discussed in [26]. As this algorithm minimises the length of the description of a random flow that circulates throughout the considered network, the resulting communities can be interpreted as groups of agents with intense information exchange. Regarding the settings of its implementation, in order to balance the weights of the three layers, a principle regarding the density of the layers is here taken into account. More specifically, for each layer the ratio between (i) the density of the most dense layer and (ii) its density⁶ is used as a multiplication factor for the weight of each connection belonging to that layer. In addition, Infomap algorithm is set to detect overlapping communities, i.e. agents can belong to more than one community. As a consequence, an interconnected structure of communities can emerge.

2.2 Step 2: Topic Modelling

An emerging technology may trigger interactions across several technological fields and geographical locations. Moreover, agents' spatial and economic process-related relations convey heterogeneous characteristics to the agents, which may be direct or indirect. The transversal distribution of direct or latent interactions, jointly with the heterogeneous profiles of agents, raises the issue of how to effectively detect the complex network of a technology, the technological subdomains, and most importantly the evolution of an emerging technology.

The convergence to a common use of a set of terms in order to communicate, creates a thematic group or topic. The agents using this topic adopt a common language to communicate and exchange information [10,17], and this creates topic relations. Thence, the topic relations can be directly apparent, when the agents collaborate to an activity and use a common set of terms to communicate, or else a topic.

The topological relations in conjunction to the topic relations of a complex network's structure, is a subject that has not been thoroughly studied. Relevant literature presents various methods and reports reasonable performance [12,35]. This final part of the suggested approach aims to associate and interpret all the aforementioned information, so as to render comprehensive information on innovative technological pathways that will allow policy makers and R&D managers to make timely and effective decisions. In order to capture the pathway of the technology, its subdomains should be identified, to assess their persistence, disappearance or emergence through interdisciplinary activities. To this end, the thematic content of the research and patenting activities that represent the interconnected system's development is considered through topic modelling.

The technology's subdomains are approached as topics, based on a number of documents that represent the activities performed by a technology. This conjecture is drawn from the idea that every collected activity, hence document, has one or multiple thematic subjects. To have an overview of the most salient thematic

⁶ In an undirected graph the density is defined as the number of connections divided by $N \cdot \frac{N-1}{2}$, where N is the number of nodes.

subjects of all the documents, these subjects should be grouped and provide the most pertinent topics of the entire collection of documents. Previously, in order to define the topics of the corpus that the documents form for technology mining, so with real-world corpora, expert assessment was invoked, which moreover requires a further refinement of the analysis [8]. To avoid this subjective approach, a generative hierarchical Bayesian mixed-membership model for discrete data is employed, to identify the topics that are generated based on the collected documents, namely the Latent Dirichlet Allocation (LDA) [5, 19]. The assertion is that based on the co-occurrence of terms - the probability distribution of terms - in a document, one or multiple topics are formed. The LDA model, as a mixed-membership model, returns the most probable topics from each document, and then from the entire corpus. So each document may be consisted of a mixture of topics, as the model allows to a document to belong to several topics simultaneously, which happens also in real-world. Respectively, a topic is a mixture of terms from several documents [4, 25, 32]. Accordingly, the hypotheses that are made are the following:

- A technology is approximated by a number of collected activities, which is assumed to be representative of the complex network of this technology.
- Each detected activity (patent, project) is represented by a document.
- Each document is represented by a set of terms (or else words, hereinafter interchangeably used).
- A mixture of terms creates a topic, regardless the document to which it belongs.
- A term may be assigned to one or more topics. This non-exclusive assignment of a term to a topic is sustained by the observation that in a natural language words do not belong to one topic only.
- The order of terms in a document and the order of documents in a corpus are disregarded (exchangeability assumption) [1, 5, 32]. So each document may be a mixture of topics, which are mixtures of words (bag-of-word-assumption) [5, 32].
- Each topic represents a combination of subdomains or one subdomain of the technology.

Probabilistic Generative Model Notation. To formalise the aforementioned, let the basic unit be a term, or word w , which is an element of dictionary, namely unique words, $W = \{w_1, \dots, w_V\}$. A document, which is a sequence of words, is denoted by $d = \{d_1, d_2, \dots, d_N\}$. A corpus, which is a set of documents, is denoted by $D = \{d_1^1, d_2^2, \dots, d_N^M\}$. The generative process for each document in a corpus provided by the LDA model, is outlined in the seminal article of Blei et al. [5].

Since a topic model assumes a latent structure to approach the entirety of set of words as a computable distribution over K topics, to identify the topic distribution of the corpus and the words assigned to each topic, the learning process of the LDA is then implemented. More specifically, for a topic index k , the goal is to compute the word-topic distribution $P(w|k)$, namely the most important words w of a topic k , and the topic-document distribution $P(k|d)$,

namely the prevailing topics of a document, so as to eventually converge to a robust set of topics in the corpus and compute the $P(k|w)$. Analytically, $P(w|k)$ is the probability of a word w_i assigned to a topic $k_i = j$, represented by a set of multinomial distributions ϕ over the W dictionary of words, so that $P(w_i|k_i = j) = \phi_w^j$. Accordingly, $P(k)$ is the probability distribution of topics within a document d , represented by a set of multinomial distributions θ over the K topics, so that a w in d is assigned to topic $k_i = j$, based on $P(k = j) = \theta_j^d$.

However, to define in a corpus the probabilities of word instances per topic, and the set of topics, involves enumerating a very large discrete state space and compute a probability distribution ($P(k|w)$) on that space, which requires the computation of $P(w|k)$, through ϕ and θ . So ϕ and θ will be estimated through LDA and prior probability distributions, or else priors. Let α and β be the parameters of the priors, called hyperparameters, each of them having a single value. The Dirichlet priors are conjugate priors to the multinomial distribution. Namely, as the prior distributions of the multinomial parameters are Dirichlet, respectively Dirichlet will be the posterior distributions $\phi \sim \text{Dirichlet}(\alpha)$ and $\theta \sim \text{Dirichlet}(\beta)$, which renders possible the computation of the joint distribution $P(w, k) = P(w|k) \cdot P(k)$ through integration of ϕ and θ separately. The integration of ϕ results to the computation of $P(w|k)$, and of θ to the $P(k)$.

The evaluation of the posterior distribution $P(k|w) = \frac{P(w, k)}{\sum_k P(w, k)}$ has to be indirectly computed, as the denominator creates a state space too large to enumerate, through sampling from the target distribution with Markov Chain Monte Carlo (MCMC) [16, 32]. The target distribution is approximated by a Markov chain, and then samples are selected from this chain through Gibbs sampling, which does not require tuning. All the conditional distributions of the target distribution are sampled and a sequence of observations used to approximate the latent variables of the LDA is returned. So finally, for each sample among the set of samples from the predicted $P(k|w)$ posterior distribution of words as mixtures of topics, the ϕ and θ can be estimated [16].

Topic Distances and Visualisation. Interdisciplinarity and relevance of the ensued technology-related topics is subsequently pursued, based on inter-topic distances. The hypothesis is that if two topics appear close in distance and their topic content is not thematically relevant, this points to a potential emergence of a new subdomain from the activities of these subdomains in a subsequent time period. The relevance of the topics is computed through a weight parameter, the topic prevalence with Principal Components Analysis (PCA), and the inter-topic distances with Jensen-Shannon divergence. The interdisciplinarity aspect points to potential generation of new subdomains in a subsequent time period, based on the common use of terms from co-participation to activities subdomains that were assigned together to a topic.

Let the *relevance metric* be the relevance of a term w to topic k given a weight parameter λ . This weight parameter determines the weight given to the probability of a term under a topic relative to its lift, on a log scale. It is computed as the ratio of a term's probability within a topic to its marginal probability within the corpus. A lift of $\lambda = 1$ results in a ranking of terms according to

their topic-specific probability, and lift of $\lambda = 0$ to a ranking solely according to their lift. The lift parameter decreases the appearance of globally (in the corpus) frequent terms in high ranking places. However, rare terms that occur to very few or a single topic could appear in high rankings, decreasing the interpretability of a topic if the term is very rare [31, 33].

3 Results and Discussion

Communities. The identification of cohesive groups of agents, namely communities, may assist in recognising recursive elements regarding agents' interactions and to shape and design policy interventions targeted to specific interest groups. The resulting communities are formed by agents more likely to share information in any of the three considered dimensions: activity, location and technology. After the analysis of Infomap output, communities *c.1* and *c.2* have been removed from the analysis as they were grouping together all the players participating in EU projects (*c.1*), and all the patenting players (*c.2*). Some descriptive statistics regarding the detected communities are presented in the Appendix A. Through the identity of the agents in each community, the community's profile is explored. In addition, the information regarding the structure of communities, i.e. how communities generate a structured network by having agents in common, is considered. Figure 2 immediately reveals which communities are the most connected and with which ones they share agents. Further analysis would entail the identification of the technological subdomains in which the communities are involved.

The community detection analysis in the multilayer network of R&D in photonics reveals the presence of over one hundred communities. Many communities are isolated, as their agents are not involved in other communities. This is the case of a large set of very small communities made of agents that patented alone, or that participated in few EU projects with low number of participants. When the communities' structure is observed from a country perspective, some elements emerge. While Japanese agents are mainly grouped just in community *c.3*, Chinese agents populate four large communities, namely *c.4*, *c.5*, *c.7* and *c.10*. Two of them (*c.5* and *c.7*) are connected (as they have agents in common), while the remaining ones are independent. The connection involving *c.5* and *c.7* reveals the presence of collaborations or technological convergence between the agents of these two communities. On the opposite, *c.4* and *c.10* do not have relevant convergences with other large communities and they have only smaller neighbouring communities. Therefore, the two countries with the highest propensity to develop patents show two community structures that are very different from each other. The intertwined single structure made of multiple communities observed in Fig. 2 is mainly populated by European agents. The communities reveal the presence of supra national patterns involving in different ways the countries of the Union. Some of these communities are mainly dominated by agents in a single country, e.g. *c.6* (almost all in France), *c.13* (the majority in United Kingdom), *c.22* and *c.24* (the majority of agents in Germany). Then,

Fig. 2. Structure of Photonics communities of agents detected by Infomap algorithm, in a Fruchterman-Reingold force-based layout. The largest components and some isolated communities are here presented. Each node represents a group of agents, i.e. a community. The size of the nodes is proportional to the number of agents belonging to the communities. The labels indicate the 24 most relevant communities (by amount of Infomap flow). Communities *c.1* and *c.2* have been removed from the analysis as they were grouping together all the players participating in EU projects (*c.1*), and all the patenting players (*c.2*). The width of the links is proportional the number of agents simultaneously belonging to the two communities involved.

there are many heterogeneous communities composed by agents from different countries. Some of them are mainly made of agents from two countries, like *c.8*, which includes nine agents from Germany and seven from Italy, *c.17*, which includes six agents from Germany and six from Italy, and *c.12*, which includes seven agents from United Kingdom and six from Italy. Some communities present a larger mix, as they include agents from many different countries. This is the case of *c.11*, a medium size community whose 21 agents participated in 14 FP7 projects, 5 H2020 projects and in 1 patent.

Topic Identification. Another potential grouping of agents is examined assuming that as the technological subdomains do not operate exclusively one from another, considering that they may exchange applications and methods, they present relationships that may create thematic topics. These topics ensuing from interdisciplinary combinations of the subdomains of the technology's system, are targeted to be identified with the topic modelling. Some of the subdomains that are found within one or more topics are for instance: fiber optics and waveguides, photonic devices, optical materials, light pulses, etc (see Fig. 3 and Appendix B). The potential overlap of topics sharing subdomains is studied with the thematic distance between the topics. This analysis provides insights about plausible emergence of new specialised topics, or about possible combinations of several topics into a single one due to the application of common methods.

Fifteen topics are identified in the photonics technological landscape over the period 2008 to 2016. The topics are mapped in an interactive representation

based on their thematic distance. A snapshot of this map is presented in Fig. 4. To inquire each topic's prevalence over the corpus of the technology's activities, all the resulting topics are ranked in descending order based on their occurrence frequency (Fig. 3). Moreover, the topics are classified according to the kind of use these topics are more associated with. The activities that characterise each topic are classified as follows: (i) only application-oriented, (ii) only research method-oriented, or (iii) both. A more detailed presentation of the contents of each topic per (i), (ii) or (iii) is provided in Appendix B. The colours of the bars are used to highlight this classification. The most frequently occurring topic is the multidisciplinary topic 2, which contains three interdisciplinary subdomains: fiber optics and waveguides, general optics, and lasers. All detected subdomains demonstrate activities in applications. It is followed by the topic 3 related to the subdomains of photonic devices (applications-oriented subdomain), and of physical foundations (methods-oriented subdomain). This reveals that the basic physical methods are intensively investigated for the development of photonic devices. The third most frequently occurring subdomain is regarding optical materials, such as display devices and LEDs, and it is identified only in topic 6. Subdomains regarding applications occur more frequently than subdomains regarding methodological aspects in the photonics interconnected system. This implies that the most significant evolution is expected mainly by the interactions within the subdomains regarding applied research. Furthermore, the detected combinations of subdomains within each topic indicate the potential upcoming paths of development for the considered technology.

Topic Distances. In Fig. 4, the topics resulted from the topic model are illustrated after a PCA analysis, in order to identify which topics, over the entire period 2008–2016 in the photonics R&D activities, are closer to each other. In case that the observed topics appear thematically very close to each other, e.g. the topic consisted of subdomains regarding photonic devices employing physical foundations methods (topic 3) and the topic on optical materials (topic 6), or in case that the topic appears to extensively thematically overlap with another, e.g. the topics of photonic devices applications employing methods from chemistry and biology (topic 13), of lasers for optical sensing employing methods from biology (topic 14) and of fiber optics and waveguides applications employing methods from quantum optics and physical foundations (topic 9), it is likely that their evolution will be intertwined. Namely, when topics are close in thematic distance, eventually they will either merge into a new single multidisciplinary topic including stronger relations between their corresponding subdomains, or they will diverge and create more specialised and independent topics, where their subdomains will not demonstrate interdisciplinary relations. In the first case, the destination becomes common, and in the second the origin remains common.

The terms of each topic are displayed on the side of the map, offering a comparative ranking of the most relevant terms within the topic and across all the topics. Hence, the terms that describe the subdomain that appear in the topic more pertinently are depicted by red bars Fig. 4, and the most relevant terms of the topic in the entire photonics domain, namely the terms that are relevant for

Fig. 3. Occurrence of photonics topics in R&D activities per type of topic, 2008–2016.

other topics too, are depicted by blue bars. The illustration is given here for the highest ranked topic (topic 2), but the observations are made for all.

There, it is possible to identify which topics, over the entire period 2008–2016 in the photonics R&D activities, are closer to each other. In case that the observed topics appear thematically very close to each other, e.g. the topic consisted of subdomains regarding photonic devices employing physical foundations methods (topic 3) and the topic on optical materials (topic 6), or in case that the topic appears to extensively thematically overlap with another, e.g. the topics of photonic devices applications employing methods from chemistry and biology (topic 13), of lasers for optical sensing employing methods from biology (topic 14) and of fiber optics and waveguides applications employing methods from quantum optics and physical foundations (topic 9), it is likely that their evolution will be intertwined. Namely, when topics are close in thematic distance, eventually they will either merge into a new single multidisciplinary topic including stronger relations between their corresponding subdomains, or they will diverge and create more specialised and independent topics, where their subdomains will not demonstrate interdisciplinary relations. In the first case, the destination becomes common, and in the second the origin remains common.

The terms of each topic are displayed on the side of the map, offering a comparative ranking of the most relevant terms within the topic and across all the topics. Hence, the terms that describe the subdomain that appear in the topic more pertinently are depicted by red bars 4, and the most relevant terms of the

Fig. 4. Inter-topic distance map for R&D activities in photonics, 2008–2016 (left). Top-30 most relevant terms in topic 2 (right). Red bars indicate frequency of occurrence of a term within the topic, and blue bars frequency of occurrence of the term across all topics of the corpus. Note: For the visualisation the *LDavis R package* is used. (Color figure online)

topic in the entire photonics domain, namely the terms that are relevant for other topics too, are depicted by blue bars. The illustration is given here for the highest ranked topic (topic 2), but the observations are made for all.

Topic Dynamics. To delineate the performance of the observed technological topics, the popularity of each topic, namely the topic’s probability of occurrence based on the technology-related collected documents, is analysed on a yearly basis over the time period from 2008 to 2016 (Fig. 5). However, given the short time period between its establishment and the end of the considered study period, definitive conclusions are not to be deduced. Moreover, it is noted that the number of collected activities during the last three years of the study period present a downward trend (ranging from 1.6 to 2.7 times less activities between 2014 and 2016, in comparison to the average number of activities between 2008 and 2013), due to delays in the registration of the patenting activities.

The popularity of the multidisciplinary topic with subdomains regarding applications of fiber optics and waveguides, general optics and lasers (topic 2) is among the highest in the photonics interconnected system. This is corroborated by the high occurrence of the topic, as mentioned in the previous indicator. Moreover, a 15% mean increase in occurrence is shown during the period 2014–2016 in comparison to 2008–2013. A similar interpretation can be proposed also considering the high occurring topic of optical materials applications (topic 6): its

Fig. 5. Topic evolution based on topic-document probability in Photonics R&D activities, 2008–2016.

mean occurrence increases of 4.8% during the period 2014–2016, after a drop in 2013. Also the topic of photonic devices and of methods of physical foundations occurred rather highly (topic 3, 11% mean occurrence during 2008–2016). In particular, stabilisation is observed after a period of fluctuation and final increase in 2012. The mean low popularity of the technological topic on quantum optics (topic 11, 3% mean occurrence during 2008–2016), after a marginal increasing peak in 2014, confirms the topic’s lower ranking in terms of occurrence that is found in the intertopic map. This variation is marginal and may conceal any trend due to the aforementioned issues of delay in registration and uptake of decisions.

4 Conclusions and Perspectives

The heterogeneity and spatial dimensions of agents and activities within emerging technologies, as well as the complex interactions that develop due to these factors, and the lack of correspondence to standard classifications (industrial or products), demand a methodology that identifies characteristics of the technology and the evolution of their relationships. The presented methodology through community detection and semantic analyses successfully detects not only the established R&D applications of the photonics technology based on collaborations of agents, but most importantly the latent relations of agents that appear in homogeneous groups due to spatial or thematic interactions. The analysis of these dynamic groups’ attributes offers valuable insights, which were not acknowledged without the proposed method. The identified technological pathway and subdomains in conjunction with the yielded observations, provide valuable suggestions for the development of timely and reliable policies.

The perspectives of this ongoing study comprise automatic detection of the corpus to improve topics' robustness, and semantic analysis within the resulted communities. In particular, the presence of multiple topics within the same community, or the presence of the same specific topic in distant and not else related communities, will be examined. Lastly, the TES methodology will be implemented for the entire complex network of the technology, so as to cover both R&D and non-R&D activities.

A Appendix: Detected Communities

In the following table, some descriptive statistics of the main detected communities are provided.

<p>Descriptive statistics regarding detected communities. For each community, four statistics are here reported:</p> <ul style="list-style-type: none"> - number of players - number of players by region (in parenthesis the country) - number of players by institutional type - number of activities by type <p>Community: c.3 <i>N. of Players:</i> 65; <i>N. of Players by Region:</i> Tokyo (JPN) 65; <i>N. of Players by Instit. Type:</i> companies 56, gov. non-profit 4, universities 5; <i>N. of Activities by Type:</i> FP7 1, patent 313.</p> <p>Community: c.4 <i>N. of Players:</i> 53; <i>N. of Players by Region:</i> Jiangsu (CHN) 53; <i>N. of Players by Instit. Type:</i> companies 27, gov. non-profit 27, public research institutes 4, universities 20; <i>N. of Activities by Type:</i> patent 193.</p> <p>Community: c.5 <i>N. of Players:</i> 48; <i>N. of Players by Region:</i> Beijing (CHN) 49; <i>N. of Players by Instit. Type:</i> companies 18, companies gov. non-profit 2, gov. non-profit 3, public research institutes 11, universities 14, unknown 1; <i>N. of Activities by Type:</i> patent 290.</p> <p>Community: c.6 <i>N. of Players:</i> 48; <i>N. of Players by Region:</i> Ile de France (FRA) 47, Prov. Vlaams-Brabant (BEL) 1; <i>N. of Players by Instit. Type:</i> companies 33, gov. non-profit 4, public research institutes 7, universities 4; <i>N. of Activities by Type:</i> FP7 86, H2020 20, patent 32.</p> <p>Community: c.7 <i>N. of Players:</i> 43; <i>N. of Players by Region:</i> Guangdong (CHN) 43; <i>N. of Players by Instit. Type:</i> companies 29, gov. non-profit 3, public research institutes 1, universities 10; <i>N. of Activities by Type:</i> patent 161.</p> <p>Community: c.8 <i>N. of Players:</i> 24; <i>N. of Players by Region:</i> Aquitaine (FRA) 1, Dresden (DEU) 1, Hannover (DEU) 5, Hovedstaden (DNK) 1, Liguria (ITA) 2, Lombardia (ITA) 2, Piemonte (ITA) 2, Prov. Hainaut (BEL) 1, Provence-Alpes-Côte d'Azur (FRA) 1, Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest (BEL) 1, Stuttgart (DEU) 2, Thüringen (DEU) 1, Toscana (ITA) 1, Zahodna Slovenija (SVN) 2, Zürich (CHE) 1; <i>N. of Players by Instit. Type:</i> companies 16, gov. non-profit 1, public research institutes 2, universities 5; <i>N. of Activities by Type:</i> FP7 15, H2020 4.</p> <p>Community: c.9 <i>N. of Players:</i> 40; <i>N. of Players by Region:</i> Daejeon (KOR) 1, Gyeonggi-do (KOR) 1, Seoul (KOR) 38; <i>N. of Players by Instit. Type:</i> companies 25, gov. non-profit 2, universities 13; <i>N. of Activities by Type:</i> FP7 1, patent 125.</p> <p>Community: c.10 <i>N. of Players:</i> 31; <i>N. of Players by Region:</i> Shanghai (CHN) 31; <i>N. of Players by Instit. Type:</i> companies 14, gov. non-profit 2, public research institutes 5, universities 10; <i>N. of Activities by Type:</i> FP7 1, patent 120.</p> <p>Community: c.11 <i>N. of Players:</i> 21; <i>N. of Players by Region:</i> Braunschweig (DEU) 1, Cataluña (ESP) 1, Centre (FRA) 1, Devon (GBR) 1, Dorset and Somerset (GBR) 1, Emilia-Romagna (ITA) 1, Gloucestershire, Wiltshire and Bristol/Bath area (GBR) 1, Lazio (ITA) 1, Münster (DEU) 1, País Vasco (ESP) 1, Poitou-Charentes (FRA) 1, Praha (CZE) 2, Prov. Vlaams-Brabant (BEL) 1, Stockholm (SWE) 1, Thüringen (DEU) 1, West Midlands (GBR) 1, Wielkopolskie (POL) 3, Wien (AUT) 1; <i>N. of Players by Instit. Type:</i> companies 12, public research institutes 2, universities 6, unknown 1; <i>N. of Activities by Type:</i> FP7 14, H2020 5, patent 1.</p> <p>Community: c.12 <i>N. of Players:</i> 22; <i>N. of Players by Region:</i> Attiki (GRC) 1, Campania (ITA) 3, Cataluña (ESP) 3, Hovedstaden (DNK) 1, Inner London - West (GBR) 7, Kentriki Makedonia (GRC) 1, Lazio (ITA) 1, Lombardia (ITA) 2, Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest (BEL) 1, Syddanmark (DNK) 2; <i>N. of Players by Instit. Type:</i> companies 12, companies gov. non-profit 2, public research institutes 2, universities 6; <i>N. of Activities by Type:</i> FP7 21, H2020 9.</p> <p>Community: c.13 <i>N. of Players:</i> 18; <i>N. of Players by Region:</i> Área Metropolitana de Lisboa (PRT) 1, Braunschweig (DEU) 1, Devon (GBR) 4, Dorset and Somerset (GBR) 4, Hamburg (DEU) 1, Hampshire and Isle of Wight (GBR) 1, Ile de France (FRA) 2, Lazio (ITA) 1, Norte (PRT) 1, Oberbayern (DEU) 1, Severovychod (CZE) 1; <i>N. of Players by Instit. Type:</i> companies 14, public research institutes 1, universities 3; <i>N. of Activities by Type:</i> FP7 15, H2020 4, patent 2.</p>	<p>Community: c.14 <i>N. of Players:</i> 30; <i>N. of Players by Region:</i> Gyeonggi-do (KOR) 30; <i>N. of Players by Instit. Type:</i> companies 23, gov. non-profit 1, universities 5, unknown 1; <i>N. of Activities by Type:</i> patent 115.</p> <p>Community: c.15 <i>N. of Players:</i> 22; <i>N. of Players by Region:</i> Ar Riyad (SAU) 1, Área Metropolitana de Lisboa (PRT) 1, București - Ilfov (ROU) 1, Comunidad de Madrid (ESP) 1, HaMerkaz (ISR) 12, Jerusalem (ISR) 2, Kentriki Makedonia (GRC) 1, Outer London - West and North West (GBR) 1, Prov. Liège (BEL) 1, Tel Aviv (ISR) 1; <i>N. of Players by Instit. Type:</i> companies 19, public research institutes 1, universities 2; <i>N. of Activities by Type:</i> FP7 12, H2020 2, patent 3.</p> <p>Community: c.16 <i>N. of Players:</i> 13; <i>N. of Players by Region:</i> Bratislavský kraj (SVK) 2, Devon (GBR) 1, Gloucestershire, Wiltshire and Bristol/Bath area (GBR) 1, Hampshire and Isle of Wight (GBR) 1, Homyer (BLR) 1, Ile de France (FRA) 1, Neuchâtel (CHE) 1, Łódzkie (POL) 1, Noord-Brabant (NLD) 2, Prov. Oost-Vlaanderen (BEL) 1; <i>N. of Players by Instit. Type:</i> companies 10, universities 3; <i>N. of Activities by Type:</i> FP7 4, H2020 2.</p> <p>Community: c.17 <i>N. of Players:</i> 16; <i>N. of Players by Region:</i> Ankara (TUR) 1, Amsberg (DEU) 5, Calabria (ITA) 2, Comunidad de Madrid (ESP) 1, Friuli-Venezia Giulia (ITA) 3, Hamburg (DEU) 1, Liguria (ITA) 1, Noord-Brabant (NLD) 1, South Western Scotland (GBR) 1; <i>N. of Players by Instit. Type:</i> companies 7, public research institutes 3, universities 6; <i>N. of Activities by Type:</i> FP7 16, H2020 3, patent 1.</p> <p>Community: c.18 <i>N. of Players:</i> 22; <i>N. of Players by Region:</i> California (USA) 22; <i>N. of Players by Instit. Type:</i> companies 20, universities 2; <i>N. of Activities by Type:</i> FP7 1, patent 95.</p> <p>Community: c.19 <i>N. of Players:</i> 14; <i>N. of Players by Region:</i> Ankara (TUR) 1, Área Metropolitana de Lisboa (PRT) 1, Breisgau (FRA) 1, Cataluña (ESP) 1, Düsseldorf (DEU) 1, Gloucestershire, Wiltshire and Bristol/Bath area (GBR) 1, Köln (DEU) 1, Latvija (LVA) 2, Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest (BEL) 1, Tel Aviv (ISR) 1, Toscana (ITA) 1, Mazowieckie (POL) 1, Wien (AUT) 1; <i>N. of Players by Instit. Type:</i> companies 10, companies gov. non-profit 1, gov. non-profit 1, public research institutes 2; <i>N. of Activities by Type:</i> FP7 2, H2020 2.</p> <p>Community: c.20 <i>N. of Players:</i> 15; <i>N. of Players by Region:</i> Andalucía (ESP) 4, Attiki (GRC) 1, Cataluña (ESP) 1, Gelderland (NLD) 5, Languedoc-Roussillon (FRA) 3, Overijssel (NLD) 3; <i>N. of Players by Instit. Type:</i> companies 8, public research institutes 1, universities 6; <i>N. of Activities by Type:</i> FP7 8, H2020 5.</p> <p>Community: c.21 <i>N. of Players:</i> 13; <i>N. of Players by Region:</i> Área Metropolitana de Lisboa (PRT) 1, Bretagne (FRA) 5, Comunidad de Madrid (ESP) 1, Düsseldorf (DEU) 1, Franche-Comté (FRA) 1, País Vasco (ESP) 2, Prov. Vlaams-Brabant (BEL) 1, Tel Aviv (ISR) 1; <i>N. of Players by Instit. Type:</i> companies 10, public research institutes 1, universities 2; <i>N. of Activities by Type:</i> FP7 4, H2020 2, patent 1.</p> <p>Community: c.22 <i>N. of Players:</i> 21; <i>N. of Players by Region:</i> Oberbayern (DEU) 22, Sachsen-Anhalt (DEU) 1; <i>N. of Players by Instit. Type:</i> companies 13, gov. non-profit 4, public research institutes 1, universities 3; <i>N. of Activities by Type:</i> FP7 52, H2020 9, patent 13.</p> <p>Community: c.23 <i>N. of Players:</i> 14; <i>N. of Players by Region:</i> București - Ilfov (ROU) 1, City of St. Petersburg (RUS) 4, Etelä-Suomi (FIN) 2, Hampshire and Isle of Wight (GBR) 1, Lazio (ITA) 1, Moskva (RUS) 1, Schleswig-Holstein (DEU) 1, Sar-Trøndelag (NOR) 2, Utrecht (NLD) 1; <i>N. of Players by Instit. Type:</i> companies 9, gov. non-profit 1, public research institutes 2, universities 2; <i>N. of Activities by Type:</i> FP7 5, H2020 1, patent 3.</p> <p>Community: c.24 <i>N. of Players:</i> 14; <i>N. of Players by Region:</i> Berlin (DEU) 3, Dorset and Somerset (GBR) 1, Mecklenburg-Vorpommern (DEU) 1, Oberbayern (DEU) 1, Salzburg (AUT) 1, Severovychod (CZE) 2, Stockholm (SWE) 1, Unterfranken (DEU) 4; <i>N. of Players by Instit. Type:</i> companies 8, public research institutes 1, universities 5; <i>N. of Activities by Type:</i> FP7 7, H2020 4.</p>
--	--

Note: The first two communities, namely "c.1" and "c.2", have been removed from the analysis, as they were basically grouping together all the players participating in EU project (the 1st community), and all the patenting players (the 2nd community).

B Appendix: Association of Subdomains and Method- or Application-Oriented Activities of the Photonics Market Found Based on R&D Activities (2008–2016)

In the following table, the description of topics and the contained subdomains is presented.

Topic index	Topic's Content (identified subdomains per topic)		Topic's top keywords
	Applications	Methods	
2	Fiber optics and waveguides; General optics; Lasers	-	photonic crystal fiber; laser beam; semiconductor laser; resonant cavity; beam splitter
3	Photonic devices	Physical foundations	etching; light emitting diode (LED); nitride; buffer; compound semiconductor
6	Optical materials		display device; liquid crystal; resin; light emitting diode (LED)
4	Fiber optics and waveguides	Lightwave communications	multiplexing; integrated circuit; optical waveguide; optical signal coupler
1	Telecommunications equipment; Fiber optics and waveguides	-	microwave antenna; optical signal; coupler; frequency band
12	Ink-jet printing for photonic devices; Fiber optics and waveguides	-	microsphere; polystyrene; photonic band gap; curing; ink
5	Light pulses; Photonic devices; Fiber optics and waveguides	-	photonic crystal fiber; coupler; fusion; broadband; bragg grating
8	Photonic devices (solar cells)	-	etching; lithography; crystalline silicon; electron beam; solar energy
11	-	Quantum optics	information processing; data storage; quantum communication; quantum cryptography
13	Photonic devices	Chemistry; Biology	optoelectronic; biotechnology; imager; terahertz; instrumentation
9	Fiber optics and waveguides	Quantum optics; Physical foundations	optoelectronic; nanophotonic; nanotechnology; compatibility; readout
10	Fiber optics and waveguides; Photonic devices	Optical metrology	microfluidic; biosensor; metrology; contamination; fluidic
14	Lasers for optical sensing	Biology	scalability; cascade; immunity; photodiode; lidar
7	Quantum optics	Lightwave communications	broadband; transceiver; signal processing; power consumption; heterogeneous integration; digital signal processing
15	Photonic devices	Medicine	power consumption; transmitter; nitride; foundry; integrated circuit; biopsy

References

1. Aldous, D.J.: Exchangeability and related topics. In: Hennequin, P.L. (ed.) *École d'Été de Probabilités de Saint-Flour XIII—1983*. LNM, vol. 1117, pp. 1–198. Springer, Heidelberg (1985). <https://doi.org/10.1007/BFb0099421>

2. Becattini, G.: The Marshallian industrial. In: *Industrial districts*, vol. 37 (1990)

3. Bedau, M.A., Packard, N.H., Rasmussen, S.: *Protocells: Bridging Nonliving and Living Matter*. MIT Press, Cambridge (2009)

4. Blei, D.M., Lafferty, J.D.: Topic models. In: *Text Mining: Classification, Clustering, and Applications*, vol. 10, no. 71, p. 34 (2009)

5. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**(Jan), 993–1022 (2003)

6. Breschi, S., Malerba, F.: Sectoral innovation systems: technological regimes, schumpeterian dynamics, and spatial boundaries. In: *Systems of Innovation: Technologies, Institutions and Organizations*, pp. 130–156 (1997)

7. Brusco, S.: The idea of the industrial district: its genesis. In: *Industrial Districts and Inter-firm Co-operation in Italy*, pp. 10–19 (1990)
8. Chuang, J., Manning, C.D., Heer, J.: Termite: visualization techniques for assessing textual topic models. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 74–77. ACM (2012)
9. Commission of the European Communities: COM(2009) 512. *Preparing for our Future: Developing a Common Strategy for Key Enabling Technologies in the EU*. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, September 2009
10. Cronen, V., Pearce, B.: *Communication, Action and Meaning: The Creation of Social Realities* (1980)
11. De Domenico, M., Lancichinetti, A., Arenas, A., Rosvall, M.: Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys. Rev. X* **5**(1), 011027 (2015)
12. Ding, Y.: Community detection: topological vs. topical. *J. Informetr.* **5**(4), 498–514 (2011)
13. Dosi, G., Freeman, C., Nelson, R., Silverberg, G., Soete, L., et al.: *Technical Change and Economic Theory*, vol. 988. Pinter, London (1988)
14. Fortunato, S., Hric, D.: Community detection in networks: a user guide. *Phys. Rep.* **659**, 1–44 (2016)
15. Freeman, C.: Networks of innovators: a synthesis of research issues. *Res. Policy* **20**(5), 499–514 (1991)
16. Griffiths, T.L., Steyvers, M.: Finding scientific topics. *Proc. Natl. Acad. Sci.* **101**(suppl 1), 5228–5235 (2004)
17. Hacklin, F.: *Management of Convergence in Innovation: Strategies and Capabilities for Value Creation Beyond Blurring Industry Boundaries*. Springer, Heidelberg (2007). <https://doi.org/10.1007/978-3-7908-1990-8>
18. Huffman, D.A.: A method for the construction of minimum-redundancy codes. *Proc. IRE* **40**(9), 1098–1101 (1952)
19. Jurafsky, D., Martin, J.H.: *Speech and Language Processing* (2014)
20. Lane, D., Maxfield, R., et al.: *Foresight, Complexity, and Strategy*. Santa Fe Institute New Mexico, Santa Fe (1997)
21. Lane, D.A.: Complexity and innovation dynamics. *Handbook on The Economic Complexity of Technological Change*, vol. 63 (2011)
22. Lane, D.A., Maxfield, R.R.: Ontological uncertainty and innovation. *J. Evol. Econ.* **15**(1), 3–50 (2005)
23. Lundvall, B.A., Dosi, G., Freeman, C.: Innovation as an interactive process: from user-producer interaction to the national system of innovation, **1988**, 349–369 (1988)
24. Malerba, F., Orsenigo, L.: Technological regimes and sectoral patterns of innovative activities. *Ind. Corp. Change* **6**(1), 83–118 (1997)
25. Papadimitriou, C.H., Raghavan, P., Tamaki, H., Vempala, S.: Latent semantic indexing: a probabilistic analysis. *J. Comput. Syst. Sci.* **61**(2), 217–235 (2000)
26. Righi, R.: A methodological approach to investigate interactive dynamics in innovative socio-economic complex systems. *Statistica Applicata - Ital. J. Appl. Stat.* **30**(1), 113–142 (2018)
27. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.* **105**(4), 1118–1123 (2008)
28. Russo, M., Hughes, T.: Complementary innovations and generative relationships: an ethnographic study. *Econ. Innov. New Technol.* **9**(6), 517–558 (2000)

29. Samoili, S., Righi, R., Lopez-Cobo, M., De Prato, G.: Modelling emerging topics in a techno-economic segment (TES) network. In: XII Workshop on Artificial Life and Evolutionary Computation (2017)
30. Saxenian, A.: Regional networks: industrial adaptation in silicon valley and route, vol. 128 (1994)
31. Sievert, C., Shirley, K.: LDAvis: a method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces. pp. 63–70 (2014)
32. Steyvers, M., Griffiths, T.: Probabilistic topic models. In: Handbook of Latent Semantic Analysis, vol. 427, no. 7, pp. 424–440 (2007)
33. Taddy, M.: On estimation and selection for topic models. In: Artificial Intelligence and Statistics, pp. 1184–1193 (2012)
34. Van de Velde, E.I.C., et al.: Feasibility study for an EU monitoring mechanism on key enabling technologies (2012)
35. Zhou, D., Manavoglu, E., Li, J., Giles, C.L., Zha, H.: Probabilistic models for discovering e-communities. In: Proceedings of the 15th International Conference on World Wide Web, pp. 173–182. ACM (2006)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.