Energy Policy 97 (2016) 60–72

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

Investigating biofuels through network analysis

Ylenia Curci [a,*], Christian A. Mongeau Ospina [b,c]

[a] BETA, CNRS-UMR 7522, Université de Strasbourg, 61 avenue de la Forêt Noire, 67085 Strasbourg Cedex, France
[b] Department of Economics, University Gabriele d'Annunzio, viale Pindaro 42, I-65127 Pescara, Italy
[c] a/simmetrie, Italian Association for the Study of Economic Asymmetries, via Filippo Marchetti 19, I-00199, Rome, Italy

HIGHLIGHTS

- Innovative effort is devoted to biofuels additives and modern biofuels technologies.
- A breaking trend can be observed from the second half of the last decade.
- A patent network is identified via text mining techniques that extract latent topics.

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form
1 May 2016
Accepted 2 July 2016

JEL Classifications:
C38
O31
Q16
Q42

Keywords:
Biofuel
Innovation
Patent data
Topic model
Text mining
Network analysis

ABSTRACT

Biofuel policies are motivated by a plethora of political concerns related to energy security, environmental damages, and support of the agricultural sector. In response to this, much scientific work has chiefly focussed on analysing the biofuel domain and on giving policy advice and recommendations. Although innovation has been acknowledged as one of the key factors in sustainable and cost-effective biofuel development, there is an urgent need to investigate technological trajectories in the biofuel sector by starting from consistent data and appropriate methodological tools. To do so, this work proposes a procedure to select patent data unequivocally related to the investigated sector, it uses co-occurrence of technological terms to compute patent similarity and highlights content and interdependencies of biofuels technological trajectories by revealing hidden topics from unstructured patent text fields. The analysis suggests that there is a breaking trend towards modern generation biofuels and that innovators seem to focus increasingly on the ability of alternative energy sources to adapt to the transport/industrial sector.

1. Introduction

The last decade has witnessed a period of intense instability in oil supply and prices, and there has been growing concern over the environmental costs of carbon emissions from fossil fuels in the transport sector (Sterner and Coria, 2012). In this intricate *status quo*, biofuels have been recognized as being one of the key answers for the transport sector and for domestic and international policies aimed at fostering alternative energy sources and empowering energy security (IEA, 2011). The number of global policy statements addressing this issue is soaring in both developed and developing countries.

At the beginning of the twenty-first century, the European Union (EU) started fostering the production and use of biofuels, and bioenergy in general, in several forms. Despite the fact that a specific emission target for the transport sector was already in place in 2003 (Eurostat, 2007), and despite GHG emissions decreased in the majority of sectors between 1990 and 2012, road transport still represented about 19% of total GHG emissions in the EU in 2012. In light of this, during the recent years, European policy makers have been setting several specific policy tools to reduce transport sector GHG emissions (EEA, 2014).

In 2010, the global production of biofuels amounted to 59,261 ktoe (kilotonne of oil equivalent) which represents around 1–2% of total fuel consumption in transportation. Projections on market

* Corresponding author.
E-mail address: curci@unistra.fr (Y. Curci).

http://dx.doi.org/10.1016/j.enpol.2016.07.001
0301-4215/

shares foresee a huge increase, reaching around 13% of global fuel consumption in 2050 (IEA, 2007).[1] By looking at the IEA forecasts, biofuels can provide up to 27% of world transportation fuel by 2050 (IEA, 2011). The scope of this increase will critically depend on the technological change rate and the diffusion rate of new technologies in the biofuels sector.

Despite the fact that a full deployment of the sector remains controversial, empirical evidence of the great potential of biofuels can be observed in several countries. In particular, the Brazilian ethanol sector and the Swedish alternative-energy mix represent two blatant examples of a successful integration of biofuels in the transport system. Elsewhere, only few examples of biofuel uptake can be observed despite the huge number of political and fiscal incentives the sector have been receiving worldwide for the last decade and despite the effort in R&D investments (IEA, 2011).

Recently, the massive promotion of biofuels has raised concerns over the effective economic, environmental and social benefits of their production and use (Carriquiry et al., 2011), concluding that the biofuels market would not exist without the political instruments (production incentives, tax exemptions, etc.) to support it. In order to be cost-effective and sustainable from an environmental and social point of view, the commercial production of biofuels has to rely on the most efficient, sustainable technological paradigms. In line with Suurs and Hekkert (2009), the biofuels sector can be defined as an emerging technological domain, meaning that well established technologies used in production process constitute only a portion of the technological route under development in the whole sector. Accordingly, as in many other "green sectors", innovation has been considered the key driver in biofuels development (Johnstone et al., 2008).

In recent years, several attempts have been made to depict technological trends and innovation features in environment-friendly fields (Johnstone et al., 2008; Krafft et al., 2011; De Freitas and Kaneko, 2011a). Much attention has been given to biotechnologies and environmentally sound technologies at large, in particular due to the difficulties in distinguishing between patents related to specific economic sectors (Costantini et al., 2015). Hence, in this paper we investigate biofuels by using an ad hoc patent dataset called BioPat.[2]

In this study we aim to investigate the technological drawbacks hampering the diffusion of biofuels by looking at the innovative trajectories that characterize the sector. In order to analyse this aspect, we need to disentangle innovative activities using an accurate classification. While specific data on R&D expenditures in biofuels are only available for the public sector and aggregated items (USDA, 2011), working with patent data allows innovative activity by both public and private firms to be considered and more specific technological domains can be distinguished.

The main analytical tools used in this work are drawn from the network analysis and text-mining fields. The proliferation of studies in these fields has led to the creation of a number of instruments that enable patent data to be visualized, processed and interpreted, allowing researchers to understand and explain the relationship between different technology fields (international patent classes, keywords, patent applicants, inventors, patent documents, etc.; Sternitzke et al., 2008).

The aim of this research is twofold: on the one hand, we develop an original method based on network analysis and text-mining to be applied to a narrowly defined technological domain. On the other, we apply this new methodology to biofuels data in order to map out connections within this technological domain.

[1] The OECD-FAO (2010) projection for 2010–2019 on bioethanol and biodiesel production pointed out that this amount is probably underestimated.

[2] For an exhaustive description of the BioPat database, see Costantini et al., 2013.

In terms of methodology, we contribute to the literature by:

- Developing a system to look at patent connections in a specific technological domain by replacing patent citations with discriminating keywords;
- Using an appropriate measurement for text similarity when comparing patent documents;
- Proposing a technique for the identification of the cut-off value in the definition of the network;
- Adopting a mechanism for the identification of sub-network topics.

The remainder of the paper falls in five sections. Section 2 describes the methodology used to build the patent sample, identify the links, compute the patent connections and cluster the network. Section 3 presents the main results which are discussed in Section 4. Finally, Section 5 concludes and draws policy implications of the paper's findings.

2. Methods

2.1. Patent analysis and network analysis

A network can be identified in different fields of analyses and can consider various analytical units: several actors interacting for a single purpose, knowledge spillovers across countries or institutions, inventions which build incremental innovations and so on.[3] In innovation economics, two main different types of networks can be analysed: "innovators networks",[4] based on innovative actors, and "technology networks", based on technology or on industrial sectors. The former includes collaboration and communication networks where organizations (firms, research centres, etc.) or individuals (investors, scientists, etc.) are the nodes. The latter analyses similarity or knowledge flows between technologies and among technological sectors. (van der Valk and Gijsbers, 2010). Our work elaborates on a "technology networks" based on inventions tout court, where the nodes are represented by patents themselves.

Patents provide a variety of information for the analysis of both innovators and inventions network. This information can be grouped into two major categories: structured data, which are characterized by unvarying format (classification codes, dates, citations and so on), and unstructured data, which are patchy text fields such as descriptions or claims. Moreover, patent data offer a great deal of information which can be used to observe links across both innovative actors and inventions. In particular, citations, inventors or assignees' nationality and classification codes[5] are extensively used to observe relationships across nodes (Daim et al., 2006; Huang et al., 2004; Narin, 2000; Camus and Brancaleon, 2003; Fattori et al., 2003).

Here we consider a technology network (nodes indicates patents) where links signal the similarity between the patents' technological content. Reference to the technological content of patents can be found in both structured data (e.g., in classification codes) and unstructured data (by looking at descriptive texts, in particular Title, Abstract, Claims and Description). The former, due to their homogeneity, are particularly easy to handle but they suffer from a few limitations in terms of explanatory and creative

[3] For an introduction to network theory, see Newman (2010).

[4] The importance of networks of innovators has been analysed, among others, by Powell and Grodal (2006).

[5] Cecere et al. (2012) uses co-occurrence of IPC classes to depict patents interactions. Different approaches are also possible (Krafft et al., 2011; Corrocher et al., 2007).