The Electricity Journal 29 (2016) 64–75

Contents lists available at ScienceDirect

The Electricity Journal

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

Ecosystem discovery: Measuring clean energy innovation ecosystems through knowledge discovery and mapping techniques

Jessica Lin[a,*], Supriya Chinthavali[b], Chelsey Dunivan Stahl[b], Christopher Stahl[b], Sangkeun Lee[b], Mallikarjun Shankar[b]

[a] ORISE Fellow for the U.S. Department of Energy, United States
[b] Oak Ridge National Laboratory, United States

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Clean energy innovation ecosystems
Natural language processing
Mapping

ABSTRACT

While the term 'innovation ecosystem' is often utilized, the concept is rarely quantified. Oak Ridge National Lab conducted a ground-breaking application of natural language processing, link analysis and other computational techniques to transform text and numerical data into metrics on clean energy innovation activity and geography for the U.S. Department of Energy. The project demonstrates that a machine-assisted methodology gives the user a replicable method to rapidly identify, quantify and characterize clean energy innovation ecosystems.

EPSA advanced a novel definition for clean energy innovation ecosystem as the overlap of five Ecosystem Components: 1) nascent clean energy indicators, 2) investors, 3) enabling environment, 4) networking assets and 5) large companies. The tool was created with the flexibility to allow the user to choose the weights of each of the five ecosystem components and the subcomponents. This flexibility allows the user to visualize different subsets of data as well as the composite IE rank. In an independent parallel effort, a DOE analyst in EPSA developed a short list of 22 top US clean energy innovation ecosystems; the Ecosystem Discovery tool was able to identify over 90% of the analyst-reported ecosystems. Full validation and calibration remain outstanding tasks.

The tool and the underlying datasets have the potential to address a number of important policy questions. The initial broad list of U.S. clean energy innovation ecosystems, with geographic area, technology focus, and list and types of involved organizations can help describe regional technology activities and capabilities. The implementation of knowledge discovery techniques also revealed both the potential and limitations of an automatic machine extraction methodology to gather ecosystem component data. The project demonstrates that a machine-assisted methodology gives the user a replicable method to rapidly identify, quantify, and characterize clean energy innovation ecosystems.

1. Introduction

1.1. Background

On Nov. 30, 2015, as part of the Paris Conference of Parties 21, the U.S. government along with 19 other countries announced that they would double their R&D budgets for clean energy innovation over the next five years under the umbrella of Mission Innovation. In this context, EPSA required a broad list of current clean energy innovation ecosystems (IEs) in the U.S. and their characteristics. Manual IE compilation can be labor-intensive and is subject to poorly characterized completeness. EPSA and ORNL jointly tested whether knowledge discovery methods would work to find information about clean energy IEs.

While there is much competing literature on the topic of their effectiveness, our intent was not to study whether or not an ecosystem is "effective." The purpose of this exercise was to demonstrate the feasibility of an automatic data ingest pipeline to perform text analysis, natural language processing, and link analysis to identify clean energy IEs in the U.S. in a replicable manner and, if possible, offer insights into their data characteristics. Far less literature is available with empirical data on the

* Corresponding author.
E-mail address: jessica.lin@alum.mit.edu (J. Lin).

http://dx.doi.org/10.1016/j.tej.2016.09.012
1040-6190/

characteristics and composition of IEs, so this work was intended to fill this gap.

2. Theory

Per the project's objective, we were looking for clean energy IEs that already exist and that have the ability to share resources within a close physical proximity. In addition to the human capital resource, clean energy innovations face barriers to finance due to the high capital intensity and long lead times compared to other sectors (Howell, 2015), so we deemed it important to include the five most critical components of clean energy innovation in our definition of ecosystems: nascent clean energy indicators, investors, a commercialization-enabling policy environment, networking assets, and large companies. At the inception of the project, we were not sure how heavily each factor would weigh in the ecosystem, but did want to be able to have measurable factors. Therefore, we defined a clean energy IE as the intersection of nascent clean energy assets with financing mechanisms and related companies located in a geographically concentrated area that has an enabling environment, or supportive clean energy policies, which encourages commercialization and networking assets to increase communication and collaboration opportunities as illustrated in Fig. 1.

Our working definition of an innovation ecosystem is derived from the cluster theory of competitive advantage (Harvard Business School, 2016), a highly regional strategy of economic development introduced by Michael Porter in 1990. Porter's framework defines clusters as geographically concentrated collections of firms, suppliers, and related industries. Today, economists still define a cluster as a concentration of interconnected companies that both compete and collaborate (Regalado, 2013). Brookings has recently expanded the definition of innovation ecosystems as a synergistic relationship between people, firms, and place that facilitates idea generation and accelerates commercialization (Katz and Wagner, 2014). Physical proximity is a key factor in this framework because proximity assists in idea formulation, refinement, and implementation by creating conditions conducive to the formation of teams, and connecting those teams to resources that support transition to market.

Critics to Porter argue that his theory and subsequent work pays insufficient attention to innovation and knowledge as inputs to cluster formation. For example, the U.S. Cluster Mapping website, a national initiative that provides open data on regional clusters and economies to support U.S. business, innovation, and policy, co-sponsored by Harvard Business School and the U.S. Economic Development Agency, have based their analysis on North American Industry Classification (NAIC) codes which rely on standard industrial classifications which were originally developed in the 1930s to promote the comparability of data describing various facets of the U.S. economy. This information is too broad and dated to characterize clean energy innovation ecosystems. To address this criticism, we focused on metrics that could potentially capture early-stage, technology-specific activities in each component of our framework in Fig. 1.

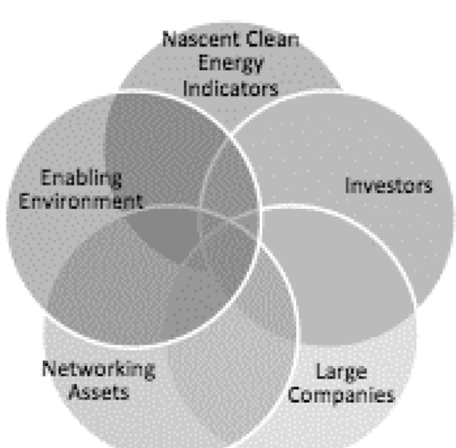

Fig. 1. Clean Energy Innovation Ecosystem.

The definitions of the five components of a clean energy IE are below.

Nascent clean energy indicators are the drivers of innovation. Depending upon the stage of the idea, these assets could range from very early stage research ideas formed in the lab or a university to a fledgling startup to an idea that was conceived in the research lab of an incumbent company. These assets are also determined by the presence of individuals who can build technologies to scale such as engineers, research scientists, and businesspeople. Nascent clean energy indicators could be measured by the presence of startup companies, patents coming out of research labs or universities or even the movement of individuals who are cited in cutting edge research reports or management teams of successful startups. Research also supports the idea that a disproportionate amount of high-technology startups in the United States are founded by highly skilled immigrants (Jackson, 2015; Economist, 2013).

Investors and financing mechanisms are key to the growth and commercialization of nascent clean energy assets. Many early stage finance companies have a geographical bias, investing locally, because their work tends to be very hands-on. Access to finance is particularly important in the clean energy space where the amount of funding needed as well as the time horizons to commercialization are typically larger than other sectors, such as IT or biotech. The funding types and levels could vary based on the size of the nascent clean energy indicators and could range from grants such as those offered by Small Business Innovation Research (SBIR) and the Advanced Research Projects Agency – Energy (ARPA-E) programs, to venture capital to crowdsourced funds. In general, these will be more risk-tolerant investing mechanisms.

The enabling environment encompasses the regulatory regime and/or the physical electricity generating characteristics of a region that would favor the growth of one sector of clean energy over another. There is evidence throughout the world that local market incentives have contributed to the diffusion of renewable technologies, and have likely enabled smaller companies to perform research in these fields (Corsatea, 2014). For example, a robust net metering program could incent distributed generating options such as solar photovoltaic (PV) technologies. The physical topology of an area could also give rise to demand for certain technologies. For example, the wind-swept plains of the Midwest or the mountain-bracketed waterways of the Pacific Northwest could create corresponding demand for wind or hydro technologies, respectively, that utilize local resources. Most of the enabling environment characteristics will be at the local or state level, rather than national and include things such as general business incentives, state and local clean energy subsidies, and other clean energy incentives.

Networking assets are events or entities that facilitate communication and collaboration among the various actors of the ecosystem. There is a growing body of research which asserts that networking assets are increasingly valuable within innovation-driven economic clusters (Katz and Wagner, 2014). Brookings characterizes an IE as the overlap of physical assets, economic assets, and networking assets where networking assets are the