Accepted version
Licence CC BY-NC-ND
Please cite as:
Marra A., Antonelli P., Pozzi C. (2016). Emerging green-tech specializations and clusters – A network analysis on technological innovation at the metropolitan level. RENEWABLE & SUSTAINABLE ENERGY REVIEWS, vol. 67, p. 1037-1046, ISSN: 1364-0321, doi: 10.1016/j.rser.2016.09.086

Emerging green-tech specializations and clusters – A network analysis on technological innovation at the metropolitan level

Alessandro Marra (University D'Annunzio, amarra@unich.it, corresponding author)

Paola Antonelli (University D'Annunzio, paola.antonelli@unich.it)

Cesare Pozzi (University of Foggia, cesare.pozzi@unifg.it)

Accepted version

Licence CC BY-NC-ND

Please cite as:

Marra A., Antonelli P., Pozzi C. (2016). Emerging green-tech specializations and clusters – A network analysis on technological innovation at the metropolitan level. RENEWABLE & SUSTAINABLE ENERGY REVIEWS, vol. 67, p. 1037-1046, ISSN: 1364-0321, doi: 10.1016/j.rser.2016.09.086

Abstract

Current models of production and energy use are promoting an effervescent green-technology (green-tech) market focused on innovation and improved environmental performance. Most green-tech firms are innovative start-ups or small and medium-sized enterprises (SMEs) characterized by large intangible assets and technological uncertainty, which makes it difficult to identify their research and development (R&D) and innovation paths. In addition, green-tech companies tend to aggregate spatially and, throughout the world, clusters are increasingly seen as a strategy to foster innovative production and R&D

activities and stimulate a sustainable energy transition. This paper investigates green-tech companies in San Francisco, New York and London to identify their specialization and underlying technological and/or market complementarities, and emerging aggregates and specific clusters. Based on information from CrunchBase and the web, we propose a network analysis using metadata on technological innovations produced by green-tech companies in these three metropolitan areas. Metadata are keywords and terms that help to describe an item, and in the database they identify the products, services and technologies driving innovation. The main metrics of the networks identified (at both node and network level) are discussed from an economic perspective. The approach helps in the design and implementation of targeted and information-based local policies and facilitates closer relationships between innovative companies, suppliers and clients, venture capitalists, large corporations and research laboratories involved in the green-tech industry.

JEL Code

L52, L86, M13, O52

Key-words

Green-technology, green clusters, green policies, network analysis, CrunchBase

1. Introduction

> Current models of production and energy use have promoted the development of an effervescent market in green-technology (green-tech) centred on innovation and improved environmental performance.

Green-tech is defined as any product, service or process that delivers value using fewer resources and producing less pollution than current standards [1]. Innovations that result in improved environmental performance fall under the green-tech umbrella: renewable energy (wind-power, solar-power, biomass, hydropower, biofuels), recycling, green transportation, green buildings, electric motors, green chemistry, lighting, grey-water, information technology, and many other energy efficient appliances [2]. These industries are sometimes very different, but all of them share a common feature: the use of new and innovative technology to create products and services with a minor impact on the environment [1].

The green-tech market is large and continues to see significant growth all over the world. The expansion of green tech industry and the success of most start-ups and small and medium-sized enterprises (SMEs) operating in the sector have been fostered mainly by changed in the market environment and costumer’s demand. Whether for e-vehicles, renewable energy or energy-efficient buildings, green technologies are becoming more and more important.

According to Advanced Energy Economy [3], in 2014 the green-tech market achieved nearly US $1.3 trillion in global revenue, the most successful result ever recorded (with an increase in global revenue of 12% compared to 2013), and almost four times more than the level achieved worldwide by the semi-conductor industry. The US innovative energy market is worth US $200 billion in revenue, an increase of 14% year on year, a rate of growth of five times the rate of the US economy overall. Green-tech is bigger than the airline industry (US $180 billion) and is nearly on a par with pharmaceuticals (US $200 billion) and consumer electronics (US $211 billion) [3]. According to the American Recovery and Reinvestment Act, passed in February 2009, an investment of approximately US $92 billion is providing direct support for green-tech in US [4][1].

Two key issues require deeper investigation. Firstly, most green-tech companies are innovative start-ups and SMEs. It has been estimated that start-ups account for over 90% of green-tech in the UK, which is similar to sectors such as information and communication technology (ICT) and biotechnology, but differs from industries such as mining where large companies dominate [6]. It is estimated also that there are about 1,400 green-tech companies around the world, which raised US$27 billion in venture capital in the period 2006-2013 [7]. Large intangible assets and technological uncertainty characterize these start-ups and SMEs, which typically involve industrial as well as individual-scale applications, and include both supply-side and demand-side technologies. Such characteristics continuously reshape the industry configuration making the identification of companies and promoting diffusion in green-tech a challenging process [8]. Obtaining updated, detailed and reliable information about the pattern of technological innovation in this

[1] Similarly, between 2006 and 2010, China invested between US $700 million and US 1.4 billion in green energy, 50% of which was focused on research and development (R&D) related to renewable energy [5].

fast growing and highly fragmented market is fundamental for both private sector players and policy makers and identifying driving products, technologies and complementarities is crucial to ensure the industry's continued growth (on the difficulties of identifying and promoting diffusion in green-tech, see [8]).

Secondly, green-tech companies tend to aggregate spatially in clusters, particularly in large urban areas, a strategy which tends to be seen as supporting innovative production and R&D activities, promoting economic spillovers and stimulating sustainable energy production (on the role played by policy in the emergence of green-tech clusters in Sweden, North Jutland, Denmark and California, USA, see [9]). Green-clusters are a widespread and successful phenomenon. For example, in 1999, the regional government in Austria decided to create a sustainable energy cluster, which so far includes 140 companies and 3,500 employees, and has a total turnover of more than €1.6 billion. The Austrian cluster includes several businesses such as solar energy (solar thermal energy, photovoltaic), wind energy, biomass and biogas, geothermal energy and heat pumps, small hydropower, energy efficiency technology and low energy buildings. D'Amico [10] and Kennedy [11] also reported about the success and the contribution of the green-tech cluster in revitalizing the economy in Worcester, the second largest city in Massachusetts, through the collaboration between academia, government and local business. McLennan [12] and Porter [13] outlined the Australian green production context, suggesting the necessity of increased private and public sector financial support for local renewable energy clusters. Normile [14] talked about Dongtan, the plan for a new eco-city on the island of Chongming in Shanghai, based on green-tech innovations, social engagements and community participation. Sustainable, renewable-based energy clusters include renewable energy technologies, smart grid technologies, clean-web, carbon trading, recycling and low-impact transportation systems [15], [13] and [9]. It is expected that green-tech companies benefit from the same agglomeration economies that have for long benefited firms in other industries, and that green-tech will generate innovation and economic spillovers that will promote business growth, market opportunity, R&D and complementary innovation.

Since green clusters encompass an amalgam of industries (e.g., electricity, transportation, building industries, software, consulting) and refer to production and R&D activities that involve multiple application pathways [16], the definition of industrial boundaries becomes more puzzling than in traditional sectors and it is not easy to define their industry boundaries: green clusters rarely conform to standard industrial classification systems (e.g., National Classification of Economic Activities, NCA; Statistical Classification of Economic Activities in the European Community, NACE) and conventional datasets and product categories have turned out to be too rigid to appreciate the changing and wide range of new technologies in the industry. Hence, it would be useful to define a detailed classification of production and R&D activities in green-tech that more accurately reflects the range of existing specializations.

The purpose of the paper is to understand the specializations and technological complementarities underlying green-tech companies, at the metropolitan level, and to detect emerging clusters, understood as geographically proximate groups of firms linked by commonalities (i.e. sharing features or attributes such as markets or technologies) and complementarities (especially complementarity between products and technologies to satisfy customers' needs).

Data are collected from CrunchBase, the world's most comprehensive database on high-tech companies, and the web. We performed text mining and web retrieval to collect metadata (keywords and terms that help to describe items, and in relation to the database reveal the markets, products and technologies driving innovation) on green-tech companies in three metropolitan areas (San Francisco, New York and London) to support a network analysis. The main metrics of the resulting networks (at both node and network level) are discussed from an economic point of view. The results have interesting implications for green policy, which are particularly useful from an operational perspective.

The paper is organized as follows. Section 2 discusses the criteria used to select the data and describes the proposed method. Section 3 compares the main metrics of the networks for San Francisco, New York and London, discusses results in an economic context and identifies emerging clusters at the metropolitan level. Section 4 highlights some operationally useful implications of the analysis and concludes with some suggestions for further research.

2. Data and Method

2.1. Dataset

CrunchBase is fully accessible through an application-programming interface. CrunchBase was established in 2007 as a crowd sourced database to track high-tech start-ups included in TechCrunch (a well regarded blog on technological innovation). Currently CrunchBase includes some 500,000 data points profiling companies, people, funds and funding, and has more than 50,000 active contributors and 2 million users accessing the data monthly. Subject to registration, users can make submissions to the database, with any suggested changes being reviewed by a moderator before final inclusion. CrunchBase editors continuously review the data to ensure their accuracy and currency.

The companies listed in CrunchBase include firms across the world, in several high-tech industries including bio-tech, green-tech, nano-tech, finance, hardware, software, mobile, e-commerce, and so on. For most start-ups and SMEs, the database includes information such as city of registration and location of operating offices, number of employees, category code, total money raised, number and timing of financing rounds, and metadata (tags) related to markets, products, services, technologies, etc. Crunchbase data are used increasingly for research. For example, Block and Sandner [17, 18] and Werth and Boert [19] employed data on the number of funding rounds and the average amount of funds to analyse, respectively, the effect of the financial crisis on the US venture capital market and the effects of business angel networks on the performance of start-up investments. Using a sample of young start-ups, Waldner [20] analysed their service-based business models to identify some of the generic characteristics of innovative business models in the recorded music industry. Data on person's profile working for a company or financial organizations have been used by Yuxian and Yuan [21] to discover if investors have a tendency to invest in companies that are socially similar to them, and by Homburg et al. [22] to demonstrate how the characteristics of the Chief Marketing Officer (CMO) might be related to the likelihood of funding. Adcock et al. [23] looked at the CrunchBase investment network and compared this network to previously studied social/information networks, to investigate the structure induced by the variations in node types, while Marra et al. [24] employed keywords on companies' products and technologies to discover specializations and technological complementarities using network analysis tools.

The observed dataset includes more than a thousand green-tech companies founded between 2001 and 2013. Data selection criteria reflect the focus on innovative and young companies and availability of information required for the descriptive purpose of the paper.

We list a couple of indicators for a subset of nearly six hundred start-ups with complete information on financing and location. Firstly, the trend in the number of green-tech start-ups and SMEs reveals that the number of companies rose substantially between 2007-2009. Secondly, for the observed subset of start-ups, the total amount raised for the years 2009-2013 was US $27.8 billion; annual financing is over US $4 billion[2].

Metadata are a reliable source of information and are consistent with other financial and geographic data provided in CrunchBase, which benefits from both a large number of contributors and the high quality of the associated technology blog. Sorting by tag relevance, the most frequent terms are generic keywords which include clean-tech, energy, solar, environment, sustainability, recycling, smart-grid, electricity, water, solar panels. Other tags represent more specialized terms and refer to emerging market niches, key technologies, and innovative products and services driving change in the industry, such as lighting control, solar water purification systems, solar monitoring software, wind-power forecasting, micro wind-turbines, micro-grid technology, wind-farms, etc.

2.2. A network analysis based on metadata

Network analysis is a well-known and popular methodology that has been developed to the point that now it is considered technically sophisticated and is used in several disciplines. Network analysis has many potential uses within the field of economics, especially to understand how networks are formed and influence economic activities [26], [27] and [28].

---

[2] The amount of funding per year of investment is consistent, in terms of trend, with other estimates (at least since 2007 when CrunchBase was launched). Data from CBInsights [25] show, for example, that the amount of funding raised by green-tech start-ups was US $8.9 billion in 2009 and US $8.3 billion in 2012.

A network is a set of nodes connected by a set of links. The nodes can be individuals, teams, firms, other organizations, sectors, patents, concepts and so on. For example, Park and Kim [29] employ network analysis to industries, Burt [30] use it in the context of product markets, and Leoncini et al. [31] and Leoncini and Montresor [32] use it for technological systems that have been disaggregated into manufacturing sectors.

Network analysis is based on the assumption of the importance of the links (edges) between nodes [33]. Several studies investigate R&D and innovation activities by referring to links between nodes and resulting networks[3].

Schrader [35] shows that the frequency of interactions among R&D employees from different firms has a positive impact on the frequency of innovation in these firms. To promote such interaction, it is important for individuals to engage in reciprocal relationships regarding the quality and quantity of knowledge being exchanged. Piore and Sabel [36] provide more direct evidence of the importance of social networks related to R&D, innovation activities and knowledge transfer. They analyse the determinants of innovative activities in industrial districts in North-eastern Italy and find supporting evidence for the hypothesis that stable networks among firms lead to both increased innovative activity and reduced transaction costs. Also in other Italian districts, Lazerson [37, 38] and Gottardi [39] achieved similar

[3] In a pioneering work on the geographical dimension of knowledge with a focus on the spatial aspects of innovation diffusion, Hägerstrand [34] examines the dissemination of innovation as a spatial process. His main hypothesis is that geographic differences in human behaviour should be analysed in terms of the knowledge available to the individual decision-maker (adopter, firm, etc.) and in terms of the networks of interpersonal communications through which the knowledge diffuses.

results, confirming the fundamental role of cohesive relationships between firms in generating knowledge, increasing efficiency and, thus, innovation. Furthermore, these facts have been abundantly demonstrated in Silicon Valley region [40], Cambridge high-tech cluster [41], Danish industrial districts [42, 43] and French ones [44, 45].

Cantner and Graf [46] apply network analysis to describe the evolution of an innovating community in Germany in a time interval of seven years, Owen-Smith et al. [47] examine R&D cooperation and compare the organization of scientific research in the US and Europe employing network analysis, Breschi and Lissoni [48] and Singh [49] find that social rather than geographical proximity is relevant for knowledge spillovers, and Balconi et al. [50] analyse inventor networks resulting from co-patenting, focusing on the specific role of academic inventors in different technological classes. Stuart and Podolny [51] provide a systematic definition of local search in the technological landscape, and firms' trajectories, Valentin and Jensen [52] suggest that the technological systems showing the best performance in the emergence of science-based technologies are those combining internal and external connections, and Paci and Batteta [53] investigate localized knowledge spillovers and examine the technological networks represented by flows of patent citations, in different sectors.

Innovation in high-tech and, more specifically, in green-tech is rapidly evolving and in order to keep up with emerging markets, products and services, technologies and changing business strategies, the proposed network analysis allows for the identification of patterns of technological change, and of existing clusters in geographically circumscribed areas. In this paper, networks are homogeneous, and links between nodes are undirected and weighted.

Based on the available data, two investigations could have been pursued:

- examination of the network of green-tech start-ups and SMEs, where the links between *i-th* start-up and *j-th* start-up result from the tags used for them both, which is based on a two-mode matrix X_c, where the rows are companies and the columns are tags;
- examination of a network of green-tech tags, where the links between *i-th* tag and *j-th* tag result from the co-existence of both tags in the same start-up, which is based on a two-mode matrix X_t, where rows are the tags and columns are the companies.

In the former case, the square matrix indicating the number of links between *i-th* start-up and *j-th* start-up is described as the adjacency matrix A_c, which is computed as the product of X_c and its transposed (X'_c).

$$X_C = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad (1)$$

$$A_C = X_C X'_C = \begin{pmatrix} - & 0 & 2 & 2 \\ 0 & - & 0 & 1 \\ 2 & 0 & - & 3 \\ 2 & 1 & 3 & - \end{pmatrix} \quad (2)$$

In this case, the approach could be described as “technological overlap” and consists of linking innovators based on their technological knowledge: the more common fields of research, the more close the relations between the innovators. In the latter case, a network of green-tech tags, the square matrix indicating the number of links between *i-th* tag and *j-th* tag is called the adjacency matrix A_t, which again is computed as the product of X_t and X'_t (equations are the same of those of adjacency matrix A_t). In the present study, we employ the second method, where the network nodes represent tags and the co-occurrence in one or more companies that use different tags is depicted along the links of the network. For example tags A and B are linked in the network if they coexist in the same company, and the weight is heavier if the number of companies in which the two tags coexist is larger. Therefore, for tags A and B, the weight of the edge A-B is 5 since they coexist in five different companies, the weight of the edge A-C is 2 since they coexist in two different companies, and so on (Table 1).

Table 1: Row data (example)

	Name	**Tags**
1	$Company_1$	A, B, C, D
2	…	A, B, E, F
3	…	A, B
4	…	A, B, C, G, H, I
5	$Company_n$	A, B, L, M, N

Stronger edges represent the links between two product and/or technological areas (tags) where innovative start-ups, usually in combination, invest tangible and intangible resources and drive their business, production activities and

R&D efforts. Weaker edges originating from a specified node provide information and can be interpreted in terms of the potential for horizontal and/or vertical expansion and collaboration based on less complementary or disruptive innovations.

To sum up, the proposed network analysis reveals links between tags (which are formed whenever these co-occur in the same company) and not between companies (which instead would be formed whenever they are active in the same market, are specialized in the same R&D domain, offer the same product, or make use of the same technology). To visualize the data, we use the open-source software package Gephi [54].

The network analysis is developed at three levels: network, node and cluster.

At network level, the main metrics used for the analysis are the average degree of all the nodes in the graph and the average weighted degree, representing respectively the total number of edges incident to each vertex and the number of edges for each vertex, ponderated by the weight of each edge. Also other metrics are employed. The density D of the network is defined as the ratio of the number of edges to the maximum number of edges possible within the network. D varies between 0 and 1; the closer it is to 1, the denser the network. This measure is sensitive to the number of nodes and can be used only to compare similar sized networks. We can also calculate the diameter of the network, as the longest among all the shortest paths in the network, while the average path length (l) is calculated by adding together the shortest path between each pair of nodes and dividing this number by the total number of pairs.

$$l_G = \frac{1}{n \cdot (n-1)} \cdot \sum_{i \neq j} d(v_i, v_j) \qquad (3)$$

where v_i and v_j are generic nodes of the network *G* and n is the total number of vertices in the network *G*. This gives the average number of steps needed to get from one network node to another.

At node level, the centrality index allows rankings that identify the most central nodes in a network. Intuitively, a node is central if it has the highest number of edges with other nodes. This concept of centrality is limited to direct links (i.e. direct connections to that node). However, centrality can also include indirect links to a specific node. For instance, in a network with a star structure, in which all nodes have ties to one central node, the centrality of the central node is equal to 1 (normalized value). While the former concept of centrality is expressed in terms of the number of nodes to which a node is connected, the latter is expressed in terms of the distances among the various nodes: two nodes are connected by a path if there is a sequence of distinct links connecting them, and the length of the path is simply the number of edges that comprise it. Betweenness centrality, $g(v)$, measures the extent to which a particular node (v) lies "between" other nodes in the network, and determines the relative importance of a node by measuring the fraction of paths connecting all pairs of nodes and containing the node of interest.

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \qquad (4)$$

where σ_{st} is the total number of shortest paths from node s and node t and $\sigma_{st}(v)$ is the number of those path that pass through v. A node with a few edges might play an important intermediary role and, thus, be very central in the network. The betweenness of a node measures the extent to which a node plays the part of gatekeeper with the potential to control many other nodes. Methodologically, measuring betweenness is complex (for more information on the algorithm, see [55]), but is intuitively meaningful.

At cluster level, modularity is a measure of network structure and determines the division of the network into clusters. High network modularity means dense connections between the nodes within modules, and sparse connections between nodes in different modules (for more information on the algorithm, see [56]). Another structural measure is calculated by averaging the clustering coefficients of all the nodes in the network. The clustering coefficient (*C*) of the *v-th* node is

$$C_v = \ 2e_v/k_v(k_v - 1) \tag{5}$$

where k_v is the number of neighbours of the *v-th* node and e_v is the number of connections between all these neighbours (for more information on the algorithm, see [57]). That is, if tags A, B, and C are all related to D, and if A is linked to B and B is linked to C, then it is highly likely that A is linked to C. When a given network's clustering measures are high, network robustness increases: in effect, in a cluster in which each tag is linked to every other tag, it is unlikely that a given tag will be essential.

3. Results and discussion

3.1. Green-tech networks comparison

Cities are often seen as incubators for young high-tech companies. San Francisco has the most of green-tech companies accounting for 7% of the start-ups in the dataset, followed by London (5.8%), New York (4.7%), Cambridge (4.6%), San Jose (3.3%), Austin (3,1%) and Boston (3,1%). California has become the epicentre of the US green-tech market with high levels of green investments and adoption of clean electricity deployment, energy efficient innovations (especially electric/hybrid vehicle adoption, smart-meter installations, solar power capacity), and green policy innovations. More than two hundred specialized green-tech companies are located in San Francisco, which receives excellent state and local financial incentives for green R&D and innovative production and has the largest concentration of green-tech venture capital investors in the US.

In the UK, London hosts several smart city and clean-web companies focusing on management construction, energy, lighting and smart transport systems to cut emissions and raise efficiency. Green-tech is related to physical (such as driverless cars and smart meters) and online (crowd-sensing apps and maps, environmental data, online marketplaces) devices.

New York is the largest end-user market in the US, and is set to become the world’s leading urban green-tech market. It has over one million buildings, nearly eight and a half million residents, US $15 billion in annual energy spending, cutting-edge sustainability policies, and growing demand for energy efficiency and green-tech products and services. New York is home to an emerging, innovation-driven green-tech segment, which uses digitally enabled solutions to address environmental and resource constraints. The next evolution of the green industry in New York will be aligned to its assets and strengths in software, finance and analytics.

The networks linking green-tech companies in San Francisco, New York and London respectively consist of 185 nodes and 782 edges, 157 nodes and 612 edges and 146 nodes and 475 edges. The magnitude of these networks is similar, permitting comparison among their metrics. Gephi allows the calculation of several descriptive statistics (Table 2). Below, we discuss the reasoning underlying our economic interpretation.

Table 2: Descriptive statistics on metropolitan green-tech networks

Metrics	**San Francisco**	**New York**	**London**
Number of nodes	185	157	146
Number of edges	782	612	475
Avg. Degree	8.5	7.8	6.5
Avg. Weighted Degree	9.9	8.4	6.7
Network Diameter	7	7	11
Graph Density	0.05	0.05	0.05
Modularity	0.7	0.8	0.7
Avg. Path Length	2.9	3.1	4.3
Avg. Clustering Coefficient	0.9	0.9	0.9

Source: Own elaboration on Crunchbase (2013)

San Francisco has the largest and most interconnected network (see Table 2).

Average degree in the San Francisco network is 8.5 (versus 7.8 for New York and 6.5 for London) and average weighted degree is 9.9 (versus 8.4 New York and 6.7 London). According to weighted degree, the highest weighted links in San Francisco are between solar-energy, solar-panels, residential-solar, solar-financing, photovoltaic-systems, pay-as-you-go-solar and solar-renewable-energy-certificates, in New York between micro-grid, hybrid-energy, solar-energy, wind-power, telecoms, green living, and in London between transport, taxis, collaborative consumption, data, analytics, smart meters and utility bills.

Network diameter for the green-tech industry is 7 in San Francisco and New York 7 and 11 in London. Graph density is 0.05 for all three cities, and modularity is 0.7 in San Francisco and London and 0.8 in New York. Average path length is 2.9 in San Francisco, 3.1 in New York and 4.3 in London. The average clustering coefficient is 0.9 in all three cities.

According to betweenness centrality, the central nodes (or major hotspots) in San Francisco are solar-energy, carbon-offset, electric-vehicles, renewable-energy-certificates, renewable-energy-credits, smart-grid and green-buildings. In New York, they are micro-grid, recycling, hybrid-energy, wind-energy, green living, solar-energy, wind-turbines and transportation. In London, they are data, smart-homes, green-buildings, electricity, waste, recycling, transport and analytics.

3.2. Focus on emerging aggregates and clusters

Green-tech is a fast growing industry, based on specific technologies and markets. According to the World Bank [58], investment in renewable power and fuels was dominated by solar in 2013, which rose from US$12.1 billion in 2004 to US$113.7 billion in 2013, with a compound annual growth rate (CAGR) of 28%, higher than any other renewable energy source (see also [59]). McKinsey [60] estimates that, since 2006, solar panel installations globally have risen by an average of 50% year on year. CBInsights [25] reports global investments in recycling increased from US$110 million to US$210 million in the period 2010-2013. Smart-grids and storage have become the preferred new energy systems. Moreover, there has been a general shift towards chemicals, transportation and energy efficiency (due to investors' preferences for capital-light deals). The companies that are likely to benefit most from this shift are those involved in water management, waste-to-energy and “clean-web” (software, applications and data analytics) technology. It results a combination of production and R&D activities that involve multiple application pathways and asks for a detailed classification that more accurately reflects the range of existing specializations in the green-tech industry. San Francisco, for example, presents a network with high numbers of products, markets and technologies, which are common to several specializations in the city.

The higher the number of nodes (i.e. the commonalities among companies), and the higher the degree of edges (in terms of degree of complementarities), the wider, more diverse and interconnected will be the resulting green-tech network. The level of interconnection of the resulting network suggests the ability of green-tech companies to exploit external knowledge for innovative activities and commercial purposes, which derives mostly from existing market and technology-specific competences, and measures the ability to assimilate innovative green products, markets and technologies [61].

An in-between perspective of analysis, beyond network and node level, could be represented by a focus on aggregates and clusters. As seen, modularity and the average clustering coefficient suggest the existence of clusters, which are groups of tags linking specializations in each metropolitan area, which allows calculation of the degree of complementarity underlying innovative production and R&D activities[4].

The San Francisco network reveals three relevant clusters: "solar energy", "carbon trading market" and "smart mobility". The first, which is related to hotspots such as solar panels, residential solar, energy management, pay-as-you-go technology, solar hot water systems, energy analytics, solar pool heating systems, photovoltaic systems, home-solar and wireless energy management, is strongly linked to real estate industry and residential buildings (Figure 1). San Francisco is promoting solar power installations through local incentive programmes, rebates for solar installations, and workforce development incentives for installers. With a generating capacity of 5 megawatts, San Francisco hosts the largest urban solar array in California. The city has stimulated private sector adoption of green building practices by

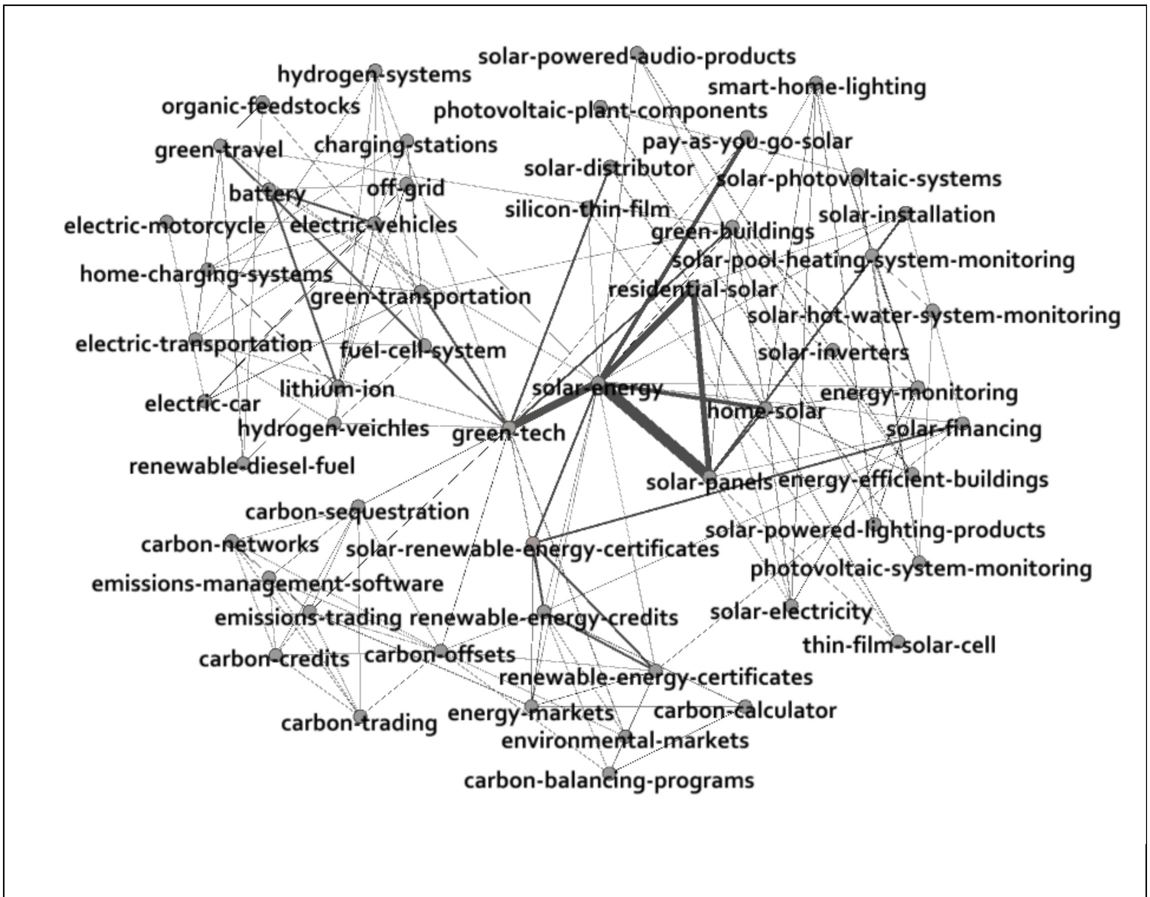
[4] Evidence on clusters at the metropolitan level is supported by diverse web sources, such as technical reports and expert analyses, case studies and industry studies by venture capitalists, market surveys on high-tech start-ups, etc.

streamlining the photovoltaic permit process, instituting a priority permit programme for private projects and establishing binding green building specifications. In 2011, San Francisco was one of the largest markets for green building opportunities and among the top ranked for sustainable buildings. Within this cluster, we can identify companies focused on more established markets such as photovoltaic and solar panels, and start-ups specialized in smart home technologies to control domestic energy use by means of digital interactive tools and energy management and analytics.

Another cluster is the “carbon trading market”, which is focused on carbon network, carbon offsets, renewable energy certificates, carbon sequestration, environmental markets, carbon balancing programmes, carbon credits and emissions management software. In 2009, San Francisco International Airport was the first airport in the US to introduce a passenger offset programme, managed by a local company, which allows them to calculate and reduce the carbon footprint of passengers' air travel. Several companies offer a wide range of business greening services including consulting, ecological foot printing, and carbon offsetting.

The “smart mobility” cluster is related to electric vehicle specialization including electric cars, green transportation, lithium-ion batteries, electric motorcycles, green travel, hydrogen technology and electric vehicle batteries. Several companies in San Francisco produce electric vehicles and related technologies and have achieved technological advances in engineering and supply of high-performance powertrains, energy storage systems, drive systems and software, and electric and hybrid vehicles. These innovations have had ripple effects on local industry, inspiring a new wave of start-ups working to design and manufacture new efficient vehicles and battery storage systems, and to develop apps to identify local charging stations, provide information on battery life, etc.

Figure 1: Green-tech clusters in San Francisco

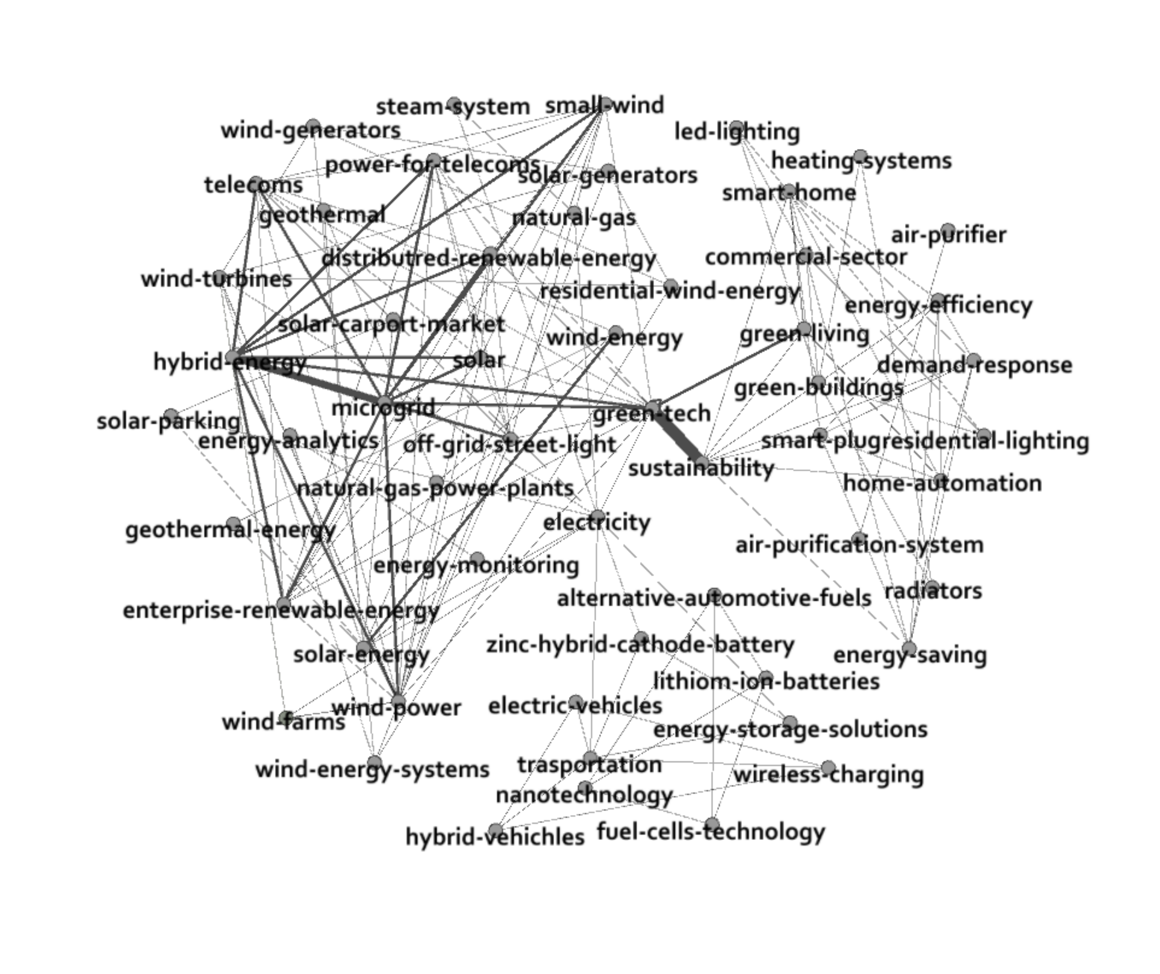


Source: Crunchbase (2013)

New York has several relevant clusters, the most relevant being related to "solar, wind and geothermal energy" (wind power, wind-farms, telecoms, hybrid energy, geothermal-energy, micro-grid, solar-energy, solar-generators, energy analytics, steam-system, and off-grid street light), which is supported by a large and diversified set of companies that apply micro-grid technology to exploit several green sources, and offer products based on various technologies (Figure 2). For example, the New York State Energy Research and Development Authority has established a Clean Energy Fund (of US $5 billion) to promote cleaner, more resilient and affordable energy. It covers a broad range of energy efficiency and renewable programmes including micro-grids to support and stimulate technology developments, particularly digital micro-grid controls, automated distribution system controls, communication protocols, data analytics and asynchronous inverters [62]. The “smart building” cluster is related to energy-saving, led-lighting, home automation, green-living, smart-plugs, commercial-buildings, chiller-plants, demand response technology, heating-systems, air-purification-systems, and smart-homes. Companies that specialize in home energy efficiency, such as innovative air conditioning technology, and heating and lighting systems dominate the green building cluster.

The third New York cluster involves “green transportation” (alternative-automotive-fuels, electric-vehicles, energy-storage-solutions, wireless-charging, nanotechnology, lithium-ion-batteries, hybrid-vehicles, fuel-cells-technology and zinc-hybrid-cathode-battery). The New York Battery and Energy Storage Technology Consortium was created in 2010 to position the city as the global leader in this field and many recent green-tech start-ups are involved in energy storage technology, with a focus on batteries and ultra-capacitors (and their components), fuel cells and systems and products that incorporate these technologies, mainly in the transportation and grid storage sectors.

Figure 2: Green-tech clusters in New York

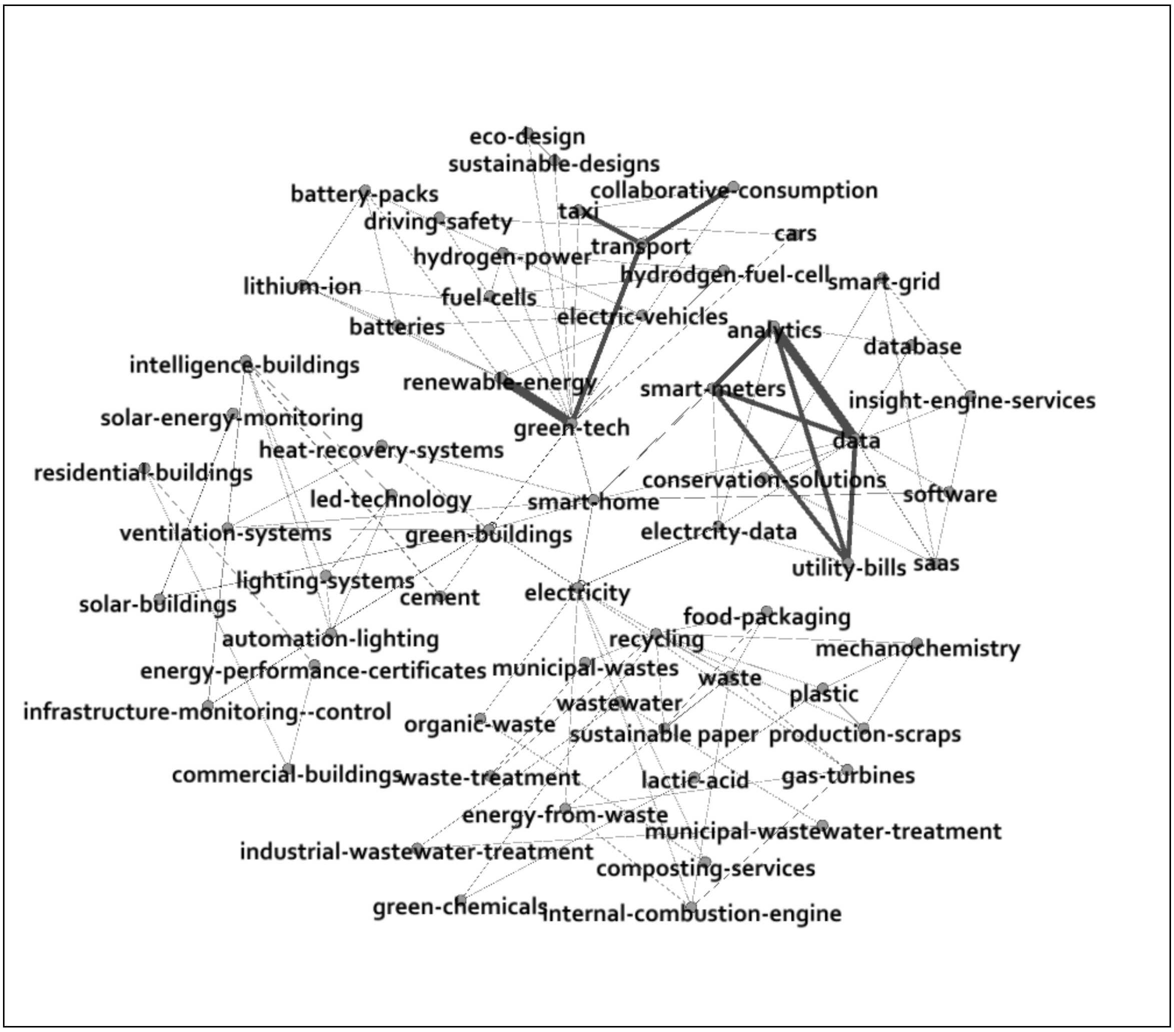


Source: Crunchbase (2013)

In London, there are three major clusters related to “green transport” (electric vehicles, taxis, hydrogen-power, eco-design, collaborative consumption, lithium-ion, cars, energy efficient, fuel-cell, battery-packs and driving-safety), “waste” (recycling, wastewater, production-scraps, municipal-wastes, industrial-wastewater-treatment, mechanic-chemistry, plastic, municipal-wastewater-treatment, food-packaging and sustainable paper) and “green buildings” (smart-home, lighting-systems, automation-lighting, intelligencebuildings, infrastructure-monitoring-control, utility-bills, electricity-data and solar-buildings). The last cluster has strong links to the clean-web sector, which includes information technology tools, big-data and analytics to reduce costs and help to reduce demand for energy and/or improve energy efficiency (Figure 3).

The number of green-tech companies in London has increased, especially in the East London area. Companies are involved in several activities such as optimization and control in efficient buildings, smart energy management and sustainable transport. The most interesting developments are clean-web and waste treatment. There are several start-ups based on energy audits, data management consultancy services and analytics to assess energy performance for both business and residential buildings, development of online apps, tools for energy saving, greener software, and so on. In the field of waste treatment, solutions with huge green potential include development/deployment of technologies to transform wastewater into a useful product such as coolant for electric power plants, and biogas generated in anaerobic digesters and used for power generation or heating. Companies in these areas are developing new generations of plants that combine wastewater treatment with biomass production, and produce various products (e.g. fuel, clean water, fertilizers).

Figure 3: Green-tech clusters in London



Source: Crunchbase (2013)

4. Conclusions and policy implications

The proposed network analysis, once replicated in the relevant geographic area, could act as an informative basis for policy makers for the formulation and implementation of local policies directed to supporting green-tech production and R&D activities. Our results show, for example, that San

Francisco has the largest and most interconnected network, compared to New York and London, in which start-ups focus mainly on solar-energy, carbon-offset, electric-vehicles, renewable-energy-certificates, renewable-energy-credits, smart-grid and green-buildings, and the most relevant clusters are those connected to solar energy, carbon trading market and smart mobility. Similar insights can be derived from the observation of networks in New York (major hotspots are micro-grid, recycling, hybrid-energy, wind-energy, green living, solar-energy, wind-turbines and transportation, and emerging clusters are related to solar, wind and geothermal energy, smart building and green transportation) and London (innovative areas include data analytics, smart-home, green-buildings, electricity, waste, recycling, transport, and the most relevant clusters are green transportation, waste industry and smart buildings). The proposed method could become a useful exploratory tool and allow addressing some key policy issues such as: What specializations exist in the area? Which green technologies should be prioritized among the many technological opportunities? Which market and/or technological complementarities are companies exploiting? Which green-tech clusters play a relevant role in the high-tech metropolitan eco-system? In our view, the analysis has been useful to answer these questions.

A synthesis at the metropolitan level of markets, products and technologies (with underlying competences and capacities in which companies invest their resources) and an appraisal of the patterns linking emerging specializations, help in the design and implementation of targeted and information-based local policies. Such policies, reflecting local specializations and business needs, might act as effective drivers of long-term innovation and success in the green-tech industry. Rather than providing broad and undifferentiated support to local start-ups, policy makers can signal advances in R&D, new products and prevailing technological trends. Secondly, they could identify and direct active industry players and financial institutions towards emerging businesses and specializations and/or products and technologies that appear complementary to create local demand for given services and to attract innovative firms, talents and competences. The promotion of a more intensive and rapid use of new green products, services and technologies over which relevant firms' clusters and aggregates already exist in the urban landscape can contribute at creating leading innovation hubs. Last, policy makers should support the creation/aggregation of core competences and facilitate closer relationships between innovative companies, suppliers and clients, venture capitalists, large corporations and research laboratories involved in the green-tech industry.

The network analysis might even be extended across industries to capture inter-industry linkages: investigating the co-existence among the tags related to all the high-tech companies in a metropolitan area would provide evidence on the existence and extent of technological complementarities among industries. Over time, green-tech clusters are being transformed into open cross-industry platforms that combine competences, specializations and capabilities from several industries, such as software, transportation, analytics, construction, biotechnology. In all three metropolitan areas emerges that the green-tech industry is linked to most other high-tech sectors with no dissimilarity in the pattern, but only in the degree. According to the average degree of all nodes in the cross-industry graphs, there are more cross-industry interconnections in San Francisco (average degree equals 31.2), compared to New York (29.7) and London (27.8), while average weighted degree is higher in New York (3.891) than in San Francisco (3.438) or London (2.628). Policy should also facilitate cluster companies' participation in green-tech cluster programmes and companies’ access into different as well as related industries and clusters.

References

[1] Cooke P. Cleantech and an Analysis of the Platform Nature of Life Sciences: Further Reflections upon Platform Policies. Eur Plan Stud 2008;16:375-393.

[2] Pernick R, Wilder C. The Clean Tech Revolution: The Next Big Growth and Investment Opportunity. New York: HarperCollins; 2007.

[3] Advanced Energy Economy. Advanced energy now. 2015 Market report. Report, available: http://info.aee.net/hs-fs/hub/211732/file-2583825259-pdf/PDF/aen-2015-market-report.pdf; 2015 [accessed 17/06/2015].

[4] Mundaca L, Richter JL. Assessing 'green energy economy'stimulus packages: Evidence from the US programs targeting renewable energy. Renew Sustain Energy Rev 2015;42:1174-1186.

[5] Liu H, Liang D. A review of clean energy innovation and technology transfer in China. Renew Sustain Energy Rev 2013;18:486-498.

[6] UK Department for Business, Innovation & Skills (UK BIS). Low Carbon and environmental Goods and Services: an Industry Analysis. Report, available: http://webarchive.nationalarchives.gov.uk/20121212135622/http://bis.gov.uk/assets/biscore/business-sectors/docs/10-795-low-carbon-environmental-goods-analysis-update-08-09.pdf; 2010 [accessed 17/06/2015].

[7] Ernst & Young. Cleantech matters - Global competitiveness. Report, available: http://www.ey.com/Publication/vwLUAssets/Global_cleantech_insights_and_trends_report_2012/$FILE/cleantech_matters.pdf; 2013 [accessed 17/06/2015].

[8] Negro SO, Alkemade F, Hekkert MP. Why does renewable energy diffuse so slowly? A review of innovation system problems. Renew Sustain Energy Rev 2012;16(6):3836-3846.

[9] OECD. Clean-Tech Clustering as an Engine for Local Development - The Negev Region, Israel. Report, available: http://www.oecd-ilibrary.org/docserver/download/5k98p4wm6kmv.pdf?expires=1434641044

&id=id&accname=guest&checksum=5D433768FE28CA120EB238649F3AF8D1; 2012 [accessed 17/06/2015].

[10] D'Amico G. Clean energy: a second cluster for Worcester. 2007 Worcester, MA.

[11] Kennedy E. Green tech cluster chats up Murray, McGovern. Worcester Business Journal, Worcester, available: http://www.wbjournal.com/article/20090219/NEWS01/302199996/green-tech-cluster-chats-up-murray-mcgovern; 2009 [accessed 17/06/2015].

[12] McLennan Magasanik Associates. Renewable Energy – A Contribution to Australia's Environmental and Economic Sustainability. Report; 2006.

[13] Porter ME, Kwek JH, O'Neil C, Satchcroft A, Vogt T. The Australian renewable energy cluster. Microeconomics of competitiveness. Harvard Business School, available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.566.5467&rep=rep1&type=pdf; 2008 [accessed 17/06/2015].

[14] Normile D. China's living laboratory in urbanization. Sci 2008;319:740–743.

[15] KanEnergi. Green energy clusters. Stimulation of renewable energy markets through the establishment of regional clusters. Skara, Sweden, available: https://ec.europa.eu/energy/intelligent/projects/en/projects/green-energy-cluster; 2007 [accessed 17/06/2015].

[16] Gawer A. Platforms, Markets and Innovation. Cheltenham: Edward Elgar Pub; 2009.

[17] Block JH, Sandner PG. What is the Effect of the Financial Crisis on Venture Capital Financing? Empirical Evidence from US Internet Start-Ups. Venture Cap An International J Entrepreneurial Finance 2009;11: 295-309.

[18] Block JH, Sandner P. Venture capital funding in the middle of the year 2011: are we back to pre-crisis boom levels?. Strategic Change 2011;20:161-169.

[19] Werth JC, Boert P. Co-investment networks of business angels and the performance of their start-up investments. International J Entrepreneurial Venturing 2013;5:240-256.

[20] Waldner F, Zsifkovits M, Heidenberger K. Emerging Service-Based Business Models in the Music Industry: An Exploratory Survey. In: Snene M, Ralyté J, Morin JH, editors. Exploring Services Science, Berlin: Springer; 2012, 321-329.

[21] Yuxian EL, Yuan STD. Where's the Money? The Social Behavior of Investors in Facebook's Small World. In: Ieee/acm International Conference on Advances in Social Networks Analysis and Mining. Istanbul, Turkey; 2012.

[22] Homburg C, Hahn A, Bornemann T, Sandner P. The Role of Chief Marketing Officers for Venture Capital Funding: Endowing New Ventures with Marketing Legitimacy. J Mark Res 2014;51:625-644.

[23] Adcock AB, Lakkam M, Meyer J. CS 224W Final Report Group 37. Stanford University; 2013.

[24] Marra A, Luciani M, Antonelli P. Data analytics nel comparto high-tech statunitense: impatti, evoluzione e tutela della privacy. L'industria 2014;35:319-338.

[25] CBInsights. Venture Capital Database; 2014.

[26] Bloch F. Group and network formation in industrial organization. In: Demange G, Wooders M, editors. Group Formation in Economics; Networks, Clubs and Coalitions, Cambridge, UK: Cambridge University Press; 2004, p. 335-353.

[27] Jackson M. The Economics of Social Networks. In: Blundell R, Newey W, Persson T, editors. The Econometric Society, Advances in Economics and Econometrics, Theory and Applications, Cambridge, UK: Ninth World Congress of the Econometric Society, Cambridge University Press; 2006.

[28] Jackson M. An Overview of Social Networks and Economic Applications. In: Benhabib J, Bisin A, Jackson MO, editors. Social Economics, Amsterdam: Elsevier; 2011, p. 511-547.

[29] Park YT, Kim MS. A Taxonomy of industries based on knowledge flow structure. Technol Analysis Strategic Manag 1999;11(4):541-549.

[30] Burt RS. Structural holes: The Social Structure of Competition. Cambridge: Harvard University Press; 1992.

[31] Leoncini R, Maggioni MA, Montresor S. Intersectoral innovation flows and national technological systems: network analysis for comparing Italy and Germany. Res Policy 1996;25:415-430.

[32] Leoncini R, Montresor S. Network analysis of Eight Technological Systems. International Rev Appl Economics, 2000;14(2):213-234.

[33] Scott J. Social Network Analysis. A Handbook. Thousand Oaks, California: Sage Pubns Ltd; 2000.

[34] Hägestrand T. Innovation Diffusion as a Spatial Process. Chicago: University of Chicago Press; 1967.

[35] Schrader S. Informal technology transfer between firms: Cooperation through information trading. Res Policy 1991;20:153-170.

[36] Piore M, Sabel C. The Second Industrial Divide. New York: Basic Books; 1984.

[37] Lazerson M. A new phoenix?: Modern putting-out in the Modena knitwear industry. Administrative Sci Q 1995;40:34-59.

[38] Lazerson M. Factory or putting-out? Knitting networks in Modena. In: Grabhner G, editor. The Embedded Firm: On the socioeconomics of industrial networks, London: Routledge; 1993, p. 203-226.

[39] Gottardi G. Technology strategies, innovation without R&D: And the Creation of Knowledge Within Industrial Districts. J Industry Stud 1996;3:119-134.

[40] Saxenian A. Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Cambridge: Harvard University Press; 1994.

[41] Garnsey E, Cannon Brookes A. The Cambridge phenomenon revisited: aggregate change among Cambridge high tech firms since 1985. Entrepreneurship Reg Development 1993;5:179–207.

[42] Maskell P. Nyetableringer i industrien og industristrukturens udvikling (New firm formation and industrial restructuring in Denmark). Copenhagen: Copenhagen Business School Press; 1992.
[43] Kristensen PH. Industrial districts in West Jutland, Denmark. Industrial Districts and Local Economic Regeneration. Geneva: International Institute for Labour Studies, ILO; 1992.
[44] Saglio J. Localized Industrial Systems in France: a particular type of industrial system. In: Storper M, Scott AJ, editors. Pathways to industrialization and Regional Development, London: Routledge; 1992, p. 230-253.
[45] Ganne B. Place et évolution des systèmes industriels locaux en France: économie politique d'une transformation. Paris: Les régions qui gagnent, PUF; 1992.
[46] Cantner U, Graf H. The network of innovators in Jena: an application of social network analysis. Res Policy 2006;35:463–480.
[47] Owen-Smith J, Riccaboni M, Pammolli F, Powell WW. A comparison of U.S. and European university-industry relations in the life sciences. Manag Sci 2002;48:24–43.
[48] Breschi S, Lissoni F. Mobility and Social Networks: Localised Knowledge Spillovers. Milan: University of Bocconi, CESPRI Working Paper; 2003, 142.
[49] Singh J. Social Networks as Drivers of Knowledge Diffusion. Boston, MA: Harvard University, Mimeo; 2003.
[50] Balconi M, Breschi S, Lissoni F. Networks of inventors and the role of academia: an exploration of Italian patent data. Res Policy 2004;33:127–145.
[51] Stuart TE, Podolny JM. Local Search and the Evolution of Technological Capabilities. Strategic Manag J 1996;17:21-38.
[52] Valentin F, Jensen RL. Reaping the fruits of science: comparing exploitations of a scientific breakthrough in European innovation systems. Econ Syst Res 2002;14(4):363-388.

[53] Paci R, Batteta E. Innovation networks and knowledge flows across the european regions. CRENS Working Paper 2003; 13.

[54] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media; 2009.

[55] Brandes U. A Faster Algorithm for Betweenness Centrality. J Mathematical Sociology 2001;25:163-177.

[56] Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Statistical Mechanics: Theory and Exp 2008;10:10008.

[57] Latapy M. Main-memory Triangle Computations for Very Large (Sparse (Power-Law)) Graphs. Theoretical Computer Sci 2008;407:458-473.

[58] World Bank. Renewables 2014 Global Status Report. Washington: World Bank. Report, available: http://www.ren21.net/Portals/0/documents/Resources/GSR/2014/GSR2014_full%20report_low%20res.pdf; 2014 [accessed 17/06/2015].

[59] Bloomberg New Energy Finance (BNEF). Global Trends in Renewable Energy Investment. Report, available: http://fs-unep-centre.org/sites/default/files/attachments/datapack.pdf; 2014 [accessed 17/06/2015].

[60] McKinsey. The disruptive potential of solar power. McKinsey Quarterly, available: http://www.mckinsey.com/insights/energy_resources_materials/the_disruptive_potential_of_solar_power; 2014 [accessed 17/06/2015].

[61] Walz R, Marscheider-Weidemann F. Technology-specific absorptive capacities for green technologies in Newly Industrialising Countries. International J Technol Globalisation 2011;5(3-4):212-229.

[62] New York State Energy Research and Development Authority,. Toward a clean energy future: A strategic Outlook. Report, available: https://goo.gl/HNXB13; 2015 [accessed 17/06/2015].