



Paper to be presented at the DRUID Academy conference in Rebild, Aalborg, Denmark on January
15-17, 2014

Exploring Technology Evolution in Danish Smart Grid Development - An NLP approach

Roman Jurowetzki
Aalborg University
Department for Business and Management
roman@business.aau.dk

Abstract

The present work utilises state-of-the-art natural language processing (NLP) tools (Applications from the NLTK python library (Bird,2009) and the vector space modeling and topic modeling toolkit gensim (Hofmann 2010). for enriching the analysis of evolutionary processes in large technological systems (LTS). Introducing these methods opens up for the exploration of large corpora of text-data as it is available from technical reports, industrial publications, online technology blogs and even public press (Henschel 2010). These sources allow to identify the different technology- and related application trajectories within the evolving LTS.

Large technological systems as defined by Hughes (1987) combine a variety of technological artefacts, organisations and other institutional components such as legislative regulations. As the system evolves new components and applications can emerge, complementing or substituting the old ones. Identifying these components is challenging, given the complexity of large systems and often the absence of clear delineation.

The current issues of clean and efficient energy and green transportation - often also referred to as the great challenges - require large scale changes at a systemic level. Studies of such processes often experience the problem of the scope at which innovation takes place in systemic setups. One particular challenge is the delineation of the systems in which change takes place and the identification of components in a context in which new technologic paths and applications emerge and discontinue as the system evolves. Research is therefore often forced to focus on single technologies or applications, where the selection of the topic is guided by data availability.

Quantitative research on the evolution of LTSs and technology forecasting has often referred to patent analysis as a method of identifying technological change. This practice is problematic as patents are granted to inventions not to innovations within or at the periphery of the dominating technological system. In the same time many important innovations do not rely on patents.

A promising approach to capture the evolution of large scope evolution in systems is inspired by the "Historical Event Analysis". The methodology used by Hekkert (2009) is built around the identification of events using industrial publications and the classification of these within the technological innovation system (TIS) functions framework. This

approach shows however two drawbacks. Firstly, it is very labor intensive as it requires manual exploration of all inputs. Secondly, the extraction and selection of used publications assumes that the researcher knows the relevant keywords and topics. In a technical field, which develops new components as it evolves this assumption might be too strong. NLP methods can be used to support such an approach by optimising keyword extraction and validation, the selection of sources and even perform some content analysis.

The present study utilises NLP for an explorative investigation the smart grid technology evolution in Denmark. Smart Grid is a not clearly delineated target-concept related to different technological additions and changes in the electrical grid. The Danish case is particularly interesting since large scale large scale wind power integration perpetuated a lot of research projects in this area, making Denmark the country with the highest amount of smart grid research in Europe. Furthermore, the development of the smart grid enjoys strong national policy support. This study reconstructs the development of the smart grid exploring three areas: (i) Policy, (ii) Research and (iii) Industry & Applications. Preliminary observation suggest that policy and industrial attention tend to increasingly concentrate around some technologies over time. This can indicate a premature picking winners strategy.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc.

Hekkert, M. P., & Negro, S. O. (2009). Technological Forecasting & Social Change. *Technological Forecasting & Social Change*, 76(4), 584-594.

Henschel, A., Casagrande, E., Woon, W. L., Janajreh, I., & Madnick, S. (2011). A Unified Approach for Taxonomy-Based Technology Forecasting. *Business Intelligence Applications and the Web: Models, Systems and Technologies*, 178.

Hughes, T. P. (1987). The evolution of large technological systems. *The social construction of technological systems: New directions in the sociology and history of technology*, 51-82.

Řehánek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora (pp. 45-50). Presented at the Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, Valletta, Malta: ELRA.

Exploring Technology Evolution in Danish Smart Grid Development

An NLP approach*

Roman Jurowetzki

*Ph.D. Student, Aalborg University, Department of Business and Management, IKE /
DRUID, e-mail: roman@business.aau.dk*

1. Introduction

Many of today's environmental but also broader social challenges stem from the miss-performance of *large technical systems* (Hughes, 1987) or - to use a STS concept *socio-technical regimes* (Geels, 2010). Examples include the agri-food sector, many types of manufacturing, transportation and more generally the energy system, which this study will focus on (Geels, 2011). Miss-performance means in this context not that the systems do not furnish the goods and services which they are meant to provide but rather that this happens unsustainably, contributes to global warming and the deterioration of natural resources and living space. The transition towards a sustainable future depends on the ability to transform these sectors, which by their very nature are resilient to change.

One of the most unequivocal descriptives of the inability to change is the notion of the *carbon lock-in* (Unruh, 2000). The lock-in is seen as a consequence of technological and institutional co-evolution, which is substantiated and stabilized by various mechanisms within and outside the systems. In a later article Unruh and Carrillo-Hermosilla (2006) argue that escaping this condition requires major systemic change driven by exogenous forces, such as radically innovative technologies that can develop and improve, their initially mediocre performance levels in *niches* outside the dominant system. Another outlined option is that a sufficiently strong social consensus can be reached through education and discourse or as a result of a *focusing event*, such as a climate catastrophe (Unruh, 2002). Similar argumentation can be also found within the transition literature (Geels, 2002; Markard et al., 2012).

The past couple of years have seen the emergence and growth of industries related to the generation of renewable energy. In many cases these developments changed the paradigm in energy generation from energy-production to energy-harvesting thus imposing new challenges on the energy transmission and consumption side (Foxon et al., 2010). While in the "old" system production would be adjusted to typical consumption

*Paper to be presented at the DRUID Academy Conference 2014, Skørping, Aalborg (Denmark)

levels, now production levels are literally depending on the weather (Mattern et al., 2010). In addition electric mobility, solar cells on rooftops and other new applications on the consumer side are altering consumption patterns (Elzinga, 2011).

The upcoming technological paradigm - in a Dosian (1982) understanding - that is evolving as a *normal solution* to cope with the new dynamics of energy generation, can be found in the development of the *smart grid*.

Smartening the electricity grid means here the installation of a number of technologies along the generation-transition-distribution-consumption chain, that will enable the grid to control and balance itself automatically given the new patterns of energy generation and consumption (Farhangi, 2010).

The analysis of this systemic development is important in order to understand the state of the overall sustainable transition, and it comes up with a number of challenges. The first and possibly most intriguing is the delineation of the *smart grid* with relating technologies and applications. Particularly, in the case of this infrastructure system, where regulations on national level have a strong influence on technological outcomes, the identification of technological trajectories within the borders of a particular country can provide valuable insights for understanding systemic change (Sawin, 2006).

This exploratory paper aims at mapping out the *smart grid field* looking at both, the research- and the industrial landscape in Denmark, a country which pioneered in the large scale integration of wind-power (Garud and Karn e, 2003) and as in 2011 was the country with the highest share of *smart grid* R&DD projects in Europe (Giordano et al., 2011).

Methodologically this research relies on the combination of methods from statistical natural language processing, vector space modelling and network analysis. The filtered data consists of two types of Danish text-documents: 99 R&DD project descriptions and 574 (initially 813) non-academic/industrial journal publications.

2. Industrial and Theoretical Background

2.1. Smart Grid technology in Denmark

The traditional architecture of the electricity grid assumes a unidirectional energy flow from centralized energy plants via the transmission and distribution grids to consumers, where energy production levels are constantly adjusted to match the over time fluctuating energy demand (Farhangi, 2010). Embracing the renewable energy paradigm, centralized energy production is gradually replaced by decentralized energy farming. The harmonization between production and consumption has to move from the traditional generation side into the transmission and consumption areas. ICT technologies will play a central role, supporting this process (Mattern et al., 2010).

In the Northeuropean setup two options are possible and currently discussed. Firstly, the construction of a European transmission super-grid, to allow for instance energy exports from Denmark to Germany in wind-peak times. Secondly, the development of a national *smart grid*, that is able transmit energy and information in both ways, thus allowing for harmonization by the means of flexible consumption. This requires the upgrade of the existing grid by adding a *layer of intelligence* - advanced measurement, communication and control technology - thus making the grid able to handle a higher share of decentralized renewable energy generation and the recently evolving

consumption patterns (Elzinga, 2011). If flexible consumption can be activated by the introduction of smart functionality, costly investments in the reinforcement of the distribution system can be moved into the future or avoided (Forskningsnetværket, Smart Grid, 2013).

Denmark is already today counting the largest amount of R&DD projects within the smart energy area in Europe (Giordano et al., 2011). The extremely high ambition of the national energy agreement, passed by the government in 2012 targets a wind-power share of 50 percent by 2020 and the more recently announced Smart Grid Strategy sees the country as a European laboratory for innovative energy solutions (KEMIN, 2013).

An important role is given to the Smart Grid Network, which informs the national grid development policies. The Network was set up in 2010, bringing together key research institutions, industrial organizations and the national transmission system operator (TSO) *Energinet.dk*. In a report from published in October 2011 the network estimated the economic potentials and drafted a roadmap for the further development of the grid infrastructure (Forskningsnetværket, Smart Grid, 2013).

An industrial Intelligent Energy association was founded in 2012 bringing together around 130 actors from business, research and policy. Overall, the condition for the development of a variety of technologies are given. Yet, there are also very obvious barriers: In a brief interview carried out in December 2012, the chairman of the organization, Morten Baadsgaard Trolle, provided some illustration for a possible lock-in in the current grid infrastructure: Currently there is a mismatch between the aspiration to promote entrepreneurship and the development of a broad variety of solutions, especially of grid hardware components such as regulation and control equipment. However, interoperability issues, complex security regulations and the lack of common standards makes it difficult for new players to enter this area. This trajectories are supported by established routines between utilities and incumbent technology producers. In addition, the formal set-up and resulting composition of research projects often favour the reproduction of existing networks of actors.

The majority of these arguments apply, however to most countries and are not specifically Danish problems. On the positive side can be noted that Danish actors have a rich experience with large-scale wind power integration and one of the world's most developed district heating infrastructures, providing plenty of opportunities for consumption flexibilization (Lund et al., 2012).

2.2. The challenge to delineate a large technical system

An energy grid is a complex system with extremely interwoven technical, economic, institutional and administrative structures and therewith a great example for a large technical system as defined by Hughes (1987). The system includes physical artefacts such as hardware components for the transmission and distribution of electricity. In addition it contains organizations, such as manufacturing firms or utility companies. All these components interact with each other, following formal, normative and cognitive rules. Since energy grids are physically connected to energy producers on the one side and users on the other, the aforementioned components also interact with artefacts and agents external to the system. Given this setup paired with the non-linear nature of innovation processes (Nelson and Winter, 1982), the transformation of such a system

is expectedly complex and difficult to predict. Work by sociologists of technology (e.g. Bijker, 1997) give highly important insights into agency and the social construction of the technology, doing so by increasing the scope and thereby the complexity. The same is true for much of the STS literature, which focuses on socio-technical transitions (e.g. Geels, 2002). While the theoretical frameworks proposed by this literature are easy to grasp, multidimensionality, the high number of actors and feedback loops make them hard to operationalise. An approach, often taken by scholars that refer to this theoretical foundation, is focusing on technology niches. The STS-Transition Literature depicts niches as protected spaces on the micro level for the development of innovation within rigid socio-technical systems. This strategy allows for a close exploration of the development in a narrow technological area and its interaction with the meso and macro levels. Yet, this approach requires *a priori* identification and to some extent dilatation of the relevant niches. This is also the case for innovation system approaches. An analysis based on the functionalistic Technological Innovation System (TIS) framework, which is often seen in energy industry studies, (Bergek et al., 2008; Hekkert et al., 2007) for instance builds on a clear delineation of the technology in question. In the present case, deciding on a particular technology focus is challenging for at least two reasons:

The large number of interacting technologies and components on the one hand and the system history aspect on the other. The latter reason relates to path dependency and institutional factors in a national setting. History matters, and therefore a *smart grid* system in country A will not build on the same technologies and components as in country B.

Yet, an assumption that can be made here is that the level of technological diversity will increase along the path from the transmission to consumption area.

2.3. Technological trajectories

A less complex approach could build on the seminal work of Dosi (1982) on technological paradigms and trajectories. The sustainable energy system can be interpreted as a (targeted) technological paradigm with the *smart grid* as an option for a technological (meso)trajectory and different components building the micro trajectory level.

Economic studies of technological trajectories mainly focus on patent and patent citation analysis, assuming that highly similar patents belong to and form a trajectory. Citations provide data to construct patent networks and model the evolution of knowledge represented by the patents (Verspagen, 2007). A recent study by Chen et al. (2012a) illustrates for instance the identification of *smart grid* trajectories combining patent citation networks and some NLP for topic detection.

While patents are a well structured and reliable source of information about theoretically available technology and its development, they do not reflect the application of the technology. Chen et al. (2012b) identify a number of credible trajectories in a comparative study of the US and the global context, employing a similar methodology as the one used in the present study. Yet, the conclusion that US patents in a certain field can be interpreted as the presence of a particular technological trajectory in the US, would rely on strong assumptions about the relation between patents and technology outcomes.

The present explorative study is an attempt to map out the present technological

trajectories within *smart grid* in Denmark, account for some of the cognitive and institutional factors, which shape the development of a technology in a national context (Kaplan and Tripsas, 2008). Therefore this paper uses two types of sources: *Smart grid* research project description and publications from national industry journals, where the former source is expected to represent the institutionally shaped technology development while the latter should provide rich information on the application, policy and business side.

3. Data overview & preparation

The analysis is based on two types of text-documents: Approximately 100 descriptions of smart grid R&DD projects and initially 813 non-academic journal articles in from Danish industrial publications¹.

Initially 102 project descriptions are obtained through www.energiforskning.dk, the joint database of publicly (co-) funded energy research in Denmark. The database contains detailed information on nearly 2100 projects, classified into 7 broad technology areas. This classification is interesting as such, since it indirectly grounds on the (partly political) decision to fund a particular activity or not. Thus, it can be assumed that this *pre-classified* data implicitly carries information about the technology perception by the public sector. The descriptions are usually 500 words long and briefly outline the background and purpose of the project, technologies used and expected outcomes. Project start dates range from 1996 until 2013, however the distribution is highly skewed towards the latest 5 years (see Figure A.1). In some few cases result descriptions are included for terminated projects, providing additional or more specific information on the technological outcomes of the particular activity. Since this research is not interested in the evolution of particular projects, but rather aims at identifying general technology trends, where available, result descriptions are appended to the initial project descriptions. The projects span from basic research to deployment activities, which is explicitly indicated in meta-data available for each activity but also can be inferred from the finance mechanism applied. That opens up for the analysis of potential technology development cycles for particular technological trajectories. Yet, this question was not central for the presented study. Project duration, budget, number and type of participants were not analysed explicitly. However, since named entities in the texts were not systematically identified and excluded, actor name appearance did certainly have an influence on document similarity calculation (see 4.2). Finally selection filters were applied to exclude too short descriptions, that indicated that a proper description is yet to be posted. A language detection module sorted out description in other languages than Danish, leaving 99 documents for the analysis.

In order to map out the scope of *smart grid* applications in a national context, non-academic industrial publications have been explored. These were retrieved from the Danish national publications database *infomedia*. A search-string was systematically build up by exploring term frequencies, n-grams and collocations within the press

¹The text-data used for this research was exclusively in Danish. Even though English descriptions are available for most of the projects, it is assumed that Danish descriptions are more accurate and rich. Text examples presented in this paper are translations by the author. Trajectory specific TFIDF-Keywords were however automatically translated relying on the Google Translate API.

releases by the Danish Smart Grid Alliance, which since its initiation in April 2012 informs about the national *smart grid* industry². An initial search returned 813 articles for the timespan 2004-2013³.

The coverage of *smart grid* related themes remains marginal until 2009. The majority of articles before 2009 are published by the engineering journal *Ingenøren*. Only starting 2009, more practically oriented periodicals take up the topic, indicating the upcoming interest for the *smart grid* outside the engineering community. Overall, articles come from 61 different periodicals largely affiliated with engineering and construction themes (see Table A.3 for more details). However, 81 percent of the publications relate to 12 journals focusing on the national energy system, appliance installation, computer – and information science, the business part of the engineering and energy industry.

Just like for the descriptions, language detection was applied to exclude non-Danish articles. Also here, too short texts (shortest percentile) were sorted out. A quick search within the retrieved documents confirmed that industrial publications often report on ongoing research projects or *smart grid* research in general. These reports are however not conducive for a detached exploration of technological trajectories in the domains of research and application. Documents that mentioned the term *project* or *research* in their titles or introductory abstracts were dropped. While most of the articles are clearly associated with national developments, it is possible that an article only covers to technology or market developments abroad. An additional filter selected out documents that wouldn't mention "Denmark" or "Danish" in any part of the document. As expected the number of in this last step excluded texts remained very low.

4. Method

For both types of documents a three step analysis was applied. Project descriptions and the text-bodies of articles underwent (1) term extraction and filtering using basic Statistical NLP techniques. Arrays of nouns and nominal expressions were analysed for semantic similarity with the help of (2) vector space modelling. Thereby obtained document similarity estimates were used to construct a document network. (3) Network analysis calculations were used to cluster the documents thematically. Finally NLP was used one more time to (4) retrieve representative keywords for particular clusters and *sub-clusters*⁴.

The following describes the above lined up steps in detail:

4.1. NLP based term extraction

The goal of the term extraction is to reduce the text-documents into bag of words (BOW) representations - an array of terms of high information content. These are usually nouns and noun phrases. Firstly, important noun phrases are selected by

²<http://www.ienergi.dk/English.aspx>

³(smartgrid OR (smart OR intelligent* OR klog*) DNEAR5 (grid OR energisystem* OR elnet*))The english translation corresponds to: (smartgrid OR (smart OR intelligent* OR clever*) DNEAR5 (grid OR energisystem* OR electricitygrid*)), the DNEAR5 command specifies that the distance between the array of adjective classifiers in the first parenthesis and the *grid*-synonyms in the second parenthesis is ≤ 5 .

⁴NLP using the NLTK package (Bird et al., 2009), vector space modelling with the GENSIM package (Řehůřek and Sojka, 2011) within IPython, *community detection* and visualization within GEPHI

identifying high-document frequency (DF) bi-grams⁵. For project descriptions, which tend to use many standardized formulations, a domain specific stopword filter was applied.

Part of speech (POS) tagging is performed using a Brill tagger, trained on the Danish Morphosyntactically Tagged PAROLE Corpus (Keson and Norling-Christensen, 1998) combined with two affix taggers. The POS identification accuracy ranges at 97 percent⁶.

Not-nouns or noun expressions, domain specific- and general Danish stopwords, and low-DF words are dropped. For the presented analysis the low-DF threshold was set at 1, thus only excluding *singletons*- spelling-errors and very rare terms that would not contribute to the classification of documents. Finally, terms are reprocessed by stemming, which once again reduces the vocabulary by approximately 23 percent⁷.

4.2. Vector space modelling & Latent Semantic Analysis

The BOW extract of the documents is transformed into sparse vectors where each term is represented as (w, c_w^d) with w being the word-id in the initially created dictionary and c_w^d the integer word count of w for the particular document. The resulting representation is then *tf-idf* (term frequency – inverse document frequency) weighted, in order to discount generic terms across documents and equivalently promote document-specific terms.

$$TFIDF[t_i^d] = TF[t_i^d] \cdot IDF[t_i] \quad (1)$$

Where TF is the term t_i frequency in a document d divided by the document length and IDF the logarithm of the number of all documents divided by the number of documents, containing the term.

The returned vectors in normalized unit length, of same dimensionality are then once again transformed into a vector space of lower dimensionality using the latent semantic indexing (LSI or LSA) algorithm (Deerwester et al., 1990; Dumais et al., 1988). See Figure 1 for a schematic representation of the process.

Following Bradford (2008) a target dimensionality of 400 is chosen, where each dimension can be interpreted as a topic inferred from the whole input BOW corpus. Each document is now represented as a 400-dimensional vector.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (2)$$

Finally, cosines between the document vectors are calculated as a similarity measure.

⁵*n-grams* are consecutive term compounds of length n . Only compounds of nouns (e.g. “grid stabilization”) and adjective noun phrases (e.g. “flexible consumption”) were detected.

⁶The evaluation is performed by training the tagger on 90 percent (39190 sentences) of the PAROLE corpus and testing it on the remaining 10 percent (1022 sentences), which represent a previously unseen text (Bird et al., 2009).

⁷Stemming algorithms determine the root of any term and return it instead of the original term: *Innovation, innovative, innovations, innovating* → *innov*.

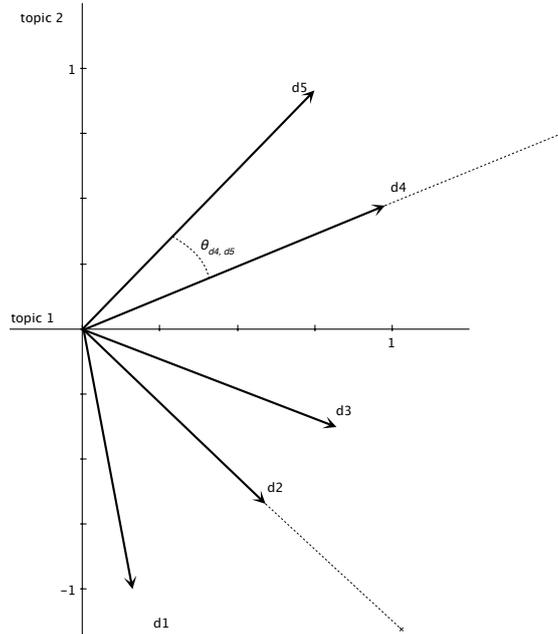


Figure 1: Schematic representation: Dimensionality reduction and similarity calculation within the LSA framework

4.3. Network analysis and community keyword extraction

The document similarity matrix is used as an input for a weighted undirected network where documents form nodes while edges are defined by the $\cos(\theta)$ values between the documents. The distribution of similarity values is highly skewed towards the lower bound. Therefore a cut-off is set at $\cos(\theta) \approx 0.1$.

For project descriptions a threshold value for edge creation is set at the 80 percentile that corresponds to a maximum cosine of 0.14, meaning that cosines below that value (80 lower percent) will not form edges. For industrial publication the percentile value was even higher (around 90 percentile). Modularity class calculation within *Gephi* is used to detect communities (Blondel et al., 2008). The basic intuition of the method is that ties within a community are more common than ties across communities. Yet, results suggest to re-run the algorithm on very large communities separately to split them up in *sub-communities*.

One observation is that industrial publications contain review-type documents that provide an overview on different technologies on the market and alike. These documents seem to be contra-productive for the clustering procedure, which implicitly assumes that each document is only associated with one particular technology or technology-area.

Therefore, for industrial publications nodes with very high degree are manually allocated in a class container [99] before modularity classification is performed.

Keywords describing the communities are extracted by calculating TF-IDF values for community-aggregated texts using the same kind of TF-IDF transformation as before on document-level. Scoring by TF-IDF values returns the most important keywords for a community (Salton and McGill, 1983). The weight can also be used for tag-cloud

style visualization.

5. Results

This section exhibits the results of the above described analysis. It starts with an overview over the main detected techno-thematic communities, looking at results from the research project analysis and the article database separately. A subset of interesting fields are thereafter discussed more in detail.

5.1. Main technological trajectories in energy research

The filtering procedure, described above, generates 99 project description extracts.⁸ Applying LSA and the modularity classification algorithm with a resolution of 0.6, considering edge weights and enabling randomization, returns a modularity value of 0.338, suggesting that the clustering is not optimal and overlapping communities might exist. Since, clustering is based on the document similarity measure and single projects can bridge over or are explicitly designed to connect different technological domains, overlapping is actually an expected feature. 12 communities are detected, with an average size of 8 nodes, the largest communities having 16 and the single smallest 1 node (see Table A.1 for more descriptives). Figure 2 shows the generated graph and indicates the thematic features of the communities in tag-cloud visualizations, where term size is determined by the respective TFIDF values.

Already the graph visualization indicates that the extent to which techno-thematic communities are clustered, varies across the the whole sample. Table A.1 in the Appendix provides a summary over the identified clusters and their features, Figure A.1 gives an overview over community sizes and start years for the constituent projects. Among the mostly identifiable technology centred communities we find projects related to the heat pump technology [4]⁹ and the large district heating cluster [11]. The later is however much less concentrated and seems to consist of two sub-communities, where one is more focused on the development of the technology itself and its systemic integration, while the other aims at connecting the technology to other technologies and applications. This community not only stands out in size but also is the oldest, meaning that its projects have the relatively earliest start-year on average.

Two further dense, yet rather application-focused clusters can be found in [8], which gathers projects that explore the flexibilisation of energy consumption and [5], which aggregates activities related to systemic integration. The extent, to which these thematic communities are diverse in their technology-composition will be discussed below. Smaller unambiguously identifiable technology-areas are electric vehicles [2] and [6], Decentralized communication technology and data security [0], and battery development [1]. An example for less identifiable communities can be found in [9]. Projects in this group are sparsely connected with each other. Also (significant) ties outside the cluster are few and weak.

⁸An automatically translated BOW extract example [u'concept', u'air air', u'heat pump', u'focus', u'climate', u'electricity consumption', u'demonstra', u'play solution', u'heat pump'...]

⁹Modularity class number interpreted as community or cluster e.g. [4]

5.2. Central technologies, themes and applications within the industrial discourse

The initial sample of 813 articles is stepwise reduced to a number of 574. Even though the filters applied might seem very conservative, the analysis shows that 32 articles remain in the sample that seem to be very project-related and have been automatically grouped into one class conditioned on the presence of “project-terminology”. 5 texts with the highest degree-scores in the sample have been manually allocated to the class container [99] before clustering. The idea here was to avoid “review” articles, that inherently relate to all possible energy technologies and applications. A detailed analysis of the texts showed, that indeed these 5 were very general in their contexts – providing an overview over the different *smart grid* technologies and applications that the future might bring. Apart from that, the modularity-class calculation with the same specifications as for the project descriptions, identifies 12 communities. A detailed overview over communities and their features can be found in table A2. As shown in Figure 3, the technological clusters (heat pumps [9], electric cars [8] and district heating [4]) that can be found in the project-analysis re-appear. They form very dense clusters, not only seen in the visualization but also in the TIFIDF-value distribution. These communities do however merely account for 22.5 percent of the sample. Another 18 percent are made up by the wind-power [10], energy optimization and installation (primarily by *Schneider Electric*) [0] and residential solar [2]. As in the project-analysis, consumption flexibilization [5] is very central. New are the areas energy business with a particular focus on export and collaboration with Asian countries [6], and the thematic policy-cluster [7]. Figure A2 summarizes the over-time development of the community sizes.

The remainder of the analysis will make further use of the thematic clustering to explore the overlapping technological area related to heat pumps more in detail. Furthermore, the business and export cluster, resulting from the publication analysis will be examined in order to evaluate the presented methods performance when applied in a less technical, and thus supposedly more abstract domain.

6. Trajectory Study: Heat Pumps

Heat pump units operate using electricity to drive compressors that concentrate and transport thermal energy. The thermal energy extracted from air or the ground can then be used for space and water heating, also the reverse process is possible to use the heat pump for cooling. In fact, using the technology for both heating and cooling simultaneously is most efficient, yet not a very common practice (Mathiesen et al., 2011a). Thermal energy can be stored for later use and pumps can also be combined with consumer side renewable energy generation units such as residential PV (Sanner et al., 2003). This options led to growing interest for this rather mature technology in the recent years, since it can potentially become an important component for efficiency increase and because of the storage option for the flexibilization of energy consumption. The latter is particularly important for the build up of a smart energy system that is able to integrate large amounts of fluctuating electricity generation, e.g. wind power (Lund et al., 2012; Mathiesen et al., 2011b).

While the first theoretical description of a heat pump as a devise for heating and

cooling, by the French officer Nicolas Leonhard Sadi Carnot, dates way back to 1824, it was first in 1948 that the technology was applied in the Equitable Building, New York. Commercial heat pump unit distribution commenced in the 1950s but did not take up until the 1970s and the Oil Crisis, when rising energy costs made electric furnaces less competitive (Hepbasli and Kalinci, 2009). Other reasons for the heat pump “boom” were the growing heat and warm-water demand and the transition from single-room to central heating. However, this was only a boom in relative terms and the technology was merely making up an average 1 percent in the European residential market share with significant difference in individual countries (Laue, 2002). Environmental awareness and efficiency considerations can be count as reasons for the European renaissance of heat pumps in the mid 1990s. Today heat pumps are seen as one of the key technologies to decrease CO_2 and other greenhouse gas emissions.

Denmark adopted the heat pump technology right after the oil crisis in the 1970s. Many producers entered the market, offering products that ranged from high quality pumps, some of which are still in use today, to products of very poor quality that casted a bad light on the industry (Poulsen, 2007). Today, the Danish Energy Agency estimates a total of 100.000 small residential pumps and 5.000 large industrial scale units installed. Even though this total number of installations is relatively low, the Danish market is catching up with 20.000 small mostly air-air units and 5.000 geothermal pumps sold every year (Frost-Knudsen, 2013). A new general tax on all heating systems to be passed by January 2014 is expected to make efficient heat pumps even more attractive.

6.1. Heat Pump development within Danish research

The project analysis selected 12 activities into the “heat pump” class. The TFIDF-Keyword-Cloud does not really provide much information about the important themes that constitute the cluster. Given a larger number of projects, a further clustering and topic modelling could be applied to identify *sub-clusters*. In this case the direct analysis of the descriptions seems most appropriate. As shown in Figure A.2 the first project in the field was started in 2009. Until 2011 the number of newly started activities went up peaking at 4 new projects. A brief qualitative analysis of the descriptions reveals two broad fields of activities among the projects that focus on the usage of the heat-pump technology within the new *smart grid* paradigm.

On the one hand there is projects that explore options for the integration of large-scale heat pump and district heating systems, by connecting central heat and power plants (CHP) with large heat pumps, in some cases also solar systems. Such combined systems can become a more centralized way to allow for more intermittent wind power in the overall energy system, while combining the efficiency and storage options of heat pumps with the existing CHP infrastructure. On the other hand there is a significant number of projects that focus on residential small-scale applications. These activities aim at developing and testing standards for *smart grid* ready plug-and-play solutions, remote control of pumps, test protocols and other technology standards. While the units by themselves seem efficient and mature, knowledge about automatizes interaction with the grid has to be developed. Virtual power plant projects for small and large scale areas combine different ethnologies within a complex system with many components.

6.2. Heat Pump technology in national industrial publications

The distribution of industrial publications over time displays some similarities to the projects with overall 45 reports selected into the thematic heat-pump cluster. First reports that mention heat pumps in a *smart grid* context can be found in 2010.

Re-running of the clustering algorithm generates 4 *sub-clusters* that can guide the exploration. Two of those are obviously related to efficient residential housing and for the most part inform on different projects within new and old constructions. Much emphasized is the interaction between different domestic appliances, the heat pumps as heating device of choice and the management by intelligent (partly remote) control systems.

One of the *sub-clusters* takes a more industrial installation-perspective on the technology. An article (nr. 314) for instance outlines the market potential and job creation opportunities once more heat pumps will be installed. It estimates up to 1 million new units until 2020. Other central topics of the *sub-cluster* are remote control standardization and the energy renovation of old buildings.

The largest *sub-cluster* is more (heat pump) technology focused. It outlines more generally the opportunities that the available technologies offer for the development of the energy grid, particularly concentrating on the energy storage options that are expected to facilitate the integration of more wind power and other fluctuating renewable energy sources. Another technological focus is the already in the research-analysis shown combination of heat pumps with solar. Furthermore it groups market analysis and policy reports covering the area.

Overall the brief analysis shows that there is a significant overlap between applications, developed in research projects and the expectations towards the technology that is expressed in industrial publications. The deployment of heat pumps as a grid stabilization technology seems clearly to be one of the dominant technological trajectories within the Danish *smart grid* development.

7. Thematic Field Study: Smart Energy Technology Business and Export

Gathering almost 20 percent of the publications, the Technology Export and Business cluster – as it can be called given the TFIDF-Keywords - is the largest thematic community. It has obviously no technological focus but some of the specific targeted markets can already be identified in the extracted keywords. The clustering algorithm is used once again to generate 6 *sub-clusters*. 5 of which make up at least 97 percent of the initial group. One of the *sub-clusters* takes up 40 percent, the other 4 around 14 percent each.

The first rather loose *sub-cluster* does not refer much to business or trade but to job creation in the Danish installation and to some extent it-industry. Opportunities arise, according to the grouped articles, from the connection of new hardware to the grid and the development of communication solutions.

The other nearby 16 percent *sub-cluster* is more dense and to a higher extent focused on the installation industry. Many of the publications present studies and estimations about the market opportunities related to the grid development and more broader the energy system transformation. In the desired case, the industry is expected to earn 5 billion DKK (900 million USD) yearly up to 2020. Other market estimations also mention the most central national technology competences within the *smart grid* area: Grid-automation, smart-home appliances and measurement technology has the potential to generate 2.500 jobs in Denmark and further 2.000 in the EU.

One smaller 12 percent *sub-cluster* summarizes more generally articles about required investments and changes in order to develop the *smart grid* in Denmark. While the creating IT and hardware solutions is important, coordination is emphasized to be key in this technological system. One important step was the initiation of the Intelligent Energy alliance in 2012 that brings together around 130 players with interest in *smart grid*. A central publication outlines the importance of incentive creation for the utility companies that are the central actors in the current grid infrastructure.

The last small *sub-cluster* focuses on growth. It brings together further market development studies, reports on policies and initiatives that (could) perpetuate growth from the energy system transformation, where *smart grid* development is always central.

The large *sub-cluster* is primarily looking on the energy technology market and export development. *Smart grid* technologies are expected to repeat the success of wind power and district heating technologies that still drive Danish energy technology export. Outside the EU, the US, with the in late 2012 announced *smart grid* strategy, represent an important market. Much more pronounced is however the Asian marketplace and here particularly Korea as both market and strategic partner in the development of technology. The Asian country embraced the green paradigm rather late but is moving fast and expects *smart grid* investments of 7.2 billion through 2030. The Global Green Growth Forum, initiated in 2011 between Denmark, South Korea and Mexico provides another platform to facilitate interaction and trade. In 2012 also Qatar, Kenya and China joined the organization. Collaboration between different actors, especially the utility companies is outlined as a key condition for success on the international markets.

8. Discussion

The depicted brief analysis demonstrates how the presented method can be used to map technology development and help structuring and describing the socio-economic environment, consisting of applications, expectations, markets, policies and organizations. A detailed exploration or the elaboration of a case study or narrative was not intended. Research project descriptions in summarized form as obtained through `energiforskning.dk` seem to be a concise and in the same time sufficiently rich and reliable source of data to map the various technology focus-points in the national research landscape. The brief summaries usually name the involved technologies, applications and the technical context in which these are embedded. Yet, they vary in length and detail. The extracted industry publications contribute with many diverse and interesting insights into the variety of applications and environmental factors. They open up for the understanding of the cognitive level, or as proposed by Kaplan and Tripsas (2008) the collective technology framing. This seems to be useful for exploratory and descriptive studies of a complex systemic field. The applied topic modelling and clustering generates a structure and indicates the most pronounced aspect of the environment. However, in comparison with the project-analysis the trade-off between rich and structured data becomes obvious. While project summaries follow a implicit structure as for instance research paper abstracts would do, articles are inherently messy. Smaller issues as for instance repeated reporting about one single event or policy, can be avoided by limiting the analysis to only one journal, perhaps even only on one particular format within the journal. In the presented case the Danish Energy Periodical (Nyhedsbladet Dansk Energi), which was the source of 44 percent of all articles could have been chosen (see Table A.3). Even though the applied filtering of extracted text-files can be called rather conservative and the clustering did generate coherent thematic communities, the multidimensionality of the technology and particularly of its surrounding environment complicates the interpretation of the grouped publications. Often it is unclear whether a cluster can be interpreted as a dimension such as policy or market, or rather represents a factor such as the heat pump as part of the technology-dimension.

In regard to the explored technological system the findings can be summarized as follows: While not yet implemented here, automatized entity recognition could make the mapping more powerful. The exclusion of actor-names from the topic modelling would decrease the impact of actual actor interaction on the thematic structuring, while actor-context linking could indicate the activity of actors in particular technological or thematic fields. This in turn could enrich the analysis of “real” networks, e.g. research collaboration networks by generating actor-covariates. Another way to optimize the method could be through implementation of automatized hierarchy/taxonomy building as suggested by Henschel et al. (2011).

Overall the exploration shows that *smart grid* technology is more than merely a combination of artefacts and services that are often mentioned under the collective notion of the *layer of intelligence*. These new ICT-heavy technologies are essential for precise controlling of the new grid. Yet, in the Danish context, mature consumption side technologies such as heat pumps and district heating systems will become evenly important since flexible energy consumption and storage is needed to integrate the growing amount of sustainable energy generation. *Smart energy* technologies are

expected to contribute to economic growth and spur energy technology export. In the international context, a particularly interesting country - also for future comparative research - is South Korea. The Asian country does not only move quickly in the development of smart grid technology - yet obviously creating different technological paradigms - but also seems to have a great interest in collaborating with Denmark.

A. Figures and Tables

Figure A.1: Projects modularity classes over time (smallest excluded)

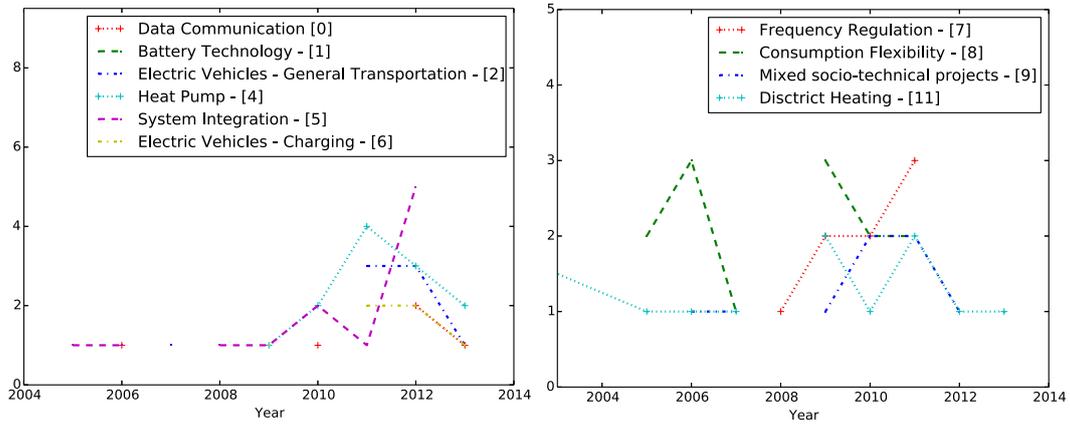


Figure A.2: Publications modularity classes over time (smallest excluded)

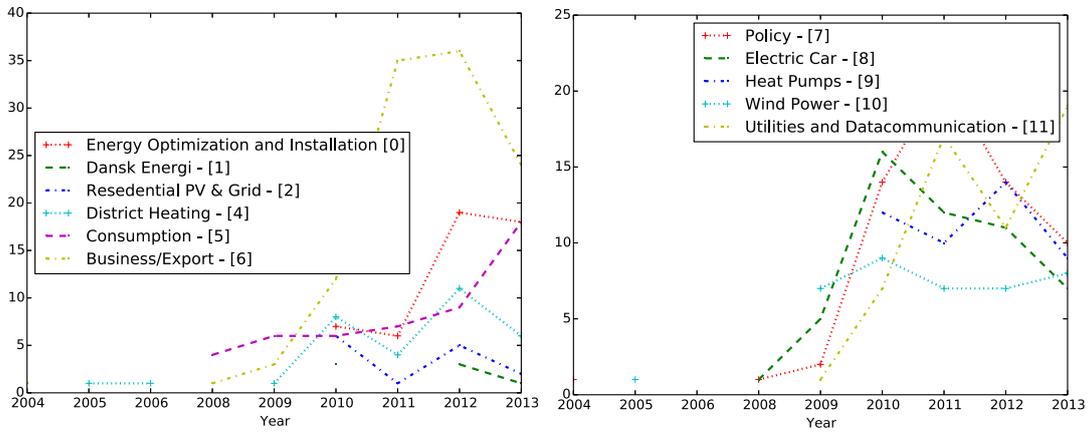


Table A.1: Descriptive statistics, Project description clustering

MC	Technology/Topic	Start Year	AVD	Size	Rel. Size
0	Data Communication	2010.6	1.96	5	5.10
1	Battery Technology	2008.3	1.38	3	3.06
2	Electric Vehicles – General Transportation	2009.9	2.26	9	9.18
3	Water/Mambrane Energy storage	2012.0	0.00	1	1.02
4	Heat Pumps	2011.3	2.95	12	12.24
5	System Integration	2008.8	2.98	13	13.27
6	Electric Vehicles – Charging	2011.3	2.04	6	6.12
7	Frequency Regulation	2008.9	1.91	9	9.18
8	Consumption Flexibility	2008.4	2.48	14	14.29
9	Mixed socio-technical projects	2009.5	1.21	8	8.16
10	Consumption & Frequency control	2008.3	1.89	3	3.06
11	District Heating	2005.4	1.28	16	16.33

Notes: Modulatory Class (MC), Technology/Topic interpretation from TFIDF-Keywords, average project start year for each modularity class, average (edge)weighted degree (AVD), Relativa Size in percent.

Table A.2: Descriptive statistics, Industrial Publications clustering

MC	Technology/Topic	Publication Year	AVD	Size	Rel. Size
0	Energy optimization and installation	2012.0	15.64	50	8.71
1	Dansk Energi (company)	2011.3	32.26	7	1.22
2	Residential PV & Grid	2011.2	23.60	14	2.44
3	Dong Energy (company)	2011.4	15.08	18	3.14
4	District Heating	2011.1	18.75	32	5.57
5	Consumption	2011.3	32.71	50	8.71
6	Technology Export & Business	2011.5	15.44	112	19.51
7	Policy	2011.1	21.60	63	10.98
8	Electric vehicles	2010.9	27.59	52	9.06
9	Heat Pumps	2011.4	27.90	45	7.84
10	Wind	2010.8	23.66	39	6.79
11	Utilities and Data-communication	2011.7	20.81	55	9.58
12	R&DD Projects	2011.2	21.92	32	5.57
99	Excluded-Too high Degree	2012.0	78.88	5	0.87

Notes: Modulatory Class (MC), Technology/Topic interpretation from TFIDF-Keywords, average publication year for each modularity class, average (edge)weighted degree (AVD), Relativa Size in percent.

Table A.3: Overview: Sources industrial publications

Source/MC	0	1	2	3	4	5	6	7	8	9	10	11	12	99	Total	%
Alt om Data				1		1	2					1			5	0.9
Bedre Hjem		2				1				1				2	6	1.0
Bo Bedre										2					2	0.3
BygTek										2					2	0.3
Byggeri	1														1	0.2
Byggeteknik					1	1		2							4	0.7
CSR	2	1					2		1						6	1.0
Computerworld				3		1	10	1	2	1	1	1			20	3.5
DI Business				1		1	8	3	1			1	6		21	3.7
DI Indsigt							1								1	0.2
DSbladet					1								1		2	0.3
Dagens Medicin						1									1	0.2
Dansk VVS	2		2				1	1		2	1				9	1.6
EksportFokus							3						1		4	0.7
El og Energi								1	1						2	0.3
Electra	3		1			4	2	4	1	1	3	1	2		22	3.8
Elektrikeren	2		1			1	4		1	1	1	1			12	2.1
Elektronik & Data													1		1	0.2
Energiwatch							1					1			2	0.3
Erhvervsbladet							1								1	0.2
Erhvervsmagasinet Installatør	2														2	0.3
Fjernvarmen					15					6		2			23	4.0
Fritidsmarkedet						2				1					3	0.5
Hvidvare-Nyt	1					1		1						1	4	0.7
Ingeniøren	1	1	4		5	12	4	7	3	8	9	5	1		60	10.5
Installatør Horisont	8	1				1	3	1		3		1	1		19	3.3
Jern Og Maskinindustrien	6				1		5	1					1		14	2.4
Jern og Maskinindustrien	2								2	1	1				6	1.0
Karrierevejviser											1		1		2	0.3
Kommunen							1						1		2	0.3
LandbrugsAvisen								1							1	0.2
Magasinet Finans												1			1	0.2
Magasinet Statsindkøb	1							2							3	0.5
Magisterbladet							1	1							2	0.3
Mandag Morgen				1			2	3			1	1	1		9	1.6
Mandag Morgen Navigation						1									1	0.2
Mandag Morgen News							3								3	0.5
Maskinmesteren		2			2		3		1		1	1	1		11	1.9
Mester & Svend										1					1	0.2
Mester Tidende	3						1		1						5	0.9
Motor-Magasinet									3						3	0.5
Natur & Miljø									2						2	0.3
Nyhedsbladet Dansk Energi	13		6	7	7	19	51	34	33	13	20	36	13	1	253	44.1
Optimering														1	1	0.2
Pack Markedet	1														1	0.2
Proces-Teknik				2											2	0.3
Prosabladet							1								1	0.2
Samdata										1					1	0.2
Teknovation				2											2	0.3
Telekommunikation				1			2								3	0.5
Tænk						3									3	0.5
byggeplads.dk	2											1			3	0.5
danskVAND										1			1		2	0.3
Økonomisk Ugebrev CFO												1			1	0.2
Totals															574	100

References

- Bergek, A., Jacobsson, S., Carlsson, B., Lindmark, S., and Rickne, A. (2008). Analyzing the functional dynamics of technological innovation systems: A scheme of analysis. *Research Policy*, 37(3):407–429.
- Bijker, W. E. (1997). *Of Bicycles, Bakelites and Bulbs*. Toward a Theory of Sociotechnical Change. The MIT Press.
- Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*. O’Reilly Media, Inc.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.
- Bradford, R. B. (2008). An empirical study of required dimensionality for large-scale latent semantic indexing applications. In *Proceeding of the 17th ACM conference*, page 153, New York, New York, USA. ACM Press.
- Chen, S.-H., Huang, M.-H., and Chen, D.-Z. (2012a). Technological Forecasting & Social Change. *Technological Forecasting & Social Change*, 79(6):1099–1110.
- Chen, S.-H., Huang, M.-H., Chen, D.-Z., and Lin, S.-Z. (2012b). Technological Forecasting & Social Change. *Technological Forecasting & Social Change*, 79(9):1705–1719.
- Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., and Harshman, R. A. (1990). Indexing by latent semantic analysis. *JASIS*, 41(6):391–407.
- Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3):147–162.
- Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., and Harshman, R. (1988). Using latent semantic analysis to improve access to textual information. *Bell Communications Research*, pages 281–285.
- Elzinga, D. (2011). Smart Grids Roadmap. Technical report.
- Farhangi, H. (2010). The path of the smart grid. *Power and Energy Magazine, IEEE*, 8(1):18–28.
- Forskningsnetværket, Smart Grid (2013). Roadmap for forskning, udvikling og demonstration inden for Smart Grid frem mod 2020. Technical report.
- Foxon, T. J., Hammond, G. P., and Pearson, P. J. G. (2010). Developing transition pathways for a low carbon electricity system in the UK. *Technological Forecasting & Social Change*, 77(8):1203–1213.
- Frost-Knudsen, M. (2013). Varmepumpers anvendelse i Danmark.

- Garud, R. and Karnøe, P. (2003). Bricolage versus breakthrough: distributed and embedded agency in technology entrepreneurship. *Research Policy*, 32(2):277–300.
- Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy*, 31(8):1257–1274.
- Geels, F. W. (2010). Ontologies, socio-technical transitions (to sustainability), and the multi-level perspective. *Research Policy*, 39(4):495–510.
- Geels, F. W. (2011). The multi-level perspective on sustainability transitions: Responses to seven criticisms. *Environmental Innovation and Societal Transitions*, 1(1):24–40.
- Giordano, V., Gangale, F., Fulli, G., Jiménez, M. S., Onyeji, I., Colta, A., Papaioannou, I., Mengolini, A., Alecu, C., and Ojala, T. (2011). Smart Grid projects in Europe: lessons learned and current developments. *European Commission, Joint Research Centre, Institute for Energy, Luxembourg*.
- Hekkert, M. P., Suurs, R. A. A., Negro, S. O., Kuhlmann, S., and Smits, R. E. H. M. (2007). Functions of innovation systems: A new approach for analysing technological change. *Technological Forecasting & Social Change*, 74(4):413–432.
- Henschel, A., Casagrande, E., Woon, W. L., Janajreh, I., and Madnick, S. (2011). A Unified Approach for Taxonomy-Based Technology Forecasting. *Business Intelligence Applications and the Web: Models, Systems and Technologies*, page 178.
- Hepbasli, A. and Kalinci, Y. (2009). A review of heat pump water heating systems. *Renewable and Sustainable Energy Reviews*, 13(6-7):1211–1229.
- Hughes, T. P. (1987). The evolution of large technological systems. *The social construction of technological systems: New directions in the sociology and history of technology*, pages 51–82.
- Kaplan, S. and Tripsas, M. (2008). Thinking about technology: Applying a cognitive lens to technical change. *Research Policy*, 37(5):790–805.
- KEMIN (2013). Smart Grid Strategy. Technical report.
- Keson, B. and Norling-Christensen, O. (1998). PAROLE-DK. *The Danish Society for Language and Literature*.
- Laue, H. J. (2002). Regional report Europe: “heat pumps—status and trends”. *International journal of refrigeration*, 25(4):414–420.
- Lund, H., Andersen, A. N., Østergaard, P. A., Mathiesen, B.V., and Connolly, D. (2012). From electricity smart grids to smart energy systems—A market operation based approach and understanding. *Energy*.
- Markard, J., Raven, R., and Truffer, B. (2012). Sustainability transitions: An emerging field of research and its prospects. *Research Policy*, 41(6):955–967.

- Mathiesen, B. V., Blarke, M., Hansen, K., and Connolly, D. (2011a). The role of largescale heat pumps for short term integration of renewable energy. Technical report.
- Mathiesen, B. V., Lund, H., and Karlsson, K. (2011b). 100growth. *Applied energy*, 88(2):488–501.
- Mattern, F., Staake, T., and Weiss, M. (2010). ICT for green. In *the 1st International Conference*, pages 1–10, New York, New York, USA. ACM Press.
- Nelson, R. R. and Winter, S. G. (1982). *An evolutionary theory of economic change*. Harvard University Press.
- Poulsen, C. S. (2007). Varmepumper – status sommeren 2007 i og uden for Danmarks grænser. Technical report.
- Řehůřek, R. and Sojka, P. (2011). Gensim—Statistical Semantics in Python.
- Salton, G. and McGill, M. J. (1983). *Introduction to modern information retrieval*. McGraw-Hill College.
- Sanner, B., Karytsas, C., Mendrinou, D., and Rybach, L. (2003). Current status of ground source heat pumps and underground thermal energy storage in Europe. *Geothermics*, 32(4-6):579–588.
- Sawin, J. (2006). Renewable Energy: A Global Review of Technologies, Policies and Markets - Google Books. *Renewable Energy A Global Review of Technologies*.
- Unruh, G. C. (2000). Understanding carbon lock-in. *Energy Policy*, 28(12):817–830.
- Unruh, G. C. (2002). Escaping carbon lock-in. *Energy Policy*, 30(4):317–325.
- Unruh, G. C. and Carrillo-Hermosilla, J. (2006). Globalizing carbon lock-in. *Energy Policy*, 34(10):1185–1197.
- Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, 10(01):93–115.