Mapping the (R-)Evolution of Technological Fields – A Semantic Network Approach*

Roman Jurowetzki and Daniel S. Hain

Aalborg University, Department of Business and Management, IKE,
Fibigerstrde 11, 9220 Aalborg, Denmark
{roman,dsh}@business.aau.dk
http://www.ike.aau.dk

Abstract. The aim of this paper is to provide a framework and novel methodology geared towards mapping technological change in complex interdependent systems by using large amounts of unstructured data from various recent on- and offline sources. Combining techniques from the fields of natural language processing and network analysis, we are able to identify technological fields as overlapping communities of knowledge fragments. Over time persistence of these fragments allows to observe how these fields evolve into trajectories, which may change, split, merge and finally disappear. As empirical example we use the broad area of *Technological Singularity*, an umbrella term for different technologies ranging from neuroscience to machine learning and bioengineering, which are seen as main contributors to the development of artificial intelligence and human enhancement technologies. Using a socially enhanced search routine, we extract 1,398 documents for the years 2011-2013. Our analysis highlights the importance of generic interface that ease the recombination of technology to increase the pace of technological progress. While we can identify consistent technology fields in static document collections, more advanced ontology reconciliation is needed to be able to track a larger number of communities over time.

Keywords: Technological change, transition, technology forecasting, natural language processing, network analysis, overlapping community detection, dynamic community detection.

1 Introduction

Understanding the pattern of technological change is a crucial precondition to formulate meaningful long-term research and industry policy. Technological change usually happens along *technological trajectories* [1] focusing its pathway within a *scientific paradigm* [2]. Apart from defining the boundaries, a paradigm

* We would like to thank Dan Mc Farland, Dan Jurafsky, Walter W. Powell, Chris Potts, all participants of the 2014 ISS Jena conference, the 2014 KID Nice workshop, and the 2014 Summer Term Stanford Network Forum for inspiration and feedback. All opinions, and errors, remain our own.

L.M. Aiello and D. McFarland (Eds.): SocInfo 2014, LNCS 8851, pp. 359–383, 2014.
© Springer International Publishing Switzerland 2014

often provides a set of generic *technology artifacts* which can be deployed along multiple trajectories [3]. Furthermore, recent trends towards modularization and the development of common interfaces have led to an increasing compatibility of technologies within and between paradigms. We argue that today we face an accelerating deterioration of burdens for technology (re-)combination through growing complementary of components [4,5]. In order to understand innovation activity in many modern technological fields, it therefore becomes pivotal to deploy conceptual frameworks, methods, and data geared towards the analysis of such dynamic and highly interdependent systems.

Common approaches to analyze technological change are yet limited to qualitative in-depth case studies [6,7], quantitative methods depending on data such as patents [8] or scientific publications [9], and more generic simulation models [10,11]. While undeniably useful, they either require massive effort to qualitatively analyze complex interaction patterns in technological space, or rely on quantitative data only available with non-negligible time delay, and only relevant for certain technology domains, often underestimating the context in which technology is used. During the last decade we have witnessed tremendous growth of freely available digital information, often in the form of unstructured text data from sources such as web-sites and blogs, written communication of communities in forums or via e-mail, and knowledge repositories (e.g. SSRN, Researchgate). The topicality and sheer amount of such data bear great opportunities for social science research in general, and particularly to timely analyze complex technological change, as we attempt to demonstrate in the following.

In this paper we present a framework and suggest a set of methods to map technological change by using large amounts of unstructured text data from various on- and offline sources. We conceptualize technological change as the reconfiguration of interaction patterns between *technology fragments*, and their clustering in space to *technological fields*, and in time to *technological trajectories.* To analyze such change, we propose the combination of techniques from the fields of natural language processing (NLP) and network analysis. We use the case of *technological singularity* to illustrate our approach graphically as well as with key measures derived from network analysis.

The remainder of the paper is structured as follows. Section 2 reviews and discusses literature and concepts of technological change, and provides a theoretical framework for our approach. In Section 3 we suggest a set of methods suitable to analyze such a framework, and illustrate it in Section 4 at the case of *singularity* technologies. Finally, Section 5 concludes, provides implications for theory, empirical research, and suggests applications for science and industry policy.

2 Conceptualization and Analysis of Technological Change

2.1 Conceptualization of Technological Change

The conceptualization of technological change has a long tradition in different academic communities. Generally, technology exists to fulfill or support some societal functions through direct application or indirectly through derived products. It is thus always embedded in and framed by a societal, political and organizational context, which co-evolves with it [12]. It is also understood as happening within broader *scientific paradigms* [2].

Scholars studying industrial dynamics further describe the development of technology as contextual to the evolution of industrial structures [1,13]. Technology is envisioned as a mean to problem solving in a particular context, which could usually be solved in various other ways using other technologies. *Technological trajectories* represent pathways spanning across the technological space delimited by the paradigm [1], focusing the problem solving process over time around one possible configuration of technologies. While this process usually unfolds gradually, sometimes significant technological discontinuities punctuate a trajectory [14]. Such disruptive change radically alters a trajectory's or even paradigm's internal logic, or completely replaces it in an act of Schumpeterian creative destruction [18]. Overall that suggests competition between substitutional trajectories. Yet, they can also be compatible and complementary to each other, since generic technological artifacts may feed the progress of multiple trajectories.

Drawing on work in theoretical biology [16], technological evolution can be conceived as a recombinatory process of novel and existing component technologies within complex adaptive systems [17]. Innovative recombinations can address fundamentally different problems from the ones that were initially targeted within the components' paradigms. This comes close to a Schumpeterian understanding, where the innovation process is envisioned as the recombination of existing resources in a novel way [18]. The result of such a development can also be envisioned as a complex system with a number of elements that collectively fulfill a single or various goals [19]. A main characteristic of such complex systems is a high degree of interdependence (or epistasis), meaning a functional sensitivity of a system to changes in constituent elements [17]. Thus, a change in one element will affect not only affect its own but also the functioning of epistatically related ones [20]. Since the complexity of the system increases with the number of elements and their degree of interdependence, in large epistatic systems one faces a *complexity catastrophe*, making it increasingly hard to find useful combinations [17].

A possible solution suggested to avoid the *complexity catastrophe* is to increase the systems modularity [4,5,21]. This approach aims at the development of standardized interfaces between more discrete elements to mediate interdependence [22], thus allowing to decrease the overall complexity while maintaining the number of possible recombinations. Modularity and common interfaces further ease the way to combine and recombine components stemming from different trajectories, perhaps even different paradigms. On a higher level, technological revolutions disrupting current techno-economic paradigms are usually accompanied by the emergence of such modules, which can be deployed in various contexts [14]. A recent and very obvious example for this development, the smartphone, is illustrated in Figure 1. The combination of voice and data communication with GPS, camera, compass and accelerometer technologies, bound together by a miniature touchscreen-computer, opened up for a uncountable number of not anticipated applications. Various standardized wireless connection technologies like bluetooth or WiFi allow for compatibility with many other external devises, thus increasing the functionality and re-purposing the phone.

We argue that today we are witnessing a rapid decline of the burdens to technology-combination through efficient modularization between components within artifacts such as the smartphone. Embracing this line of thought, we aim to develop a framework and methodology geared towards the analysis of evolving interdependent technology systems. Such a framework has to be able to capture the ongoing incremental adjustment of interaction pattern between its components (*technological evolution*) as well as disruptive changes fundamentally altering the systems logic (*technological revolution*).

2.2 Measurement and Analysis of Technological Change

Existing empirical research on technological change can broadly be divided in three fields. Work from scholars associated with the Science, Technology and Society (STS) tradition mainly relies on detailed ethnographic studies of the complex multidimensional setup around technological systems, and sheds light on the variety of factors that influence and shape its development [23,24,25].

A stream of more positivistic research in the fields of industrial economics and sciencometrics is primarily based on patent and scientific publication data as an approximation for technological development. Research so far mostly incorporates patent data as aggregated numbers to explain differences in scale [26], or in a network representation to explain structural differences [8,27] in the development of technologies across countries and industries. Patent data has also been used to study invention as a recombination process [17,28,29].[1]

Most recently, social scientists have also started to deploy methods from the fields of computational linguistic and NLP to advance empirical research on the development of science and technology [32,33,34,35]. In their essence, such linguistically informed methods are capable of identifying patterns of language usage in large bodies of text and communication. They range from simple measures of word co-occurrence across documents, corpora and over time [36], to complex linguistically informed probability model [37,38,39].

We perceive the latter as a fruitful way to analyze technological change, implicitly accounting for the socio-economic context in which it is embedded. Such an approach integrates the broad multidimensional perspective of qualitative researchers, that very importantly emphasizes the role of technology users, organizations and governments in innovation processes, with quantitative objectivity given by the machine learning based methodology.

[1] However, besides its merits and easy accessibility, there are widely recognized limits in the use of patent data [30,31] such as the high variation of importance across industries and countries, and over time and the long delay between the time research is conducted and the corresponding patent publication.

2.3 Technology Evolution as Structural Network Change

We conceptualize technology as a system of interdependent components [40] within their respective trajectories of development [1]. Representing such systems of interacting elements as networks has brought fresh perspectives and insights to the analysis of complex phenomena from the biological to the social sciences [41]. Embracing this approach, we attempt to analyze technological change as the ongoing structural reconfiguration of interaction between elements in a technology network, which allows us to deploy the rich set of network analysis.

On the lowest level of aggregation in a network representing a technological system, one finds what we call *technology fragments*. They represent atomic, non-reducible repositories of scientific/technological knowledge needed to fulfill certain narrow tasks. Scientific, technological and industrial applications such as machines, software and other devices (which we call *technological artifacts*) combine *technology fragments* in a functional relationship to produce some output. In our previous example, GPS devices, touchscreens and WiFi receivers represent *technology fragments*, which combined in a functional relationship can resemble the smartphone, a *technological artifact*. On a higher level, sets of complementary and substitutional artifacts form a*technological field* (which could be, let's say *mobile applications and devices*). Over time, such fields develop along *technological trajectories*, where accumulated sets of common configuration patterns reproduce over time and set the foundation for further combinations. Again, fragments and artifacts originating from one field might be reconfigured and redeployed in a different field to fulfill the same or even a different purpose. Furthermore, fragments as well as artifacts might not even mainly belong to one field, but be equally employable across multiple fields.

In summary, our conceptualization of technological change, and the suggested methods to analyze it, is based on the following assumptions:

Assumption 1: *Knowledge fragments are atomic, non-reducible repositories of scientific/technological knowledge*

Assumption 2: *Technology fragments can be arbitrary combined and recombined to resemble functional technological artifacts of varying quality*

Having clarified the elements (or edges) in such a network, one has to decide how to measure the functional relationships between them. In our case, identifying technology fragments in unstructured text data, we have to add the following assumption:

Assumption *3: Co-location of technology fragments in documents imply a functional relationship between them*

3 Analyzing Technology Evolution: Dynamic Semantic Network Approach

After providing a conceptual framework to analyze technological change, in this section we suggest a set of methods to empirically study such changes. A illustration of the method pipeline is provided in Figure 5.

3.1 From Unstructured Text to Technology Fragments: Entity Extraction

First obvious choice to be made is which corpus of technology related text documents one wants to analyze. Such a corpus should optimally (i.) consist of technology related writings (ii.) ranging equally distributed over a time sufficient to observe technological change, and (iii.) not be biased towards particular technologies within the system. Examples for such data are scientific publications, patent descriptions, articles in industry journals, but also online sources such as collections of tech-blogs. In Section 4 we illustrate how to generate an online data corpus with socially enhanced web scraping techniques.

In a next step, it is necessary to convert the unstructured text documents to a machine readable representation.[2] For our means, the goal is to reduce each document to the contained technological concepts. Instead of using a probabilistic approach that stepwise excludes text-elements that are definitely not a technology, we try to detect mentioned technologies in the data. This task falls into the category of *named entity extraction*, which typically relies on tagged dictionaries and string-matching rules to identify the required concepts.

A number of applications related to this development target the identification of different concepts in unstructured text, among others technological and industrial terms. The advantage of these semantic web tools is that they are supported by large, centralized, constantly updated and optimized dictionaries and intelligent disambiguation functions. The result of a successful entity extraction returns a collection of documents that only contain the mentioned technology terms and their document appearance frequency. Referring to our conceptual framework in Section 2, the extracted technology term resemble the elements (nodes) in our technological system, which we label as *technology fragments*.

3.2 From Technology Fragments to a Network: Vector Space Modelling

After having defined the nodeset in our network of *technology fragments*, we have to create weighted edges between them, representing their technological relatedness and interaction. In a first step we construct a (hierarchical) 2-mode network between *technology fragments* and the corresponding documents they occur in. We weight the edges by the pairwise cosine similarity between the vectors of the *technology fragment* and document within a vector space, which we define by by training a Latent Semantic Indexing (LSI) model [42,43] on the

[2] Typically, this takes the format of a bag of words (BOW), a line-up of thematically relevant keywords, usually nouns and bi-gram noun phrases. The key assumption of this type of NLP applications is that statistically significant co-occurrence patterns of concepts across the corpus is indicative for actual association between them.

full corpus of documents.[3] Thus, our measure of edge weight indicates to which extent the term representing the *technology fragment* is semantically close to the entirety of other terms contained by the document (see. Section 5). To map technological change over time, we do this separated for every observation period.

While the entirety of *technology fragments* is stable over time, documents obviously experience a 100% turnover in population every observation period. To coerce a stable nodeset, we project the 2-mode to a weighted 1-mode network in technology space. Again, the underlying rationale is based on the assumption that co-occurrence in documents - at least on an aggregated level - also corresponds to a functional relationship between *technology fragments*. However, on a document level that will not always be true. While some documents may discuss technology in the realm of one particular *technological fields*, others might serve more as an overview on industry or research of a broader context, hence contain a collection of *technological fragments* from many otherwise distinct fields. Thus, we penalize documents containing more technology fragments in a similar spirit as the method used by [44], represented by the following equation [45]. Here w_{ij} represents the edge-weight between node i and j, and p the corresponding documents.

$$w_{ij} = \sum_{p} \frac{w_{i,p}}{N_p - 1} \tag{1}$$

We end up with a one-mode network of *technology fragments* connected by the pairwise projected semantic similarity values, associated with the corresponding period. Figure 3 illustrates these nodeset properties in dynamic networks.

Identifying Technological Fields: Overlapping Community Detection. We depict technological change as the structural reconfiguration of micro level interactions between *technology fragments*. When analyzing the structure, function, and dynamics of networks, it is extremely useful to identify sets of related nodes, known as communities, clusters, or partitions [46]. Such communities of closely connected technologies resemble what we call a *technological field*, a set of complementary or substitutional technologies following one *technological trajectory*, and clustering over time around a common objective. Therefore, we attempt to identify *technological fields* using a community detection algorithm of choice.[4]

[3] Before training the model, we apply TF-IDF weights to all terms within the documents. This appreciates the value of particularly important terms for the single document, while depreciating the value of generic terms that often occur across the corpus. Here we have chosen the established LSI algorithm for training the vector space model but other algorithms e.g. Latent Dirichlet allocation (LDA) or Random Projections would also be feasible to calculate pairwise cosines.

[4] An alternative approach would be to use to identify technological fields by the using topic modeling, an approach that lately started to gain traction in social science [35,34,37], create a two-mode network of terms and topics, and project it to an one-mode network of terms. However, for reasons described we here want to offer an alternative, where the topics are already identified using the powerful community detection methods offered by network analysis.

Early clustering and community detection algorithms, in network analysis and elsewhere, usually assumed that the membership of entities to one distinct group. However, depending on the meaning of edges and nodes, many real life networks show a high overlap of communities, where nodes at the overlap are associated with multiple communities. This especially tends to happen when relationship of different quality are projected in a one-mode network [47]. Ones' social interaction network for instance may consist of family members, work colleagues, members of the same karate club or other associations. The more diverse interests such a person has, the more different communities this person will be assigned into. In the same way, the more generic the nature of a *technology fragment* or artifact, the more technological fields will it have functional relationships with. Some *technological artifacts* (and the *technology fragments* resembling them) are that pervasive, they facilitate almost all other technologies in the way they work, such as by its time steam-power or nowadays semiconductors [14]. Embracing that line of thought, researchers recently stated to develop community detection algorithms able to cope with overlapping and nested community structures [48,49], which can be deployed to properly delimit interdependent *technological fields*.

Identifying Technological Trajectories: Dynamic Community Detection. *Technological fields* do not spontaneously appear and reassemble in a vacuum. They gradually change, grow or decline in an cumulative manner, following a historical *technological trajectory* which connects them over time. However, in times of disruptive technological change, former technology interaction pattern might completely reconfigure, particular new configurations might spin-off a main trajectory and so forth. Owing respect to the evolutionary nature of technology, we want to identify communities which are somewhat stable and thus to be found in multiple observation periods, but also allow *technological fields* to experience disruptive key-events in their life-cycle. Besides helping us linking changing communities over time, the identification of such effects in itself represents an interesting information. We consider the following significant events a community might experience during its evolution, also illustrated in Figure 4:

- Birth & Death: The first time a community C_i^t (which are the representation of a *technological field*) is observed and not matched with an already existing community C_j^{t-1}. This community, however, does not have to be stable over time. We in fact expect a substantial share of communities to only appear in on period but not sustain.
- Pause: Communities might be more stable than the reporting on them in the corpus.Thus, allowing them to pause for a period might smoothen birth & death dynamics.
- Merge: In case two communities develop substantial functional interdependence, the main interaction with the rest of the system only happens between them. Thus they merge and form a new community consisting of both

sets. Technically that happens when two or more different communities are matched with one dynamic community D_j in the previous period.
– Split: In the same manner, communities can also separate into independent disciplines. Technically a split occurs when one community C_i matches with two or more dynamic communities in the previous period.

We do so by applying a simple but effective heuristic threshold-based method allowing for many-to-many mappings between communities across different observation periods proposed by [50]. Here we compare an identified community C_i^t in observation period t with the set of dynamic communities in the previous period $\{C_1^t, \ldots, C_J^{t-1}\}$ by employing the widely adapted Jaccard coefficient J_{ij}^t, calculated as follows:

$$J_{ij}^t = sim(C_i^t, C_j^{t-1}) = \frac{|C_i^t \cap C_j^{t-1}|}{|C_i^t \cup C_j^{t-1}|} \tag{2}$$

If the similarity exceeds the defined matching threshold $\theta \in [0, 1]$, both communities are added to the dynamic community D_i. Using this has the advantage that is independent of the (static) community detection method of choice in the observation periods, hence represents a somewhat modular approach. It can also handle overlapping as well as (with some minor adjustments) weighted communities. A major advantage of this approach is the separation of static and dynamic community detection is the high flexibility in the choice of suitable algorithms.

4 Demonstration Case

In the following section we demonstrate the capabilities of our approach to deliver insightful results, and provide some illustrative examples of measures and graphical representations that can be used to gain further insights. We intended to find an empirical case of technological development that would combine a large number of components from traditionally disconnected *technological fields*. Additionally, the *technology field* in focus should be yet in a formative stage and have a potentially strong and broad social impact to generate enough attention and thus reporting texts online. We decided to explore the field of *singularity*. Rather then a clearly delineated *technological field*, singularity represents a future scenario and an umbrella term that summarizes a number of developments in areas as diverse as neuroscience and 3D printing. Based on the context of the technology under study and the characteristics of the corpus, we provide examples how to calibrate the techniques used in the different stages of or method pipeline.

4.1 Empirical Setting: The *Singularity* Case

Technological Singularity as a term has gained momentum since the publication of Ray Kurzweil's book in 2005 [51]. Observing various measures of technological progress over time, he argues that most technologies improved their performance exponentially and therefore it is only a matter of a few decades until we will have reached a point in history when artificial intelligence will supersede human intelligence. The most powerful technological advancement of the 21^{th} century will happen when robotics, nanotechnology, genetic engineering and artificial intelligence reach a certain level of development and can be combined, what will have disruptive consequences for society, culture and the human nature.

Recently, *singularity* entered the European technology policy context, as a technological field within the Horizon 2020 programming. Since 2012, the Directorate General for Communications Networks, Content and Technology (DG CONNECT) is undertaking a foresight process to inform the ICT related programming of research to be financed under Horizon 2020, where *singularity* was identified as one of the 10 central technological fields. It is currently being examined closer to capture early signals and anticipate beneficial trends that should be supported within public research funding schemes.

4.2 Data Mining and Corpus Generation

Researchers, organizations and science journalists are increasingly using social media and the blogosphere to communicate findings and developments, far ahead of journal publication or conference proceedings. This makes microblogging platforms and in particular Twitter with over 200 million monthly active users (Feb. 2014) a valuable source of data. We now describe our data mining approach aiming at selecting relevant twitter updates by relevant users. Instead of using already available corpora to study technological change in *singularity*, such as patent description, scientific publications and industry journals, we choose to create an own out of a variety of online available technology relevant text documents, including publications, tech-blogs *et cetera*. Since *singularity* is a recent and very heterogeneous movement spanning various scientific, industries and tech-communities with distinct routines for communicating and publishing findings and progress, our final corpus therefore is supposed to be unbiassed towards a particular discipline.

To identify relevant documents, we employ a socially-enhanced search routine based on twitter tweets. Twitter's graph structure, built on followship links, is similar to citation networks in academic publications. This enables the construction of large directed graphs and allows applying network analysis methods, to identify central actors for a particular field or topic. For this study we constructed a large followship graph around the - somewhat arbitrarily selected - account *Singularity Hub*, which is an online news platform that actively reports on the topic. The initial *snowballed* network has 49,574 accounts. Using eigenvector centrality, we identify the most influential users and then manually reduce the number of nodes down to 34 twitter accounts that indicate an interest for the area in their profile.[5] Figure 6 shows a central fragment of the network.

[5] This selection is very restrictive but is likely to make the final corpus less noisy. Alternatively the manual reduction can be skipped and a corpus filtering built in, at a later stage.

Coloring represents communities, detected by the Louvain algorithm [53], merely for illustration. We can see that the red cluster seems to contain all the central organisations that are present on twitter and focused on singularity and transhumanism like the H+ movement, KurzweilAI, David Orban and more. The green cluster is mostly populated with users that are related to robotics and the violet to software architecture. An overview of the selected user accounts can be found in Table 1.

Micro-blogged tweets (status updates) by these actors often contain links to research papers, popular media articles or blog entries that the selected user considers as worth communicating. For each of these accounts we extract up to 3,200 status updates starting with the most recent, 63,000 in total. We discard all updates that do not carry a link. Relevant tweets were then identified using a vector space model powered semantic search. The text content behind the embedded links - outside of Twitter is then extracted and processed, and finally represents our document corpus for further analysis.

4.3 Identification of Technology Fragments: Entity Extraction

The documents in our corpus discuss technology from very different angles. Some talk about state-of-the art research in certain university labs, while others review the allocation of public research grants or venture capital investment strategies. When attempting to uncover functional relationships between technology fragments, it is crucial to avoid false positives caused by relationships that are non-technical in nature, such as *being funded by the same investor*, or *developed in the same country*. We rely on entity extraction when condensing documents to BOW representations. In the particular case we use OpenCalais, a free web service that performs entity identification across 39 different concepts within submitted text data. The great advantage of *cloudsourcing* in this case is given by the fact that the centralized machine learning algorithms of OpenCalais are trained on a very large amount of natural text and its dictionaries are constantly updated and optimized. An offline solution would hardly be able to compete in terms of performance and topicality.[6] When inspecting the results we find clear technology terms such as *dna profiling*, *robotic surgical systems*, *clinical genomics* or *regenerative stem cell technologies*, which come fairly close to how we understand technology fragments. These terms narrowly describe technology deployed for a fairly delimited task. However, we also find boarder technology terms such as *stem cells genomics*, which span across a somewhat larger field of applications and are likely to include some of the aforementioned terms, and on an even more generic level terms such as *biotechnology* or *robot*.While this clearly diverts from our theoretical framework, where we find on node level only functional interac-

[6] For an overview and performance evaluation of available systems see [52]. In addition, OpenCalais provides ontology reconciliation and disambiguation. Identified entities are in many cases enriched with metadata (e.g. profession for persons, ticker symbols for companies and geospatial coordinates for locations). Other detected entity types are not used in this analysis.

tion of atomic *technology fragments*, we do not consider that as worrisome for the analysis to come.

4.4 Network Generation, Technological Field and Trajectory Identification

For a very first inspection and illustration of the nodeset we create a simple static network of all documents connected by their similarity in terms of containing *technology fragments*, cluster them by applying the common Louvain algorithm [53], and plot them in Figure 7. For the three main communities detected we provide a tag-cloud, weighted by the fragments' TF-IDF scores. One can see at first glance that our *singularity* corpus very broadly consists of three fields, where the biggest is centered around robotics, and the two others around (stem) cell and brain research, or to be more interpretative: Robotics, biotechnology and neuroscience. Table 2 provides some key statistics on the networks, communities, and their development. While subject to some fluctuation, the networks seem to develop from many to less nodes and edges, and to less but denser communities. This might indicate *singularity* after an initial phase of experimentation to mature and establish more delimited fields and sub-disciplines, as life-cycle theories might suggest.

We now construct a set of two-mode networks between this nodes and the documents in our corpus,[7] containing only documents published in the corresponding observation period, which we choose to be half a year.[8] Finally, we project this structures on one-mode networks between technology fragments.

Now we identify *technological fields* with the link community detection algorithm proposed by [48], which is able to detect communities with highly pervasive overlap by clustering links between the nodes rather than the nodes themselves.[9] Each node here inherits all memberships of its links and can thus belong to multiple, overlapping communities (*technological fields*). By doing so, we owe respect to the overlapping and nested structure of technology, and are able to identify key *technological fragments* interacting with multiple distinct fields.We first run the community detection separated for every time step independently. We do not *a-priori* define a fixed amount of communities, but rather set the cutoff at the point where the average community density is optimized in every observation period.

Table 3 plots the network of *knowledge fragments* and their membership to *technological fields* for every timestep. Again, what can be seen is that *singularity*

[7] Vector space modeling is performed with the GENSIM package [54] within IPython, using LSI and a 400 dimensional model as suggested by [55].

[8] This choice has to be made according to the properties of the data to be analyzed, since best results can be achieved when the network structure shows some gradual change between the observation periods, but no radical turnover suggestion complete discontinuity. This corresponds roughly to a Jaccard index of the two networks between 0.2 and 0.8.

[9] We use the implementation of the link-community approach provided by [56] as package for the statistical environment R.

appears to develop from a broad area without clear boundaries and high interconnectedness towards clearly delimited *technological fields*. However, we also find first hints that over time some very generic technologies such as *smartphones* and *artificial intelligence* appear to develop towards a very central position, where they serve as common interface between most other fields. While it seems unlikely that *smartphones* (as we understand them today) will be around for much longer then a decade, their centrality in the singularity discussion can be understood as the importance of mobile devices that enhance our by nature limited interaction range. In fact, *smartphones* became a rapidly adopted human enhancement device and currently a number of different wearable technologies are entering the mainstream markets. We also see the generic *artificial intelligence*, which is at the very core of the singularity debate, in a very central position as interface or generic technology between *technological*.

We now perform a threshold-based dynamic community detection[10], where we besides an immense turnover of briefly appearing and disseminating short-term trends indeed find identify a set of persistent *technological trajectories*. Table 4 illustrates the composition of some selected communities which proves to be somewhat stable over time.[11] The tag-cluster are a good way to visualize the interaction between the actual technologies, principal applications and challenges. The first cluster suggests for instance that an important area of application for biometric technologies in conjuncture with machine learning will be found within law enforcement. The second cluster addresses advancements in the area of augmented reality and connections to existent social network structures using primarily mobile devices.

5 Summary and Conclusion

The aim of this paper was to provide a framework and novel methodology geared towards mapping technological change in complex interdependent systems by using large amounts of unstructured data from various recent on- and offline sources. We combine techniques from the fields of NLP and network analysis. Our approach is based on the following steps:

- Using entity recognition techniques we identify technology related terms in the text document of our corpus, which resemble *technology fragments*.
- In a first step, using vector space modeling, we construct an undirected two-mode network between technology fragments and corpus documents for every observation period, where the edges are weighted by the pairwise cosine similarities between documents and terms.
- After projecting this network in technology space, we end up with an undirected one-mode network of technology fragments connected by their weighted co-occurrence in documents of the corresponding observation period.

[10] We use a C++ implementation provided by [50].

[11] For the sake of clarity, the technology fragments are weighted by their within-cluster centrality.

- To delimit *technological fields* in every observation period, we use overlapping community detection techniques, owing respect to the interdependent and nested nature of technology.
- To identify *technological trajectories*, we link *technological fields* between observation periods over time using

As empirical example we use the broad area of *Technological Singularity*, an umbrella term for different technologies ranging from neuroscience to machine learning and bioengineering which are seen as main contributors to the development of artificial intelligence and human enhancement technologies. We extract 1,398 relevant text documents all over the internet, using a social search routine that we built around the followship structure within the microblogging service twitter. Using entity recognition tools from the semantic web area, we reduce documents to technology-term representations and finally generate a semantic timestep network of technology fragments. Our community detection exercise identified many coherent technological fields within each community. Already the static clustering provides valuable insights in the emergence of new technological fields and applications for existing technologies. Overlapping community detection, allowed us also to identify certain *general* technologies that work as hubs between other technologies, stemming from a large number of different domains.

Yet, we find the results of the community-tracking over time unsatisfactory. The obstacle are *false negatives* that obstruct the identification of similar communities over time. Our language is full of synonyms, metaphors and unregulated terminology. The reader of this article has no difficulty comprehending that we use the terms *clusters* and *communities* interchangeably, a computer would not. While we are (yet) unable to *teach* the algorithm a deep understanding of ontology, we can try to normalize the terminology as far as possible. This future measure should increase the number of identical terms over time. Furthermore, there seem to be a trade-off between the thematic scope of a given corpus and the resolution of the analysis. Therefore, a broader corpus is most suitable for creating a broad-brush picture of technological change.

We believe a major advantage of our approach is that it conveys text data into a network representation suitable for a dynamic analysis of technology. It proves to be more flexible with respect to the corpus than other semantic or n-gram based methods in natural language processing. Furthermore, for subsequent quantitative analysis and graphical representation one can now draw from the large toolkit of powerful methods available for network analysis. The here performed dynamic community detection is one example, but other methods such as blockmodeling appear to be promising to gain further insights into the evolution of technology. Finally, networks are well established in many areas of social science and thus a representation of semantic features as networks is likely to help bridging the gap between scholars in computer and social science.

References

1. Dosi, G.: Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. Research Policy 11(3), 147–162 (1982)
2. Kuhn, T.S.: The Structure of Scientific Revolutions. University of Chicago Press (1962)
3. Timothy, F.: Bresnahan and Manuel Trajtenberg. General purpose technologies engines of growth? Journal of econometrics 65(1), 83–108 (1995)
4. Baldwin, C.Y., Clark, K.B.: Design Rules: The power of modularity. MIT Press (2000)
5. Schilling, M.A.: Toward a general modular systems theory and its application to interfirm product modularity. Academy of Management Review (2000)
6. Davies, A.: Innovation in large technical systems: The case of telecommunications. Industrial and Corporate Change 5(4), 1143–1180 (1996)
7. Hekkert, M.P., Negro, S.O.: Functions of innovation systems as a framework to understand sustainable technological change: Empirical evidence for earlier claims. Technological Forecasting & Social Change 76(4), 584–594 (2009)
8. Verspagen, B.: Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. Advances in Complex Systems 10(01), 93–115 (2007)
9. Wagner, C.S., Leydesdorff, L.: Network structure, self-organization, and the growth of international collaboration in science. Research Policy 34(10), 1608–1618 (2005)
10. Dawid, H.: Agent-based models of innovation and technological change, vol. 2, ch. 25, pp. 1235–1272. Elsevier (2006)
11. Lopolito, A., Morone, P., Taylor, R.: Emerging innovation niches: An agent based model. Research Policy 42(6), 1225–1238 (2013)
12. Kaplan, S., Tripsas, M.: Thinking about technology: Applying a cognitive lens to technical change. Research Policy 37(5), 790–805 (2008)
13. Hain, D.S., Jurowetzki, R.: Incremental by design? on the role of incumbents in technology niches - an evolutionary network analysis. In: Conference Proceeding, 6th Academy of Innovation and Entrepreneurship Conference, Oxford, UK (2013)
14. Perez, C.: Technological revolutions and techno-economic paradigms. Cambridge Journal of Economics, bep051 (2009)
15. Schumpeter, J.A.: A Theory of Economic Development. Harvard University Press, Cambridge (1911)
16. Kauffman, S.A.: The Origins of Order. Self-Organization and Selection in Evolution. Oxford University Press (1993)
17. Fleming, L., Sorenson, O.: Technology as a complex adaptive system: evidence from patent data. Research Policy 30(7), 1019–1039 (2001)
18. Schumpeter, J.A.: Capitalism, Socialism and Democracy. Harper, New York (1942)
19. Simon, H.A.: The Sciences of the Artificial. MIT Press (1969)
20. Frenken, K.: A fitness landscape approach to technological complexity, modularity, and vertical disintegration. Structural Change and Economic Dynamics 17(3), 288–305 (2006)
21. Ethiraj, S.K., Levinthal, D.: Modularity and innovation in complex systems. Management Science (2004)
22. Langlois, R.N.: Modularity in technology and organization. Journal of Economic Behavior & Organization (2002)

23. Bijker, W.E., Hughes, T.P., Pinch, T., Douglas, D.G.: The Social Construction of Technological Systems. In: New Directions in the Sociology and History of Technology. MIT Press (2012)
24. Bijker, W.E.: Of Bicycles, Bakelites and Bulbs. In: Toward a Theory of Sociotechnical Change. The MIT Press (1997)
25. Hughes, T.P.: The Evolution of Large Technological Systems, pp. 51–82. The MIT Press (1987)
26. Pavitt, K.: R&D, patenting and innovative activities: a statistical exploration. Research Policy 11(1), 33–51 (1982)
27. Fontana, R., Nuvolari, A., Verspagen, B.: Mapping technological trajectories as patent citation networks. an application to data communication standards. Economics of Innovation and New Technology 18(4), 311–336 (2009)
28. Fleming, L., Sorenson, O.: Science as a map in technological search. Strategic Management Journal 25(89), 909–928 (2004)
29. von Wartburg, I., Teichert, T., Rost, K.: Inventive progress measured by multi-stage patent citation analysis. Research Policy 34(10), 1591–1607 (2005)
30. Griliches, Z.: Patent statistics as economic indicators: a survey. In: R&D and Productivity: The Econometric Evidence, pp. 287–343. University of Chicago Press (1998)
31. Pavitt, K.: Patent statistics as indicators of innovative activities: possibilities and problems. Scientometrics 7(1), 77–99 (1985)
32. Mohr, J.W., Bogdanov, P.: Introduction – topic models: What they are and why they matter. Poetics 41(6), 545–569 (2013)
33. Ramage, D., Rosen, E., Chuang, J., Manning, C.D., McFarland, D.A.: Topic modeling for the social sciences. In: NIPS 2009 Workshop on Applications for Topic Models: Text and Beyond, vol. 5 (2009)
34. McFarland, D.A., Ramage, D., Chuang, J., Heer, J., Manning, C.D., Jurafsky, D.: Differentiating language usage through topic models. Poetics 41(6), 607–625 (2013)
35. DiMaggio, P., Nag, M., Blei, D.: Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of us government arts funding. Poetics 41(6), 570–606 (2013)
36. Chen, S.-H., Huang, M.-H., Chen, D.-Z., Lin, S.-Z.: Technological Forecasting & Social Change 79(9), 1705–1719 (2012)
37. Hall, D., Jurafsky, D., Manning, C.D.: Studying the history of ideas using topic models. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 363–371. Association for Computational Linguistics (2008)
38. Ramage, D., Manning, C.D., McFarland, D.A.: Which universities lead and lag? toward university rankings based on scholarly output. In: Proc. of NIPS Workshop on Computational Social Science and the Wisdom of the Crowds (2010)
39. Nallapati, R., Shi, X., McFarland, D.A., Leskovec, J., Jurafsky, D.: Leadlag LDA: Estimating topic specific leads and lags of information outlets. In: ICWSM (2011)
40. Hughes, T.P.: The evolution of large technological systems. The social construction of technological systems: New directions in the sociology and history of technology, pp. 51–82 (1987)
41. Newman, M., Barabási, A.-L., Watts, D.J.: The Structure and Dynamics of Networks: Princeton University Press (2006)
42. Deerwester, S.: Improving information retrieval with latent semantic indexing. In: Proceedings of the 51st ASIS Annual Meeting (ASIS 1988), vol. 25 (1988)
43. Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A.: Indexing by latent semantic analysis. JASIS 41(6), 391–407 (1990)

44. Newman, M.: Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. Physical Review 64(1), 16132 (2001)
45. Opsahl, T.: Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. Social Networks 35(2), 159–167 (2013)
46. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D.: Defining and identifying communities in networks. Proceedings of the National Academy of Sciences of the United States of America 101(9), 2658–2663 (2004)
47. Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A.: Multilayer networks. arXiv preprint arXiv:1309.7233 (2013)
48. Ahn, Y.-Y., Bagrow, J.P., Lehmann, S.: Link communities reveal multiscale complexity in networks. Nature 466(7307), 761–764 (2010)
49. Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., Onnela, J.-P.: Community structure in time-dependent, multiscale, and multiplex networks. Science 328(5980), 876–878 (2010)
50. Greene, D., Doyle, D., Cunningham, P.: Tracking the evolution of communities in dynamic social networks. In: Proc. International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2010 (2010)
51. Kurzweil, R.: The Singularity Is Near. When Humans Transcend Biology. Penguin (2005)
52. Rizzo, G., Troncy, R.: NERD: evaluating named entity recognition tools in the web of data. In: Workshop on Web Scale Knowledge Extraction, WEKEX 2011 (2011)
53. Vincent, D.: Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008(10), P10008 (2008)
54. Řehůřek, R., Sojka, P.: Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks (2010)
55. Bradford, R.B.: An empirical study of required dimensionality for large-scale latent semantic indexing applications. In: Proceeding of the 17th ACM Conference, p. 153. ACM Press, New York (2008)
56. Kalinka, A.T., Tomancak, P.: linkcomm: an r package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. Bioinformatics 27(14) (2011)

Appendix

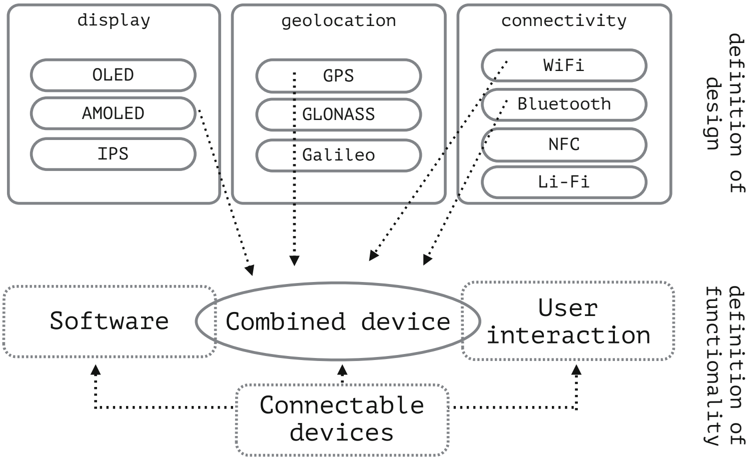
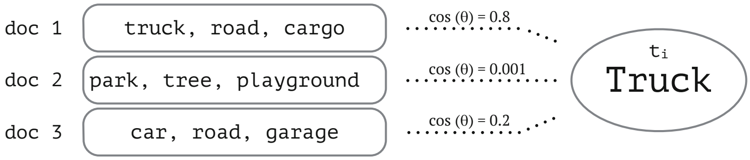


Fig. 1. Illustrative combination of technology components from different trajectories

Fig. 2. Example of pairwise semantic similarity between terms and documents



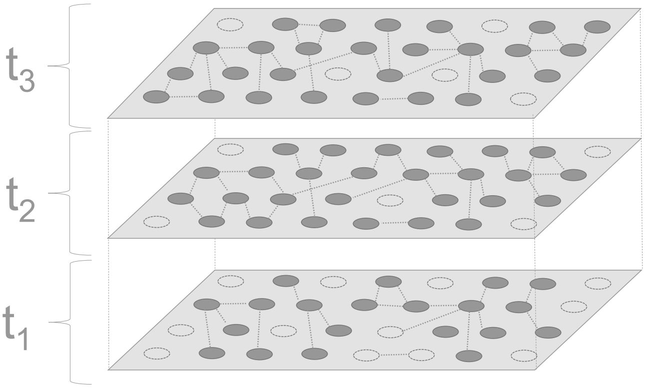


Fig. 3. Illustration of the development of a nodeset over time

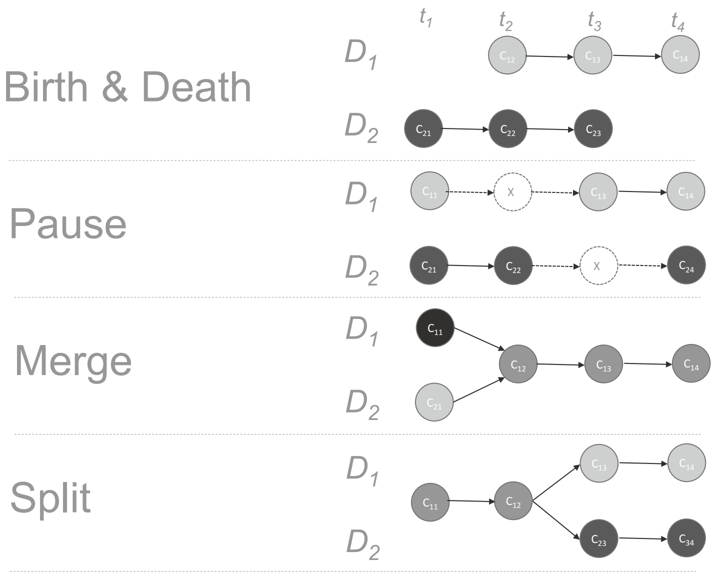


Fig. 4. Illustration of significant events in the evolution of communities, adopted from [50]

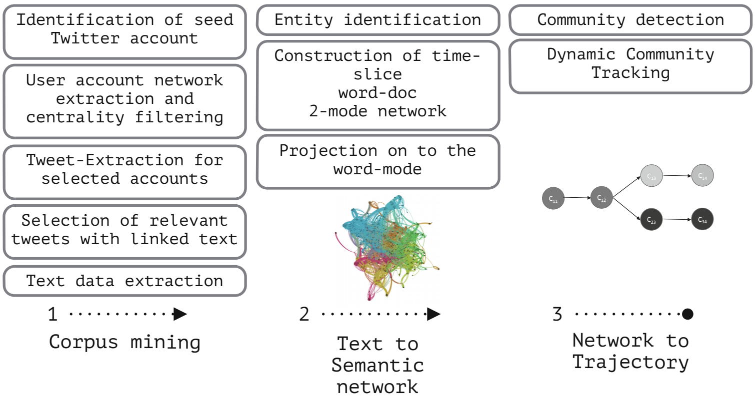


Fig. 5. Illustration of the method pipeline

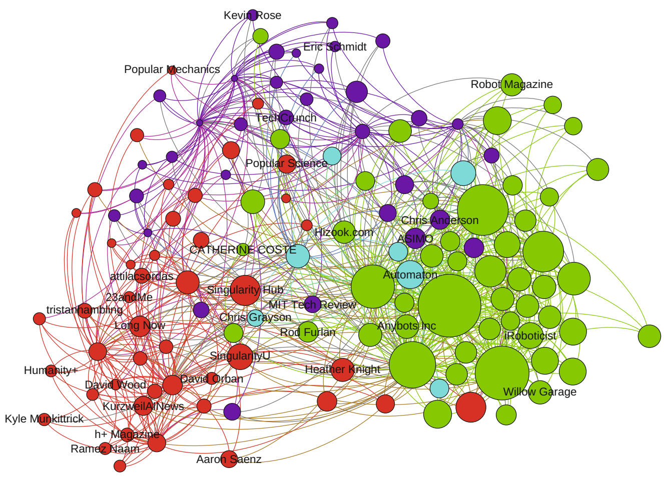


Fig. 6. The central fragment of the twitter account network with the finally selected profiles for text-extraction

Table 1. Overview over the "expert" Twitter-accounts that were used for the text extraction

Twitter-id	Name	Location	Description
121684992	CATHERINE COSTE	Genomic Entertainment	MIT certificate in Genomics. Genomic & Precision Medicine UCSF. Blog: Ethics, Health & Death 2.0 - DTC Genomics
16870421	SingularityU	NASA Moffett Field, CA	Silicon Valley's leading experts on exponential technology. Follow @singularityhub @singularitylabs @suglobal @exponentialmed
6044272	Ramez Naam	Seattle	Author: Nexus / Crux / More Than Human / The Infinite Resource. Formerly a computer scientist at Microsoft. Interested in everything.
18705065	Humanity+	Global	Humanity+ is dedicated to promoting understanding, interest and participation in fields of emerging innovation that can radically benefit the human condition.
95661007	Kyle Munkittrick	Denver, NYC, San Fran	Bioethics: the unholy union of science, medicine, and philosophy. Blame no one but myself for what you find here.
16352993	Heather Knight		CMU Roboticist with a soft spot for interactive art & live robot performance: Founder @MarilynMonrobot, Director @robotfilmfest, Robo-Tech @RobotCombatSyFy!
28132585	Aaron Saenz		Writer for Singularity Hub, former Physics dude, Improv Comedian, Nomad
2443051	attilacsordas	Cambridge, UK	bioinformatician, EBI, regular Hadoop & R tinkerer, personal proteomics instigator, ex mitochondrial-stem cell biologist driven by healthy lifespan extension
15249166	Singularity Hub	NASA Moffett Field, CA	News network covering science, technology & the future of humanity. Follow @singularityu — Become HUB Member: http://t.co/wXGCjvICbk
16838443	KurzweilAINews	California/Mass	KurzweilAI (http://t.co/KD0Hq6D66p) is a newsletter/blog covering nano-bio-info-cogno-cosmic breakthroughs in accelerating intelligence
7445642	Chris Grayson	New York City / San Francisco	#Wearables / Advisor: http://t.co/ks23gJj30J / Prior ECD: http://t.co/Q4pSuEbFs4 / Events: http://t.co/bqLtGO3G5o & http://t.co/GrmHeG38Ir
16934772	tristanhambling	New Zealand	Tracking future, tech, nano, bio, neuro, info stuff, and anything new that scans past my event horizon. http://t.co/7aJFwAjkv7 also @futureseek
19004791	David Wood		Chair of London Futurists. Writer & consultant. PDA/smartphone pioneer. Symbian co-founder. Formerly at Psion and Accenture. Collaborative Transhumanist
15410587	Rod Furlan	Vancouver, BC	Artificial intelligence researcher, quant, Singularity University alum, Google Glass Explorer, serial autodidact, science lover & soon-to-be-robot
23115743	h+ Magazine	USA	h+ Magazine covers technological, scientific, and cultural trends that are changing human beings in fundamental ways.
743913	David Orban	New York, NY	CEO, Dotsub / Advisor & Faculty, Singularity University Analyzing and applying cycles of accelerating technological change. Flowing in wonderment.
19748200	Gizmag		I am a website about emerging technologies.
19722699	Popular Science	New York	Science and technology news from the future! Tweets from @RosePastore
138222776	Neurotechfuture	Boston, MA, USA	The future of life, humanity, and intelligence rests in the minds and hands of the innovators who envision, guide, and build it.
594718367	Grishin Robotics	New York	Everything about consumer robotics, connected devices & IoT. Published by the first robotics investment company. Founder - @dgrishin, feed editor - @Valery_Ka.
86626845	Eric Topol	La Jolla, CA	Cardiologist, geneticist, digital medicine aficionado, Editor-in-Chief, Medscape, author of The Creative Destruction of Medicine
15808647	MIT Tech Review	Cambridge, MA	We identify important new technologies deciphering their practical impact and revealing how they will change our lives.
101775759	Hizook.com	San Jose, USA	Robotics News for Academics & Professionals by Travis Deyle
44910688	Robot Magazine	Ridgefield, CT USA	The latest in hobby, science and consumer robotics.
16695266	ChiefRobot	Boston	Your daily dose of robots.
151648741	RobotoWear		Clothing for humans, inspired by robots. Robot t-shirts, hats, polos and hoodies.
87468736	Eric Tatro	Chicago, IL	Tweets about transhumanism, the singularity, AI, nanotech, biotech, robotics, life extension and human enhancement. All tweets and opinions are my own.
103516873	Willow Garage	Menlo Park, CA	Helping to revolutionize the world of personal robotics
6778032	Robert Oschler	Idaho	Artificial Intelligence and smart phone developer, currently focusing on speech recognition and natural language understanding applications and robotics.
18066713	robots_forever	Tokyo, Japan	Robot news, robotics research, combat and humanoid robot events, and other robot coverage from Japan.
8125922	Alexander Kruel	Germany	Transhumanist, atheist, vegetarian interested in math, programming, science fiction, science, language, philosophy, consciousness, the nature of reality...
22910080	Rob Spence Eyeborg	Toronto, Canada	We've built a wireless video camera eye. Tweets about privacy, cyborgs, prosthetics, eyepatches, Star Trek, The Bionic Man, and Augmented Reality.
7796912	Transhumanists	New York, NY	Singularity, Transhumanism, Artificial Intelligence, Human Enhancement, Stem Cells, Nanotechnology, Renewable Energy
15784353	Sensium		Revolutionary body monitoring for healthcare: wireless, intelligent, continuous, low-cost.
23116280	Popular Mechanics	New York City	The best in tech, science, aerospace, DIY and auto news. Customer Service: http://t.co/rYWTFWzg2R

Notes: Data extracted using the Twitter API in May 2014. Accounts can be freely accessed using https://twitter.com/intent/user?user_id=[insert here the twitter id]

Table 2. Network and community statistics over time

	2011, 2nd	2012, 1st	2012, 2nd	2013, 1st	2013, 2nd
N nodes	320	293	341	163	233
N edges	3,979	2,579	3,445	1,105	1,752
N communities	74	49	66	30	36
Max. community density	0.58	0.77	0.63	0.75	0.71
Max. nodes community	54	34	28	21	26

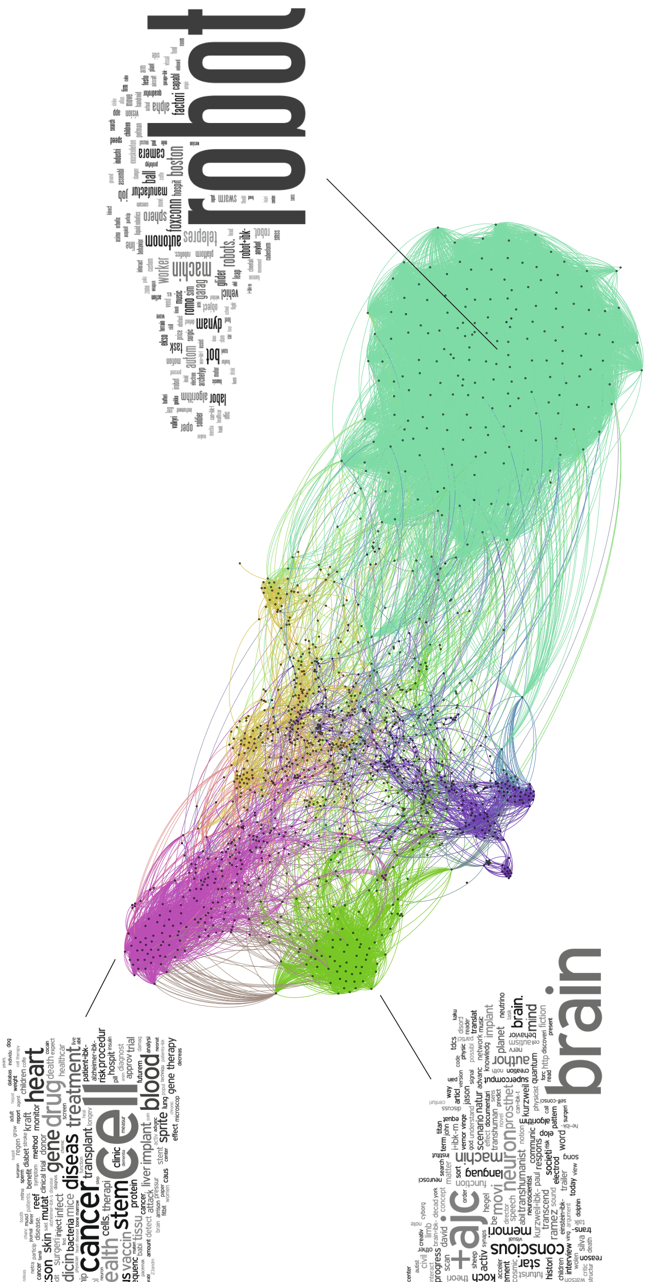
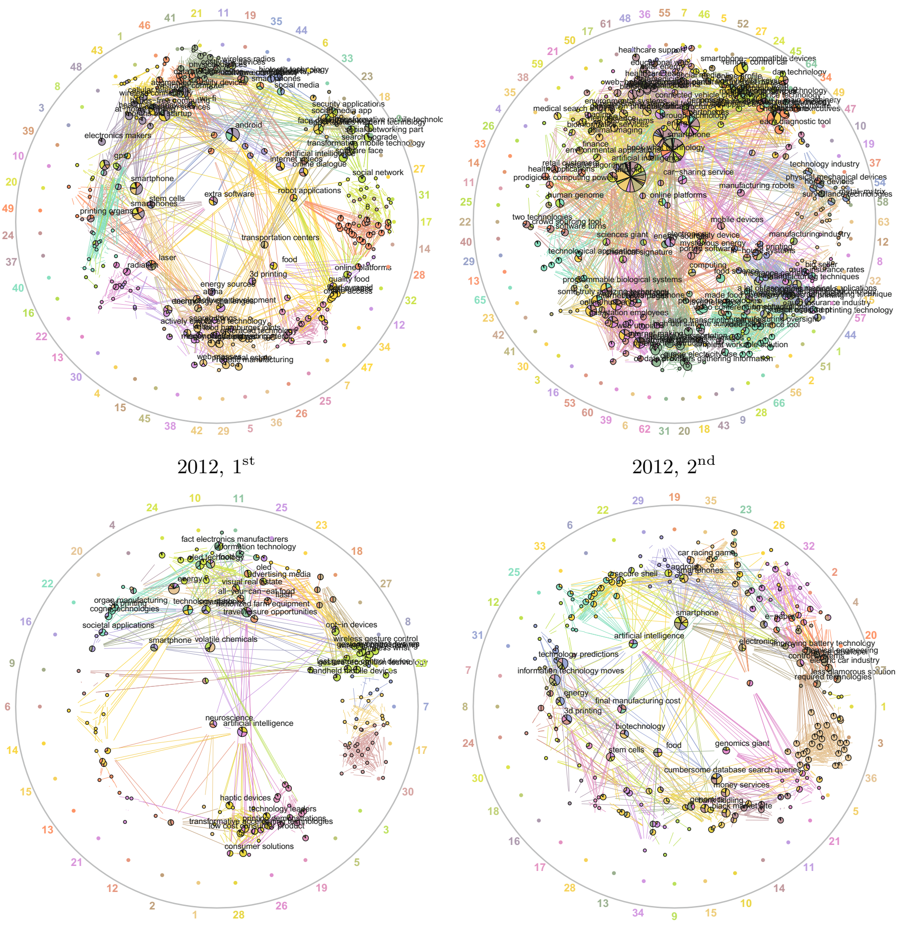


Fig. 7. Static Community Detection: Document similarity network of the whole corpus

Table 3. Network of Knowledge Fragments per Period after Overlapping Community Detection



2012, 1st

2012, 2nd

2013, 1st

2013, 2nd

Nodes are aligned according to their main community, represented by the number outside the circle. Node size is scaled by number of communities the node belongs to. Multi-community membership is also indicated by multiple node color

Table 4. Exemplary identified technological fields and their knowledge fragments

Biometrics & Law Enforcement

Ubiquity & Social Networks

Genomics

3D Printing

Nodes term representing the name of the technology fragment represented as tag-cloud. Size weighted by the nodes within community degree centrality.