

## Communities and central nodes in the mobility network of US inventors

István Márton Kiss<sup>1</sup>, Norbert Buzás<sup>1,2</sup>

<sup>1</sup>Knowledge Management Research Center, University of Szeged, Szeged, Hungary

<sup>2</sup>Department of Health Economics, Faculty of Medicine, University of Szeged, Szeged, Hungary

{kiss, buzas}@kmcenter.szte.hu

**Abstract.** This study presents the basic features and community structure of the US inventor mobility network between 1999 and 2010, based on an analysis of patent documents. Since mobile inventors have proved to be among the most effective knowledge mediator entities, this mobility network can be seen as a knowledge diffusion network among innovative companies. During the investigation, we identified the basic features of the network, such as short effective diameter and scale-free degree distributions, and we also demonstrated the central nodes, community structure, and hidden core of the network. Our results indicate that there is a small number of nodes that can effectively absorb knowledge from the network and pool it. We also find that this core mostly consists of IT and semiconductor companies as well as the largest universities in the USA.

**Keywords.** Intellectual Capital, Knowledge Management, Industrial Structure, Knowledge Economy, Company Research, Diffusion of Innovation.

### 1 Introduction

According to Gassmann and Bader (2006), innovation and technologies are responsible for half of the economic growth in developed and industrialized economies, while leading innovative companies realize more sustainable profits than imitators and trend followers. However, R&D costs increase steeply due to shortening innovation cycles and the growing number of imitators, with empirical evidence indicating a positive correlation between the success of a company and the strength of its intellectual property, R&D, and patent portfolio (Gassmann and Bader, 2006). In innovation-driven economies, it is an obvious scientific question how knowledge as an essential resource of R&D activities is diffused among business entities. It is difficult to find an ultimate indicator for company innovativeness, but the most frequently used measure for it is patenting because patents provide a monopoly for their owners for a space in the technology arena. Approximately two-thirds of the market value of large companies in the USA can be traced to intellectual property, especially patents and trademarks (Shapiro and Pham, 2007).

Patents grant their owners monopoly rights over novel technologies. However, ideas about research directions or developments can be diffused from firm to firm and from inventor to inventor without infringing patent rights. The platform for knowledge flows can be formal agreements, collaborations, and informal social ties among researchers

(Freeman, 1991; von Hippel, 1987; Powell and Grodal, 2005). Mobility is also one of the possible ways in which such flows are realized; however, as has been pointed out by Breschi and Lissoni (2009), it has been proved to be one of the most effective mediators of knowledge and plays the most important role in knowledge spillover. In this paper, we investigate the exchange of inventors among firms, research institutions, and universities in the United States from 1999 to 2010. We can track the migration of inventors and knowledge by analyzing patent documents and recording which inventor created and patented technology, where, and when. On the basis of network analysis (Wasserman and Faust, 2009), we explore the structural properties, the organizing principle, and the community structure of this network. Our study is based on the transformed Harvard University's Patent Network Dataverse (Lai et al., 2011). In the following section, we summarize the existing literature on the economic impacts of innovation and patents, innovation networks, inventor mobility, and knowledge diffusion. This is followed by a description of the database we used and the basic properties of the mobility network. Then, we highlight which firms, research institutions, or universities play the most important role in the cohesion and shaping of the US mobility network. In parallel, we investigate the community structure of the organizations and the network core consisting of nodes with the greatest network power.

## **2 Related work and research conception**

### **2.1 Patents in the economy**

From the corporate to the national level, performance greatly depends on the stock of knowledge which can produce intellectual assets, such as patents, software, and organizational structures. A large proportion of R&D expenditures covers the wages of highly educated and skilled white-collar labor, an investment which sees returns in the form of these assets (OECD, 2006).

IP-intensive industries accounted for about \$5.06 trillion in added value, or 34.8 percent of US gross domestic product, and for 40 million jobs in 2010. In the post-recession economy, employment in the IP-intensive industries is showing considerably faster growth than in non-IP industries (ESA and USPTO, 2012).

Intellectual properties are mainly embodied in patents and trademarks (Shapiro and Pham, 2007). According to the EPO, "A patent is a legal title granting the holder the executive right to make use of an invention for a limited area and time by stopping others from, among other things, making, using or selling it without authorization". This is an accurate definition; however, it only captures one aspect of patenting. In addition to the important function of protection and commercial exploitation of new technologies, patents provide important information for others on the market. According to Granstrand (1999), the contract between the patenting firm and society grants a temporary monopoly, on the one hand; on the other hand, it works as an information system that provides ideas about recent research directions for others. The patent system thus simultaneously stimulates invention and investment in R&D and public disclosure of technical information. Finally, it fosters technological progress and competition after patent protection has ceased (Granstrand, 1999).

## 2.2 Networks of innovation

The previous section shows the importance of intellectual property, i.e., a patent portfolio and the knowledge behind it, both for organizations and the economic system as well. However, knowledge and skills as sources of innovation and patents cannot be materialized; they are in the heads of the inventors and managers.

From the market participants' point of view, innovative ideas and knowledge can be considered a scarce resource. By pooling knowledge, firms become able to utilize complementary skills which are otherwise inaccessible and create new technologies or overcome resource restrictions (Penrose, 1959). Networks of innovation are the spaces where common knowledge creation and utilization take place among firms. The sharing of firms' resources leads to a decrease in the risk that the development, introduction, or application of a new technology entails (Freeman, 1991). Furthermore, it creates an opportunity to combine and access knowledge which would otherwise be impossible for individual organizations to attain (Freeman, 1991; Knell, 2011). Formal and informal networks play an equal role in the creation, transfer, and absorption of new knowledge and technology (Powell and Grodal, 2005). Formal networks are established based on collaborative innovation, in which case organizations utilize their resources in a cooperative manner to achieve their aims (Freeman, 1991). By today, these formal networks have developed into global cooperative systems and serve common knowledge-creating goals in several ways, e.g., research collaboration, joint ventures, technical assistance programs, and technological licensing agreements (Knell, 2011). Informal network ties are seen as undeclared platforms for knowledge flow, and they function via various social interactions among the company's employees; in many cases, they lay the groundwork for the development of formal network ties or enhance their sustainment (Powell and Grodal, 2005). Using the steel mini-mill industry in the United States as an example, von Hippel (1987) demonstrates the way in which the personal network among engineers and the norms of the professional community facilitated a flow of technical knowledge among rival companies.

Although networks can produce a significant knowledge surplus, companies have to face the geographical (Jaffe et al., 1993) and technological (Rosenkopf and Almeida, 2003) limits of the accessibility of external sources of knowledge. The path dependency characteristic among companies derives from the fact that the knowledge search often happens in the local network, and thus it is only the knowledge capital of geographically proximate companies that can be utilized. On the other hand, the absorption capacity of a particular company is basically determined by its technological portfolio; therefore, organizations are not able to utilize any of the complementary sources of knowledge (Rosenkopf and Almeida, 2003).

In the present study, we consider knowledge flow via researcher mobility a special case of informal (and, to a certain extent, formal) network relationships because with mobility the company utilizes external skills, usually without any formal agreement. However, it is important to note that the flow of researchers among companies is possible in the case of different strategic alliances, which makes such phenomena the outcome of formal agreements.

### **2.3 The impact of mobility on organizations and inventors**

Rosenkopf and Almeida (2003) found that inventor mobility significantly fosters inter-organization knowledge transfer, while such an effect could not be detected in the case of strategic alliances. This result stresses the importance of mobility for knowledge transfer. Other studies have also investigated the impact of mobility on organizations in many ways. Employee mobility and employee enticement from rival companies generate a serious existential risk for the parent company; on the other hand, they create important developmental potential for the progeny company (Phillips, 2002). Newly established and less embedded companies are able to increase their innovative capacity and overcome resource restrictions by attracting talent (Rao and Drazin, 2002). The reason why mobility can have such an impact on knowledge flow is that during an employee move employees take not only their human capital away but also their social capital, along with company routine and practices (Pennings and Wezel, 2007). Nevertheless, the utilization of knowledge and routines strongly depends on the structure of the new company since operating practice is not an individual-level task but a group-level one (Phillips, 2002; Rao and Drazin, 2002; Pennings and Wezel, 2007). Breschi and Lissoni (2006) argue that it is social distance – direct and indirect ties among inventors – that is of the greatest importance in knowledge spillover. The greater and more diverse a researcher's social capital is, the more easily he or she can access external or new knowledge through personal contacts. These crucial interpersonal links are mostly established by mobile inventors who move from firm to firm and sell their brainpower to various business entities (Breschi and Lissoni, 2006; Moen, 2000; Lamoreaux and Sokoloff, 2009). This increased social capital among mobile inventors is one of the key factors in knowledge transfer (Breschi and Lissoni, 2009).

Moreover, mobility influences researcher productivity as well. The greater the size of the organization, the more productive its inventors are due to the availability of a significantly greater number of resources for R&D projects and the lower the perceived risk of company failure (Hoisl, 2007; Kim et al., 2004). Mobile researchers are more productive than their immobile fellows; nevertheless, increasing productivity decreases willingness for mobility due to effective research practice (Hoisl, 2007). Researcher productivity also increases if mobility happens in the direction of a less path-dependent organization and if the inventor possesses unique knowledge compared to the existing knowledge base of the particular company (Song et al., 2003).

### **2.4 Patent analysis methods in knowledge diffusion studies**

Various studies using a variety of methods have relied on patent data to investigate knowledge spillover and the knowledge diffusion process. Numerous studies focus on patent co-citations, which represent the links between older (cited patent) and novel (citing patent) inventions and can therefore trace how existing knowledge affects new knowledge. A pioneer investigation in this field was carried out by Jaffe et al. (1993) in the area of patent co-citation and local knowledge spillover. The authors revealed the importance of localization in knowledge flows by finding evidence that with co-cited patents the cited patent is more likely to come from the same state and from the same metropolitan area than randomly selected patents in the same field of technology.

Wang et al. (2011) identified fields of technology among leading organizations by analyzing patent co-citations among Fortune 500 companies. Petruzzelli et al. (2015) also measured the inter- and intra-firm and -industry impact of biotechnology innovations through patent co-citations.

Xuefeng et al. (2012) investigated Chinese and American co-patenting companies based on USPTO datasets. Co-patenting in this sense denotes the sharing of developmental expenditures and rights of a patent between firms from the USA and China as well.

Breschi and Lissoni (2009) analyzed a co-invention network of inventors to investigate knowledge flow processes. Co-invention occurs when inventors jointly file at least one patent application. In that study, mobile inventors are defined as individuals who move across companies and file patent applications with different assignees. The authors found that co-citations of patents occur more frequently when inventors are closer to each other in the co-invention network. These close ties between inventors are established by mobile inventors; they can therefore be seen as key figures in the knowledge spillover process (Breschi and Lissoni, 2009). This approach is incorporated into the present study. Although we describe the knowledge flow network based only on inventor mobility, we do not measure the co-citation and co-invention properties of the network in this study.

## **2.5 Concept of the mobility network of US inventors**

Since inventor mobility has proved to be a key factor in inter-organizational knowledge transfer, tracing movements enables us to draw up the organization-level mobility network among inventors. The present study aims to detect and investigate researcher mobility and the network of inter-organizational knowledge flow through an examination of patent documents. Our approach identifies the mobility of inventors who file patents with changes in assignees over time. Thus, the network that we aim to investigate is that of companies with patent rights, while the knowledge flow among them is demonstrated through the continuous patenting activity of mobile inventors. Movements identified in this way do not necessarily mean that the researcher has physically changed jobs. However, it can be considered mobility in the sense that he or she utilizes his or her mental capacity and technological and organizational knowledge acquired at another company; hence, we see complementary skills and a flow of knowledge among organizations. It is important to emphasize the fact that this network is only one of the possible approaches to the entire knowledge flow among organizations. In this case, manager movements cannot be detected, nor can those cases when ideas flow via interpersonal relations such as through advice from researchers working for different companies. It is also impossible to detect knowledge flow which is generated by the mobility of inventors whose innovation is not manifested in patents. Furthermore, the flow of knowledge based on existing patents also remains invisible for this method. However, as we have seen in the introduction, protecting valuable knowledge with patents bears great economic significance; thus, researchers that create patents are considered key figures in knowledge creation, and the knowledge flow generated by their mobility is seen as a substantial aspect of the total knowledge transfer.

We believe that our study represents the first attempt to describe a network of inventor

mobility in the United States and, thereby, a network of knowledge flow among organizations. Therefore, the first aim of our study is to explore the basic properties of this network. Our second aim is to investigate the community structure of the network, seeking answers to the following questions: Which organizations tend to interact with each other and why? What is the organizing principle of the communities? How do path dependency and technological distance influence the flow of knowledge and which organizations are able to overcome them?

### 3 Dataset and methods

#### 3.1 Source data and transformation

Our investigation is based on Harvard University's Patent Network Dataverse (Lai et al., 2011). This is a cleaned and disambiguated dataset based on the NBER database and the US Patent and Trademark Office weekly publications from 1975 to 2010. In their study, the authors propose a disambiguation algorithm which is an application and further development of the existing *Author-ity* approach developed by Torvik and Smalheiser (2009). The primary input data of the disambiguation algorithm is a cleaned and formatted version of the data sources noted above. The processed dataset consists of the name of the inventor, the patent, the assignee of the patent, the technology class of the patent, and the location of the inventor. This data preparation process defines inventor-patent instances, which form the units of analysis. The algorithm applies a comparison function which returns the similarity vector of inventor-patent instance pairs. The dimensions of the similarity vector, and the scale of the similarity are the following: first name [0..4], middle name [0..3], last name [0..5], co-author [0..6], technology classes [0..4], assignee [0..6], and location [0..5]. Zero values were assigned when the variable values were completely different and maximum values were assigned when they were identical in the case of inventor-patent instance pairs. To decide whether the given inventor-patent instance pairs match or not, the algorithm uses an iterative blocking scheme. As a result of the classification process in the disambiguated database, the inventors receive unique ID numbers to distinguish them from others with very similar attributes. The primary aim of this dataset is to investigate the co-authorship and collaboration networks among inventors who have registered patents in the USA. The network files are split into three-year intervals, so the ties in particular files only contain collaborations during that specific time interval.

In the original network files, the nodes represent the inventors. Every inventor has the following attributes: *inventor ID*, *assignee*, *assignee ID*, *first name*, *last name*, *city*, *state*, *country*, and *patents*. If some of the attributes of the inventors change, one can trace them as a difference in attributes for the same inventor among network files applied to different time intervals. For example, if an inventor works for IBM and takes a patent out on some technology between 1999 and 2001, his or her assignee attribute in the network file will be "IBM". If this inventor later moves from IBM to Microsoft, his or her attribute in the corresponding network file will change to "Microsoft". We used the 1999 to 2010 time interval, because a longer interval would present us with older, obsolete edges in our network, which may represent knowledge flows which are no longer relevant. Although 1999 to 2002 seems remote from 2010, that period

coincided with the dot-com boom, which indicated the growing importance of the IT industry in the economy. We therefore considered this interval an important turning point in technology development which created an industrial environment that is relevant even in our days.

The differences in the inventors' attributes among the network files gave us the idea of transforming the collaboration network to the network of firms, where the edges represent the migration of inventors detected through temporal changes of the *assignee* attribute in the network data. This network is directed, meaning we can name the source and target of the knowledge flow represented by the migration. The attributes of the nodes in our network are the following: *assignee*, *assignee ID*, *community*, *in-degree*, *out-degree*, *betweenness centrality*, and *multi-edge node pairs*. The assignee and assignee ID variables signify the name and the unique identifier of organizations that hold the patent rights. Inventors with no assignee have been excluded from our network. The other attributes of the nodes will be described later. On the edges of the network, we defined one attribute, *weight*. Weight denotes the number of mobile inventors that move from the source company to the target company. We wanted to focus on US inventors; hence, we viewed mobility as knowledge flow only in the case of US residents. This filtering was based on the country attribute of the inventor nodes from the Patent Network Dataverse. Although international innovation networks play an important role in knowledge creation (Knell, 2011), our assumption is that foreign inventors with patent applications in the USA tend to work for multinational corporations, while smaller foreign firms cannot file their patents abroad. If foreign inventors were present in the network, it would over represent multinational corporations. In spite of the fact that US firms hire inventors with foreign residency, we assume that restricting the mobility to US residents gives us a better and less biased picture of knowledge flow patterns in the USA.

### 3.2 Definitions and variables

In the following part of this section, we explain the basic definitions, which are crucial for an understanding of our results.

A network  $G$  consists of a set of nodes,  $N=\{n_1, n_2, \dots, n_g\}$ , and a set of edges,  $E=\{e_1, e_2, \dots, e_l\}$ . Nodes represent the entities we examine, while edges indicate the presence or absence of some relation between them (Wasserman and Faust, 2009). In our case, as we have already noted, companies, institutions, and universities are the nodes, and the migration of mobile inventors among them are the edges that we consider as knowledge flow. Our network is directed; therefore, we can count in-degrees and out-degrees on the nodes. In-degree is how many companies sent inventors to a particular node, and out-degree is how many companies received inventors from it.

Geodesic distance: the length of the shortest path between two nodes in the network. If a company shares an edge with another company, the geodesic distance between them is one. If a company has no edge with the other company, but they have a common neighbor, the geodesic distance is two. Breschi and Lissoni (2009) showed that short geodesic distance between inventors greatly increases the probability of knowledge flows. We also assume that the closer the organizations are in the mobility network, the more chance they have to reach each other's knowledge base.

**Network diameter:** the shortest path between the two furthest nodes; in other words, this is the longest geodesic distance in the network. Large network diameter may present a less effective feature of knowledge diffusion since there are nodes which are inaccessibly far away from others. This can happen when there is almost no knowledge exchange among industries or among strategic alliances.

**Characteristic path length:** the average geodesic distance between any nodes in the network. This is a very important feature of the mobility network because the smaller the characteristic path length is, the greater the chance for knowledge diffusion among nodes, and the more effective the network is.

**Betweenness centrality:** one of the most often used measures of the centrality indices besides the in-degree and out-degree values. Betweenness centrality indicates the proportion of pathways passing through the given node. If the betweenness centrality is high, the node serves as a mediator of information in the network (Freeman, 1979). In the case of the mobility network, nodes with large betweenness centrality values represent the backbone of knowledge diffusion. An enormous amount of knowledge is accessible for and is mediated by such organizations.

**Multi-edge node pairs:** the number of mutually connected neighbors. In our case, the number of neighbors who also send and receive knowledge from a given node. Organizations with multi-edge connections use each other's knowledge base. The parts of the network where multi-edge connections frequently occur can be seen knowledge pools.

**Communities (or clusters):** densely connected parts of the network, in which the nodes probably share some common property or play similar roles (Fortunato, 2010). Communities are densely connected internally and sparsely connected externally. In our case, we expect that communities will be based on industrial similarities, where nodes from the same industry exchange knowledge with each other more often than with other communities based on another industry. The community attribute of a node indicates the number of the community to which the organization belongs.

**Modularity:** a measure of how densely communities are connected to each other and how appropriate the network is to find communities in it. The modularity score represents the strength of the communities. The stronger the community structure of a network, the more edges exist inside those communities and the fewer exist outside. Strong communities in the case of a mobility network means that the potential circle of organizations is small for an inventor who would like to move from one to another. It also indicates the overall path dependency of the organizations since it restricts the circle of companies with which they can exchange knowledge. The modularity value varies between 0 and 1. The 0 value indicates a network without communities, and 1 characterizes graphs with perfect communities (where edges exist only inside the communities). The modularity score of real world-based networks often varies between 0.3 and 0.7 (Fortunato, 2010; Newman and Girvan, 2004).

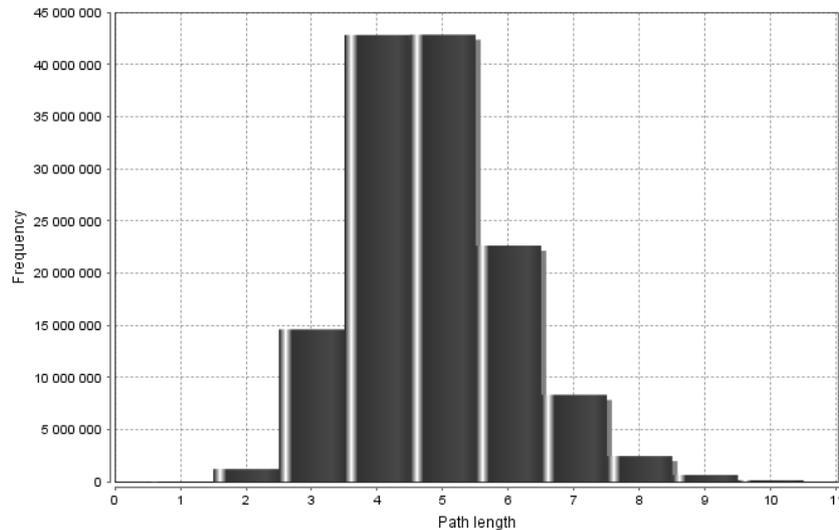
## **4 Basic features, network structure, and centralities**

### **4.1 Network coherency and statistics**

On the time horizon of the investigation, we found 28,695 companies, universities, and institutions involved in inventor mobility. Among these, there exist 50,170 paths, where inventors move from one to the other. This network of inter-firm inventor mobility is an exceedingly sparse network with only a 0.0001 network density value. In other words, only 0.01% of the possible ways exist among the nodes. Network density value would be 1 if all the firms in the network were connected to all the others. On the pathways, we detected 83,640 inventors who changed their position between 1999 and 2010.

The biggest component in the mobility network contains a large number of nodes. 20,998 organizations out of 28,697 are connected to a giant component. Next to 73% of the nodes, this coherent subgraph contains 45,707 edges, 91% of all of them. This interconnected component is the core of the knowledge flow. Many of the firms are linked to this network by almost all of the mobility paths. This is the main platform where the entities compete for inventors' knowledge. If a company is tied to this network by a mobile inventor, it means that it transmitted or received knowledge from this common platform for knowledge exchange. Due to this connectedness, knowledge and experience can be accumulated in the network, since it is available for the interconnected nodes. We can consider this knowledge system as the space for knowledge recombination and for the acquisition of social capital.

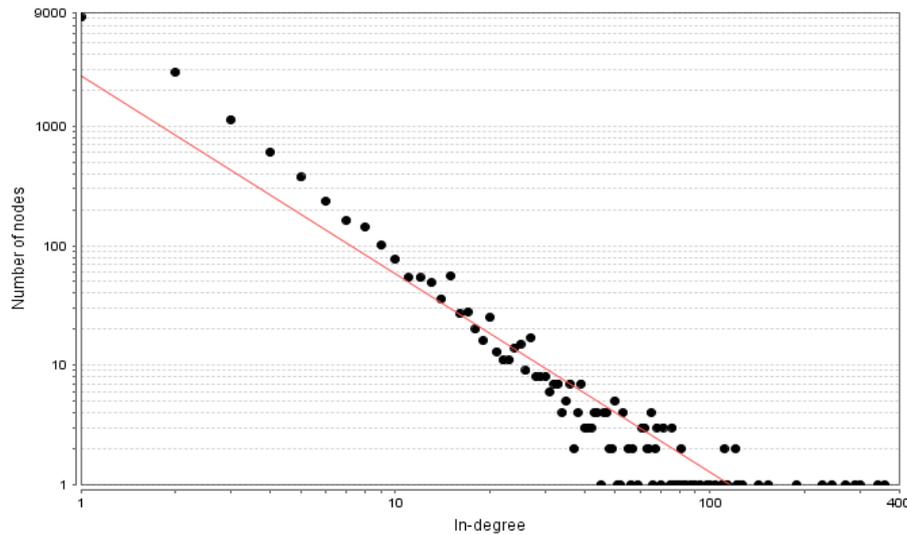
The network diameter is 17, so seventeen hops separate the furthest companies from each other. Nonetheless, the characteristic path length is just 4.81, which means the average distance between any two firms in the network is slightly less than five paths (Fig. 1). This indicates a small world property, where despite the sparseness and size of the network, the indirect links between firms are quite short (Watts and Strogatz, 1998). Firms competing for the same inventors are similar in their knowledge needs. As we have noted, the closer firms are in the mobility graph, the more similar the knowledge required for their innovation practice. Therefore, it is striking that the characteristic path length is less than five among patenting firms in the USA, regardless of the location, industry, or size of the companies. This indicates that there must be a set of companies or technology fields which are less path-dependent with their considerable absorption capacity and are responsible for making the network with such diverse nodes so "small".



**Fig. 1.** The Frequency Distribution of the Shortest Paths in the Network

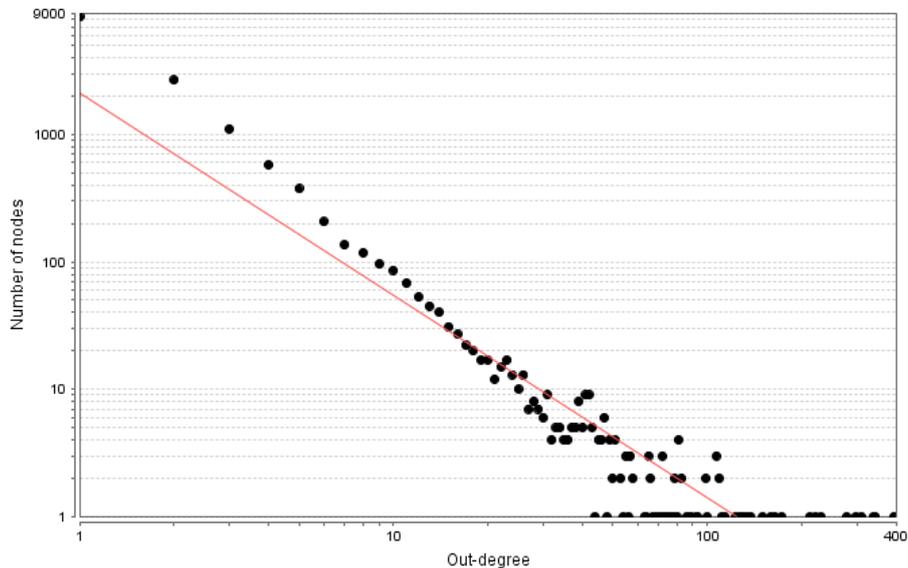
#### **4.2 Degree distributions and correlation of network variables**

In our first attempt to unravel the mystery, we investigate centrality indices for the firms. One of the most important traits of a network is the degree distribution or, in other words, the distribution of the edges among the nodes. This is a highly informative property of the graph; hence, it can ultimately describe the structure of the network. If the degree distribution follows a normal or Gauss distribution, the nodes in the network will have a typical value of edges connecting them to others. In the case of such a distribution, extremely large values are rare or not present and the average degree characterizes the biggest proportion of the nodes in the network. However, if the degrees follow a power-law distribution (i.e., the network is scale-free), there is no characteristic value at the edges. Many nodes have just a few links – and most have only one link – to other nodes, while a small but considerable group of them has a high or extremely high number of direct paths to others. The power-law distribution can be described with its exponent, which is -1.656 for in-degrees and -1.585 for out-degrees in our case. In classical studies of networks, for example, the World Wide Web, the exponents have been between -2 and -2.5 (e.g., Barabási, 2011). Where the exponent is closer to zero, there is a bigger chance of there being nodes with values higher than one.



**Fig. 2. In-degree Distribution of the Nodes on a Log-log Scale**

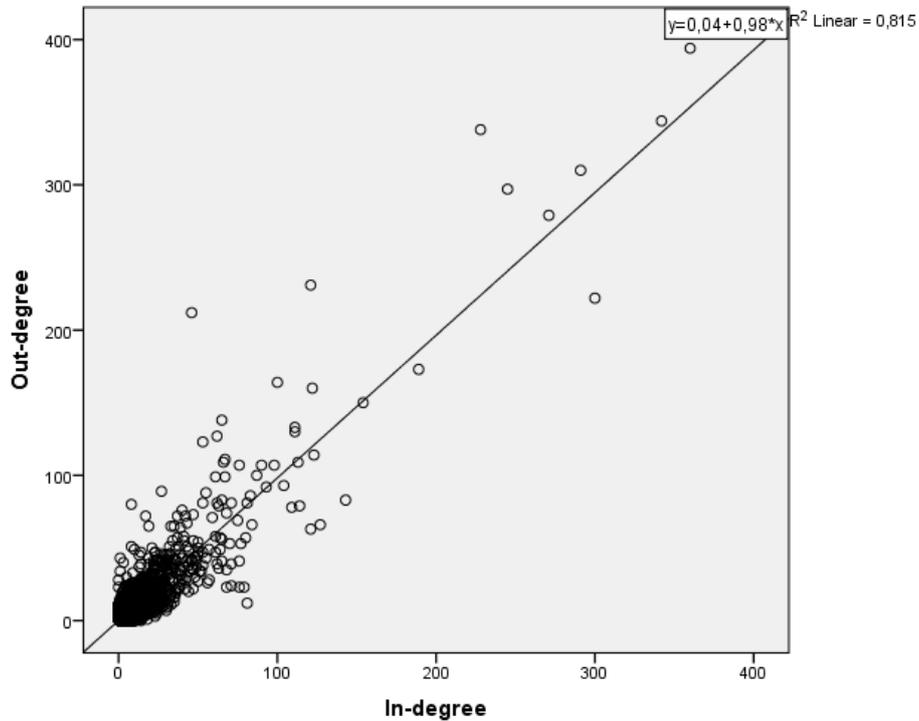
As we can see in Fig. 2, the in-degree distribution of the mobility network is unbalanced. Instead of a characteristic or average edge value, many nodes (8153) have managed to obtain only one edge, but a small minority can absorb connections more frequently. Some of the nodes have more than 100 acquaintance organizations (the maximum is 360) that transmit knowledge to them directly. These powerful companies in the network are the greatest beneficiaries of the knowledge flow. Due to the fact that they are the most attractive for inventors, they represent the core of the competences and accumulated experience in the network. It is likely that the knowledge and social capital of the researcher at these centers grow multiple times compared to those at the peripheries. It seems these nodes serve as “black holes” in the knowledge network and can absorb a great amount of – and maybe more diverse – knowledge from it. The next question is whether they put something back into the common pot or not?



**Fig. 3.** Out-degree Distribution of the Nodes on a Log-log Scale

The function of out-degree distribution (Fig. 3) is quite similar to that of the previous one. Hence, there are firms, institutions, or universities in the network that do the opposite of the knowledge “black holes”. The tail of the distribution, namely, the nodes with the greatest out-degrees, transmit knowledge to the network and continuously lose their researchers and knowledge base. The greatest giver in this case had to let its patenting inventors go to 394 different organizations. How could a firm survive such great losses? Maybe these great knowledge-providers went bankrupt or downsized.

As we can see in Fig. 4, there is an undoubtedly strong positive linear correlation between the in- and out-degrees of the nodes. In this sense, both the great knowledge takers and givers are the same entities. These strong nodes simultaneously serve brain drain and brain gain functions in the network. They function as cores of knowledge creation, accumulation, and distribution. These are junctions, the most frequent platforms where inventors with various corporate histories can meet and increase their professional skills and social capital. However, these organizations also mediate knowledge to the network through departing researchers with their rich human and social capital. This high equality of knowledge absorption and loss is also a feature of the less frequented nodes. Due to this strong linearity in the degree distributions, there is a low number and extent of outliers. A further interesting property of the network is the fact that the slope of the curve is very close to 1. This indicates that in general the number of organizations from which the company can recruit inventors tends to be equal to the number of organizations which can recruit researchers from that same company.



**Fig. 4.** Linear Regression of the In-degree and Out-degree variables. The regression is significant at the 0.01 level.

Although there are a few cores that frequently absorb and transmit knowledge, it is doubtful how effectively they do so. It is possible that they simply recruit from and hand over inventors and knowledge to a clique of nodes, and therefore the knowledge of this subset of companies circulates within a relative small space. Table 1 shows the correlation matrix of the main centrality indices and the variable of multi-edge node pairs, which is the measure of mutually connected neighbors of any node.

**Table 1.** Person Correlation Matrix. All the correlations are significant at the 0.01 level.

Person Correlation Matrix				
	In-degree	Out-degree	Betweenness Centrality	Multi-Edge Node Pairs
In-degree	1			
Out-degree	0.903	1		
Betweenness Centrality	0.936	0.94	1	
Multi-Edge Node Pairs	0.936	0.926	0.943	1

According to the correlation matrix, betweenness centrality and multi-edge node pairs also show a surprisingly strong correlation with in-degree and out-degree scores. The betweenness centrality score indicates the proportion of pathways passing through the particular node. If the betweenness centrality is high, the node serves as a mediator of information in the network. The strong correlation between the degree centrality variables and betweenness shows that the cores of the mobility have far-reaching ties in the network, therefore functioning as the backbone of knowledge diffusion. On the other hand, the correlation between the three centrality indices and the multi-edge node pairs variable means that the higher number of edges a given company has, the more likely it is to send and receive knowledge from entities of similar size.

#### 4.3 The top 30 central nodes in the network

Table 2 contains a list of the top 30 central nodes with rank scores of in-degree, out-degree, and betweenness centrality variables. In this case, the value “1” represents the highest value of a given variable. The table is sorted by in-degree ranks. Although the identity of the companies outlines the fact that the most IP-intensive industries are pharmaceuticals, communications equipment, and semiconductors (Shapiro and Pham, 2007), it is the dominance of IT and IT-related sectors that emerges from the data, while members of the pharmaceutical industry are at the back of the pack. It is a pleasant surprise that the four universities managed to make their way to the top 30. However, it is not surprising that they form a small elite of the world’s leading universities. In the upper right quartile in Fig. 4, we can see seven dots separating them from the others. These are the most powerful nodes in the network. With one exception, they are the top-ranked entities in all three categories.

**Table 2.** The Most Central Nodes in the Network Sorted by In-degree Ranks

<b>The Most Central Nodes In The Network Sorted By In-degree Ranks</b>			
Name	In-degree Rank Number	Out-degree Rank Number	Betweenness Rank Number
IBM	1	1	1
US NAVY	2	2	2
MICROSOFT	3	8	5
INTEL	4	4	4
GENERAL ELECTRIC	5	6	7
HEWLETT PACKARD	6	5	3
UNIVERSITY OF CALIFORNIA	7	3	6
CISCO TECHNOLOGY	8	10	10
HONEYWELL INTERNATIONAL	9	13	8
BROADCOM	10	34.5	27
APPLE	11	53.5	25
3M INNOVATIVE PROPERTIES	12	19	11
TEXAS INSTRUMENTS	13	12	12

MOTOROLA	14.5	7	9
QUALCOMM	14.5	59	36
BOEING	16	41.5	15
E I DU PONT	17	21.5	13
APPLIED MATERIALS	18.5	15	18
PROCTER & GAMBLE	18.5	16	14
BOSTON SCIENTIFIC SCIMED	20	43	22
BRISTOL MYERS	21	29	26
MIT	22	11	16
UNIVERSITY OF TEXAS	23	24	20
MICRON TECHNOLOGY	24	30	38
STANFORD UNIVERSITY	25	24	32
KODAK	26	26	21
WYETH	27	53.5	43
MEDTRONIC	28	33	29
ORACLE	29.5	37.5	41
GENERAL MOTORS	29.5	528	156

## 5 Community structure and the core of the network

In the previous section, we highlighted the basic properties of the inventors' mobility network. However, the main organizing principle of the network ties has still remained hidden! From whose knowledge base do companies tend to absorb external knowledge and why?

### 5.1 Modularity score and community structure of the mobility network

In this section, we investigate whether the network contains communities in which organizations tend to develop more ties with their community members than with outer nodes. Community structure can provide us with a better understanding of the organizing principle of the knowledge flow network. In order to identify these communities, we used the *fastgreedy* algorithm proposed by Clauset et al. (2004). Based on this algorithm, we found that the modularity value of the network is 0.72, which suggests an exceedingly strong community structure. The modularity value in real world networks very rarely reaches such a high modularity score. This indicates that many of the nodes keep in touch with restricted types of organizations with whom they are willing or able to exchange inventors. Presumably, this is the system-level outcome of the path dependency of individual organizations. When organizations absorb knowledge from others, the closer that knowledge is to their existing knowledge base, the easier it is to successfully achieve the absorption. Therefore, the strong cluster structure and high modularity value stress the fact that many of the nodes in the network

must be highly path-dependent with a very restricted possible circle of knowledge absorption. It also predicts the importance of industrial differences in the knowledge flow system.

Although the fastgreedy algorithm found 268 clusters, the six largest of them contain more than 50% of the nodes. These six clusters were examined in this study for clear interpretability. We analyzed them as independent graphs; hence, only the intra-community edges are present in the statistics. Table 3 contains the number of nodes (Nodes), number of edges (Edges), characteristic path length (Cp), and network diameter (Nd) values of the communities as well as a list of the five most central companies per community.

All the communities can be seen as significantly separated subgraphs of the whole mobility graph. Surprisingly, their basic properties are very similar to the features of the parent graph. While not presented in the table, these subgraphs are also scale-free networks with power-law degree distribution. On the other hand, these communities can easily be characterized by their central nodes. Community No. 1 is led by pharmaceutical companies, the second consists of firms in the household, fashion, and cosmetics industries, and the third is made up of medical technology-related companies. The IT and communication industries, regardless of the hardware or software feature, form the biggest community with a very short diameter and characteristic path length. In the fifth cluster, we can find the US Navy and the prestigious universities, while the sixth cluster consists of the semiconductor industry. It seems that the main organizing principle of the network is the industrial structure of the economy. This is not surprising since every industrial area has specific knowledge needs. Although it is more striking that despite the strong impact of industrial knowledge needs on the network structure, the whole mobility network has almost as short a diameter and characteristic path lengths as the clusters in it do. It seems that, despite the strong community structure, there are companies which can overcome path dependency and that this ability increases the effectiveness of knowledge transfer on the system level.

**Table 3.** The Six Biggest Communities in the Mobility Network

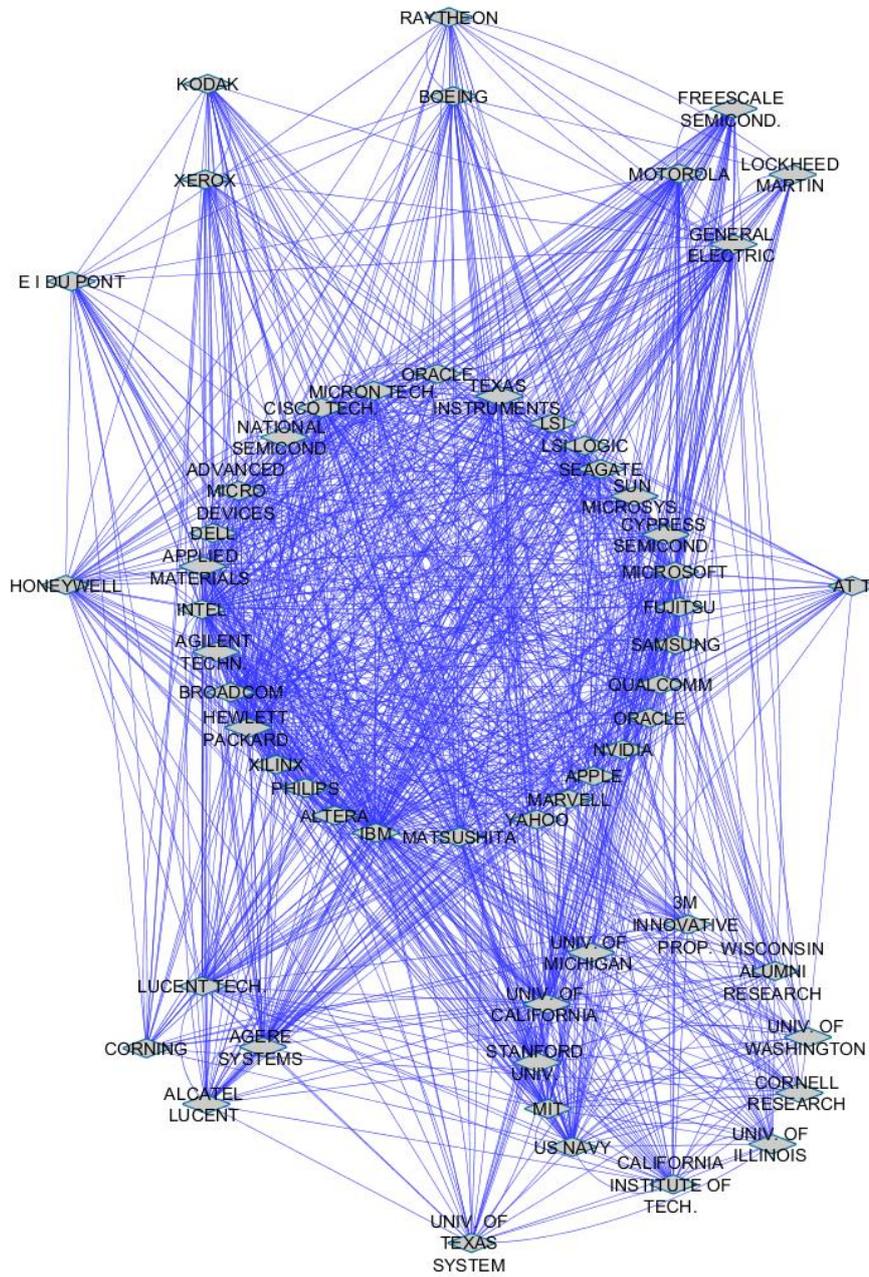
<b>The Six Biggest Communities in the Mobility Network</b>						
<b>Community No. 1</b>		<b>Community No. 2</b>		<b>Community No. 3</b>		
Nodes	1626	Nodes	2322	Nodes	1437	
Edges	3351	Edges	3600	Edges	2331	
Cp	4.46	Cp	5.49	Cp	5.169	
Nd	11	Nd	15	Nd	16	
Central Nodes	BRISTOL MYERS SQUIBB	Central Nodes	PROCTER AND GAMBLE	Central Nodes	BOSTON SCIENTIFIC SCIMED	
	PFIZER		SUNBEAM		MEDTRONIC	
	WYETH		S C JOHNSON SON		ETHICON ENDO SURGERY	
	MERCK		NIKE		SCIMED LIFE SYSTEMS	
	AMGEN		PEPSICO		ETHICON	

Community No. 4		Community No. 5		Community No. 6	
Nodes	2998	Nodes	2252	Nodes	922
Edges	6354	Edges	3658	Edges	1221
Cp	3.8	Cp	4.06	Cp	3.86
Nd	10	Nd	11	Nd	12
Central Nodes	IBM	Central Nodes	US NAVY	Central Nodes	GENERAL ELECTRIC
	INTEL		UNIVERSITY OF CALIFORNIA		MOTOROLA
	MICROSOFT		MIT		LOCKHEED MARTIN
	HEWLETT PACKARD		3M INNOVATIVE PROPERTIES		SIEMENS MEDICAL SOLUTIONS
	CISCO TECH.		CALIFORNIA INSTITUTE OF TECH.		FREESCALE SEMICONDUCTOR

As we can see, the most central nodes in the whole network are the leaders of the communities as well. Moreover, according to the results in Table 1, which have been examined previously, the more central a particular node, the more mutual ties it has with other central nodes. Therefore, next to the strong cluster structure of the network, community leaders must have frequent contacts not just with their community members but potentially with other community leaders as well. It is possible that despite the strong modularity of the network, network leaders have formed a sort of elite club with dense and mutual ties.

## 5.2 The core of the mobility network

To examine the core of the mobility network postulated above, we used the *k-core algorithm* proposed by Batagelj and Zaversnik (2002), which finds the most densely connected and interconnected subgraph in the network. Fig. 5 shows a graphic representation of the result attained with this algorithm. The core of the mobility network consists of 58 organizations with at least 23 ties to other core members. The total number of edges in this subnetwork is 1112, which indicates high density since the maximum number of possible edges is 3306. The circular layout and spatial proximity help to identify common community memberships. Parallel to the size of the communities, the IT industry – the biggest circle of nodes – is by far the most overrepresented cluster in the core with 31 companies. Besides hardware and software companies, the dominant community in the core is that of the universities with 10 institutions of higher education, supplemented by the US Navy and 3M Properties. Semiconductor companies are also frequently represented in the core graph, but the pharmaceutical industry is not included.



**Fig. 5.** The core of the US inventor mobility network with 58 organizations and 1112 ties

## 6 Conclusions and avenues for further research

### 6.1 Main conclusions

Our study has aimed to explore the mobility network of US inventors. Based on the transformed Harvard University's Patent Network Dataverse, the present paper has highlighted the knowledge flow network among US firms, institutions, and universities tracked by mobile inventors from 1999 to 2010.

Studies conducted so far in this area have investigated specific economic segments, such as the semiconductor and pharmaceutical industries. To the best of our knowledge, this is the first attempt to examine the inventor mobility phenomenon on a nationwide, multi-industrial level. Our findings may therefore provide a wider perspective on the inter-organizational knowledge flow processes via inventor mobility, such as inter-industry knowledge transfer, system level innovation communities, and identification of community and network leader organizations. Consistent with previous studies, our results underline the importance of knowledge pooling as well as informal innovation networks and researcher mobility in knowledge transfer. We have also pointed to the strong impact of path dependency on inventor mobility, but at the same time we found evidence of a previously hidden coherent network, where some of the nodes which manage to overcome path dependency can achieve significant advantages in knowledge absorption.

We have demonstrated that this system is a scale-free network with a short effective diameter and small characteristic path length, where central nodes simultaneously engage in both brain drain and brain gain functions. Giant communities in the graph reflect the set of the most frequent patenting industries, plus the unique cluster of the Navy and universities. Though the community structure is extremely strong in this network, community leader entities hold important intra- and inter-cluster ties as well.

According to recent outcomes, three main lessons can be learned from the core structure of the network. First, the IT and semiconductor industries have emerged as organizers of the mobility network with a high absorption capacity and low path dependency. The leaders in these industries are the far more central and interconnected nodes in the network. Their high betweenness enables these companies to search effectively in the network. The literature (Breschi and Lissoni, 2009) has highlighted the fact that mobile inventors' personal networks play a significant role in knowledge spillover and knowledge diffusion. With their far-reaching ties, these central firms are able to absorb knowledge from other communities and expand the search horizon to the social ties of new employees. On the other hand, scientists who leave these organizations bring valuable intellectual and social capital to periphery firms and mediate knowledge to them. The knowledge-emitting function of the core is supported by the fact that inventors were obtained from the 58 organizations noted above as follows: they were received by 1210 nodes out of a total 8153 nodes with 1 in-degree, 705 nodes out of a total 2833 nodes with 2 in-degrees, and 436 nodes out of a total 1139 nodes with 3 in-degrees.

The second remarkable consequence of the results is the central position of universities. It seems that they are no longer "ivory towers" of science but proactive creators and mediators of knowledge not only in educating people but also in patenting and

exchanging inventors with others. Nonetheless, they are not just isolated points in the system, but leading and organizing entities within one of the largest communities of knowledge transfer.

Third, the dense mobility ties between rival firms underline the importance of the knowledge transfer represented by inventor mobility. It is very likely that companies must continuously search and obtain skilled inventors from rivals in order to keep up with rapid technological development and with the expanding knowledge base of their competitors. This frequent inventor exchange gives rise to an effective and coherent, but invisible and informal knowledge network among innovative organizations in the United States. Furthermore, in the center of the network, the dense core of the biggest companies form and run a valuable knowledge pool. It is possible that this pool serves a dual function. First, it levels knowledge among these firms, preventing individual organizations from breaking away and enabling them to use the most advanced technologies and ideas. Second, it preserves the technological advantage of the core members against the periphery.

## **6.2 Limitations**

One of the greatest drawbacks of our work is the temporality feature of the source files, since the original Dataverse network files are split into three-year intervals. The disadvantage of this is if an inventor changes assignees more than once in three years, only one assignee can be indicated for him or her in the network file. A side-effect of this is that our analysis underestimates frequency of mobility in the graph. The other limitation factor of our study lies in the noise in the assignee names. Although inventor names are disambiguated, unfortunately, in some cases, the misspelling of assignee names creates false nodes in the network. The third limitation we faced during our research is associated with the size of the network. Since nearly twenty-one thousand organizations are represented in our network, we could not involve a further set of control variables in the investigation, only the network-based ones. We were not able to examine such properties as size, market value, profile, or type of organization (e.g., profit-oriented company, governmental institution, university etc.). In the case of the central nodes, we deduced these variables from the names of the organizations, but overall statistics cannot be displayed. We also could not filter mergers, acquisitions, and parent company and subsidiary relations.

## **6.3 Managerial implications**

Our findings represent further evidence of the importance of inventors' mobility in technology development and knowledge transfer. As we have seen, large successful companies maintain a knowledge pool, which is a place for effective knowledge recombination. We also presented the knowledge absorption and emission feature of this core, where IT-related companies are overrepresented compared to other industries. Consistent with recent findings, acquiring external knowledge in order to catch up with advanced organizations on the market by hiring inventors from core organizations in the network can foster technology development, especially if the inventors come from the IT industry.

#### 6.4 Avenues for further research

We believe that this approach toward the mobility of inventors and knowledge diffusion processes offers great potential for further research. One of the possible options for future studies is to analyze the evolution of the network. This could reveal how and when ways of knowledge diffusion have changed over time and how successful companies (such as Microsoft and Apple) have risen in the network. It may provide a better understanding of how knowledge flow supports emerging organizations. Another possible direction is to examine the career of inventors by measuring their productivity compared to the network properties of companies where they invent technology or to the diversification of the classes of technology to which their inventions belong. The third most promising path for future research is to analyze the specific features of individual communities. The remarkable importance of universities in the knowledge flow network, for example, raises the question of how academic inventors move between the academic and the business sector, or what specific features universities show in the network. Fourth, it would also be interesting to compare the mobility network of inventors with the patent co-citation network. Once we identify the similarities and differences between these two networks, we could gain an understanding of the different forms and mechanisms of inter-organizational knowledge transfer.

## 7 References

- Barabási, A. L. (2011). *Linked: The new science of networks*. Budapest: Helikon.
- Batagelj, V., & Zaversnik, M. (2002). *An  $O(m)$  Algorithm for Cores Decomposition of Networks*. Accessed 20th February 2015. <http://arxiv.org/pdf/cs/0310049v1.pdf>
- Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography*, 9(4), 1-30.
- Breschi, S., & Lissoni, F. (2006). *Mobility of inventors and the geography of knowledge spillovers. New evidences on US data*. Accessed 1st March 2015. [http://www.researchgate.net/profile/Francesco\\_Lissoni/publication/5010491\\_Mobility\\_of\\_inventors\\_and\\_the\\_geography\\_of\\_knowledge\\_spillovers.\\_New\\_evidence\\_on\\_US\\_data/links/0fcfd50e1a7ca5dc9f000000.pdf](http://www.researchgate.net/profile/Francesco_Lissoni/publication/5010491_Mobility_of_inventors_and_the_geography_of_knowledge_spillovers._New_evidence_on_US_data/links/0fcfd50e1a7ca5dc9f000000.pdf)
- Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), 066111.
- ESA, & USPTO (2012). *Intellectual Property and the U.S. Economy: Industries in Focus*. Accessed 12th May 2015. [http://www.uspto.gov/sites/default/files/news/publications/IP\\_Report\\_March\\_2012.pdf](http://www.uspto.gov/sites/default/files/news/publications/IP_Report_March_2012.pdf)
- Fortunato, S. (2010). *Community detection in graphs*. Accessed 24th February 2015. <http://arxiv.org/pdf/0906.0612.pdf>
- Freeman, C. (1991). Networks of innovators: a synthesis of research issues. *Research Policy*, 20(5), 499-514.

- Freeman, L.C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215-239.
- Gassmann, O., & Bader, M. A. (2006). *Intellectual Property Management in Inter-firm R&D collaborations*. Accessed 12th March 2015. <https://www.alexandria.unisg.ch/export/DL/47921.pdf>
- Granstrand, O. (1999). *The Economics and Management of Intellectual Property: Towards Intellectual Capitalism*. Northampton: Edward Elgar Pub.
- Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3), 577-598.
- Hoisl, K. (2009). Tracing mobile inventors: The causality between inventor mobility and inventor productivity. *Research Policy*, 36(5), 615-636.
- Kim, J., Lee, S.J., & Marschke, G. (2004). *Research Scientist Productivity and Firm Size: Evidence from Panel Data on Inventors*. Accessed 17th August 2015. <http://www.albany.edu/economics/research/workingp/2004/scientprod.pdf>
- Knell, M. (2011). *Global network of innovators: A synthesis of research issue*. Accessed 15th August 2015. [http://final.dime-eu.org/files/Knell\\_C3.pdf](http://final.dime-eu.org/files/Knell_C3.pdf)
- Lai, R., D'Amour, A., Yu, A., Sun, Y., & Fleming, L. (2011). *Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)*. Accessed 25th January 2015. <http://hdl.handle.net/1902.1/15705>
- Lamoreaux, N., & Sokoloff, K. L. (1999). The Geography of the Market for Technology in the Late Nineteenth and the Early Twentieth Century United States. In G. D. Libecap (Ed.), *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth*, (pp. 67-121). Stamford: JAI Press.
- Moen, J. (2000). *Is mobility of technical personnel a source of R&D spillovers?* Accessed 30th March 2015. <http://www.nber.org/papers/w7834.pdf>
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E* 69(2), 026113.
- OECD (2006). *Creating value form Intellectual Assets: Meeting of the OECD Council at Ministerial Level 2006*. Accessed 4th May 2015. <http://www.oecd.org/sti/inno/36701575.pdf>
- Pennings, M. J., & Wezel, F. C. (2007). *Human Capital, Inter-firm Mobility and Organizational Evolution*. Cheltenham: Edward Elgar Publishing.
- Penrose, E. (1959). *The theory of the growth of the firm*. Oxford: Oxford University Press.
- Petruzzelli, A. M., Rotolo, D., & Albino, V. (2015). Determinants of patent citations in biotechnology: An analysis of patent influence across the industrial and organizational boundaries. *Technology Forecasting & Social Change*, 91, 208-221.
- Phillips, D.J. (2002). A Genealogical Approach to Organizational Life or Chances: The Parent-Progeny Transfer among Silicon Valley Law Firms, 1946-1996. *Administrative Science Quarterly*, 47(3), 474-506.
- Powell, W.W., & Grodal, S. (2005). Networks of innovations. In J. Fagerberg, D. C.

- Mowery, R. R Nelson (Ed.), *The Oxford Handbook of Innovation*, (pp. 57-85). Oxford: Oxford University Press.
- Rao, H., & Drazin, R. (2002). Overcoming resource constraints on product innovation by recruiting talent from rivals: a study of the mutual fund industry, 1986–94. *Academy of Management Journal*, 45(3), 491–507.
- Rosenkopf, L., & Almeida, P. (2003): Overcoming Local Search Through Alliances and Mobility, *Management Science*, 49(6), 751-766.
- Shapiro, R. J., & Pham, N. D. (2007): *Economic Effects of Intellectual Property-Intensive Manufacturing in the United States*. Accessed 10th May 2015. [http://www.sonecon.com/docs/studies/0807\\_thevalueofip.pdf](http://www.sonecon.com/docs/studies/0807_thevalueofip.pdf)
- Song, J., Almeida, P., & Wu, G. (2003). Learning-by-Hiring: When is Mobility More Likely to Facilitate Interfirm Knowledge Transfer? *Management Science*, 49(4), 351-365.
- Torvik, V., & Smalheiser, N. (2009). Author Name Disambiguation in MEDLINE. *ACM Transactions on Knowledge Discovery from Data*, 3(3), 1-30.
- Xuefeng, W., Jie, R., & Youngo, W. (2012). Co-inventor Analysis on China's international technology collaboration in US patent activities: 1976-2010. *Procedia Engineering*, 37, 314-322.
- von Hippel, E. (1987). Cooperation between rivals: Informal knowhow trading. *Research Policy*, 16(6), 291–302.
- Wang, X., Zhang, X., & Xu, S. (2011). Patent co-citation networks of Fortune 500 companies. *Scientometrics*, 88(3), 761-770.
- Wasserman, S., & Faust, K. (2009). *Social network analysis*. New York: Cambridge University Press.
- Watts, D., & Strogatz, S. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684), 440-442.