UNIVERSITÀ CATTOLICA del Sacro Cuore

DIPARTIMENTO DI POLITICA ECONOMICA

Leveraging Knowledge Networks: Rethinking Technological Value Distribution in mRNA Vaccine Innovations

Rossana Mastrandrea

Fabio Montobbio

Gabriele Pellegrino

Massimo Riccaboni

Valerio Sterzi

Working Paper n. 47 - March 2025

VP VITA E PENSIERO

Università Cattolica del Sacro Cuore

DIPARTIMENTO DI POLITICA ECONOMICA

Leveraging Knowledge Networks: Rethinking Technological Value Distribution in mRNA Vaccine Innovations

Rossana Mastrandrea

Fabio Montobbio

Gabriele Pellegrino

Massimo Riccaboni

Valerio Sterzi

Working Paper n. 47 - March 2025

VP VITA E PENSIERO

Rossana Mastrandrea, Department of Management, University of Turin, Torino, Italy

✉ rossana.mastrandrea@unito.it

Fabio Montobbio, Department of Economic Policy, Università Cattolica del Sacro Cuore, Milano, Italy – ICRIOS, Bocconi University, Milano, Italy – BRICK, Collegio Carlo Alberto, Torino, Italy

✉ fabio.montobbio@unicatt.it

Gabriele Pellegrino, Department of Economic Policy, Università Cattolica del Sacro Cuore, Milano, Italy

✉ gabriele.pellegrino@unicatt.it

Massimo Riccaboni, IMT School for Advanced Studies, Lucca, Italy – IUSS, Pavia, Italy

✉ massimo.riccaboni@imtlucca.it

Valerio Sterzi, Bordeaux School of Economics (BSE), University of Bordeaux, CNRS, UMR 6060, Bordeaux, France

✉ valerio.sterzi@u-bordeaux.fr

Dipartimento di Politica Economica
Università Cattolica del Sacro Cuore – Largo A. Gemelli 1 – 20123 Milano
Tel. 02-7234.2921

✉ dip.politicaeconomica@unicatt.it

https://dipartimenti.unicatt.it/politica_economica

ISBN digital edition (PDF): 978-88-343-5991-4

www.vitaepensiero.it

PROGETTO DI RICERCA DI RILEVANTE INTERESSE NAZIONALE 2022

The Role of the Public and Private Sectors in Pharmaceutical Breakthrough Innovations (3PBI)

PNRR per la Missione 4, Componente 2, investimento1.1. Avviso 104/2022 Finanziato dall'Unione europea - Next Generation EU. 3PBI - Prot. 2022S4EAS9 - CUP J53D23004900008

Abstract

This study examines the roles of public and private sector actors in the development of mRNA vaccines, a breakthrough innovation in modern medicine. Using a dataset of 151 core patent families and 2,416 antecedent (cited) patents, we analyze the structure and dynamics of the mRNA vaccine knowledge network through network theory. Our findings highlight the central role of biotechnology firms, such as Moderna and BioNTech, alongside the crucial contributions of universities and public research organizations (PROs) in providing foundational knowledge. We develop a novel credit allocation framework, showing that universities, PROs, government and research centers account for at least 27% of the external technological knowledge base behind mRNA vaccine breakthroughs—representing a minimum threshold of their overall contribution. Our study offers new insights into pharmaceutical and biotechnology innovation dynamics, emphasizing how Moderna and BioNTech's mRNA technologies have benefited from academic institutions, with notable differences in their institutional knowledge sources.

Keywords: breakthrough innovation, innovation networks, patent analysis, mRNA vaccines, COVID-19.

JEL No. I10,I18,L65,O31,O34

Funding declaration: R.M, F.M.,G.P., M.R. acknowledge support from the PRIN 2022 project "The Role of the Public and Private Sectors in Pharmaceutical Breakthrough Innovations (3PBI)" (CUP 2022S4EAS9). R.M. and M.R. acknowledge support from the project "THE - Tuscan Health Ecosystem" (CUP:D63C22000400001). Conflict of Interests: The authors declare that they have no conflicts of interest.

1 Introduction

The timely development of vaccines against COVID-19 is an extraordinary scientific achievement and a fundamental step in the fight against the pandemic. While some vaccines (such as Vaxzevria from AstraZeneca and the University of Oxford and JNJ-78436735 from Johnson & Johnson) were based on relatively established technologies, the most effective vaccines, still in use, come from a new mRNA vaccine platform developed in record time. The mRNA technology is transforming vaccinology and opening up new opportunities for the development of new therapies in many therapeutic areas such as oncology, cardiovascular diseases, cystic fibrosis (Androsavich, 2024; Pardi et al., 2018; Sahin et al., 2014). This rapid development of mRNA vaccines would not have been possible without prior investment in basic research on mRNA, in methods to improve the efficiency of mRNA delivery (e.g. lipid nanoparticles) and in pharmacological modifications to reduce its instability and innate immunogenicity (Pardi et al., 2018; Martin and Lowery, 2020; Slaoui and Hepburn, 2020). As governments and international institutions have faced the challenge of ensuring access to vaccines for the entire global population, understanding the processes and actors that have driven this important biomedical innovation offers important insights for policy beyond the COVID-19 pandemic (Agarwal and Gaule, 2022; Florio, 2022; Florio et al., 2023; D'Souza and Snyder, 2024).

This paper analyses the different roles of the public and private sectors as a source of breakthrough innovation in the life sciences, with a focus on the development of mRNA vaccines for COVID-19. Such breakthrough innovations are fundamental to our society as they drive new developments in science and technology and provide important opportunities to increase social welfare and economic growth (Hall et al., 2005b; Phene et al., 2006; Kaplan and Vakili, 2015; Kuhn, 1962). Typically, breakthrough innovations challenge the existing technological landscape, open up new technological trajectories enabling business growth and new business development.

In order to gain a comprehensive understanding of the role played by the different actors in the development of mRNA vaccines, three different research objectives are pursued in this article. The first objective is to assess the role of different types of organizations within the mRNA patent landscape. We investigate how universities, research centers, biotech companies and pharmaceutical companies have contributed to the development of mRNA vaccines. We distinguish between organizations that have contributed directly to the mRNA vaccine platform ("Core") and those that have developed enabling technologies that support the development of mRNA vaccines ("Antecedent").

The second objective is to identify the key players in the development of the mRNA vaccine platform by relying on network analysis. We construct a comprehensive network of relationships between different types of organizations based on backward patent citations, the so-called *mRNA vaccine knowledge network*. We use centrality measures - such as *in-* and *out-strength*, *betweenness centrality* and *hub & authority scores* - to assess the importance of different actors in the mRNA vaccine knowledge network. This analysis allows us to identify the most influential actors and understand their importance based on local and global network characteristics.

Finally, the third objective is to explore how credit for mRNA vaccine discovery could be redistributed by leveraging the structural properties of the mRNA vaccine knowledge network. We introduce a novel network measure that allows us to assess the contribution of different actors to the development of specific technologies. With this measure, we can fairly distribute the merits in mRNA vaccine discovery by considering the contributions of cited patents as the main sources of knowledge. In this way, we aim to create a more nuanced understanding of how knowledge networks can be used to identify the most important contributions to the development of breakthrough innovations.

We leverage a new and carefully selected large-scale dataset with a Core set of 151 patent families covering the mRNA vaccine platform (published from 2006 to 2020) and their citations to earlier patents, totaling 3 625 patent documents (2 416 patent families). For each patent document, we performed a manual check of the name of the assignee to ensure harmonization where applicable and to identify the type of assignee (e.g. pharmaceutical company, biotech firm, public research organization). Patent applications in the field of mRNA vaccine technology have increased significantly in the last seven years, while the basic knowledge cited in the prior art dates back to the mid-1980s. We find that biotech company patents dominate both patents directly related to mRNA vaccine technology (Core) and patents that enabled the development of the mRNA vaccine platform (Antecedent). However, universities also hold a significant share of patents in both groups, about 24% in the Core group and 19% in the Antecedent group. Remarkably, pharmaceutical companies are practically not represented in the Core group (4.6%) and hold 17.5% of patents in the Antecedent group.

Based on the network analysis, we find that universities and research centers such as MIT, California University, British Columbia University and Max Planck play an important role as 'authorities' and 'bridges' (authority and betweenness centrality) and biotech companies such as Curevac and Moderna have a particularly high hub centrality. Finally, our analysis shows that the main contribution to the discovery of Moderna and BioNTech vaccines comes from the biotech sector: 42.7% and 45.2% respectively. This is followed by universities, which are credited with 19.3% and 21.7% of the contribution of the knowledge network to the discovery of the Moderna and BioNTech vaccines respectively. This distribution underlines the importance that both the private and public sectors play in the mRNA knowledge landscape.

This paper adds to the existing literature by providing a comprehensive analysis of the roles played by various entities—universities, research centers, Biotechnology firms, and pharmaceutical companies—in the development of mRNA vaccine technology. Building on prior studies that have examined the dynamics of breakthrough innovations, we extend this research by integrating patent analysis with network theory to map the complex relationships that underpinned the mRNA vaccine platform. While earlier research has emphasized the role of academic research and public funding in driving technological breakthroughs (Mansfield, 1991; Sampat, 2009), our study provides new insights into how these contributions are distributed across various actors.

In doing so, we introduce a novel approach to redistributing credit for technological innovations, addressing a significant gap in the existing literature on intellectual property rights and credit allocation. Traditional metrics, such as patent counts or simple citation numbers, often fail to capture the nuanced contributions of different entities in complex innovation ecosystems. By leveraging the structural properties of the mRNA vaccine knowledge network, we propose a more equitable method for allocating credit. This approach accounts for the significant, yet often underappreciated, role of public institutions in generating the foundational knowledge on which private firms rely. This aligns with the broader discourse on how innovation policy should more fairly distribute the economic and reputation rewards of scientific discovery, particularly in fields where public and private sectors both play essential roles (Mowery et al., 2004; David et al., 2001).

The rest of the paper is organized as follows. Section 2 outlines the background of this study by introducing the specific characteristics of the mRNA ecosystem and discussing the contribution of academic research to the breakthrough innovation. Section 3 presents the data and key concepts of complex network theory and describes the methodology of the study. Section 4 presents the results. In the final section, we draw a conclusion and discuss some policy implications.

2 Literature and Background

Our contribution in this paper is based on three main strands of research. The first strand deals with the network origins of innovation in the biotechnology sector (Orsenigo et al., 2001). The second research strand deals with the emergence of breakthrough innovations as a recombination process (Trajtenberg et al., 1997; Fleming, 2001). The last line of research provides a historical reconstruction of the most important inventions that paved the way for the discovery of mRNA vaccines. This literature is mainly descriptive and collects evidence for the key discoveries that enabled the mRNA revolution.

2.1 The network of R&D collaborations in the bio-pharmaceutical sector

In biotechnology, the development of breakthrough innovations depends on close collaboration between companies and the scientific community, as well as complex interactions between public and private institutions (Owen-Smith et al., 2002). These interactions have been analyzed in an extensive literature, which shows that the link between innovation breakthroughs and scientific activities is facilitated, for example, by the engagement of star scientists and geographical proximity. Several studies also show that co-location of scientists working in both research institutions and companies significantly improves collaboration and research productivity (Phene et al., 2006; Cockburn and Henderson, 1998; Cockburn et al., 2000; Gambardella et al., 1995; Zucker et al., 1998, 2002; Powell et al., 1996; Orsenigo et al., 2001; McKelvey et al., 2003; Pammolli et al., 2021). This evidence is provided in the context of the introduction of biotechnology as a fundamental platform for drug discovery that emerged from the groundbreaking advances in genetics and molecular biology in the 1970s and 1980s (referred to as the new "genetics paradigm") (Gittelman, 2016; Ng, 2004). The development of biotechnology paralleled the shifts in industrial organization and the establishment of new institutions. Key aspects of this model include the strengthening of intellectual property rights that facilitate technology transfer and the creation of new companies (e.g. Bayh-Dole Act), the rise of start-ups that use venture capital to drive early-stage research and are often driven by prominent academic entrepreneurs (Zucker et al., 2002; Franzoni et al., 2022), and the role of large pharmaceutical companies that license new targets and compounds from smaller companies and focus their resources on applied research, regulatory approval, manufacturing and commercialization.

There are significant direct effects when public research institutions are directly involved in the invention and hold patents, and significant indirect effects when private companies rely on innovative research in conjunction with publicly funded research programs. In the new genetic paradigm, public investment has been critical from the beginning. The development of rDNA technology and the mapping of the human genome are two compelling examples of how substantial public investment over many years has generated a wealth of commercial opportunities, technology transfer, licensing, and significant entry of new companies into the industry (Hughes, 2001; Gittelman, 2016).

The available empirical evidence, which includes both direct and indirect influences, supports the notion that public sector research has a significant impact on drug development (Mansfield, 1991, 1995, 1998; Toole, 2007, 2012; Narin and Olivastro, 1992; Narin et al., 1997; Cockburn and Henderson, 1998; Franzoni et al., 2022). Basic research in academic institutions or public research centers provides the scientific basis for drug discovery by uncovering disease mechanisms, therapeutic strategies, drug targets and prototype compounds. In addition, the public sector makes an important contribution at the interface between the academic and commercial sectors through direct collaborations and part-

nerships with industry and through efforts to promote technology transfer and translational science in the public sector (Owen-Smith et al., 2002).

There is compelling evidence that basic research funded by the National Institutes of Health (NIH) has an economically and statistically significant impact on the development of new drugs in the US (Stevens et al., 2011; DiMasi et al., 2003; Kaitin et al., 1993). Sampat (2009) shows that 7.7% of all U.S. Food and Drug Administration (FDA) approvals and 10.6% of NMEs are based on academic patents. In addition, Sampat and Lichtenberg (2011) analyze 379 drugs approved between 1988 and 2007 and find that 48% were associated with a patent based on prior art generated in the public sector, indicating a significant indirect influence of government funding. They emphasize that while nine percent of the drugs had a direct public sector patent, a much larger proportion had patents that cited either a public sector patent or a government publication. They also show that the influence of the public sector was more pronounced for the most innovative drugs (Patridge et al., 2015; Galkina Cleary et al., 2018).

Assessing the role of the public sector in drug development is particularly important in the context of mRNA vaccines, as they have a significant impact on health. Pfizer's Comirnaty, for example, generated nearly $40 billion in sales in 2022, representing 38% of the company's total sales. There are also relevant oppositions and infringement actions (Montobbio et al., 2024)[1] and important patent disputes between NIH and Moderna[2] and between NIH and Pfizer[3]. The question is whether governments should try to recoup the profits from the commercialization of drugs they have helped to develop. This policy question, which dates back to the debate surrounding the Bayh-Dole Act in the 1980s, concerns whether and to what extent the government should claim a portion of the profits from drugs for which government patents exist or that were developed through government-funded research and development. If the sources of new medicines are significantly publicly funded, it can be argued that private companies should not receive the lion's share of the profits. There is also a risk that taxpayers pay twice for the new drugs: once through taxes with publicly funded R&D and another time through high market prices driven by market power.

2.2 At the origin of breakthrough innovation

A breakthrough innovation is traditionally defined as a rare and disruptive technological change that can lead to a radical shift in the prevailing technological paradigm (Kuhn, 1962; Dosi, 1982). To operationalize the concept of breakthrough innovation, researchers have developed a set of metrics based on patent citations (Phene et al., 2006). A first approach is to consider the most frequently cited patents as breakthrough innovations (Phene et al., 2006; Ahuja and Morris Lampert, 2001). More recently, backward citations have been used to identify patents that recombine basic and distant knowledge as potential breakthrough innovations (Dahlin and Behrens, 2005; Verhoeven et al., 2016; Silvestri et al., 2018). Although they are a reliable measure of technological radicalness, patent-based measures have some well-known pitfalls when it comes to identifying breakthrough innovations (Capponi et al., 2022). Some much-cited patents do not lead to effective innovation from an economic and social point of view. Citations are also an imperfect measure of technological relevance, as citations can be distorted by competitive and social forces. Kaplan and Vakili (2015) depart from a citation

[1] e.g. https://www.iam-media.com/article/gsk-becomes-latest-mrna-patentee-enter-covid-ip-wars-increasing-pressure-pfizerBioNTech

[2] https://www.nytimes.com/2021/11/09/us/moderna-vaccine-patent.html

[3] https://www.iam-media.com/article/biontech-facing-fresh-mrna-patent-spat-in-vaccine-royalties-dispute

based metric of the innovation value and use topic modeling as a method to identify breakthrough ideas. In this article, we take a different approach by looking at mRNA vaccines as a prototypical example of a breakthrough innovation with enormous societal impact. Starting from an undeniable breakthrough innovation, we use patent citations to trace the foundations of mRNA vaccines as a recombinant innovation and the contribution of different actors to the discovery of mRNA technology.

2.3 Patent Citations Networks and Credit Allocation

This paper exploits the assumption that patent citations reveal how new inventions build upon the previous ones. By examining which patents cite which others, you can trace the evolution of mRNA vaccine technology and identify the key contributions. The network of patent citations can highlight the most influential innovators in the field. The existing economics literature provides well-documented evidence that the number of citations often correlates with the value of patents (Jaffe and de Rassenfosse, 2017; Hall et al., 2005a). In addition, judges and legal experts often use patent citations in patent litigation to assess the value and impact of patents.[4]

In addition network theory provides a natural approach to describe complex systems with many interacting actors such as the ecosystem of mRNA vaccine innovation (Barabási, 2016; Newman, 2018). There is a long tradition of studies using network analysis to study technological change (see e.g. Orsenigo et al., 2001). Network analysis has also been successfully used to study citation networks in various contexts (Hummon and Dereian, 1989; Kajikawa et al., 2007; Ding et al., 2009; Radicchi et al., 2011; Wallace et al., 2012; Golosovsky and Solomon, 2017; Guan et al., 2017) including the patent landscape (Li et al., 2007; Cho and Shih, 2011; Érdi et al., 2013; Van Raan, 2017; Mariani et al., 2019; Chakraborty et al., 2020). The use of patent citation networks has been validated in different papers to address different issues. For example, Barberá-Tomás et al. (2011) use patent citations to analyze the technological trajectories in surgical prostheses. In the wind power technology, a patent citation network is used to evaluate how a product's design affects the inventive activity (Huenteler et al., 2016). Malhotra et al. (2021) study the lithium-ion battery technology and show how the emergence of new use environments shape knowledge generation and product innovation. Finally Iori et al. (2022) use patent citation networks to analyze how the role of government grants and patents by Federal Agencies and State Departments has influenced the development of artificial intelligence.

Building on this literature, we follow a different approach to analyze mRNA vaccines, focusing on the contribution of different (types of) organizations. In particular we leverage the structural property of the mRNA knowledge network and present a new method for evaluating contributions to a breakthrough technological innovation. In so doing we provide an attempt to improve on credit attribution research. Conventional measures, like patent counts or raw citation frequencies oversimplify the complex dynamics within innovation ecosystems, failing to accurately reflect the diverse contributions of various actors. This aligns with broader discussions regarding the fair distribution of economic and reputational benefits arising from scientific discovery, especially in sectors like vaccine development where both public and private investment are critical (Mowery et al., 2004; David et al., 2001).

[4]For example, in 2010, Oracle sued Google for alleged infringement of patents and copyrights related to application programming interfaces. The defendant's damages expert used Patent Citation Analysis to evaluate and rank the value of three of the patents at issue in comparison to 22 other patents (Malaspina, 2019); in "Finjan v. Blue Coat Systems, 2013", the court expressly recognized that "[...] qualitative analysis of asserted patents based upon forward citations may be probative of a reasonable royalty in some instances" (Malaspina, 2019).

2.4 The mRNA Innovation Ecosystem

The mRNA revolution in vaccinology has attracted the attention of the scientific community and is considered a prime example of a breakthrough innovation. There are a number of extremely interesting research articles and inspiring reports on how different scientific trajectories have progressed independently and contributed to the development of the new mRNA vaccines (Sahin et al., 2014; Pardi et al., 2018; Dolgin, 2021a,b; Zuckerman, 2021; Fauci, 2021; Veugelers, 2021; Barbier et al., 2022; Franzoni et al., 2022).[5]. This detailed account of technological and scientific developments confirms that the main features of the mRNA ecosystem are consistent in that the vaccine is the result of the convergence of three major scientific pathways:

1. **The discovery of mRNA in 1961**: This discovery provided the impetus for a branch of research that remained a scientific backwater for decades. The main problems were the instability of mRNA and innate immunogenicity. A team of scientists at the Pennsylvania University led by Katalin Karikò and Drew Weissman found a way to take up the mRNA into the cells without triggering an immune response.[6]

2. **Improving mRNA delivery efficiency**: Small biotech companies (such as Inex, Protiva, Tekmira, Arbutus and many others, – often based in Vancouver) developed methods to treat the fatty coats to protect the delivery of genetic material. A major advance was made in 2004 when, after eight years of research, Ian MacLachlan found a suitable mixture of lipids to form a nanoparticle that would protect the genetic material and prevent it from escaping. This innovation was suitable for manufacturing and enabled drug manufacturers to scale up production.

3. **The U.S. government's effort to find a vaccine to prevent AIDS**: At the Vaccine Research Center (VRC) of the U.S. National Institute of Allergy and Infectious Diseases, a group led by Peter Kwong has been trying to target the "spikes" on H.I.V. viruses that allow them to invade cells. Although H.I.V. vaccines were not successful, this work paved the way for research into the spike protein of the coronaviruses that cause Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). Another team led by VRC scientist Barney Graham focused on the development of a vaccine against the respiratory syncytial virus.[7]

Using Anthony Fauci's words: "So, when the genetic sequence of the SARS-CoV-2 became available, Graham's team lost no time in joining their long-time collaborators at Moderna to develop an RNA vaccine using a stabilized, prefusion spike protein as the immunogen. Pfizer and BioNTech, where Karikó was working, also used the RNA platform that she and Weissman had perfected and the immunogen designed by Graham to develop an RNA vaccine. Additional companies also used Graham's immunogen in other vaccine platforms that had been evolving for years, to make SARS-CoV-2 vaccines." (Fauci, 2021), p.109.

[5] https://www.nytimes.com/2022/01/15/health/mrna-vaccine.html;https://www.ft.com/content/b2978026-4bc2-439c-a561-a1972eeba940;https://www.forbes.com/sites/nathanvardi/2021/08/17/\covids-forgotten-hero-the-untold-story-of-the-scientist-whose-breakthrough-made-the-vaccines-possible/

[6] Their work (Karikó et al., 2005) was initially rejected by two prominent journals - Nature and Science - and finally published in a relatively lower impact journal, Immunity. Two fundamental milestones in the development of the mRNA vaccine are: (a) a patent on the "prefusion of Coronavirus spike protein" (by the Department of Health and Human Services, The Scripps Research Institute, and Dartmouth College) and (2) a patent by Katalin Kariko and Drew Weissman at the University of Pennsylvania (the research was funded by the National Institute of Health) on "RNA containing modified nucleosides and methods of use thereof". The 2023 Nobel Prize in Physiology or Medicine was awarded jointly to Katalin Karikó and Drew Weissman for their discoveries that enabled the development of these effective vaccines.

[7] Kirchdoerfer et al. (2016) and Pallesen et al. (2017) are major breakthroughs in this direction.

3 Data and Methodology

3.1 mRNA Vaccine Patents

Data collection & organization

Our analysis begins by defining a *core* set of patent families related to the mRNA vaccine platform. First, we consider 113 patent applications identified by Martin and Lowery (2020) and published between January 1, 2010 and April 1, 2020. We also consider 86 patents identified by Gaviria and Kilic (2021), which examined the patent portfolios of eight organizations involved in mRNA vaccine technology research: BioNTech, Moderna, Curevac, Arbutus, Acuitas, University of British Columbia (UBC), University of Pennsylvania and Arcturus. After excluding ten patents without additional information, our final dataset (referred to as the Core dataset) comprises 189 patents corresponding to 151 different DOCDB families. Using PATSTAT, we create an additional dataset (the so-called Antecedent dataset), which contains information on all backward citations of the Core dataset and comprises a total of 3 625 patent documents (2 416 DOCDB families). In addition, we supplement both datasets with further bibliographic information from PATSTAT, including the filing dates and the names of the applicants. The latter information is particularly important to identify all public and private contributors to the advancement of mRNA technology. A notable limitation of many patent records arises from inconsistencies in the naming of companies and organizations in patent documents. Due to the discretion given by patent offices to applicants in specifying the name of the patent assignees, there are significant differences in the presentation of company names. Although PATSTAT provides a standardized version of patent assignee names, there are still significant discrepancies in the names provided. Combined with the lack of a unique identifier for the assignee in the patent data, this makes it difficult to categorize patents belonging to the same organization.

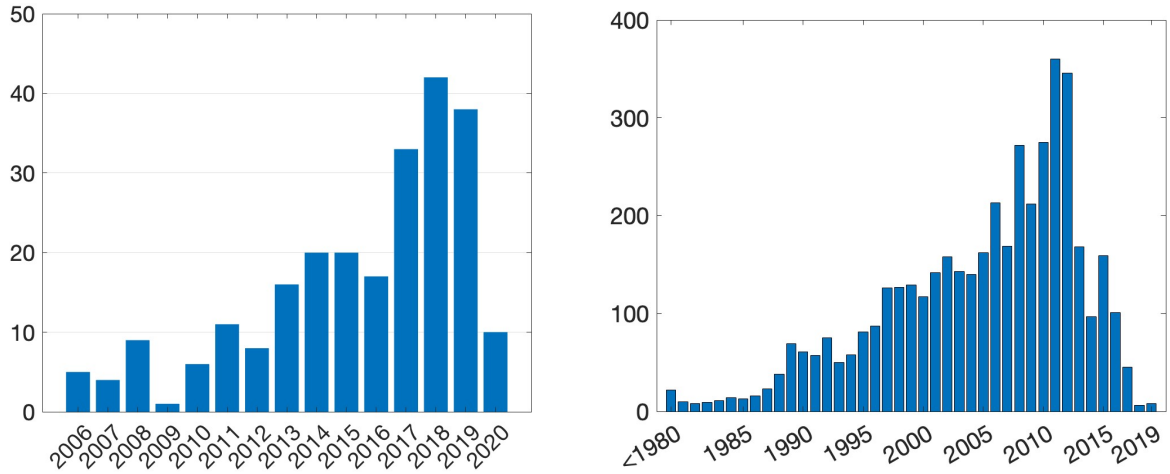
To solve this problem, we apply different strategies. First, we apply proven methods from previous studies (see e.g. Bessen, 2009) to standardize frequently occurring terms in company names. Common examples are typical acronyms attached to company names, such as INC, AG, Ltd, GmbH, Co Kg and others.In addition, we carry out manual checks of the assignee's name to ensure harmonization where necessary. This harmonization and disambiguation process results in a list of 1 173 distinct assignees, which are then categorized according to the type of institution or organization they represent. Identified categories include "University" (including hospitals), "Research Center", "Government", "Public Research Organizations - PRO", "Firm: Biotechnology", "Firm: Pharma", "Firm: Other Med", "Firm: Other non-Med", and "Individual". We divide the time span into three periods of five years according to the registration date of the citing patent to characterize the technological progress at regular intervals: $2006-2010$, $2011-2015$, $2016-2020$.

Data description

Figures 1 (a)-(b) show the number of patents by priority year in the Core and Antecedent datasets. The number of patent applications in the field of mRNA vaccine technology has increased significantly in the last seven years, while the knowledge cited in the prior art goes back further, to the mid-1980s. Figure 1 (a) shows that the Core dataset collects information on patents up to 2020.[8]

[8]In Section 2.3 we show that when the genetic sequence of SARS-CoV-2 became available, the mRNA technology to produce the vaccine was ready. Therefore, we are interested in patents up to 2020. Nevertheless, the patent race for mRNA vaccines has intensified in recent years. This is confirmed by a keyword search in the Patent Lens database: 228 patents filed between 2021 and 2022, for example, contain the keywords 'mRNA AND vaccine' in their abstracts.

The compiled dataset allows us to depict the types of assignees involved in mRNA vaccine technology. The pie chart in Figure 2 shows how the Core and Antecedents patents are distributed across the nine different categories, with the industrial patents (on the right) separated from the non-industrial patents by a dotted line. A look at the Core dataset shows that the largest share of patents (56%) was filed by biotech companies. This is not surprising, as the biotech sector plays a central role in the realization of mRNA vaccine technology. On the other hand, only a minority of patents (4.5%) were filed by pharmaceutical companies. Looking at the research institutions, the important role of universities is striking, as around 24% of all patents were filed by an academic institution, 7% by research centers and 2% by public research institutions. Finally, around 3% of the patents included in the Core dataset appear to have at least one assignee as solo inventor.[9]



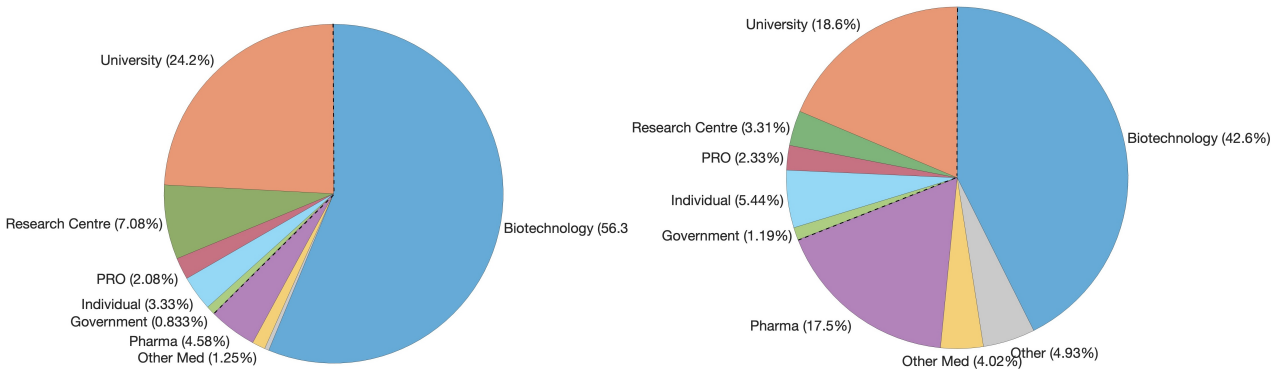
(a) Core (b) Antecedents

Figure 1: **Number of patents by priority year** (a) Patents belonging to the Core mRNA vaccine collection; (b) patents classified as Antecedents.

When looking at the Antecedents dataset, some differences are noticeable compared to the Core dataset. In particular, the contribution of pharmaceutical companies seems to be more important, as they produced about 17.5% of the total patents in this sample, while biotech companies also have a relevant presence in this group (43%). We can also note that the number of patents originating from industry is only slightly higher in the Antecedent patents (69%) than in Core patents (62.5%).

The Tables 1 and 2 show the number and corresponding percentage of companies and institutions belonging to the nine categories of organizations in the three periods analyzed (for the group of Antecedents, the first period includes patents filed before 2010). As expected, the Core size increases while the number of Antecedents decreases over time. The presence of patents originating from biotech companies dominates in both groups, especially in the last period for both the Core and Antecedent group (2006-2010). It is worth noting that in the first period in the Core database, universities are the category with the highest number of patents. Also in the Antecedent database the weight of universities is higher in the first period. It is also interesting to note the decreasing presence of patents for pharmaceutical Antecedents from the first to the last period.

[9]The most frequent assignees include Sahin Ugur (founder of BioNTtech), Hoerr Ingmar (founder of Curevac) and Pieter Cullis, Michael Hope and Thomas Madden, who have been working together for many years on the development of LNP systems that enable RNA- and DNA-based drugs.



(a) Core (b) Antecedents

Figure 2: **Patents sector pie-charts.** The pie charts show the percentage of Core (a) and Antecedents (b) patents by type of organization.

Table 1: Sector: Core dataset

Sector	***2006-2010***	***Percent***	***2011-2015***	***Percent***	***2016-2020***	***Percent***
Biotechnology	9	36%	37	49%	89	64%
Government	0	0%	0	0%	2	1%
Individual	0	0%	2	3%	6	4%
Other	0	0%	0	0%	1	1%
Other Med	2	8%	0	0%	1	1%
PRO	1	4%	1	1%	3	2%
Pharma	2	8%	2	3%	7	5%
Research Centre	1	4%	8	11%	8	6%
University	10	40%	25	33%	23	16%
Total	25	100%	75	100%	140	100%

Table 2: Sector: Antecedents dataset

Sector	***Before 2010***	***Percent***	***2011-2015***	***Percent***	***2016-2020***	***Percent***
Biotechnology	1177	38%	567	50%	121	76%
Government	35	1%	12	1	5	3%
Individual	194	6%	43	4%	1	1%
Other	180	6%	35	3%	1	1%
Other Med	153	5%	20	2%	3	2%
PRO	74	3%	25	2%	2	1%
Pharma	554	18%	201	18%	12	8%
Research Centre	100	3%	41	4%	4	2%
University	620	20%	185	16%	11	6%
Total	3087	100%	1129	100%	160	100%

3.2 Methodology

The mRNA vaccine knowledge network

A network is completely described by a set of *nodes* connected by *edges* that indicate the existence of interactions between them. In addition, a *weight* can be assigned to each link to quantify the intensity of the connection. In the case of the mRNA knowledge network, the nodes can represent patents, entities owning them[10], and categories to which these entities belong (such as different types of organizations in the private and public sectors). In the first case, the links simply show the direction of citations between patents, whereas when looking at entities and sectors, the edges are weighted by

[10]It can be a private or public entity as we explain in the section "Data collection and organization".

the total number of backward or forward citations. The mRNA vaccine knowledge network consists of 1 184 organizations, 5 168 links and a total volume of 81 358 citations.

We investigate basic network properties (see Appendix A) such as the *density* (i.e., the ratio between the total number of node connections and all possible connections), the *degree* (i.e., the total number of edges pointing to/from a node), the *strength* (i.e., the total number of backward/forward citations of a node) to elucidate the structural organization of the mRNA vaccine network. Moreover, we focus on some centrality measures - such as *betwenness centrality*, *hub & authority score* - to capture information about the role of nodes and identify key players based on local and global network properties. Lastly, we conduct a community detection analysis to uncover groups of nodes that are more densely connected to each other than to the rest of the network according to the amount of exchanged citations. This approach is particularly well-suited for characterizing the various technological trajectories that encompass the mRNA technology landscape (details can be found in Appendix B).

Credit allocation

Using network properties, we propose a method to identify key players that contributed to breakthrough innovations among the different types of actors involved in mRNA knowledge networks. We investigate how the potential value of the innovation breakthrough, such as the profits of companies and institutions involved in the commercialization of COVID-19 mRNA vaccines, would be redistributed using the structural properties of the network and the relative technological credit assigned to a specific player. This proposed distribution takes into account the contributions of the cited patents that serve as sources of knowledge for the innovation and development of these vaccines. The motivation behind this approach is to measure the efforts of different (types of) actors in innovating and creating useful step-stones in this field over the years.

It is important to clarify that public research organizations and universities contribute to the innovation process in various direct and indirect ways. For instance, they promote the creation and dissemination of basic knowledge, develop human capital and provide important research infrastructures. These contributions significantly drive innovation breakthroughs. However, our credit allocation method, which is based on the network of patent citations, only captures the direct technological contributions to the field. Consequently, the broader role of universities and public institutions in this context is significantly underestimated and should be considered as a floor or minimum threshold of their overall contribution.

Our methodology incorporates information on the relevance of different patent assignees with respect to (i) the whole knowledge network and (ii) the company that developed and marketed mRNA vaccines by taking into account the volume of citations. Practically, we focus on centrality measures of nodes both at global and local level involving edge-weights, geodesic distances and iterative attribution of importance. First, we introduce a damping factor β^d with $\beta \in (0, 1)$ and d representing the distance from node i (i.e., out-going direction) to discount the effect of contributions in terms of forward citations. Second, we compute node importance using three different measures (details can be found in Appendix C) :

1. *Markov*, the probability of observing a path of length d that starts from node i and reaches node j, calculated using the concept of Markov chains and the powers of the transition probability matrix applied to the knowledge network ;

2. *Markov + Katz*, the probability introduced at point 1. multiplied by the normalized in-Katz

centrality of node j at the global level;

3. *Markov + PageRank*, the probability introduced at point 1. multiplied by the normalized PageRank centrality of node j at the global level.

The three approaches to allocating credit[11] to a node differ in the network properties that are considered to identify key contributors to breakthrough innovations. The Markov rule can be seen as a local measure of the importance of a node, as it depends only on the distance between node i and node j, weighted by the proportion of backward citations, i.e. the ratio between the edge weight between the two nodes and the total number of backward citations of the initial node. In other words, the distance between the two nodes takes into account not only the number of paths, but also the proportion of forward citations along the path. On the other hand, the rules *Markov+Katz* and *Markov+PageRank* also contain a measure for the global importance of the node. While the in-Katz centrality assigns importance to a node if it is linked from other important nodes or if it is strongly linked (i.e. it receives several citations from a node), the PageRank dilutes the importance that emanates from other important nodes according to their out-degree. In particular, important nodes with high out-degree transfer less importance to their links than nodes with low out-degree. However, both approaches combine local importance (through the weighted distance of nodes i via the random walker) and global importance considering all links in the network (via Katz or PageRank centrality). It's worth noting that this method can be applied to any network G and any node i for which credit redistribution is required.[12]

At the end of the procedure, the percentages assigned to each node (summing to 100) can be interpreted in terms of its relative technological and scientific contribution to node i's patents. To figure out how this method can be used to assign the technological credit, assume that we consider the final products of Moderna and Biontech. Suppose also that the value of these vaccines can be measured by these companies' profits. It is then possible to calculate the ideal share of profit of each contributor, based on the share of technological credit. This could be done by multiplying the profit for this ideal share.[13]

4 The mRNA knowledge network: descriptive evidence

In this section, we present the results of the analysis of the patent citation network of the mRNA vaccine. In the first part (4.1) we present the aggregate results showing how the different types of actors contribute to the mRNA knowledge network. In the second part (4.2) of the analysis we move to the the micro level property of the network and show how specific firms and institutions shape the development of the network.

4.1 Sectors

First, we consider the network of patent citations at sector level: nodes represent the 9 groups introduced before (Biotechnology, Government, Individuals, Other, Other Medical, PRO, Pharma,

[11] From now on, we'll use the term "credit" to highlight the technological and scientific contributions of a company, paving the way for real recognition—whether it's in the form of acknowledgment or even monetary rewards—for their role in driving innovation forward.

[12] The method is additive, it is possible to compute the credit allocation starting from each node of interest separately and then sum the different allocations to obtain the overall credit assigned to a specific node.

[13] In Appendix C we show an example to illustrate how the approach could be used in this persepective.

Research Centre, University) and edges are weighted by the total number of backward/forward citations among them. In Figures 3 (a)-(c) we show the snapshots of the networks in the three periods. Node size is proportional to out-strength (i.e. number of backward citations), while colors indicate the different sectors; link arrow shows the direction of citations (from source to target), while edge thickness is proportional to the total number of citations and its color is the same of the citing node (i.e., source).

Network density goes from 0.83 in the first period to 0.96 in the last period, showing its maximum value in the second period (0.97). The greatest node is represented by Biotechnology followed by University and Pharma (reaching the same size in the third period). The thickest links starts from Biotechnology and are directed to the node itself (*self-citations*), University and Pharma companies. The network is thus heterogeneous, primarily driven by the patenting activity of the Biotechnology companies, which often involves self-citations, as well as from Universities and Pharmaceuticals companies. It is worth noticing the absence of connections from Research Centres, PROs and Individuals to the Government sector and from Pharma and Individual to Research Centres; while in the last period there are no citations from Government to Research Centres.

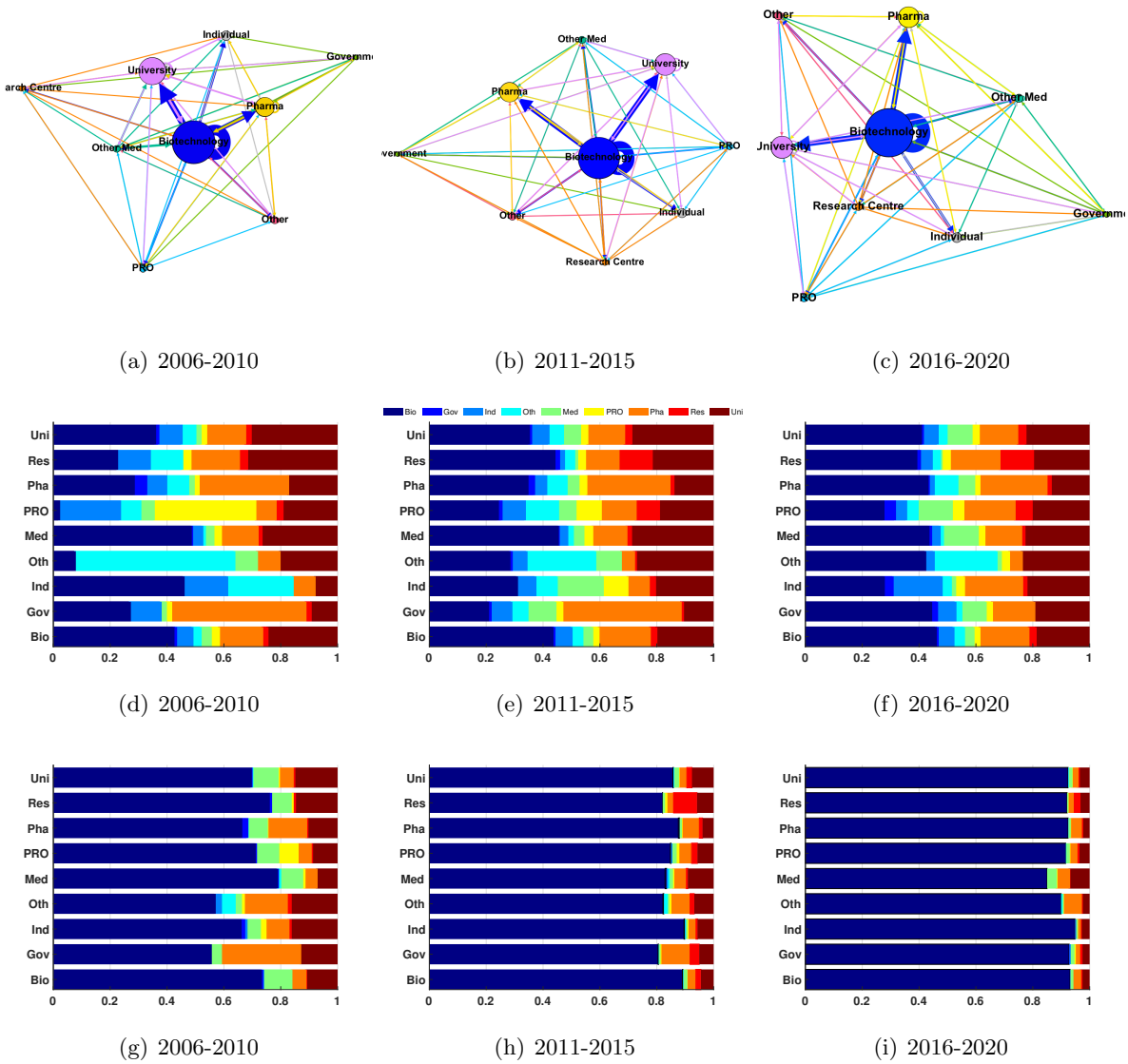
To investigate the heterogeneous distribution of citations more thoroughly, in Figures 3 (d)-(f) we show for each sector (indicated in the rows) the fraction of backward citations to the other sectors. We can see that the number of backward citations from the Biotechnology sector ($\sim 40\%$) to itself almost doubles the quantity of citations to University and Pharma ($\sim 20\%$), even increasing in the third period. In general, we observe a certain heterogeneity, with a clear abundance of citations directed to the Biotechnology sector, except for PRO and Other in the first period. Interestingly, in the initial period, Biotechnology companies seem to be less significant as a sources of knowledge compared to subsequent periods. However, Universities and Pharmaceutical companies also serve as key sources of citations from other sectors (especially for the Government sector in the first two periods).

In Figures 3 (g)-(i) we show for each sector (indicated in the rows) the fraction of forward citations coming from the other sectors. We observe a significant presence of citations from Biotechnology companies across all categories (with notable contributions from Pharma, University and Medical sector), especially in the last two periods. This trend reflects what shown in Table 1 and Table 2, where it emerges a substantial impact of Biotechnology patents, especially in the last period.

4.2 Companies and institutions

In Table 3 we report the main properties of the mRNA vaccine patent network in the three periods under study (see Appendix A for the detailed definition of the indexes). The number of entities (nodes) in the network tripled from the first to the second period, revealing the presence of novel players in the field. The increase in the number of actors leads to a noticeable rise in network sparseness, with the density decreasing from 0.01 to 0.004 in the second period and further to 0.003 in the third period. This outcome is confirmed by the relatively stable average degree, with increased volatility in both in-degree and out-degree, as evinced by the standard deviation nearly doubling from the first to the second period.

The total number of backward citations per node appears generally more variable than the number of its forward citations in all periods. Indeed, the maximum observed out-degree is five times bigger than the maximum in-degree in the first period, fifteen times in the second and almost twenty times in the third period. The number of out-degree is highly heterogeneous with a big amount of zeros (i.e., companies just receiving but not sending citations) increasing over time ($68\%, 67\%, 77\%$, respectively),



(a) 2006-2010 (b) 2011-2015 (c) 2016-2020

(d) 2006-2010 (e) 2011-2015 (f) 2016-2020

(g) 2006-2010 (h) 2011-2015 (i) 2016-2020

Figure 3: **Macroscale network.** (a)-(c) show the network of patent citations among the nine sectors indicated by the different colours: Biotechnology, Government, Individuals, Other, Other Medical, PRO, Pharma, Research Centre and University. Node size is proportional to the total number of backward citations; link thickness is proportional to the edge-weight (number of backward/forward citations) while its color is the same of the source node. (d)-(f) report the total number of backward citations starting by the sector indicated by the row and reaching the other sectors (indicated by the different colors) normalized by the total number of backward citations; (g)-(i) report the total number of forward citations pointing to the sector indicated by the row and starting from the other sectors (indicated by the different colors) normalized by the total number of forward citations.

and a few nodes citing a high number of different companies. In the first period, only 3% of organizations cited at least 10% of the total organizations (Alnylam Pharmaceuticals, British Columbia University, Cellscript Inc, Curevac, Epicentre, Inex-Tekmira-Arbutus, Protiva Biotherapeutics, Proviva, and the University of Pennsylvania). In the second and third periods, this percentage drops to 1%. The organizations citing at least 10% of the total organizations in the second period are Alnylam Pharmaceuticals, BioNTech, British Columbia University, Curevac, Inex-Tekmira-Arbutus, Moderna, and Tron. In the third period, they are British Columbia University, Curevac, Inex-Tekmira-Arbutus, and Moderna.

	2006-2010	2011-2015	2016-2020
Number of nodes	318	1005	1064
Number of links	1152	3937	3520
Volume of citations	7935	63411	63059
Density	0.01	0.004	0.003
Average in-degree (std)	3.6 (3.9)	3.9 (6.2)	3.3 (5.1)
Average out-degree (std)	3.6 (13.7)	3.9 (27.9)	3.3 (26.7)
Max in-degree	22	50	41
Max out-degree	212	782	783
Average in-strength (std)	25 (78)	63 (221)	59 (254)
Average out-strength (std)	25 (85)	63 (1334)	59 (1445)
Max in-strength	719	3500	4982
Max out-strength	2954	42138	46951

Table 3: Network binary properties

The global volume of citations significantly increases from the first to the second period, while slightly reduces in the third one. This reflects the large expansion of the network after 2010 and the fact that more recent patents (third period) receive fewer citations because of truncation. Both the distributions of the forward and backward number of citations appear heterogeneous - especially for the former, showing very high standard deviations - with a few nodes sending/receiving most of the citations.

To better quantify the heterogeneity in the distribution of forward citations per node, we introduce the Herfindahl-Hirschman Index (HHI) as $h_i = \sum_{j=1}^{k_i}(w_{ji}/s_i^{in})^2$. In doing so, we first compute the share of forward citations for each node as the ratio between edge-weights of its incoming links and their sum (i.e., the node in-strength), then we sum up their squares and normalize the result in order to have $h_i \in [0,1], \forall i = 1,\dots,N$. Hence, $h_i = 0$ implies that weights are equally distributed among links, while $h_i \to 1$ signals a high concentration of weight associated to one link. In fig.4 we report the frequency of HHI values divided into ten ranges of length 0.1, along with the frequency of in-degree equal to one (i.e., $h_i = 1$ by definition). It highlights noticeable differences, particularly between the first and the subsequent two periods with: (i) an increasing number of nodes with an in-degree of one (i.e, nodes being cited only by one node) over time; and (ii) a rise in higher HHI values, indicating a significant concentration of citations among a few other entities (i.e., higher concentration of forward citations).

In summary, the number of actors in the network grows substantially from the first to the second period, revealing the emergence of new actors in the field. Concurrently, the flow of knowledge grows after 2010, accompanied by a diverse and expanding set of actors. At the same time, the knowledge network shows that a restricted number of entities are responsible for a substantial amount of backward citations, suggesting they are central in aggregating knowledge from different institutions. In fact, there is a noticeable concentration in forward citations, indicating that certain actors play a crucial role in the innovation process. Consequently, the next section will explore more in depth the centrality of the different entities within the network.

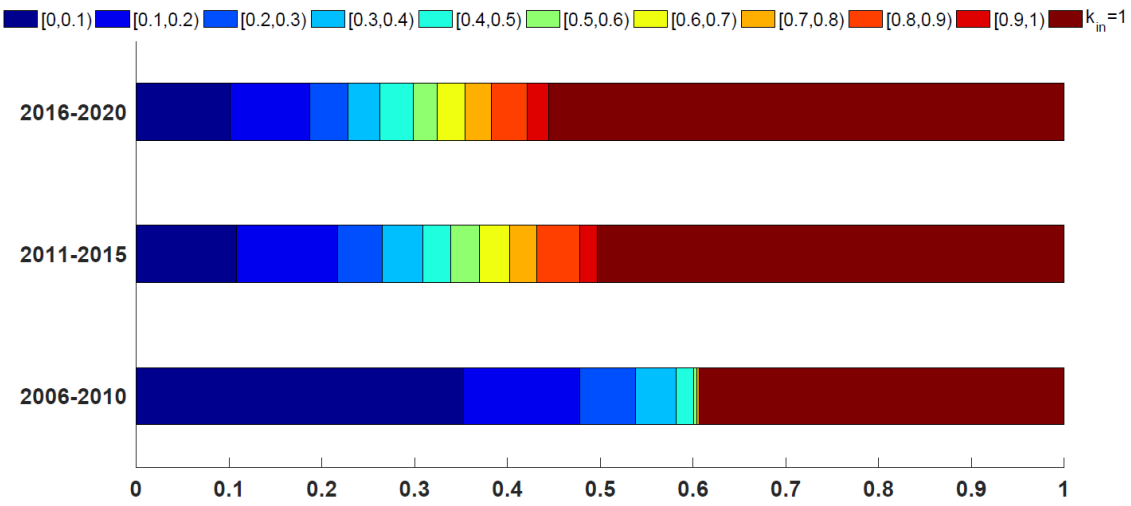
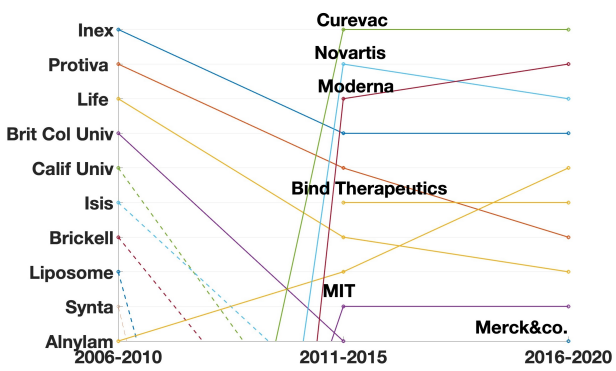


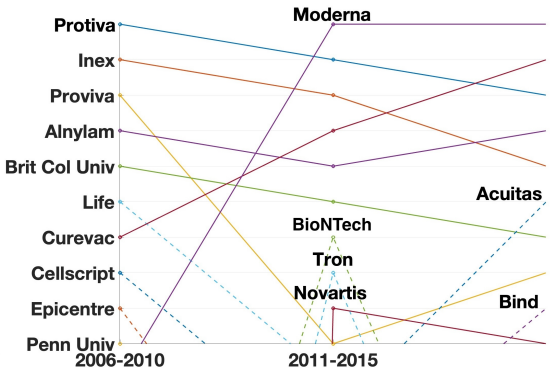
Figure 4: **Herfindahl Index of node in-strength**. Frequency of HHI values to quantify how the total number of forward citations is distributed among all incoming links of a node.

Centrality measures We examine the roles of patent assignees by analyzing centrality measures to identify the key players during each of the three periods under study. We choose four different quantities: (i) in-strength; (ii) out-strength; (iii) betweenness centrality; (iv) hub & authority scores (see Appendix A for more details). They highlight different properties of nodes in assigning the level of importance. The *in-* and *out-* strength are local properties focusing on the total number of forward and backward citations; they measure the total influence a node receives from other nodes in the network (in-strength) and the total influence a node provides to other nodes in the network (out-strength). The betweenness centrality looks at the global network organization taking into account weights, link directions and shortest paths lengths; it measures the importance of a node in terms of its ability to act as a bridge or intermediary in the network. Finally the hub & authority scores compute the centrality of a node in a recursive way taking into account the global network organization and the link directionality. Hub scores measure the value of a node as a source of information. In other words, a strong hub is a node that points to many other nodes that are considered authoritative. Authority scores measure the value of a node as a reliable source of information. In other words, a strong authority is a node to which many hubs point.

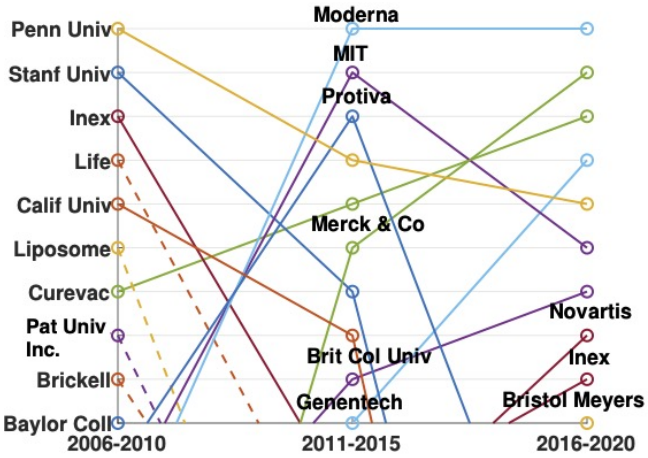
In Figure 5 we show the top ten nodes for the first three centrality measures. The rankings for the first period differ significantly from those of the subsequent two periods across the three centrality measures. For the in-strength centrality, we observe that some actors disappear from the ranking while new ones emerge. Notably, some universities played a prominent role as a source of knowledge in the first period, whereas Moderna, Curevac and Novartis gained importance after 2010. In terms of out-strength centrality, there is stability in terms of actors involved, though their position change over the three periods. The betwenness centrality ranking show a certain degree of variation in both the actors listed and their positions. Universities and research organizations play a relevant role: Pennsylvania University, Stanford University and California University are prominent in the first period, while MIT and British Columbia University, become significant from the second period onward. Additionally, some Pharma companies, such as Merck, Novartis and to a lesser extent Bristol Meyers increase their betweenness centrality in the second and third periods. It is worth noting that after 2010, Moderna and, to a lesser extent, Curevac exhibit very high centrality across various measures in the mRNA knowledge network.



(a) In-strength ranking



(b) Out-strength ranking

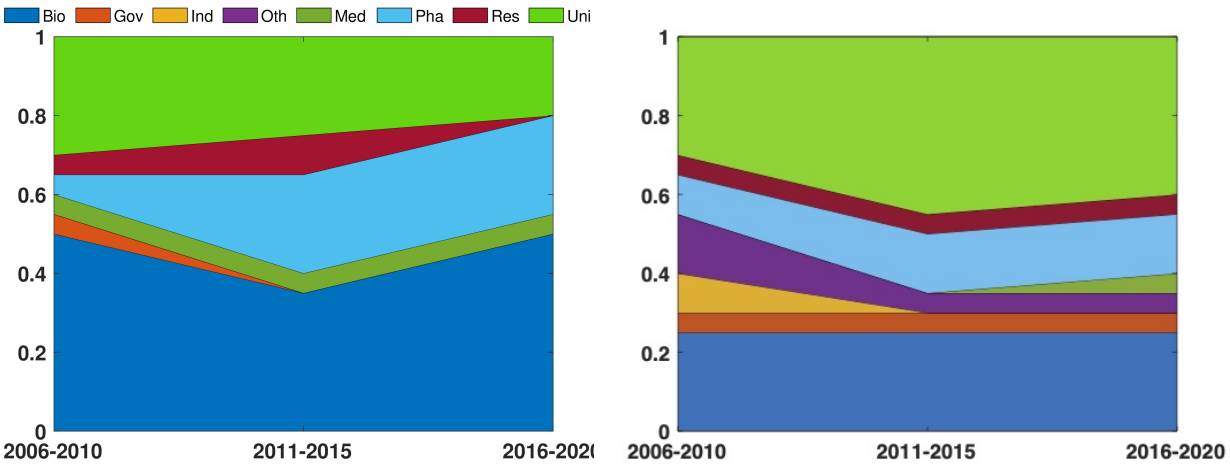


(c) Betweenness centrality

Figure 5: **Top ten rankings for centrality measures** (a) In-strength; b) out-strength; (c) betweenness centrality. The dotted lines indicates entities that appear only in one ranking.

In Figure 6, we select the top 20 entities according to their Hub and Authority scores and we categorize them by sector. The figure illustrates the proportion of each sector by showing the share of entities from each sector across the three periods. Table 4 lists the names of the entities that rank in the top ten for these measures. Remarkably, Universities and Public Research Organizations (combined into a single sector labeled as University) have a significant presence in the top 20 ranking for Authority scores, with their representation increasing over time. Biotechnology companies also maintain a steady share in this ranking. Conversely, the Hub ranking is predominantly led by the Biotechnology sector, followed by the Pharma sector, which exhibits an upward trend, and Universities, which show a decreasing trend.

From Table 4 several notable patterns and trends emerge, highlighting the evolving roles of different types of entities, including Biotechnology companies, pharmaceutical firms, universities, and PROs. Universities and PROs consistently feature prominently in the Authority rankings, emphasizing their role as key sources of foundational knowledge in mRNA research. Specifically, institutions like MIT, CNRS, Max Planck, California University, and Massachusetts University frequently appear in the Authority rankings. This trend underscores the increasing importance of academic research in driving innovation in the mRNA field. In the last period (2016-2020), MIT, California University, British Columbia University and Max Planck have notably maintained their presence in the Authority rankings, indicating their sustained impact on the field.



(a) Hub score

(b) Authority score

Figure 6: **Hub and Authority scores.** Frequency of company type in the top 20 ranking for the Hubs and Authority score (PRO and Universities have been merged in University).

Biotechnology firms dominate the Hub rankings across all three periods, reflecting their pivotal role in actively developing and citing crucial patents. Companies such as Inex-Tekmira, Alnylam, Protiva, and Curevac appear consistently in the Hub rankings. Moderna is a key Hub after 2010. The presence of Biotechnology firms in the Authority rankings, particularly in the later periods, suggests their dual role in both generating and disseminating critical knowledge. Notably, Inex-Tekmira and Alnylam are listed in both the Hub and Authority rankings in the 2016-2020 period. British Columbia University is the most important Hub among universities and PROs.

Figure 6 and Table 4 show that pharmaceutical companies are less dominant than Biotechnology firms. The presence of Novartis, Life and Takeda, particularly in the latest period, in the Hub and Authority rankings highlights their role in leveraging and building upon existing research to develop commercial applications.

Finally, it is important to highlight that some entities exhibit high centrality in either Hubs or Authorities exclusively. For example, Curevac, Moderna, Inex-Tekmira and Proviva are prominent Hubs, reflecting their role in citing and linking various patents. Conversely, CNRS, Max Planck, California University, and Massachusetts University are primarily Authorities, indicating their role in generating highly influential patents.

2006-2010		**2011-2015**		**2016-2020**	
Hub	**Auth**	**Hub**	**Auth**	**Hub**	**Auth**
Brit Col Univ	Isis	Moderna	MIT	Moderna	Inex-Tekmira
Inex-Tekmira	Life	Brit Colu Univ	Brickell	Curevac	MIT
Alnylam	MIT	Inex-Tekmira	Wisc Res Found	Brit Col Univ	Brit Col Univ
Protiva	Cnrs	Curevac	Penn Univ	Inex-Tekmira	Protiva
Proviva	Max Planck	Alnylam	Max Planck	Alnylam	Brickell
Curevac	Calif Univ	BioNTech	Inex-Tekmira	Protiva	Alnylam
Life	Pat Univ Inc.	Tron	Mass Univ	Novartis	Novartis
Epicentre	Liposome	Protiva	Us Depa Health	Proviva	Calif Univ
Cellscript	Mass Univ	Proviva	Protiva	MIT	Life
MIT	White Bio Res	Mainz Univ	Life	Takeda	Max Planck

Table 4: Hub and authorities top ten rank for the three period under study

Analysis of communities. Finally we have performed a community detection analysis. The details of the analysis can be found in Appendix B. The analysis of citation networks, over three distinct periods, reveals the following patterns. In the first period (2006-2010), the landscape is characterized by smaller, more focused communities. Key players include Cellscript, Epicentre, University of Pennsylvania (first community), and Alnylam (second community). These entities are instrumental in laying the groundwork for mRNA technology, often collaborating on fundamental research and early technological developments.

As the field matures (2011-2015), larger communities emerge. Moderna and Curevac take center stage, driving innovation in mRNA vaccines and therapeutics. Moderna, in particular, becomes a key player, with a distinctive periphery of cited entities. Meanwhile, a community focused on LNP technology, involving British Columbia University, Inex, Alnylam, Protiva, and Proviva, highlight the importance of efficient delivery mechanisms for mRNA therapeutics.

In the growth phase (2016-2020), the organization of the communities remains similar and the network expands. The first community includes beside Moderna also Novartis and the Univesity of Pennsylvania. The second community includes the LNP network, with British Columbia University, Inex, Alnylam, Protiva, and Proviva.

Finally, the German hub of innovation, centered around BioNTech, Tron, and Mainz University, continues to grow and new players like Novartis and Merck entered the scene.

The evolution of mRNA patent communities from 2006 to 2020 reflects a trajectory of foundational innovation, consolidation of leadership, and significant expansion. Over nearly two decades, the technological focus has shifted from basic research and initial applications to vaccine developments, bringing together different technological trajectories from mRNA syntesis and stabilization methods to LNPs and efficient delivery. This evolution has been driven by an heterogeneous set of companies and universities. After 2010, the mRNA patent landscape saw a consolidation, with companies like Moderna and Curevac strengthening their leadership positions and new global Pharma players emerging. Geographically, the network has expanded from being predominantly North American to a more globally distributed landscape, with significant hubs in both North America (US and Canada) and Europe (mainly Germany).

5 The mRNA knowledge network: Credit allocation

In the previous sections, we have shown that the mRNA knowledge network is characterized by a dynamic interplay between universities, PROs, Biotechnology firms, and Pharmaceutical companies. Biotechnology firms play a crucial role in the creation and diffusion of important innovations. Universities and PROs also make significant contributions, especially in the early phase. This interaction between different institutions has significantly advanced mRNA innovation and led to major breakthroughs such as the mRNA COVID-19 vaccines.

Our citation network analysis reveals that Biotechnology companies, which have significantly benefited from the commercialization of mRNA vaccines against COVID-19 (in particular BioNTech and Moderna), have thrived within a dynamic and innovative ecosystem. This ecosystem includes a variety of organizations that have played crucial roles. In this section, we propose a method to determine the relative contributions of different types of organizations to the final discovery. Our aim is to suggest an ideal redistribution of the hypothetical value of the innovation and acknowledge the technological sources (based on cited patents) that have been instrumental in the development of these vaccines over

time. Our method consists of a two-step process that combines information on the number of forward citations and the network distances of Moderna and BioNTech, the first two companies responsible for delivering the mRNA vaccines.

In doing so, we consider the three key features highlighted by the three approaches presented in Section 3.2 and detailed in Appendix C. In assigning credit from node i to node j Rule 1 (*Markov*) considers the *local* importance of node j in relation to node i, measured by citations and shortest paths; Rule 2 (*Markov+Katz*) considers the *local* importance outlined by Rule #1, along with the *global* importance of node j, determined by citations received from multiple or significant nodes; Rule 3 (*Markov+PageRank*) considers the *local* importance described in Rule #1, combined with the *global* importance of node j, assessed by citations received from multiple or important nodes discounting the effect of numerous connections. The primary difference of Rule 1 when compared to the other two rules is that it does not consider the general significance of the node in the overall progress of mRNA technology but only the importance in the progress of mRNA technologies used by Moderna (or Biontech). For this reason, in what follows we focus mainly on the first approach.

In Figures 7 we show the total (Figures (a),(c)) and average (Figures (b),(d)) credit allocated to each sector by using Rule 1 with respect to Moderna and BioNTech. This outcome could be interpreted in terms of the technological/scientific contribution given by the different sector to the development of the mRNA Covid-19 vaccines commercialized by Moderna and BioNTech [14]. From Figures 7 (a),(c) it is evident that most of Moderna and BioNTech's credit allocation is directed towards the Biotechnology sector (42.7% and 45.2%, respectively), followed by Universities (19.3% and 21.7%, respectively). Summing up Universities, PROs, Research centers, Individuals and Government, these shares go up to 30% and 37% respectively. This highlights once again the significant contributions of both the private and public sectors to the mRNA knowledge landscape. Furthermore, it is interesting to observe that Pharmaceutical companies rank third in both rankings.

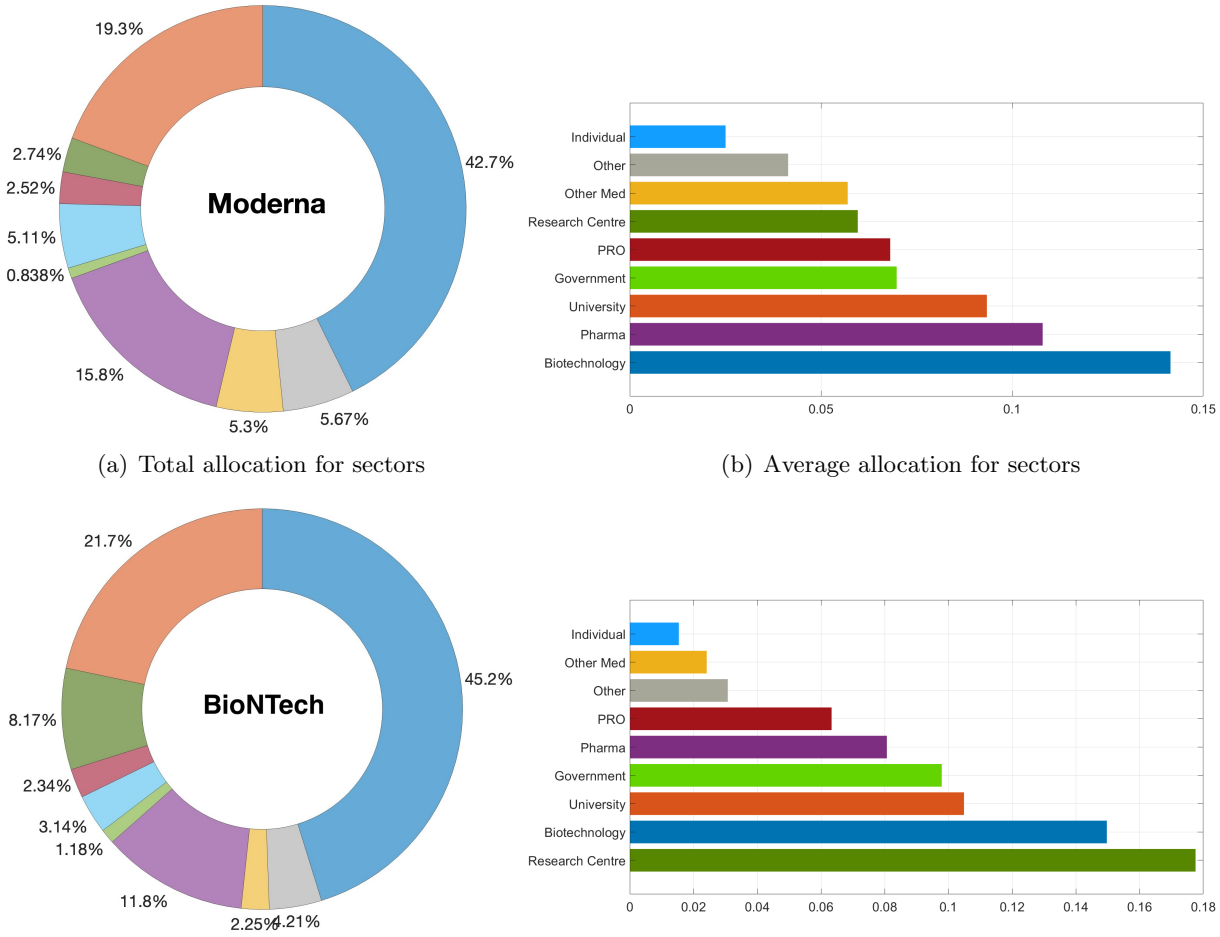
These results capture two key components: the contribution of each individual patent and the overall number of patents within a given sector. Our approach also allows us to assess a potential size effect by determining whether some sectors appear to have made a limited contribution simply because they have fewer cited patents. To address this, we measure the contribution of each individual patent within its respective sector.

To account for sector size, we report the average credit allocation per patent within each sector (Figures 7 (b),(d)). Notably, in the case of BioNTech, Research Centers receive the highest average credit allocation, followed by the Biotechnology sector, Universities, and Government institutions. This suggests that, despite having relatively few patents, research centers played a significant role in BioNTech's COVID-19 knowledge network. In the case of Moderna, the Biotechnology sector holds the highest average credit allocation, followed by the Pharmaceutical sector, Universities, and Government institutions.

Going into detail, in Table 5 and Table 6 we show the top ten main contributors to Moderna's and BioNTech's inventions, respectively, expressed as a percentage of the final technological value of the two companies' vaccine, according to three rules described before.

Table 5 shows the top ranking for credit allocation for Moderna. In case of *Markov* rule, Curevac will be assigned the 6.2% of Moderna's technological credit, and Moderna itself will receive the 5.2% from its credit redistribution. Furthermore, independently of the approach used, Curevac and Novartis appear in the top 3 ranking for credit allocation. Recalling the different properties emphasized by

[14] Results for the other two rules are qualitatively similar and are available upon request.



(a) Total allocation for sectors

(b) Average allocation for sectors

(c) Total allocation for sectors

(d) Average allocation for sectors

Figure 7: **Moderna and BioNTech's credit allocation.** Technological credit allocation of Moderna's (top) and BioNTech's (bottom) for sectors according to rule 1). (a),(c) Total credit allocated to each sector; (b),(d) average credit allocated to each sector.

Table 5: Moderna's redistribution

Markov	**Share**	***Markov+Katz***	**Share**	***Markov+PageRank***	
Curevac	6.2 %	Curevac	8.4 %	Curevac	9.3%
Moderna	5.2%	Novartis	7.1 %	Genentech	8.6%
Novartis	4.2%	Moderna	6.3%	Novartis	5.8%
Bind Th.	2.6%	Patent Univ Inc.	4.5%	Bioject	5.2%
Patent Univ Inc.	2.1%	Brickell	4.5%	Anti Gene	4.1%
Bioject Inc.	2.0%	Genetech	3.9%	Patent Univ Inc.	4.1%
Brickell	1.6%	MIT	3.1%	Merck&co.	3.3%
Genentech	1.6%	Inex*	2.8%	Brickell	3.3%
Alnylam	1.5%	Merck&co.	2.8%	Moderna	3.0%
Inex*	1.4%	WARF**	2.7%	Calif Univ	2.6%

The table displays the percentage of the final technological value of Moderna's mRNA vaccine, based on three allocation rules outlined in the paper, assuming that 50% of the value is directly assigned to Moderna. According to the Markov Rule, CureVac's contribution is estimated at 3.1% of the total value. (*Inex/Tekmira/Arbutus; **Wisconsin Alumni Research Foundation)

the three approaches (see section 3.2), it becomes evident that the two companies are significant not only for the role they played with respect to Moderna, but also for their broader impact on the entire network. This highlights their global centrality within the mRNA knowledge ecosystem. This is confirmed by the fact that these two companies appear also as main contributors of Biontech's

Table 6: BioNTech's redistribution.

Markov	**Share**	***Markov+Katz***	**Share**	***Markov+PageRank***	
Moderna	11.1 %	Moderna	13.1 %	Moderna	7.4%
Curevac	4.5%	Brickell	7.3 %	Curevac	7.0%
BioNTech	4.3%	Curevac	5.8%	Brickell	6.3%
Mainz Univ	3.9%	Novartis	4.8%	Genentech	5.3%
Tron	3.7%	WARF**	4.5%	Novartis	4.3%
Novartis	3.1%	BioNTech	3.4%	WARF**	4.1%
Brickell	2.5%	Liposome	3.8%	Merck&co.	3.6%
DFCI*.	1.8%	Merck&co.	2.6%	Liposome	3.5%
WARF**	1.7%	Genentech	2.5%	BioNTech	2.7%
Liposome	1.2%	DFCI*	2.4%	DFCI*	2.6%

The table displays the percentage of the final technological value of BioNTech's mRNA vaccine, based on three allocation rules outlined in the paper, assuming that 50% of the value is directly assigned to BioNTech. According to the Markov Rule, CureVac's contribution is estimated at 2.2% of the total value. (*Dana-Faber Cancer Institute;**Wisconsin Alumni Research Foundation)

mRNA vaccine technology (See Table 6). Finally, it is worth mentioning that Moderna is present in all the rankings (although only in position number 9 with rule #2), suggesting that Moderna appears significant both locally (due to its large number of self-citations) and through its connections to other key nodes. On the other hand, Biontech is absent from the list, indicating that its role in the advancement of Moderna's mRNA vaccine might have been relatively minor.

Notably, four academic institutions emerge as a significant contributors to Moderna's mRNA technology: University Patents Inc., MIT, University of California and Wisconsin Alumni Research Foundation (WARF). University Patents Inc. was a company in the U.S. dedicated to technology transfer for American universities. Our data includes ten patents assigned to University Patents, Inc., with nine originating from researchers at the University of Colorado. These patents are all part of the Antecedent dataset. Interestingly, seven of these patents have faced litigation at least once, indicating their relevance in the advancement of mRNA vaccine technology. MIT and the University of California are mentioned only in rules #2 and #3, indicating that their role extends beyond merely supporting the development of Moderna's technology and encompasses a broader contribution to advancing mRNA vaccine technology as a whole.

Table 6 presents the leading rankings for credit distribution concerning Biontech. Under the *Markov* rule, Moderna receives 11.1% of the technological credit from Biontech; Curevac follows as another significant contributor with 4.5%, affirming its pivotal role in advancing mRNA vaccine technology. According to the remaining two criteria, Moderna retains its leading position, with Brickell climbing into the top three.

Notably, four scientific institutions emerge as a significant contributors to Biontech's mRNA technology (under rule #1): Mainz University, Tron, Dana-Faber Cancer Institute, and Wisconsin Alumni Research Foundation. This is in line with the finding in Figure 7 (d), where individual patents from Research Centers appear to have been particularly important in the Biontech Case.

Finally, one key insight of our analysis displayed in Tables 5 and 6 is that Moderna's advancements in mRNA technology appear to rely less on Biontech's mRNA developments compared to the reverse situation. This result may have some important implications on the current patent litigation that sees Moderna arguing that BioNTech/Pfizer's vaccine infringes on Moderna's patents granted prior to the pandemic.

6 Conclusion

A central tenet of science and innovation policy is that public investment in R&D underpins breakthrough innovations with significant societal benefits. Unlike private R&D, which is often closely tied to corporate products and patents, publicly funded research is typically disseminated more broadly. This makes it difficult to track its use and determine who benefits from it. Additionally, publicly funded research can have applications that extend far beyond its original scope, potentially taking years or even decades to materialize. This complicates the task of drawing direct links between public sector research activities and commercial outcomes. Given that public investment in science and technology often serves as the basis for important breakthrough innovations — with varying time lags and spillover effects in different areas — a key challenge lies in accurately attribute its contribution to these advances.

In this study, we have addressed this challenge by analyzing the development of mRNA vaccines against COVID-19, which represents a groundbreaking achievement of modern science and technology and embodies the rapid translation of basic research into life-saving innovations. Based on a sample of 151 patent families and exploiting their citations to 2,416 patent families, this paper uses network theory to analyze the innovative ecosystem that fostered this breakthrough innovation and, in particular, the dynamics and key players involved in the development of the mRNA vaccine platform, offering insights into the interplay between different companies and institutions.

We found that universities and public research organizations (PROs) played a role in the development of the mRNA vaccine platform. Prior to 2010, their patents accounted for 44% of our Core Dataset and 23% of the Antecedent Database (which contains the backward patent citations of the Core Dataset). In addition, these institutions consistently ranked at the top of the authority ratings, highlighting their continued importance as a source of fundamental discoveries. Notable contributions came from the University of Pennsylvania, MIT, Max Planck, and the University of British Columbia, underscoring the essential role of academic research in providing the foundational knowledge necessary for mRNA vaccine development.

Based on this evidence, we have proposed a hypothetical redistribution of the innovation breakthrough's credit from the primary beneficiaries of mRNA COVID-19 vaccines — BioNTech and Moderna. Using the structural properties of the citation network, we found that the universities account for about 19.3% and 21.7% of an idealized value, respectively. These shares increase to 30% and 37% when PROs, research centers, government institutions and individuals are included.

It is important to emphasize that PROs and universities contribute to the innovation process through a variety of direct and indirect means, including the generation and dissemination of foundational discoveries, the development of human capital, and the provision of research infrastructure. Therefore, our results, based on patent citations, represent a significant yet conservative estimate of the overall contribution of the public sector to the successful development of mRNA vaccine platform.

This paper also provides a detailed description of the structure and dynamics of the COVID-19 knowledge network. Our investigation of patent assignees and their citation networks highlights the importance of interactions and knowledge flows between different types of organizations, with Biotechnology firms playing an important and ubiquitous role. Our results reveal the heterogeneous character of the mRNA knowledge network, which is characterized by a dense and increasingly interconnected structure over time. The temporal analysis shows that the number of players and the complexity of interactions increased significantly after 2010, corresponding to the years of intensified research efforts and investments in mRNA technology. During this period, new market players have also emerged

and existing market leaders have consolidated their presence, indicating a maturing of the mRNA innovation landscape.

Biotechnology companies, especially in recent years, have emerged as major players, with a significant increase in Core patents and self-citations (e.g. Moderna, Curevac and companies specializing in efficient LNPs). These companies have played a central role in the development and dissemination of mRNA technologies and are involved in extensive citation networks. The centrality measures have further highlighted the importance of Biotechnology companies and underlined their influence and leadership in the field of mRNA.

Our analysis provides a nuanced understanding of the innovative ecosystem of the pharmaceutical and Biotechnology industries (as discussed in Section 2), particularly in relation to the discovery of major breakthroughs. The findings have several policy implications. First, they underscore the critical role of publicly funded research institutions in fundamental scientific discoveries. Policy makers should prioritize funding for these institutions and universities while supporting a broad range of research trajectories. It is noteworthy that for many years mRNA research was considered a niche area with limited immediate applications. This underlines the importance of maintaining a broad spectrum of scientific research without immediate returns.

The innovative ecosystem is characterized by a multitude of actors, and collaboration between the public and private sectors has proven to be essential for translating scientific research into practical applications. To facilitate this, policy should encourage and support such partnerships and ensure that public research is effectively leveraged for innovation (especially in Biotechnology companies). This includes implementing programs to streamline and facilitate licensing and creating mechanisms to mitigate patent disputes, especially in areas with a dense patent landscape and potentially overlapping claims.

Finally, our study suggests that the development of long-term impact assessment methods is necessary to fully appreciate and demonstrate the value of public R&D investments. Such methods would help to quantify the broader societal benefits of public research beyond its immediate commercial applications and provide a more comprehensive view of its contributions. In this paper, we present a new methodological approach to trace back the underpinnings of breakthrough innovations using credit allocation schemes based on knowledge network analysis.

CRediT authorship contribution statement

Rossana Mastrandrea: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review and editing. **Fabio Montobbio:** Conceptualization, Methodology, Writing – original draft, Writing – review and editing. **Gabriele Pellegrino:** Conceptualization, Data Curation, Methodology, Writing – original draft, Writing – review and editing. **Massimo Riccaboni:** Conceptualization, Methodology, Writing – original draft, Writing – review and editing. **Valerio Sterzi:** Conceptualization, Data Curation, Methodology, Writing – original draft, Writing – review and editing.

Acknowledgments

R.M, F.M.,G.P., M.R. acknowledge support from the PRIN 2022 project "The Role of the Public and Private Sectors in Pharmaceutical Breakthrough Innovations (3PBI)" (CUP 2022S4EAS9). R.M. and M.R. acknowledge support from the project " THE - Tuscan Health Ecosystem" funded by the European Union - Next Generation EU program, in the context of the Italian National Recovery and

Resilience Plan , Investment 1.5: Ecosystems of Innovation (CUP:D63C22000400001). R.M. is member of GNAMPA (Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni) at INdAM (Istituto Nazionale di Alta Matematica).

Data Availability

Data available upon request.

References

Agarwal, R. and P. Gaule (2022): "What drives innovation? Lessons from COVID-19 R&D," *Journal of Health Economics*, 102591.

Ahuja, G. and C. Morris Lampert (2001): "Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions," *Strategic management journal*, 22, 521–543.

Androsavich, J. R. (2024): "Frameworks for transformational breakthroughs in RNA-based medicines," *Nature Reviews Drug Discovery*, 1–24.

Barabási, A. (2016): "Network Science. Cambridge University Press, Cambridge," .

Barberá-Tomás, D., F. Jiménez-Sáez, and I. Castelló-Molina (2011): "Mapping the importance of the real world: The validity of connectivity analysis of patent citations networks," *Research policy*, 40, 473–486.

Barbier, A. J., A. Y. Jiang, P. Zhang, R. Wooster, and D. G. Anderson (2022): "The clinical progress of mRNA vaccines and immunotherapies," *Nature biotechnology*, 40, 840–854.

Bessen, J. (2009): "NBER PDP Project User Documentation: Matching Patent Data to Compustat Firms," Working Paper, Unpublished working paper, Boston University.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008): "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, 2008, P10008.

Brin, S. and L. Page (1998): "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, 30, 107–117.

Capponi, G., A. Martinelli, and A. Nuvolari (2022): "Breakthrough innovations and where to find them," *Research Policy*, 51, 104376.

Chakraborty, M., M. Byshkin, and F. Crestani (2020): "Patent citation network analysis: A perspective from descriptive statistics and ERGMs," *Plos one*, 15, e0241797.

Cho, T.-S. and H.-Y. Shih (2011): "Patent citation network analysis of core and emerging technologies in Taiwan: 1997–2008," *Scientometrics*, 89, 795–811.

Cockburn, I. M. and R. M. Henderson (1998): "Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery," *The journal of industrial economics*, 46, 157–182.

Cockburn, I. M., R. M. Henderson, and S. Stern (2000): "Untangling the origins of competitive advantage," *Strategic management journal*, 21, 1123–1145.

Dahlin, K. B. and D. M. Behrens (2005): "When is an invention really radical?: Defining and measuring technological radicalness," *Research policy*, 34, 717–737.

David, P. A., D. C. Mowery, and W. E. Steinmueller (2001): "Public Policy and Knowledge-Based Economy in the United States," in *Innovation Policy and the Economy, Volume 1*, ed. by A. Jaffe, J. Lerner, and S. Stern, Cambridge, MA: MIT Press, 103–142.

DiMasi, J. A., R. W. Hansen, and H. G. Grabowski (2003): "The price of innovation: new estimates of drug development costs," *Journal of health economics*, 22, 151–185.

Ding, Y., E. Yan, A. Frazho, and J. Caverlee (2009): "PageRank for ranking authors in co-citation networks," *Journal of the American Society for Information Science and Technology*, 60, 2229–2243.

Dolgin, E. (2021a): "How COVID unlocked the power of RNA vaccines." *Nature*, 189–191.

——— (2021b): "The tangled history of mRNA vaccines," *Nature*, 597, 318–324.

Dosi, G. (1982): "Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change," *Research policy*, 11, 147–162.

D'Souza, A. and C. M. Snyder (2024): "Can Operation Warp Speed Serve as a Model for Accelerating Innovations Beyond Covid Vaccines?" *NBER Chapters*.

Érdi, P., K. Makovi, Z. Somogyvári, K. Strandburg, J. Tobochnik, P. Volf, and L. Zalányi (2013): "Prediction of emerging technologies based on analysis of the US patent citation network," *Scientometrics*, 95, 225–242.

Fauci, A. S. (2021): "The story behind COVID-19 vaccines," *Science*, 372, 109–109.

Fleming, L. (2001): "Recombinant uncertainty in technological search," *Management science*, 47, 117–132.

Florio, M. (2022): "To what extent patents for Covid-19 mRNA vaccines are based on public research and taxpayers' funding? A case study on the privatization of knowledge," *Industrial and Corporate Change*, 31, 1137–1151.

Florio, M., S. Gamba, and C. Pancotti (2023): "Mapping of long-term public and private investments in the development of Covid-19 vaccines," *Eur Parliament COVI Committee*, 1–95.

Franzoni, C., P. Stephan, and R. Veugelers (2022): "Funding risky research," *Entrepreneurship and Innovation Policy and the Economy*, 1, 103–133.

Galkina Cleary, E., J. M. Beierlein, N. S. Khanuja, L. M. McNamee, and F. D. Ledley (2018): "Contribution of NIH funding to new drug approvals 2010–2016," *Proceedings of the National Academy of Sciences*, 115, 2329–2334.

Gambardella, A. et al. (1995): *Science and innovation: The US pharmaceutical industry during the 1980s*, Cambridge University Press.

Gaviria, M. and B. Kilic (2021): "A network analysis of COVID-19 mRNA vaccine patents," .

Gittelman, M. (2016): "The revolution re-visited: Clinical and genetics research paradigms and the productivity paradox in drug discovery," *Research Policy*, 45, 1570–1585.

Golosovsky, M. and S. Solomon (2017): "Growing complex network of citations of scientific papers: Modeling and measurements," *Physical Review E*, 95, 012324.

Guan, J., Y. Yan, and J. J. Zhang (2017): "The impact of collaboration and knowledge networks on citations," *Journal of Informetrics*, 11, 407–422.

Hall, B., A. Jaffe, and M. Trajtenberg (2005a): "Market value and patent citations," 36, 16–38, type: Journal Article.

Hall, B. H., A. Jaffe, and M. Trajtenberg (2005b): "Market value and patent citations," *RAND Journal of economics*, 16–38.

Huenteler, J., J. Ossenbrink, T. S. Schmidt, and V. H. Hoffmann (2016): "How a product's design hierarchy shapes the evolution of technological knowledge—Evidence from patent-citation networks in wind power," *Research Policy*, 45, 1195–1217.

Hughes, S. S. (2001): "Making dollars out of DNA: the first major patent in biotechnology and the commercialization of molecular biology, 1974-1980," *Isis*, 92, 541–575.

Hummon, N. P. and P. Dereian (1989): "Connectivity in a citation network: The development of DNA theory," *Social networks*, 11, 39–63.

Iori, M., A. Martinelli, and A. Mina (2022): "The direction of technical change in AI and the trajectory effects of government funding," .

Jaffe, A. B. and G. de Rassenfosse (2017): "Patent citation data in social science research: Overview and best practices," 68, 1360–1374, place: Hoboken Publisher: Wiley

WOS:000401545600002.

Kaitin, K. I., N. R. Bryant, and L. Lasagna (1993): "The role of the research-based pharmaceutical industry in medical progress in the United States," *The Journal of Clinical Pharmacology*, 33, 412–417.

Kajikawa, Y., J. Ohno, Y. Takeda, K. Matsushima, and H. Komiyama (2007): "Creating an academic landscape of sustainability science: an analysis of the citation network," *Sustainability Science*, 2, 221–231.

Kaplan, S. and K. Vakili (2015): "The double-edged sword of recombination in breakthrough innovation: The Double-Edged Sword of Recombination," *Strategic Management Journal*, 36, 1435–1457.

Karikó, K., M. Buckstein, H. Ni, and D. Weissman (2005): "Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA," *Immunity*, 23, 165–175.

Kirchdoerfer, R. N., C. A. Cottrell, N. Wang, J. Pallesen, H. M. Yassine, H. L. Turner, K. S. Corbett, B. S. Graham, J. S. McLellan, and A. B. Ward (2016): "Pre-fusion structure of a human coronavirus spike protein," *Nature*, 531, 118–121.

Kleinberg, J. M. (1999): "Hubs, authorities, and communities," *ACM computing surveys (CSUR)*, 31, 5–es.

Kuhn, T. S. (1962): "The Structure of Scientific Revolutions." .

Li, X., H. Chen, Z. Huang, and M. C. Roco (2007): "Patent citation network in nanotechnology (1976–2004)," *Journal of Nanoparticle Research*, 9, 337–352.

Malaspina, P. A. (2019): "Patent citation analysis and patent damages," *Chi.-Kent J. Intell. Prop.*, 18, 232.

Malhotra, A., H. Zhang, M. Beuse, and T. Schmidt (2021): "How do new use environments influence a technology's knowledge trajectory? A patent citation network analysis of lithium-ion battery technology," *Research Policy*, 50, 104318.

Mansfield, E. (1991): "Academic research and industrial innovation," *Research Policy*, 20, 1–12.

——— (1995): "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing," *The Review of Economics and Statistics*, 77, 55.

——— (1998): "Academic research and industrial innovation: An update of empirical findings," *Research policy*, 26, 773–776.

Mariani, M. S., M. Medo, and F. Lafond (2019): "Early identification of important patents: Design and validation of citation network metrics," *Technological forecasting and social change*, 146, 644–654.

Martin, C. and D. Lowery (2020): "mRNA vaccines: intellectual property landscape." *Nature Reviews Drug Discovery*, 19, 578–579.

McKelvey, M., H. Alm, and M. Riccaboni (2003): "Does co-location matter for formal knowledge collaboration in the Swedish biotechnology–pharmaceutical sector?" *Research policy*, 32, 483–501.

Montobbio, F., G. Pellegrino, and V. Sterzi (2024): "Patent landscape analysis of mRNA COVID-19 Vaccine Technology: Examining the role of biotech companies, public research and universities," in *Handbook of Innovation and Intellectual Property Rights. Evolving Scholarship and Reflections.*, ed. by W. Park, Edward Elgar, chap. 24, 415–437.

Mowery, D. C., R. R. Nelson, B. N. Sampat, and A. A. Ziedonis (2004): *Ivory Tower and Industrial Innovation: University-Industry Technology Transfer Before and After the Bayh-Dole*

Act, Stanford, CA: Stanford University Press.

Narin, F., K. S. Hamilton, and D. Olivastro (1997): "The increasing linkage between U.S. technology and public science," *Research Policy*, 26, 317–330.

Narin, F. and D. Olivastro (1992): "Status report: linkage between technology and science," *Research policy*, 21, 237–249.

Newman, M. (2018): *Networks*, Oxford university press.

Newman, M. E. (2006): "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, 103, 8577–8582.

Ng, R. (2004): *Drugs: From Discovery to Approval.*, New Jersey: John Wiley.

Orsenigo, L., F. Pammolli, and M. Riccaboni (2001): "Technological change and network dynamics: lessons from the pharmaceutical industry," *Research policy*, 30, 485–508.

Owen-Smith, J., M. Riccaboni, F. Pammolli, and W. W. Powell (2002): "A comparison of US and European university-industry relations in the life sciences," *Management science*, 48, 24–43.

Pallesen, J., N. Wang, K. S. Corbett, D. Wrapp, R. N. Kirchdoerfer, H. L. Turner, C. A. Cottrell, M. M. Becker, L. Wang, W. Shi, W.-P. Kong, E. L. Andres, A. N. Kettenbach, M. R. Denison, J. D. Chappell, B. S. Graham, A. B. Ward, and J. S. McLellan (2017): "Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen," *Proceedings of the National Academy of Sciences*, 114, E7348–E7357.

Pammolli, F., M. Riccaboni, and A. Spelta (2021): "The network origins of Schumpeterian innovation," *Journal of Evolutionary Economics*, 31, 1411–1431.

Pardi, N., M. J. Hogan, F. W. Porter, and D. Weissman (2018): "mRNA vaccines—a new era in vaccinology," *Nature reviews Drug discovery*, 17, 261–279.

Patridge, E. V., P. C. Gareiss, M. S. Kinch, and D. W. Hoyer (2015): "An analysis of original research contributions toward FDA-approved drugs," *Drug discovery today*, 20, 1182–1187.

Phene, A., K. Fladmoe-Lindquist, and L. Marsh (2006): "Breakthrough innovations in the US biotechnology industry: the effects of technological space and geographic origin," *Strategic management journal*, 27, 369–388.

Powell, W. W., K. W. Koput, and L. Smith-Doerr (1996): "Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology," *Administrative Science Quarterly*, 41, 116.

Radicchi, F., S. Fortunato, and A. Vespignani (2011): "Citation networks," *Models of science dynamics: Encounters between complexity theory and information sciences*, 233–257.

Sahin, U., K. Karikó, and Ö. Türeci (2014): "mRNA-based therapeutics—developing a new class of drugs," *Nature reviews Drug discovery*, 13, 759–780.

Sampat, B. N. (2009): "Academic patents and access to medicines in developing countries," *American Journal of Public Health*, 99, 9–17.

Sampat, B. N. and F. R. Lichtenberg (2011): "What are the respective roles of the public and private sectors in pharmaceutical innovation?" *Health affairs*, 30, 332–339.

Silvestri, D., M. Riccaboni, and A. Della Malva (2018): "Sailing in all winds: Technological search over the business cycle," *Research Policy*, 47, 1933–1944.

Slaoui, M. and M. Hepburn (2020): "Developing safe and effective Covid vaccines—Operation Warp Speed's strategy and approach," *New England Journal of Medicine*, 383, 1701–1703.

Stevens, A. J., J. J. Jensen, K. Wyller, P. C. Kilgore, S. Chatterjee, and M. L. Rohrbaugh (2011): "The role of public-sector research in the discovery of drugs and vaccines,"

New England Journal of Medicine, 364, 535–541.

Toole, A. (2007): "Does Public Scientific Research Complement Private Investment in Research and Development in the Pharmaceutical Industry?" *The Journal of Law and Economics*, 50, 81–104.

——— (2012): "The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry," *Research Policy*, 41, 1–12.

Trajtenberg, M., R. Henderson, and A. Jaffe (1997): "University versus corporate patents: A window on the basicness of invention," *Economics of Innovation and new technology*, 5, 19–50.

Van Raan, A. F. (2017): "Patent citations analysis and its value in research evaluation: A review and a new approach to map technology-relevant research," *Journal of Data and Information Science*, 2, 13–50.

Verhoeven, D., J. Bakker, and R. Veugelers (2016): "Measuring technological novelty with patent-based indicators," *Research policy*, 45, 707–723.

Veugelers, R. (2021): "mRNA vaccines: a lucky shot?" Tech. rep., Bruegel Working Paper.

Wallace, M. L., V. Larivière, and Y. Gingras (2012): "A small world of citations? The influence of collaboration networks on citation practices," *PloS one*, 7, e33339.

Zucker, L., M. Darby, and M. B. Brewer (1998): "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review*, 88, 290–306.

Zucker, L. G., M. R. Darby, and J. S. Armstrong (2002): "Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology," *Management science*, 48, 138–153.

Zuckerman, G. (2021): *A Shot to Save the World: The Inside Story of the Life-or-Death Race for a COVID-19 Vaccine*, United States: Penguin Publishing Group.

ONLINE APPENDIX

Leveraging Knowledge Networks: RethinkingTechnological Value Distribution in Vaccine Innovations

Rossana Mastrandrea Fabio Montobbio Gabriele Pellegrino Massimo Riccaboni Valerio Sterzi

Appendix A Complex Network Theory

A network is mathematically described by $G = (V, E)$, where V is the set of all nodes with size $|V| = N$ and E the set of all edges with size $|E| = L$. The ratio between the observed number of links and the potential ones ($N(N-1)$), namely *density*, quantifies its cohesiveness level. A graph is uniquely defined by the adjacency matrix $A \equiv (a_{ij})_{1 \leq i,j \leq N}$ with $a_{ij} = 1$ if and only if it exists a link between node i and node j. It is worth noticing that for the patent citations network generally holds $a_{ij} \neq a_{ji}$, i.e. the network is directed/asymmetric. For the mRNA knowledge network we can consider different level of aggreagation: (i) network of patent citations , whose nodes represent patents and edges indicate the existence of a citation between them; (ii) network of entities, whose nodes are entities (companies, research and government institutions) and edges are weighted by the total number backward/forward citations among them; (iii) network of sectors, whose nodes are sectors and links are weighted according to the total number of backward/forward citations among them. In the cases (ii) and (iii), the network is defined also by a weighted adjacency matrix, $W \equiv (w_{ij})_{1 \leq i,j \leq N}$ with generally $w_{ij} \neq w_{ji}$.

We first investigate some basic properties (summarized in Table A1) to evaluate the network global organization and the role of nodes. We introduce the node in/out *degree* and *strength*: two local properties computing the total number of incoming/outcoming links of a node and the total number of forward/backward citations of a node, respectively. The distribution of node degrees allows to understand how connections are organized/placed in the network, for example through the presence and number of hubs (i,e., node with many links); while the strength distribution offers some insights about the heterogeneity of citing and cited entities. We also computed an higher-order property, the *in-clustering coefficient*: it counts the number of closed triangles with two links pointing to node i divided by the total number of triplets involving it. This measure sheds light on the tendency of the system to form specific clusters.[15]

Node degree and strength can also be thought as local centrality measures (when normalized by the network size and total weight) as they quantify the importance of a node in terms of number of links or weighted links. However, the importance of nodes can been explored also by different quantities: (i) the betweenness centrality; (ii) the in-Katz centrality; (iii) the PageRank centrality; (iv) the Hub&Authority scores. The betweenness centrality assigns higher value to nodes behaving as bridges helping communication flow in the network. Indeed, it counts how many shortest paths connecting any two nodes in the network pass through node i with respect to all possible shortest paths existing between any pair of nodes in the web. The in-Katz centrality is based on the idea that node importance depends upon the centrality of neighbours pointing to it in a backward fashion. In other terms, it computes the relative influence of a node coming not only from its immediate neighbours, but also through indirect and longer connections in an iterative way. Intuitively, we can imagine that the closest connected nodes have more influence over the node than more distant ones. Thus when

[15] Other clustering coefficients could be computed, according to the directions of the involved links. However, we selected the most relevant one for our scope.

combining paths of all lengths, one can introduce an attenuation factor to give more importance to shorter walks with respect to longer ones.

Let $A = (a_{i,j})$ be the adjacency matrix of a directed graph. The Katz centrality of node i is given by:

$$z_i{}^{16} = [(\alpha^0 A^0 + \alpha A + \alpha^2 A^2 + \cdots + \alpha^k A^k + \dots)\mathbf{1}]_i = \left[\sum_{k=0}^{\infty} (\alpha^k A^k)\mathbf{1}\right]_i \tag{A1}$$

where α is a constant and $\mathbf{1}$ is the vector $(1, 1, \dots, 1)$.

The series in (A1) converges when $\alpha < 1/\rho(A)$, where $\rho(A)$ is the maximum eigenvalue of the adjacency matrix A, and in that case the Katz centrality in matrix form reads:

$$z_i = [(I - \alpha A)^{-1}\mathbf{1}]_i \tag{A2}$$

The Katz status of a node is defined as the number of weighted paths reaching the node in the network: a generalization of the degree measure which counts only paths of length one. Notice that long paths are weighted less than short ones by exploiting the attenuation factor α. For small (close to 0) values of α the contribution given by paths longer than one rapidly declines, and thus Katz scores are mainly influenced by short paths (mostly in-degrees). When the damping factor is large, long paths are devalued smoothly, and Katz scores are more influenced the by endogenous topological part of the system (it is recommended to choose α between 0 and $1/\rho(A)$).

A potential problem with the Katz centrality is that if a node with high centrality links many others then all those others get high centrality. The centrality gained by an incoming link from an important node should be *diluted* if the important vertex is for example an hub.

PageRank is an adjustment of Katz centrality taking into account this need of "diluting" the importance received from a node if it is for example an hub (Brin and Page, 1998). Indeed, the centrality derived from a node' neighbors is now proportional to their centrality divided by their out-degree. According to this procedure, vertices pointing to many others transfer only a small amount of centrality to their contacts.

The PageRank[17] of node i is given by:

$$p_i = [(I - \alpha D^{-1} A)^{-1}\mathbf{1}]_i \tag{A3}$$

with α a constant and D a diagonal matrix with i-th diagonal element equal to node i out-degree, d_i. The damping factor α has the same role seen for Katz centrality. In particular, α should be chosen between 0 and $1/\rho(D^{-1}A)$, the maximum eigenvalue of the matrix $D^{-1}A$.

The Hub and Authority scores (Kleinberg, 1999) are recursively computed, such that a vertex has high authority centrality if it is pointed to by many hubs, i.e., by many other vertices with high hub centrality, while a vertex with high hub centrality points to many vertices with high authority centrality. Specifically, the Hyperlink-Induced Topic Search (HITS) is a link analysis algorithm created by Jon Kleinberg (Kleinberg (1999)) to evaluate web pages. In the Internet terminology, a good hub is a page that links to numerous other pages, while a good authority is a page that is linked by many

[16] In the general formulation the Katz centrality is expressed in an iterative way as $z_i = \alpha \sum_k a_{i,k} z_k + \beta$, where β is a constant representing some exogenous factors. Some simple computations show that two formulations are equivalent; moreover, as we are not interested in the absolute magnitude of the centrality, but in the ranking of node importance, we can assume without losing of generality $\beta = 1$.

[17] Also in this case, the general formulation is given by $p_i = \alpha \sum_k \frac{a_{k,i}}{d_k} p_k + \beta$.

different hubs. In other words, nodes with high authority scores can be considered the ones containing important information about a topic, while hubs are relevant because they point to them. For this reason, in the context of paper citations, hubs are generally referred as reviews and authorities as "relevant papers" for the topic. In our context, we can interpret as authorities such entities that released important patents and were therefore cited by several entities behaving as hubs. Of course, mixed-cases are possible: nodes showing both high authority and hub scores. Starting from this idea, the HITS algorithm assigns two scores to each page: the authority score, which measures the value of the page's content, and the hub score, which measures the value of its links to other pages. Generally speaking, the authority score of a node is the sum of the hub scores of all nodes that point to it; the hub score of a node is the sum of the authority scores of all nodes it points to. Mathematically:

$$au_i = \sum_j hu_j, \quad hu_i = \sum_j au_j \tag{A4}$$

where au_i is the authority score of node i, hu_i is the hub score of node i. Practically, after having initialized both scores to 1 for all nodes, the iterative update prescribes to compute the hub/authority score of each node according to (A4) and then normalize to prevent overflow and ensure convergence. (typically done by dividing each score by the Euclidean norm of the vector of scores). The procedure continues till to reach a stable set of values.

Network property	**Formula**	**Description**
Density	$\delta = \frac{L}{N(N-1)}$	Number of observed edges divided by the total potential connections
In Degree	$k_i^{in} = \sum_{j=1}^{N} a_{ji}$	Number of in-coming links
Out Degree	$k_i^{out} = \sum_{j=1}^{N} a_{ij}$	Number of out-going links
In Strength	$s_i^{in} = \sum_{j=1}^{N} w_{ji}$	Number of forward citations
Out Strength	$s_i^{out} = \sum_{j=1}^{N} w_{ij}$	Number of backward citations
In-clustering coefficient	$\frac{A^T A^2}{k^{in}(k^{in}-1)}$	Number of closed triangle with two links pointing to node i divided by the total number of triplets involving it

Betweenness Centrality	$b_i = \sum_{s \neq t \neq i} \frac{\sigma_{st}^i}{\sigma_{st}}$	Number of shortest paths connecting any pair of nodes and passing through node i divided by the the total number of shortest paths
Authority centrality	$x_i = \alpha \sum_j a_{j,i}\, y_j$	Defined recursively as the sum of the hub score (y_j) of nodes pointing to node i. α is a constant
Hub centrality	$y_i = \beta \sum_j a_{j,k}\, x_j$	Defined recursively as the sum of the authority score (x_j) of nodes pointed by node i. β is a constant
In-Katz centrality	$z_i^{in} = \mathbf{1}^T[(I - \alpha A)^{-1}]_i$	Number of weighted paths reaching the node i discounted by a dumping factor α. With $\alpha < 1/\rho(A)$, $\rho(A)$ maximum eigenvalue of the adjacency matrix A;$\mathbf{1}$ the unit vector
Page-Rank centrality	$pr_i = [(I - \alpha D^{-1} A)^{-1}\mathbf{1}]_i$	Variation of the Katz centrality: the centrality derived from node's i neighbors is proportional to their centrality divided by their out-degree. α is a constant, D is a diagonal matrix with $d_{jj} = k_j^{out}, 1 \leq j \leq N$

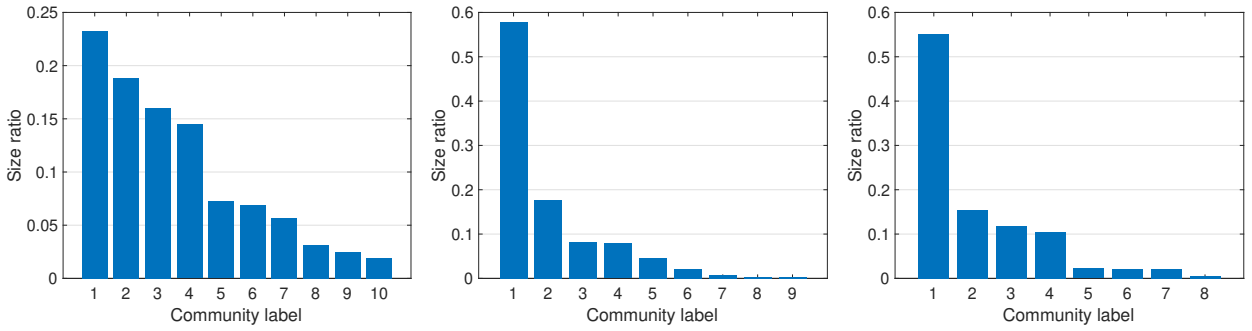
Table A1: Network properties

Appendix B Community detection

We perform a community detection analysis to identify groups of nodes more densely connected to each other than to the rest of the network. One of the possible approaches aims to maximize the *modularity* associated to network partitions (Newman, 2006) by comparing the number of edges within communities to the expected number of edges if nodes were randomly connected preserving some local constraints (generally node degree or strength). In this context, we opted for a weighted local constraint, node strength. There is a huge number of algorithms solving the maximization problem, we chose the popular and efficient "Louvain method" (Blondel et al., 2008). It is worth noting that the modularity score is strongly dependent on the network size, therefore we cannot compare it over the three periods. Moreover, this would be out of the scope, as we are interested more in understanding the organization of the citation networks over time after having identified the best partition according to the Louvain algorithm.

Across the three periods analyzed, we identified ten, nine, and eight communities, respectively. Figure 10 displays the ratio of communities to the overall network sizes. It highlights a significant difference between the first period and the following two periods, characterized by a large community that encompasses half of the nodes.

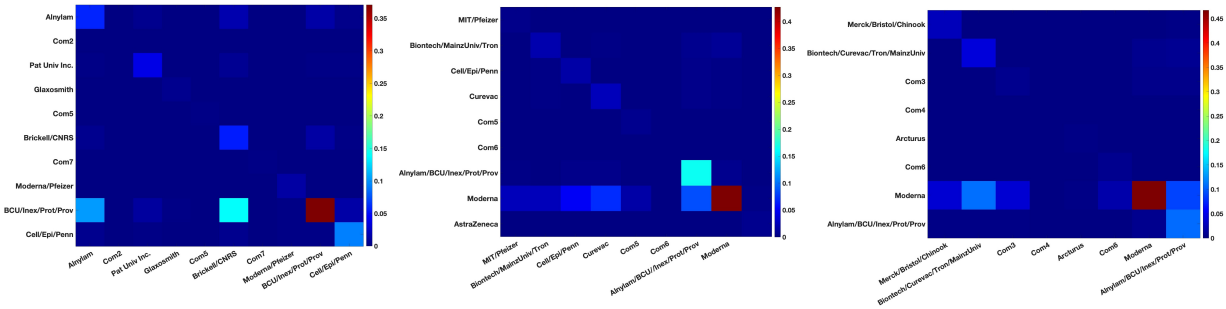
In Figure 9 we show within and between densities of the communities subnetworks for the three



(a) 2006-2010 (b) 2011-2015 (c) 2016-2020

Figure 8: **Community detection.** Ratio between the communities (identified with the Louvain algorithm) and the network sizes for the three periods under study

periods under study.



(a) 2006-2010 (b) 2011-2015 (c) 2016-2020

Figure 9: **Communities densities** Total number of citations within and between communities normalized with respect to the total number of citations for the specific period.

In what follows we investigate the structure and composition of the communities in the three periods, focusing on the role played by specific actors.

In the first period, the largest community contains Cellscript, Epicentre and University of Pennsylvania (fig.10 (a)). This community exhibits an average in-group clustering coefficient of 0.53, which is higher than the global one of 0.4, indicating strong internal cohesion. It suggests collaborative efforts focused on fundamental research and early technological developments. The presence of the University of Pennsylvania, alongside companies like Cellscript and Epicentre, highlights the importance of academic-industry partnerships in laying the groundwork for mRNA technology.

The second largest community shows a star-shaped configuration centered in Alnylam (citing 93% of the nodes in the group), which has specialized in RNA interference (RNAi) and therapeutic applications. In fact, the average clustering coefficient appears now to be smaller than the global one (0.19).

The third largest community (Fig.10 (b)) shows a pronounced star-like organization centered in Curevac (citing the 98% of nodes in the group, whose 67% has in degree equal to 1, i.e. it is cited only by Curevac) as confirmed by the very small average in-clustering coefficient (0.02). Curevac's dominant position in its community, with a nearly absolute star-like structure, underscores its pioneering efforts in mRNA vaccine technology. The small in-clustering coefficient of Curevac's community further highlights its role as a key innovator.

The fourth largest community comprises British Columbia University, Protiva, Proviva and Inex;

(fig.10 (b)). The diverse composition of this community highlights collaborative efforts focused on LNP technology. British Columbia University and Biotechnology companies like Protiva, Proviva and Inex were instrumental in developing LNPs, which are crucial for the delivery of mRNA into cells.

In the second period, post 2010, the largest community encompasses 58% of entities and presents a star-like organization, with Moderna pointing to the 97% of nodes in the group, followed by Novartis pointing to only 6% of them (Fig.10 (c)). Moderna tends to have a distinctive periphery of cited entities, with 73% of its neighbors having an indegree of one in the entire network, meaning they are cited only by Moderna. The average in-clustering coefficient of the community is 0.09, much lower than the global average of 0.3. During this period, Moderna emerges as the central innovator, reflecting its effort in developing mRNA-based vaccines and other therapeutic advancments. The second-largest community comprises 18% of the nodes and has a different organizational structure (fig.10 (d)). The top five entities with the highest out-degrees within this community are British Columbia University, Inex, Alnylam, Protiva and Proviva. The average of the in-clustering coefficient of this community is 0.43, greater than the global average. The structural organization of this community suggests focused advancements in LNP technology and RNA-based therapeutics. These companies work closely to enhance delivery mechanisms crucial for effective mRNA therapeutics and vaccines. It also highlights a regional concentration of expertise and collaboration in LNP technologies within North America, in Canada in particular, contributing to global mRNA delivery innovations.

The third largest community contains the 8% of actors and demostrates high clustering, with an average in-clustering coefficient of 0.79, much hogher than the global average. The top citing entities are BioNTech, Tron and Mainz University, all based in Germany, idicating a strong regional hub of innovation in mRNA vaccines within the country. Finally, another community exhibits a quasi star-like configuration centered in Curevac (also located in Germany) and includes highly clustered groups such as Cellscript, Epicenter, and the University of Pennsylvania, with an average in-clustering coefficient equal to 0.76.

In the third period (2016-2020) the largest community has the 55% of entities and includes beside Moderna also Novartis and University of Pennsylvania (fig.10 (e)). The organization remains very similar to the second period, with 78% of nodes cited only by Moderna, and the average in-clustering coefficient at 0.06, significantly lower than the global average of 0.25. The second largest community contains 16% of actors and resembles the second largest community of the second period (fig.10 (f)), with the top five citing entities being British Columbia University, Inex, Alnylam, Protiva and Proviva, showing high cohesion (the average in-clustering coefficient is , 0.51, doubling the global average). The third largest component contains the 12% of nodes and appears as a combination of two previously observed groups: a star-like organization centered in Curevac and a more clustered set containing BioNTech, Tron and Mainz University, with an average in-clustering coefficient slightly below the global average at 0.15. Finally, the fourth largest community contains the 10% of actors, with central nodes never observed in the previous periods: Bristol Meyers Squibb, Chinook Therapeutics and Merck.&co. having an average in-clustering coefficient in line with the global average at 0.22 (fig.10 (f)). This community could suggest an expansion of the field, with large pharmaceutical companies indicating a more global reach, potentially encompassing entities from North America and beyond. This could mark the entry of traditional large pharma companies into the mRNA space, diversifying the technology's applications.

The Sankey plot in Figure 11 shows the "movements" of nodes between communities from one period to the next. Each community is labelled with the name of the key-actors (as highlighted

in the community description). The thickness of flows between communities is proportional to the number of nodes, while the vertical bars represent the size of the community. Notably, nodes that appar exclusively in one community during a certain period are excluded from the visualization. Only the most relevant communities are shown: eight over ten in the first period, representing 92% of nodes; seven over nine in the second period representing 99% of nodes; four over eight in the third period representing 93% of nodes. The figure sheds light on the complex landscape of companies and institutions citations and their changes over time.

It is interesting to notice not only the dominant role of Moderna's community from the second period (as already observed before in terms of sizes) but also the composition of this group, which absorbed nodes from the Biontech/Mainz University/Tron community (from the first to the second period), while losing some companies in favour of Merck&co group and the German community around Mainz University. Additionally, the Alnylam and British Columbia University/Inex/Protiva/Proviva communities merged from the first to the second period, with their composition remaining rather stable in the third period. Finally, the group centered in Cellscript, Epicentre and Pennsylvania University appears stable in the first and second periods but it is absorbed by Moderna's community in the last one.

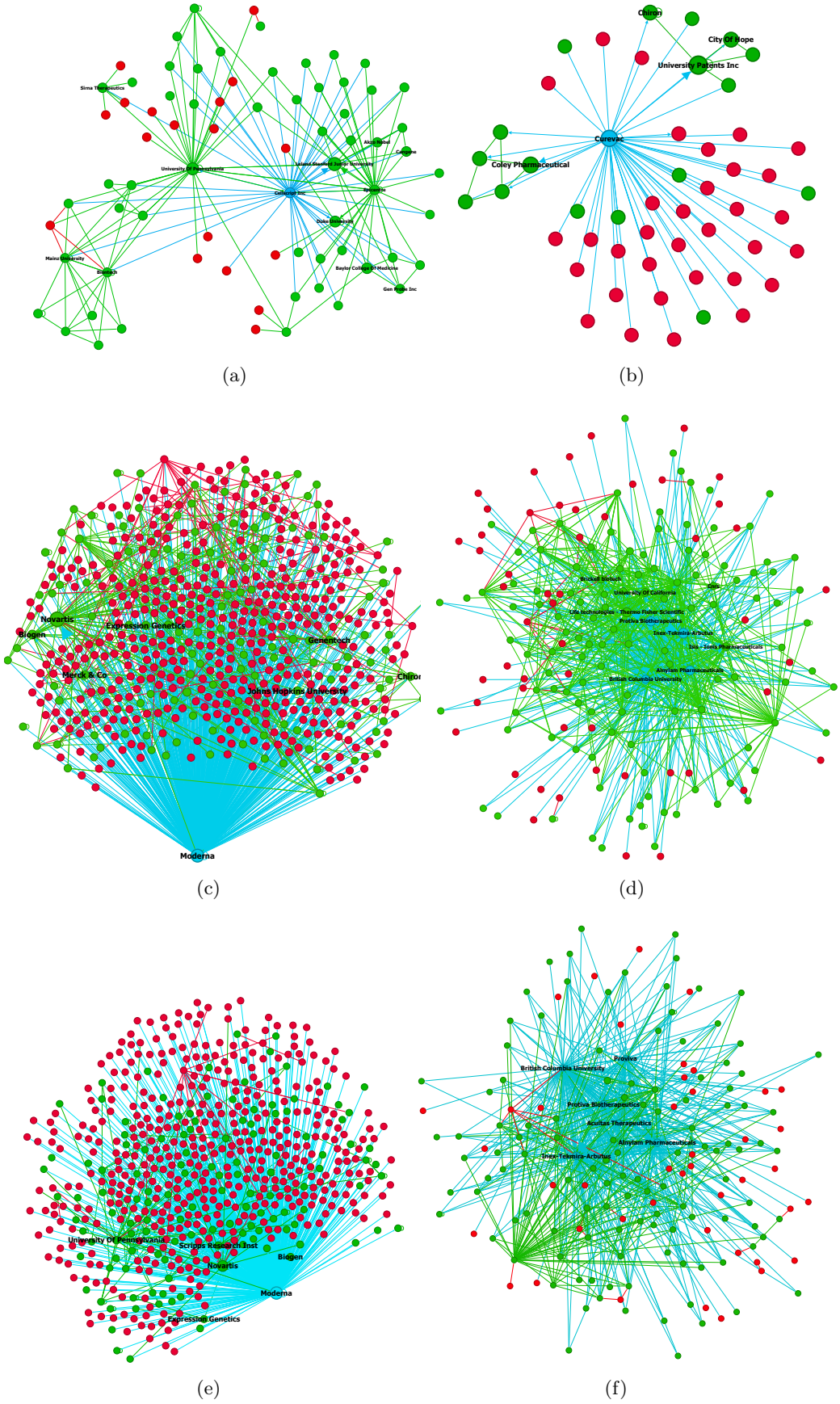


Figure 10: **Community detection analysis** Two examples of communities for the three periods under study. (a)-(b) First period; (c)-(d) second period; (e)-(f) third period. Node size is proportional to the number of forward citations; nodes color indicates: nodes with the highest out-strength (light blue); nodes receiving only one citation in the whole network (red); nodes receiving more than one citation in the whole network (green).

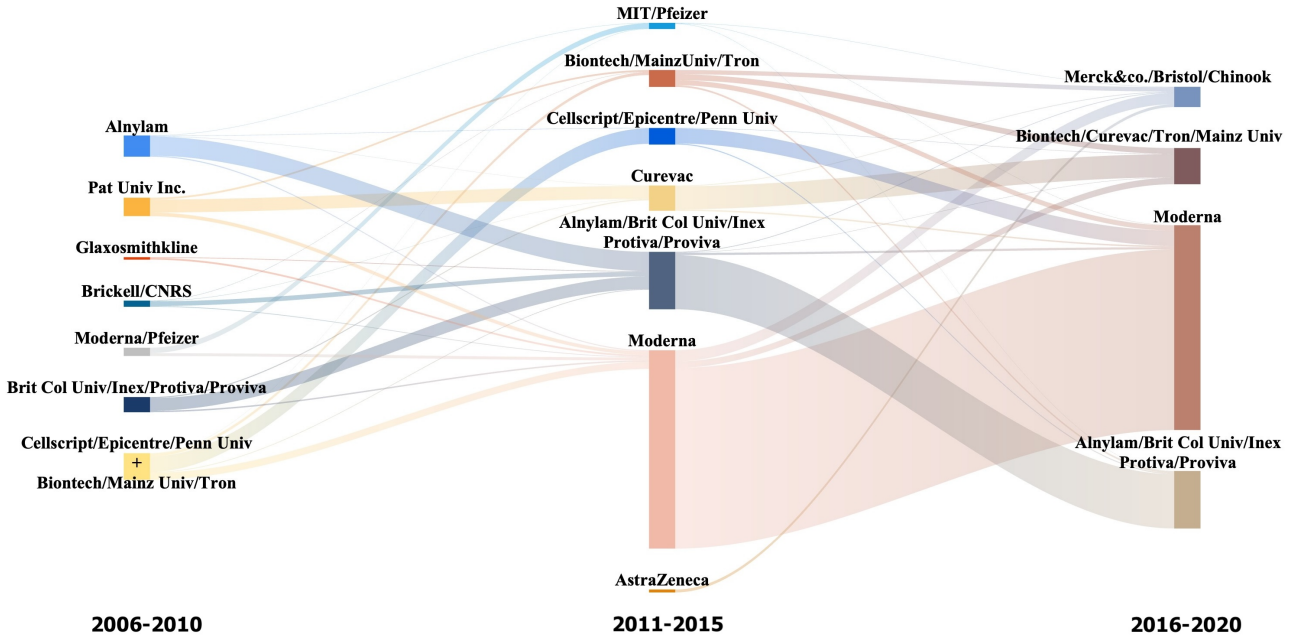


Figure 11: **Sankey plot.** "Movements" of nodes between communities from one period to the next. Each community is labeled with the names of key-actors (as highlighted in the community description). The thickness of the flows between communities is proportional to the number of nodes, while the vertical bars represent the size of each community.

Appendix C Credit Allocation

In this section we describe the procedure to allocate credit coming from main actors of mRNA vaccine commercialization to companies and institutions according to their role in the mRNA knowledge network.

We consider the weighted adjacency matrix associated to the mRNA knowledge network aggregated over the whole period under study, $W_{ij} \equiv (w_{ij})_{1\leq i,j\leq N}$; then, we compute the transition probability matrix associated to it, $M \equiv (m_{ij})_{1\leq i,j\leq N}$, and its dth powers:

$$m_{ij} = \frac{w_{ij}}{s_i^{out}}, \quad M^d = (m_{ij,d})_{1\leq i,j\leq N}^{d=1,2,...} \tag{C1}$$

where $m_{ij,d}$ represents the probability to reach node j in d steps for a random walker placed on node i . For $d = 1$ the probability associated to each link out going from node i simply corresponds to its weight normalized by the node i out strength (as in (C1)): the higher the number of backward citations the higher the probability to be the destination of the random walker first step. For $d = 2$, $m_{ij,2} = \sum_l m_{il}m_{lj}$ gives the sum of all weighted paths of length 2 going from node i to node j and so on for $d > 2$. It is worth noticing that since all directions where a link is present are allowed, the same company can be assigned several shares of node i credit during the redistribution process. The same holds for node i itself, as self loops are not excluded from the system. Once the transition probability matrix and its powers have been computed, its combination with the dumping factor β allows to compute the credit assigned to each node according to three different rules:

1. Markov approach

$$\pi_j^d = \left(\frac{m_{ij}^d}{\sum_{h\in V_i(d)} m_{ih}^d} \right) \beta^d \tag{C2}$$

   where $V_i(d)$ is the set of all nodes reachable from node i in d steps. It is worth noticing that, except for $d = 1$, the rows of M^d do not sum to 1. However, it is necessary to normalize them as in (C2) to properly distribute the share of node i profit assigned at distance level d.

2. Markov and in-Katz centrality

$$\pi_j^d = \left(\frac{m_{ij}^d}{\sum_{h\in V_i(d)} m_{ih}^d} \right) \left(\frac{z_h}{\sum_{h\in V_i(d)} z_h} \right) \beta^d \tag{C3}$$

   where z_i is the in-Katz centrality computed on the whole network according to (A2)

3. Markov and PageRank centrality

$$\pi_j^d = \left(\frac{m_{ij}^d}{\sum_{h\in V_i(d)} m_{ih}^d} \right) \left(\frac{pr_h}{\sum_{h\in V_i(d)} pr_h} \right) \beta^d \tag{C4}$$

   where pr_i is the PageRank centrality computed on the whole network according to (A3).

Example: profit redistribution

Let us assume that node i has profit $\Pi_i = 100$ (coming for example from the commercialization of the mRNA vaccine) and let us assume that half of this profit remains to node i (as the patent's owner and the one who succeeded in the commercialization), while the remaining profit will be redistributed among the other patents' assignees according to their technological and scientific contributions to the development of the vaccine. These quantities can be computed from the mRNA knowledge network using one of the three procedures introduced before (see tables 5, 6 for the main contributors of Moderna and BioNTech vaccines development). Hence, we can introduce a parameter α such that the credit that remains to node i is $\alpha\Pi_i = 50$, while $(1-\alpha)\Pi_i = 50$ is the credit to be allocated. Table C1 reports the credit to be divided among all nodes at distance $d \in \{1, 2, 3, \dots\}$ from node i, according to our choices of the parameters $\alpha = \beta = 0.5$.

Distance value (d)	**Credit to allocate**
1	25
2	12.5
3	6.25
4	3.125
...	...

Table C1: Credit share assigned to each distance value, d, and to be allocated among all nodes at such distance from node i.

In other words, the 25% of Π_i will be redistributed among all nodes at distance 1 from node i (i.e., directly cited by it), the 12.5% will be redistributed among all nodes at distance 2 from node i and so on. Generally speaking, the percentage of node i credit to be allocated at distance d from node i is calculated as $\pi_d = \alpha\beta^d\Pi$ (such that $\sum_d^\infty \pi_d = \sum_d^\infty \alpha\beta^d\Pi_i = \Pi_i$), with $0 < \alpha, \beta < 1$. Coming back to our example, all nodes at distance $d = 1$ from node i, will be allocated a share of the credit assigned to this "level", i.e. 25, according to their importance computed with one of the three aforementioned quantities. Hence, node j directly pointed by node i

1. will receive $p_{j,1} = 25 * (w_{ij}/s_i^{out})$
2. will receive $p_{j,1} = 25 * (w_{ij}/s_i^{out}) * z_j^{norm}$
3. will receive $p_{j,1} = 25 * (w_{ij}/s_i^{out}) * pr_j^{norm}$

where z_j^{norm} and pr_j^{norm} simply indicates the in-Katz and the PageRank centrality of node j, respectively, normalized with respect to the centrality values of all nodes at distance d from node i ($d = 1$ in this example).

Working Papers

Dipartimento di Politica Economica

1. *Innovation, jobs, skills and tasks: a multifaceted relationship.* M. Piva, M. Vivarelli. Vita e Pensiero, maggio 2018 (ISBN 978-88-343-3654-0)
2. *A bridge over troubled water: Interdisciplinarity, Novelty, and Impact.* M. Fontana, M. Iori, F. Montobbio, R. Sinatra. Vita e Pensiero, settembre 2018 (ISBN 978-88-343-3793-6)
3. *Concordance and complementarity in IP instruments*. M. Grazzi, C. Piccardo, C. Vergari. Vita e Pensiero, gennaio 2019 (ISBN 978-88-343-3879-7)
4. *Sustainable finance, the good, the bad and the ugly: a critical assessment of the EU institutional framework for the green transition*. L. Esposito, E.G. Gatti, G. Mastromatteo. Vita e Pensiero, febbraio 2019 (ISBN 978-88-343-3892-6)
5. *Technology and employment in a vertically connected economy: a model and an empirical test.* G. Dosi, M. Piva, M.E. Virgillito, M. Vivarelli. Vita e Pensiero, giugno 2019 (ISBN digital edition [PDF]: 978-88-343-4008-0)
6. *Testing the employment impact of automation, robots and AI: A survey and some methodological issues.* L. Barbieri, C. Mussida, M. Piva, M. Vivarelli. Vita e Pensiero, settembre 2019 (ISBN digital edition [PDF]: 978-88-343-4052-3)
7. *A new proposal for the construction of a multi-period/multilateral price index*. C.R. Nava, A. Pesce, M.G. Zoia. Vita e Pensiero, ottobre 2019 (ISBN digital edition [PDF]: 978-88-343-4114-8)
8. *Lo Stato Sociale: da "lusso" a necessità*. L. Campiglio. Vita e Pensiero, febbraio 2020 (ISBN digital edition [PDF]: 978-88-343-4184-1)
9. *Robots and the origin of their labour-saving impact*. F. Montobbio, J. Staccioli, M.E. Virgillito, M. Vivarelli. Vita e Pensiero, marzo 2020 (ISBN digital edition [PDF]: 978-88-343-4196-4)
10. *Business visits, technology transfer and productivity growth*. M. Piva, M. Tani, M. Vivarelli. Vita e Pensiero, marzo 2020 (ISBN digital edition [PDF]: 978-88-343-4210-7)
11. *Technology, industrial dynamics and productivity: a critical survey*. M. Ugur, M. Vivarelli. Vita e Pensiero, settembre 2020 (ISBN digital edition [PDF]: 978-88-343-4406-4)
12. *Back to the past: the historical roots of labour-saving automation*. J. Staccioli, M.E. Virgillito. Vita e Pensiero, novembre 2020 (ISBN digital edition [PDF]: 978-88-343-4473-6)
13. *The present, past, and future of labor-saving technologies.* J. Staccioli, M.E. Virgillito. Vita e Pensiero, dicembre 2020 (ISBN digital edition [PDF]: 978-88-343-4479-8)
14. *Why Do Populists Neglect Climate Change? A Behavioural Approach*. L.A. Lorenzetti. Vita e Pensiero, dicembre 2020 (ISBN digital edition [PDF]: 978-88-343-4483-5)
15. *Relative wages, payroll structure and performance in soccer. Evidence from Italian Serie A (2007-2019)*. C. Bellavite Pellegrini, R. Caruso, M. Di Domizio. Vita e Pensiero, gennaio 2021 (ISBN digital edition [PDF]: 978-88-343-4490-3)
16. *Robots, AI, and Related Technologies: A Mapping of the New Knowledge Base*. E. Santarelli, J. Staccioli, M. Vivarelli. Vita e Pensiero, gennaio 2021 (ISBN digital edition [PDF]: 978-88-343-4499-6)
17. *Detecting the labour-friendly nature of AI product innovation*. G. Damioli, V. Van Roy, D. Vertesy, M. Vivarelli. Vita e Pensiero, aprile 2021 (ISBN digital edition [PDF]: 978-88-343-4600-6)
18. *Circular Economy Approach: The benefits of a new business model for European Firms*. C. Bellavite Pellegrini, L. Pellegrini, C. Cannas. Vita e Pensiero, luglio 2021 (ISBN digital edition [PDF]: 978-88-343-4817-8)
19. *The impact of cognitive skills on investment decisions. An empirical assessment and policy suggestions.* L. Esposito, L. Marrese. Vita e Pensiero, luglio 2021 (ISBN digital edition [PDF]: 978-88-343-4822-2)
20. *"Thinking of the end of the world and of the end of the month": the Impact of Regenerative Agriculture on Economic and Environmental Profitability*. L.A. Lorenzetti, A. Fiorini. Vita e Pensiero, ottobre 2021 (ISBN digital edition [PDF]: 978-88-343-4898-7)

21. *Labour-saving automation and occupational exposure: a text-similarity measure*. F. Montobbio, J. Staccioli, M.E. Virgillito, M. Vivarelli. Vita e Pensiero, novembre 2021 (ISBN digital edition [PDF]: 978-88-343-5089-8)

22. *Climate reputation risk and abnormal returns in the stock markets: a focus on large emitters*. G. Guastella, M. Mazzarano, S. Pareglio, A. Xepapadeas. Vita e Pensiero, novembre 2021 (ISBN digital edition [PDF]: 978-88-343-5092-8)

23. *Carbon Boards and Transition Risk: Explicit and Implicit exposure implications for Total Stock Returns and Dividend Payouts*. M. Mazzarano, G. Guastella, S. Pareglio, A. Xepapadeas. Vita e Pensiero, novembre 2021 (ISBN digital edition [PDF]: 978-88-343-5093-5)

24. *Innovation and employment: a short update*. M. Vivarelli. Vita e Pensiero, gennaio 2022 (ISBN digital edition [PDF]: 978-88-343-5113-0)

25. *AI technologies and employment. Micro evidence from the supply side*. G. Damioli, V. Van Roy, D. Vertesy, M. Vivarelli. Vita e Pensiero, gennaio 2022 (ISBN digital edition [PDF]: 978-88-343-5119-2)

26. *The Effect of External Innovation on Firm Employment*. G. Arenas Díaz, A. Barge-Gil, J. Heijs, A. Marzucchi. Vita e Pensiero, febbraio 2022 (ISBN digital edition [PDF]: 978-88-343-5146-8)

27. *The North-South divide: sources of divergence, policies for convergence*. L. Fanti, M.C. Pereira, M.E. Virgillito. Vita e Pensiero, maggio 2022 (ISBN digital edition [PDF]: 978-88-343-3524-4)

28. *The empirics of technology, employment and occupations: lessons learned and challenges ahead*. F. Montobbio, J. Staccioli, M.E. Virgillito, M. Vivarelli. Vita e Pensiero, novembre 2022 (ISBN digital edition [PDF]: 978-88-343-5383-7)

29. *Cognitive biases and historical turns. An empirical assessment of the intersections between minds and events in the investors' decisions*. L. Esposito, L. Malara. Vita e Pensiero, gennaio 2023 (ISBN digital edition [PDF]: 978-88-343-5420-9)

30. *Interaction between Ownership Structure and Systemic Risk in the European financial sector*. C. Bellavite Pellegrini, R. Camacci, L. Pellegrini, A. Roncella. Vita e Pensiero, febbraio 2023 (ISBN digital edition [PDF]: 978-88-343-5446-9)

31. *Was Robert Gibrat right? A test based on the graphical model methodology*. M. Guerzoni, L. Riso, M. Vivarelli. Vita e Pensiero, marzo 2023 (ISBN digital edition [PDF]: 978-88-343-5457-5)

32. *A North-South Agent Based Model of Segmented Labour Markets. The Role of Education and Trade Asymmetries*. L. Fanti, M.C. Pereira, M.E. Virgillito. Vita e Pensiero, maggio 2023 (ISBN digital edition [PDF]: 978-88-343-5529-9)

33. *Innovation and the Labor Market: Theory, Evidence and Challenges*. N. Corrocher, D. Moschella, J. Staccioli, M. Vivarelli. Vita e Pensiero, giugno 2023 (ISBN digital edition [PDF]: 978-88-343-5580-0)

34. *The Effect of Economic Sanctions on World Trade of Mineral Commodities. A Gravity Model Approach from 2009 to 2020*. R. Caruso, M. Cipollina. Vita e Pensiero, dicembre 2023 (ISBN digital edition [PDF]: 978-88-343-5686-9)

35. *Education and Military Expenditures: Countervailing Forces in Designing Economic Policy. A Contribution to the Empirics of Peace*. A. Balestra, R. Caruso. Vita e Pensiero, gennaio 2024 (ISBN digital edition [PDF]: 978-88-343-5757-6)

36. *Vulnerability to Climate Change and Communal Conflicts: Evidence from Sub-Saharan Africa and South/South-East Asia*. S. Balestri, R. Caruso. Vita e Pensiero, maggio 2024 (ISBN digital edition [PDF]: 978-88-343-5829-0)

37. *Assessing changes in EU innovation policy programs: from SME instrument to EIC accelerator for start-up funding*. M. del Sorbo, C. Faber, M. Grazzi, F. Matteucci, M. Ruß. Vita e Pensiero, luglio 2024 (ISBN digital edition [PDF]: 978-88-343-5860-3)

38. *AI as a new emerging technological paradigm: evidence from global patenting.* G. Damioli, V. Van Roy, D. Vertesy, M. Vivarelli. Vita e Pensiero, settembre 2024 (ISBN digital edition [PDF]: 978-88-343-5873-3)

39. *The KSTE+I approach and the AI technologies*. F. D'Alessandro, E. Santarelli, M. Vivarelli. Vita e Pensiero, settembre 2024 (ISBN digital edition [PDF]: 978-88-343-5880-1)

40. *Quo Vadis Terra? The future of globalization between trade and war*. L. Esposito, E.G. Gatti, G. Mastromatteo. Vita e Pensiero, settembre 2024 (ISBN digital edition [PDF]: 978-88-343-5895-5)

41. *The Agents of Industrial Policy and the North-South Convergence: State-Owned Enterprises in an International-Trade Macroeconomic ABM.* L. Fanti, M.C. Pereira, M.E. Virgillito. Vita e Pensiero, ottobre 2024 (ISBN digital edition [PDF]: 978-88-343-5909-9)

42. *The impact of US elections on US defense industry: Firm-level evidence from 1996 to 2022*. A. Balestra, R. Caruso. Vita e Pensiero, January 2025 (ISBN digital edition [PDF]: 978-88-343-5937-2)

43. *Forecasting the Impact of Extreme Weather Events on Electricity Prices in Italy: A GARCH-MIDAS Approach with Enhanced Variable Selection*. M. Guerzoni, L. Riso, M.G. Zoia. Vita e Pensiero, January 2025 (ISBN digital edition [PDF]: 978-88-343-5938-9)

44. *The Theoretical Properties of Novel Risk-Based Asset Allocation Strategies using Portfolio Volatility and Kurtosis*. M.D. Braga, L. Riso, M.G. Zoia. Vita e Pensiero, January 2025 (ISBN digital edition [PDF]: 978-88-343-5939-6)

45. *Sustainable Finance in the New Geo-Political Era: A Difficult Balancing Act*. L. Esposito, M. Cocco. Vita e Pensiero, February 2025 (ISBN digital edition [PDF]: 978-88-343-5940-2)

46. *New technologies and employment: the state of the art*. M. Vivarelli, G. Arenas Díaz. Vita e Pensiero, March 2025 (ISBN digital edition [PDF]: 978-88-343-5941-9)

47. *Leveraging Knowledge Networks: Rethinking Technological Value Distribution in mRNA Vaccine Innovations*. R. Mastrandrea, F. Montobbio, G. Pellegrino, M. Riccaboni, V. Sterzi. Vita e Pensiero, March 2025 (ISBN digital edition [PDF]: 978-88-343-5991-4)