

Multilayer patent citation networks: A comprehensive analytical framework for studying explicit technological relationships

Kyle Higham^{*1}, Martina Contisciani^{†2}, and Caterina De Bacco^{‡2}

¹Institute of Innovation Research, Hitotsubashi University, Japan

²Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen, Germany

Abstract

The use of patent citation networks as research tools is becoming increasingly commonplace in the field of innovation studies. However, these networks rarely consider the contexts in which these citations are generated and are generally restricted to a single jurisdiction. Here, we propose and explore the use of a multilayer network framework that can naturally incorporate citation metadata and stretch across jurisdictions, allowing for a complete view of the global technological landscape that is accessible through patent data. Taking a conservative approach that links citation network layers through triadic patent families, we first observe that these layers contain complementary, rather than redundant, information about technological relationships. To probe the nature of this complementarity, we extract network communities from both the multilayer network and analogous single-layer networks, then directly compare their technological composition with established technological similarity networks. We find that while technologies are more splintered across communities in the multilayer case, the extracted communities match much more closely the established networks. We conclude that by capturing citation context, a multilayer representation of patent citation networks is, conceptually and empirically, better able to capture the significant nuance that exists in real technological relationships when compared to traditional, single-layer approaches. We suggest future avenues of research that take advantage of novel computational tools designed for use with multilayer networks.

^{*}higham@iir.hit-u.ac.jp

[†]martina.contisciani@tuebingen.mpg.de

[‡]caterina.debacco@tuebingen.mpg.de

1 Introduction

Patent citations have found successful application in a wide swathe of contexts, from understanding knowledge spillovers (Jaffe et al., 1993, Sorenson et al., 2006, Jaffe and de Rassenfosse, 2017, Berkes and Gaetani, 2021) to the characterization of technological change (Fleming, 2001, Choi and Park, 2009, Huenteler et al., 2016). The vast majority of this research is conducted using only patent data from a single jurisdiction and often ignores important citation context. However, as innovation and patent filings become increasingly global endeavours (Fink et al., 2016, Danguy, 2017), there are many situations where it is important to think of ‘the patent system’ as a set of quasi-coordinated processes operating across jurisdictional boundaries (Petit et al., 2021).

This coordination is desirable because the same invention can be patented in multiple jurisdictions; there are clear efficiency gains to be made if information discovered or produced during the patent prosecution process can be shared between jurisdictions (Chun, 2011). Patent families arise because these related applications, which simultaneously progress through multiple patent offices, are legally linked through their first filing. The Paris Convention¹ allows applicants to apply in multiple jurisdictions and claim the filing date of the first application, known as the priority date, as the effective filing date for subsequent applications (provided these occur within 12 months of the priority date). Information sharing between offices creates, by design, some redundancy in the information generated for family members across offices, but not enough that a complete picture can be pieced together from a single jurisdiction’s data. The existence of patent families provides the opportunity to form the most complete set of information about a particular invention that can be obtained from patent data and allows us to link metadata across jurisdictional boundaries (Nakamura et al., 2015). In this work, we demonstrate the utility of these linkages in the context of patent citation networks.

The family-level view would suggest that only using data from a single jurisdiction leaves a lot of potentially relevant information unexamined (Bakker et al., 2016). In the more and more common scenario where multiple family members exist across multiple jurisdictions, citations will often only be made to one family member.² As such, the citation network that is obtained from any single jurisdiction necessarily represents a subset of the complete network for the set of inventions under examination. While information sharing between offices will increase the amount of overlap between these networks, it does not make family-level analyses redundant, for two reasons. First, the amount of information sharing, and the modes for doing so, between patent offices has changed significantly in recent years.³ In particular, advances in information technology allow

¹Paris Convention for the Protection of Industrial Property (1883).

²Search reports will often list equivalents of the prior art that is cited, however, this additional information is not explicitly included in the associated data sets.

³See, e.g., <https://www.wipo.int/case/en/>.

patent offices to coordinate much more effectively than they did 20 years ago. Yet, some patents filed 20 years ago are only expiring now, so these patents can still be important sources of information when studying contemporary innovation. Second, many patents are *not* filed in multiple offices, and some applicants only select a few strategically important jurisdictions where they would like to protect their intellectual property. That is, the nodes and links in the citation networks of each jurisdiction are unique, and so each network contains a huge amount of potentially pertinent information that is unique to that jurisdiction. In the empirical sections of this work, we take a very conservative approach and only consider nodes that are shared across jurisdictions, as described in detail in Section 2.3.

Even for shared nodes (defined here as patents granted in multiple jurisdictions), however, the sets and types of citations made by each patent office can differ greatly, as shown in Figure 1. The primary reason for this disagreement is that different jurisdictions abide by different legal guidelines that describe when and how citations should be made. These sets of guidelines are not without strong similarities, however, and a careful reading offers pathways towards sensible aggregation or comparison of these sets of citations (Higham and Yoshioka-Kobayashi, 2022). This has become particularly feasible in recent times as more and more offices now provide metadata about citation context, such as whether the cited patent was so similar to the application as to render the latter unpatentable, or whether the cited patent was added to simply define the state of the art. A secondary reason for disagreement between jurisdictions is, in fact, a commonality: examiners in all jurisdictions are humans with limited time to examine any particular patent (see, e.g., Frakes and Wasserman, 2017). Often, it is simply not possible to find every relevant piece of prior art, particularly when language barriers are taken into consideration. Indeed, in combination with simple differences of opinion, this limitation means it is unlikely that two examiners in the same patent office would find exactly the same set of prior art (Wada, 2016). Therefore, using family-level information gives us the search result of more examiner-hours as well as the multiple opinions of what should be considered relevant prior art.

As citations made by different offices are made according to different sets of guidelines, treating these citations as equally informational may lead to misleading results. Indeed, some suggest that citations of the *same* type have become less informative over time (Kuhn et al., 2020). It is therefore important to aggregate family-level information sensibly. We propose that a multilayer network framework provides a natural representation of the patent citation network that readily incorporates differences in citation type. After all, multiple networks anchored by common nodes is the very definition of a multilayer network (De Domenico et al., 2013, Kivelä et al., 2014, Porter, 2018).

Within the multilayer framework, described in more detail for our chosen context in Section 2.2, each layer of the network represents a single link type, each node represents

a patent family (which may exist in multiple layers), and each link represents a citation between families (of the type defined by the layer). As such, the global patent citation network is an inherently multilayer system; no abstraction is required. Further, this framework is particularly flexible. For example, layers can represent jurisdictions, and links within each jurisdictional layer can represent the citations found on the front page(s) of the family member(s) granted by that jurisdiction. From this point, it is possible to layer as many jurisdictions as desired onto the network, provided there are family linkages existing between the layers. It is also possible to split these layers further, according to citation metadata that inform us of the reason for, or source of, a particular citation. This flexibility is particularly valuable when certain types of citations are irrelevant, or may even be considered pure noise, with respect to a particular research question. For example, one studying knowledge flow within a multilayer framework may not wish to consider citations discovered by the examiner, and may even want to add an additional layer for citations found in the patent specification (Verluisse et al., 2020).

However, it is not clear, a priori, whether a multilayer framework adds any information over and above that which can be found in ‘flattened’ family citation networks wherein citation context is disregarded and only link existence is examined (Nakamura et al., 2015). Thus, in order for the multilayer framework to be feasible as a research tool, it is important to first demonstrate a significant gain in information content relative to the flattened, global family citation network, or even the more commonly-used jurisdictionally restricted citation networks. To this end, we explore the information content of the triadic patent family network, wherein all layers contain the same set of nodes. This set consists of families containing at least one member granted in each of the triadic patent offices: the United States Patent and Trademark Office (USPTO), the European Patent Office (EPO), and the Japan Patent Office (JPO). The triadic offices have historically granted the majority of patents globally and contain rich and accessible citation information. Specifics about the data used in this work can be found in Section 2.3.

In this work, we first construct a multilayer family-family triadic citation network, wherein layers can be separated by jurisdiction and citation context (such as whether the citation was added by an applicant or examiner). In practice, the appropriate set of contexts can be selected based on the use-case; in this work, the additional context we consider is whether a citation was likely to have been found by the examiner or by the applicant (which is not always explicit), as we expect the differing motivations for citation between these groups affect the nature of the technological relationships reflected by these citations. We then conduct an interdependence analysis to check for redundancy of information content between the layers, finding that significant complementary information exists between jurisdictions. A community detection procedure is then conducted on the multilayer network and two comparison networks: the flattened multilayer network containing the same set of links but without information about jurisdiction or

citation context, and the US-only subset of the citation network, also flattened. The former comparison tests the role of citation context, while the latter is included as the most commonly used patent citation network in prior research. We observe, graphically, nuanced differences in inferred community structure between the multilayer network and the comparison networks.

To add colour to these differences, we examine the relationships between inferred community partitions and the technology classes of the families that comprise them. For the multilayer network communities and those of the two flattened comparison networks, we project the bipartite community-class network onto the class nodes and directly compare these projections with established class-class networks (co-classification and inter-class citation linkage) with known-node-correspondence methods. We are also able to directly measure the diversity of communities, and the spread of classes between communities to inform our interpretation of the direct network comparisons.

When compared to the other two networks, we find that the multilayer case produces communities that more closely reflect the known technological relationships implied by the established class-class networks, at both micro- and meso-scales. Further, while technological classes are more splintered across communities in the multilayer case, the internal diversity of communities is lower than the comparison networks once we account for the known technological similarity of classes. These results suggest that, even within our conservative empirical framework, citation context is an important source of information about the nature and importance of the particular technological relationships codified by citation linkages, and that examination of multilayer citation networks using novel computational techniques is an exciting and relevant avenue for future research.

The rest of the paper is structured as follows. Section 2 introduces both patent families and multilayer networks and discusses how the former naturally forms the latter in the context of citation networks. Section 2.3 describes the data we use in this work, how this forms the multilayer networks and why specific subsets of families and citations are selected for analysis. Section 3 describes the empirical procedures that we use to test and compare the information content of the multilayer citation network relative to single-layer networks and describes the results obtained. Lastly, Section 4 concludes and discusses the limitations and extensions of this research.

2 Multilayer patent networks

2.1 Patent Families

The rights bestowed by patents are only enforceable in the jurisdiction in which the patent was granted. To obtain these rights in more than one jurisdiction, an applicant first files in a single jurisdiction (often their local patent office), starting the clock on the

period during which they can file for the same invention in other jurisdictions. For the next 12 months, all subsequent filings can ‘claim priority’ from this initial application and inherit the latter’s filing date as its own for the purposes of examination (provided the same content is covered in the application).

There are two primary modes through which an invention can claim priority from an earlier application: the Paris Convention and the Patent Cooperation Treaty (PCT). The former lays down the guidelines for the treatment of foreign patent applications among the contracting parties, including the time limit on priority claims as described above. The latter, for our purposes here, is effectively an attempt to streamline and harmonise the process of patenting in multiple jurisdictions.⁴ This process does not result in a patent, but rather a preliminary prior art search report, and allows the applicant to nominate the jurisdictions to which they would like to apply for a patent without having to apply at each office separately. Priority can be claimed from a PCT filing, and PCT filings can themselves claim priority from an earlier filing at a local office.

After a patent application has reached a local office, the applicant may want to fine-tune their claims or even be asked to split the described invention into two separate patent applications.⁵ The inventor is not able to disclose new information during this process, and thus the claims made by the ‘new version’ of the application must be contained within the scope of the initial disclosure. These subsequent filings may claim the priority date of the initial filings and are referred to as ‘continuing applications’.

Patent families, in general, link patents and applications through their priority filing. The resulting ‘family trees’ can be complex and, as such, several types of families exist (Martinez, 2010, Martínez, 2011). ‘Simple’ patent families (as defined by the EPO for their DOCDB database) each consist of a set of patents and applications that are all linked to the same priority filing. This type of family is the one on which we focus in this work, and we will henceforth drop ‘simple’. As such, families can be made up of sets of documents from several jurisdictions, each of which may contain multiple documents. Other families may only consist of a single application in a single jurisdiction.

Families are the unit of analysis for the current work for two reasons. First, they generally align with what one usually thinks of as a single ‘invention’ (Martínez, 2011) and it is the relationships between inventions that we usually aim to capture with citation data. Second, they link inventions across jurisdictions, and therefore allow the alignment of jurisdiction-specific citation networks and, therefore, the introduction of the multilayer network as a potentially useful analytical, conceptual, and mathematical tool to study technological relationships.

From the perspective of data availability, detail, and volume, the obvious choice of data set for testing the utility of the multilayer framework are those patents granted

⁴<https://www.wipo.int/pct/en/>.

⁵See, e.g., Paris Convention for the Protection of Industrial Property (1883), Article 4G.

by the three (historically) largest patent offices, also known as the triadic offices: the USPTO, EPO, and JPO. Further, we wish to take a particularly conservative empirical approach to these initial explorations of the multilayer citation network. To do this, we only consider patent families that have granted members in all layers of interest (‘triadic families’) and only consider citations among these families.⁶

Theoretically, each office examining these triadic applications has access to the same information regarding prior art, and they share much of what they find with the other two offices, directly or indirectly (Wada, 2020, Petit et al., 2021). For this reason, granted members have all had the same opportunities to link to other (older) families in each layer, maximising potential redundancies between layers in the citation network. The exclusion of families that are not triadic, therefore, is why we think of this analysis as likely to produce very conservative results when compared to those that may be obtained for a network without such exclusions.

We also note triadic patent families are often used as a binary indication of a ‘high-quality’ invention (de Rassenfosse and van Pottelsberghe, 2009, Tahmooresnejad and Beaudry, 2019); after all, the applicants thought it was worth the time and money to patent their invention in three of the largest markets in the world. By this logic, our multilayer network consists exclusively of ‘high-quality’ patent families⁷ and excludes much controversial subject matter that are not universally patentable (Biddinger, 2000).

A simplified diagram of a multilayer network of citations between triadic families is shown in Figure 1, with full details of the families included in this example described in Appendix C. Note that the multilayer network that we analyse in this paper treats sub-layers, such as whether a citation has been used in a rejection decision (shown in red in Figure 1), as distinct layers. This results in seven layers in total, as the EPO also provides information about whether a citation originated from the international search report (conducted outside the EPO) or the local search report.

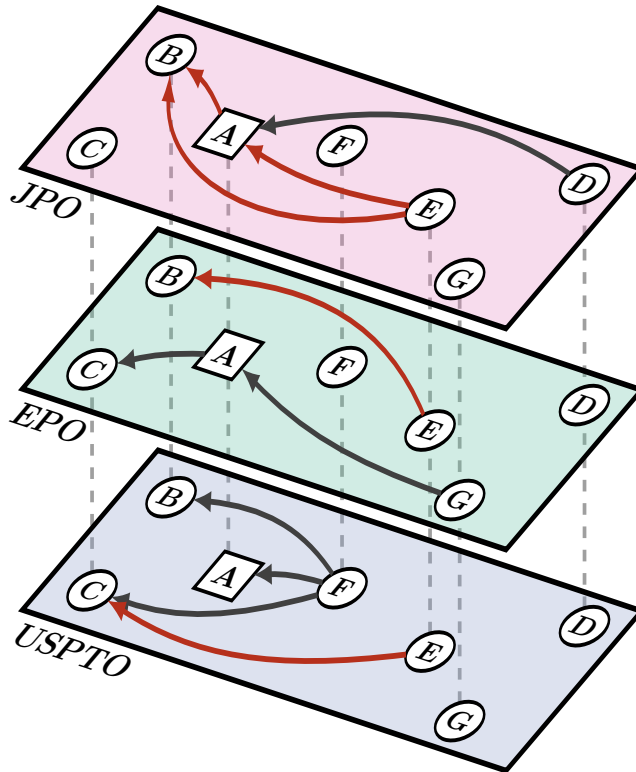
2.2 Multilayer networks

Multilayer networks have received particular attention in the past decade (De Domenico et al., 2013, Kivelä et al., 2014, Boccaletti et al., 2014, Cimini et al., 2019), and the development of mathematical and computational tools for their analysis, as well as their timely application, remains a very active field of research across many domains (Gallotti et al., 2016, Vaiana and Muldoon, 2020, Harvey et al., 2021, Yuvaraj et al., 2021, van der Marel et al., 2021). In this work, we not only suggest that patent citation networks are naturally multilayered, but aim to introduce the multilayer framework to the innovation studies community to promote the timely application of novel computational tools that

⁶The applicants to these offices, however, may be based outside these jurisdictions.

⁷Note that this is a very narrow view of patent quality. For a comprehensive discussion, refer to Higham et al. (2021).

Figure 1: **Exemplar subset of the multilayer patent citation network.** A multilayer representation of a typical subset of the inter-family patent citation network we consider in this work. Nodes and links comprise the multilayer ego network of patent family **A**, the USPTO equivalent of which is “Power source apparatus” (US6819081B2), initially filed in January 2002 by Sanyo Electric Co., Ltd. at the JPO. Each layer represents the inter-family citations made by a different patent office, and red links are those used to justify a (non-final) rejection of the application that was examined in that layer. All data represented here is subject to the restrictions described in Section 2.3 and is, therefore, an extremely simplified version of the complete ego network. Details of the families represented can be found in Appendix C.



are currently being developed.

To date, the vast majority of the studies that explicitly place patent citation data into a network setting use a single-layer framework (Von Wartburg et al., 2005, Valverde et al., 2007, Clough et al., 2015, Nakamura et al., 2015, Funk and Owen-Smith, 2017, Wu et al., 2019, Higham et al., 2019, Mariani et al., 2019). That is, there is only one type of link (i.e., a citation) between nodes in the network. This approach often makes practical sense, such as when one lacks citation metadata that may be used to distinguish or ‘colour’ the links, or if only one link type is of interest. However, a multilayer network framework is able to naturally incorporate citation metadata, if it exists, into the network structure.⁸ As an analogy, let us consider the public transport network of a large city

⁸In a related work, also using triadic patents, Morrison et al. (2014) use a multiplex PageRank to assess the centrality of technology classes where layers are defined by inventor location. However, citation source and context are not considered.

containing several different forms of transport, each with its own network of routes and stations. There are usually many points of overlap between these network layers to allow passengers to transfer between modes of public transport, such as a bus stop at a train station. These transfer points link the different network layers together. From both mathematical and computational perspectives, this kind of network is fundamentally different from single-layered networks, particularly when the different layers are defined by links with very different properties (Aleta et al., 2017, Ibrahim et al., 2021). In the public transport context, these properties can be straightforward, such as speed, price, comfort, or environmental harm, or more computationally complex, such as sensitivity to link removal and amenability to rerouting (De Domenico et al., 2014).

In the domain of patent citation networks, each jurisdiction has a set of applications and patents that each contain a set of citations made to other documents. Each of those citations comes with context (Higham and Yoshioka-Kobayashi, 2022). This context can be whether the prior art was discovered by the examiner, the justification for its addition to the document, the relationship between the citing and cited firms, or any other citation metadata that may be obtained or constructed. For many research questions that rely on information derived from the citation network, this information is important to retain, just as it is important to know whether two nodes in a transport network are connected by a bus, an airplane, or a ferry.

At the same time, every patent is part of a family (even if there is only one member). When families contain members filed in multiple jurisdictions, the citation networks associated with each jurisdiction can be linked, just as a bus may stop at a train station, or a train may stop at an airport. Of course, patent applicants are under no obligation to file for a patent on the same invention in multiple jurisdictions. That is, a node (patent family) may not exist in all layers of the network. Not every bus stop is a train station, nor vice versa. The full patent citation network is a true ‘multilayer’ network in this sense. In this work, however, we focus on the subset of nodes that exist across all three layers of interest (the triadic offices). The justifications for this choice are discussed in Section 2.3. The network we define in this work, therefore, is a special case of a multilayer network wherein the layers are node-aligned (Kivelä et al., 2014). Extensions of this work to a more general multilayer framework are discussed in Section 4.

Multilayer networks share many characteristics of interest that are found in single-layer networks; indeed, much of the early research on multilayer networks involved adapting concepts from single-layer networks to this new framework (Berlingerio et al., 2011, Bródka et al., 2012, De Domenico et al., 2013, Battiston et al., 2014). For our purposes, in order to demonstrate the utility of the multilayer framework, it is necessary to compare the network properties derived in this setting to those obtained from the equivalent, flattened single-layer network, wherein citation metadata is ignored (partially or wholly).

The domain within which we choose to explore differences between the multilayer

and single-layer frameworks, in the patent citation context, is community detection. The natural grouping of nodes is one of the characteristic features of real-world networks and plays a significant role in describing the structure of the network at scales between node-level and global-level network statistics (Wasserman et al., 1994, Newman and Girvan, 2004, Fortunato, 2010). Often, innovation researchers are interested in the composition of, and interaction between, close-knit groups of meso-scale objects such as groupings of similar technologies (Lee et al., 2015, Alstott et al., 2017, Balland and Rigby, 2017, Yan and Luo, 2017, Mejia and Kajikawa, 2020), and the application of community detection to the multilayer citation network leaves room for direct comparison between our results and these objects that we usually work with. Lastly, community detection can be applied to both multilayer and single-layer networks, which will allow for comparisons between the resultant communities.

2.3 Data

The multilayer citation network we construct is generated by citations made by triadic patents and only includes those made to and by triadic families. For the purposes of the current work, *triadic patents* are patents granted by one of the triadic offices that have family members, or equivalents, granted by the other two triadic offices. *Triadic families*, on the other hand, will refer to the full set of documents belonging to a family that contains triadic patents.⁹ These sets include both applications and patents and may be filed at or granted by offices outside the three triadic offices (provided that they are within a family containing triadic patents).

There are several reasons for choosing this subset of nodes and links to define our network, beyond the aforementioned desire to be conservative in our empirical design. The first is that we require well-defined layers. By restricting the citing patents to those granted by the triadic offices, the links (and, therefore, network layers) are defined by the citation context (e.g., the jurisdiction where it was made and the reason it was added), which isn't available for many offices. Second, restricting the cited families to those that are also triadic means that there are no cross-layer citations, which significantly simplifies the network from a mathematical perspective. For example, a US triadic patent citing a pre-grant publication that was only filed at the Japan Patent Office would be a cross-layer citation, as the latter node does not exist in the US layer. If, however, this Japanese publication was part of a triadic patent family, we can 'redirect' this US-originating citation to the US-granted family member, as this patent covers the same

⁹Note that this definition is slightly different to that used in previous work, notably Dernis and Khan (2004). Until the year 2000, applications to the USPTO were not published, so it was generally impossible to know whether equivalents were filed in all jurisdictions. This led to a slightly awkward definition (families with equivalents granted by USPTO and applied to EPO and JPO) that was in wide use until sufficient time had passed for USPTO application data to accumulate. A common definition in use currently is those families with equivalents filed at the triadic offices; however, as US applications do not list citations, we restrict this definition further to require a grant at each office.

technical content, and the citation can remain within the US citation network layer where it was generated. Third, all triadic offices provide detailed citation data. There is no theoretical reason why citation network layers associated with other countries cannot be added if the data exists, but we deemed the triadic offices to be the best starting point to demonstrate the use of the multilayer framework due to their existing popularity among both applicants and researchers.

In this work, we also wish to demonstrate the importance of citation source and context. During the application and examination process, citations that reach the front page of the patent may be added by one of several parties for a variety of reasons. One problem inherent in this citation metadata is that different offices have different examination guidelines and legal frameworks that inform how prior art is cited (Higham and Yoshioka-Kobayashi, 2022). Further, the way that these differences manifest themselves in the metadata that researchers can access is not consistent across offices or, indeed, across time. For some of the analyses in this work, we broadly group citations at each office into two groups: those that were likely found by the examiner and those that were likely found by the applicant. While these groups are far from perfect,¹⁰ we do so to illustrate the flexibility of the multilayer network approach—the citations that comprise each layer can be filtered based on the research purpose. This flexibility is discussed in more detail in Section 4. One minor restriction that accompanies this approach is that we require citation metadata to exist for all citing patents. The USPTO only started to include this metadata for granted patents from the start of 2001, so the triadic families we consider in this work are those for which the first US grant was in 2001 or later. All families considered in this work have all of their triadic members granted before April 2020. A histogram of the priority dates of the families that comprise the networks we consider in this work is displayed in Fig. 2.

Most of the data used in this work were obtained from Google Patent Public Datasets.¹¹ However, noting that, at the time of data collection, that data was not complete for citations between Japanese publications (notably, Japanese patents citing published Japanese applications), this data was supplemented by data supplied by the Intellectual Property Institute’s Patent Database.¹² We also make use of Cooperative Patent Classifications (CPCs); for consistency, we assign each family the classifications associated with their first US member, as determined by the USPTO. This data was obtained from PatentsView.¹³

To reduce the computational complexity associated with large networks, we prefer to work with a subset of the whole patent family network that nonetheless resembles the

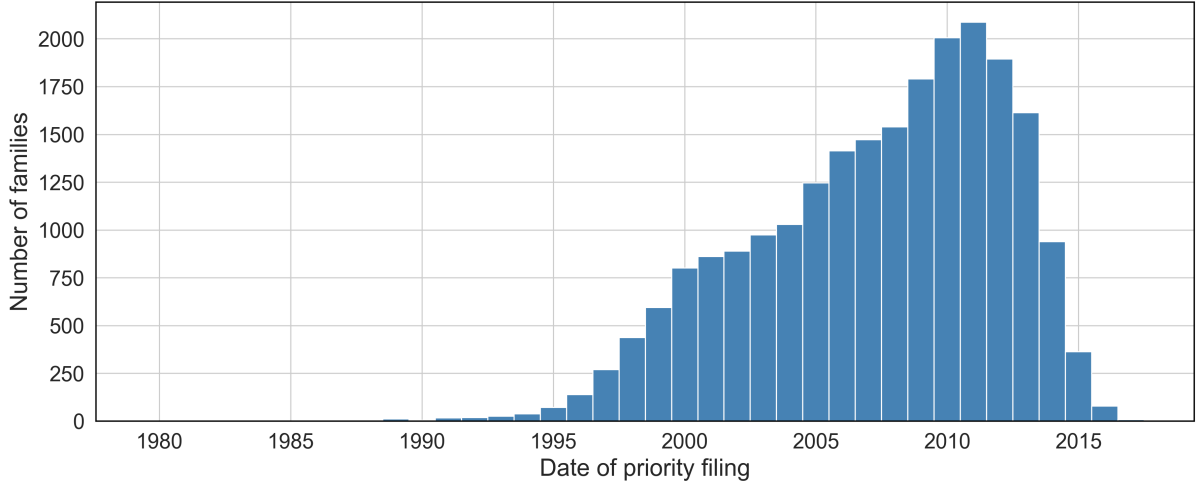
¹⁰This is particularly true for the JPO. However, there is suggestive evidence that applicant citations are more likely to be background art than art that could lead to a rejection of the application (Okada et al., 2018).

¹¹<https://tinyurl.com/googlepatentdata> (accessed 25/10/2021).

¹²www.iip.or.jp/e/patentdb/index.html (accessed 25/10/2021).

¹³<https://patentsview.org/> (accessed 25/10/2021).

Figure 2: **Family priority dates.** A histogram of the priority dates of the triadic families considered in this work, subject to the restrictions laid out in Section 2.3. All families have their US member granted in 2001 or later, but the earliest filing date can be considerably earlier.



structure of the whole. Using a set of obviously technologically related families such as those in a specific technology class or filed by firms in a specific sector may not satisfy this requirement, given the known differences in citation patterns across fields (Alcácer et al., 2009, Higham et al., 2017). To remedy this, we choose the subset of patents assigned to CPC class Y02: “technologies or applications for mitigation or adaptation against climate change.” The Y02 class is always a secondary classification and can be added to patent families from a broad set of technologies, from those aimed at reducing drag on airplanes to those aimed at treating diseases whose impact may be exacerbated by climate change (Veefkind et al., 2012, Haščič and Migotto, 2015). This class (and its subclasses) are commonly used as filters to study patented technological developments within specific domains related to both the mitigation of climate change, such as cleaner transport (Aghion et al., 2016, Barbieri, 2016) and energy production (Sun et al., 2021, Persoon et al., 2020), and our adaptation to the inevitable and wide-ranging environmental challenges we will face in the near future (Dechezleprêtre et al., 2020, Hötte et al., 2021). As such, we believe this technology class comprises a suitable microcosm within which we can effectively demonstrate the application of multilayer network methods to patent citation networks.

The resulting data set consists of a well-defined set of citing families, their CPC classifications, the citations they make,¹⁴ and the jurisdiction and context of each citation. A description of the layers considered in this work (which can be aggregated for specific empirical tests) can be found in Table 1.

¹⁴We exclude very rare citation types, such as those originating from third parties.

Table 1: **Layer descriptions.** Descriptions of the layers considered in this work, alongside their abbreviations and the number of links found within them. All layers contain 22653 nodes, and there are a total of 63916 citations in the multilayer network (MULTI) that is comprised of the layers described in the first seven rows. The last two rows are single-layer networks obtained by flattening the two USPTO layers (US-AGG) and all seven layers (ALL-AGG), respectively.

Layer	Citing party	Abbreviation	Description	Links
USPTO	Examiner	US-EXM	Cited by examiner during patent prosecution	15607
USPTO	Applicant	US-APP	Cited by applicant through an Information Disclosure Statement and unused by examiner	23145
EPO	Applicant	EP-APP	Cited by applicant, in the patent text or otherwise	4326
EPO	Examiner	EP-ISR	Cited by examiner in an international search report	5732
EPO	Examiner	EP-SEA	Cited by examiner in an EPO search report	5206
JPO	Examiner	JP-REJ	Cited by examiner as justification for application rejection	4612
JPO	Examiner	JP-BCK	Cited by examiner as background information	5288
USPTO	All	US-AGG	Cited by anyone (USPTO patents)	38752
All	All	ALL-AGG	Cited by anyone (all triadic patents)	63916

3 Methods and Results

3.1 Interdependence

Before a detailed examination into the *kind* of information that may be extracted from the multilayer network that is not accessible when using a single layer, it is first important to assess whether there is new information in the multilayer network at all. That is, if there is a high level of redundancy between the information contained in each network layer, then the case for using a multilayer framework is weakened. At the same time, if the layers contain very different structural patterns, then a multilayer framework may not be ideal, and more informative results may be obtained if they are treated as individual single-layer networks instead.

One way of assessing these properties is by measuring the interdependence of each layer, or set of layers, relative to the information that can be found elsewhere in the network. Several measures of interdependence have been proposed in the past (Parshani

et al., 2011, Morris and Barthelemy, 2012, Nicosia et al., 2013), many of which take a random walk approach to the level of layer interdependence or ‘coupling’ of layers in the network. In this work, at a high level, we are instead interested in the degree to which the information contained in one network layer can inform us about the information contained in another layer.

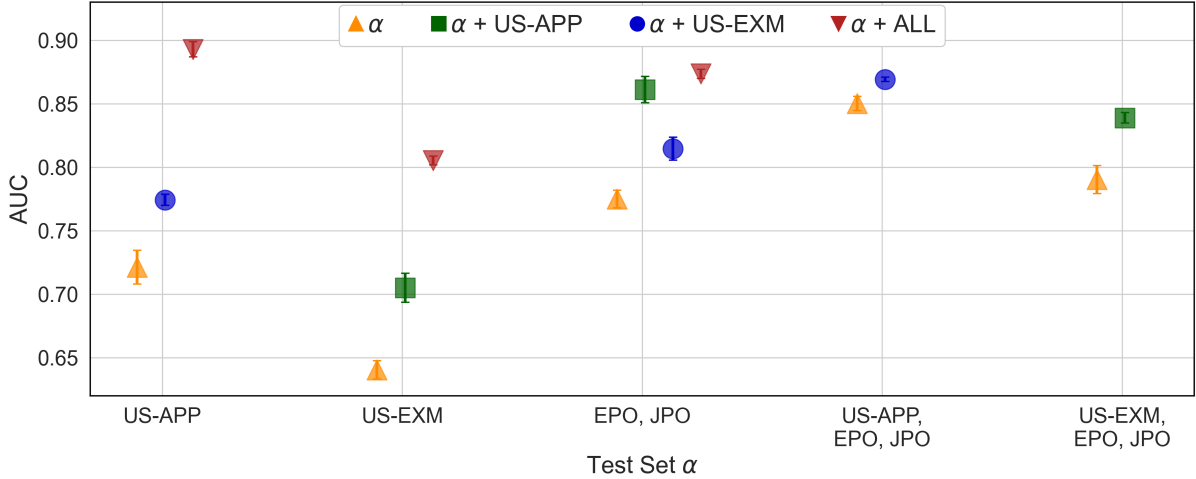
To this end, we employ the method introduced by De Bacco et al. (2017) and described in detail for our case in Appendix A.2. This method is a link prediction exercise, whereby a randomly-selected portion of the target layer or layers α have their link information removed and the remaining information in the network may be used to predict the existence of links. As a baseline, the remaining portion of the α is used as the training set, the receiver operating characteristic (ROC) curve is calculated, and the area under this curve (AUC) is computed. We can then introduce sets of other layers, β , into the training set, and compare results obtained by adding this information to those of the baseline. If the predictive power (as measured by the AUC) of this augmented set $\alpha + \beta$ is not significantly larger or smaller than the baseline predictive power, then β does not contain useful information over and above that contained in α . If, however, we note a significant increase in predictive power relative to the baseline, then β contains complementary information that cannot be extracted from what remains of α .

Much information can be garnered from comparisons of the change in predictive power when α and β are interchanged. For example, when the links in one layer are a subset of links in another, then we expect the change in predictive power to be asymmetric when we swap α and β — adding the subset to try to predict links in the full set will likely produce *worse* results than if only the full set was used for training the model. The information in the subset is redundant and could even mislead the model.

When two layers contain complementary information we would expect increases in predictive power regardless of the layer comprising the test set. This complementarity can arise in several ways, such as through similar community structure despite large differences in the specific links that produce these structures. A significance decrease, on the other hand, would indicate that β contains information that is irrelevant for the prediction task and actually added noise; this could occur, for example, if the link generation mechanisms were independent of the node properties, or were driven by different node properties in different layers.

Figure 3 shows the results of the interdependence analysis for various α and β sets in which we are interested. For graphical simplicity, we focus on the sublayers generated by the USPTO and the complete JPO and EPO layers (where the latter two always include all of their sublayers listed in Table 1). This is done to demonstrate, compactly, the complementarity of information across jurisdictions as well as that of their sublayers, with the most commonly utilised sublayers in the literature (US applicant and examiner citations) as exemplars for the latter calculations.

Figure 3: **Layer interdependence.** The x -axis presents different target layers α , and the y -axis shows the AUC obtained through 5-fold cross-validation for measuring layer interdependence. Orange results refer to the baseline AUC, where the algorithm is only given access to that target layer. Green and blue markers show the increase in the AUC for the α set when the algorithm is given access to the US-APP or the US-EXM, respectively. The red points refer to the AUC obtained by giving access to all other layers in the network. The results displayed are averages and standard deviations over the 5 folds.



The results displayed in Figure 3 show that adding more layers increases predictive power across all combinations of α and β we considered. This outcome suggests that, while they differ by the amount of unique complementary information they contain, each layer nonetheless contains information that is not available in the other layers. Specifically, information about the missing values in α is more accurately predicted when layers that are not already in α are included in the training set, relative to the sole use of the information that remains in α . This is to be expected, as examiners at each office conduct much of their prior art search independently.

A prime example of complementarity is displayed by the US sublayers (US-APP and US-EXM). These layers are almost mutually exclusive,¹⁵ but predictive power for links in one layer is significantly boosted when the other layer is added, regardless of which is the test set. That is, there is very little overlap in these layers, and yet one can be successfully used to predict the links in the other, likely due to the similarity of mesoscale network communities within each of these layers.

That some citation types add more information than others is also expected. After all, information sharing occurs regularly between offices (Wada, 2020), and this process leads to the duplication of citations between specific layers. While this sharing happens

¹⁵Copying occasionally happens due to the recycling of citations for continuing patent applications.

increasingly through direct collaboration between offices examining equivalents,¹⁶ most of the citations we consider here were made before these formal programs were launched. As such, for much of the time period we consider, the information ‘sharing’ likely takes place indirectly, through applicants. For example, the EPO produces a search report for the applicant to consider before a substantial examination takes place. Under their duty of disclosure obligations at the USPTO, it is considered good practice to pass this information on to the USPTO if an equivalent is being examined there simultaneously (which will usually be the case for triadic patent families). This information is submitted via an information disclosure statement and the USPTO examiner then assesses the relevancy of the prior art that is listed on the search report. When it happens at all, only a small percentage of citations from the EPO search report will be used to justify rejection and be recorded as examiner citations, while the remainder will be recorded as applicant citations. As such, the EPO search report is a non-obvious mechanism through which citations are duplicated from EP-SEA citations to US-APP citations (and sometimes to US-EXM citations).

Similarly, while there is a knowledge disclosure obligation at the JPO, the incentives for complying are very weak relative to the USPTO (Nakamura and Sasaki, 2016). However, applicants to the JPO often use in-text citations to make a case for patentability, and perhaps much more so than the typical applicant to the USPTO or EPO. As such, it is plausible that, for triadic patents, these citations are included in-text in other equivalent applications and are therefore easily accessible to examiners in all jurisdictions. If these citations are deemed relevant by multiple examiners, these citations might also appear to be duplicated across network layers.

3.2 Community detection

Having found that the different layers likely contain complementary information, we now investigate the patterns extracted from a multilayer network approach and compare them with those extracted from single-layer networks that exclude citation context. Specifically, we wish to detect communities of triadic families that are similar in their citation patterns. These communities represent mesoscopic structural patterns contained in the networks that are not objectively or directly observed, but can be inferred from the data.

To this end, we apply a community detection algorithm to three networks: i) the (seven-layer) multilayer network containing the EPO, JPO, and USPTO layers (MULTI); ii) the network obtained by flattening all the layers in (i) into a weighted single-layer network and ignoring citation origin and context (ALL-AGG); iii) the weighted single-layer network obtained by flattening the USPTO examiner and applicant layers only (US-AGG). Each link is weighted by the sum of link weights across all layers we consider; that is, if the same family-family citation exists once in each of n layers, then the link is

¹⁶<https://www.wipo.int/case/en/>.

assigned weight n . While rare, link weights greater than one can occur within sublayers; for example, when a divisional makes the same type of family-family citation as its parent, the link weight corresponding to this link will be two.

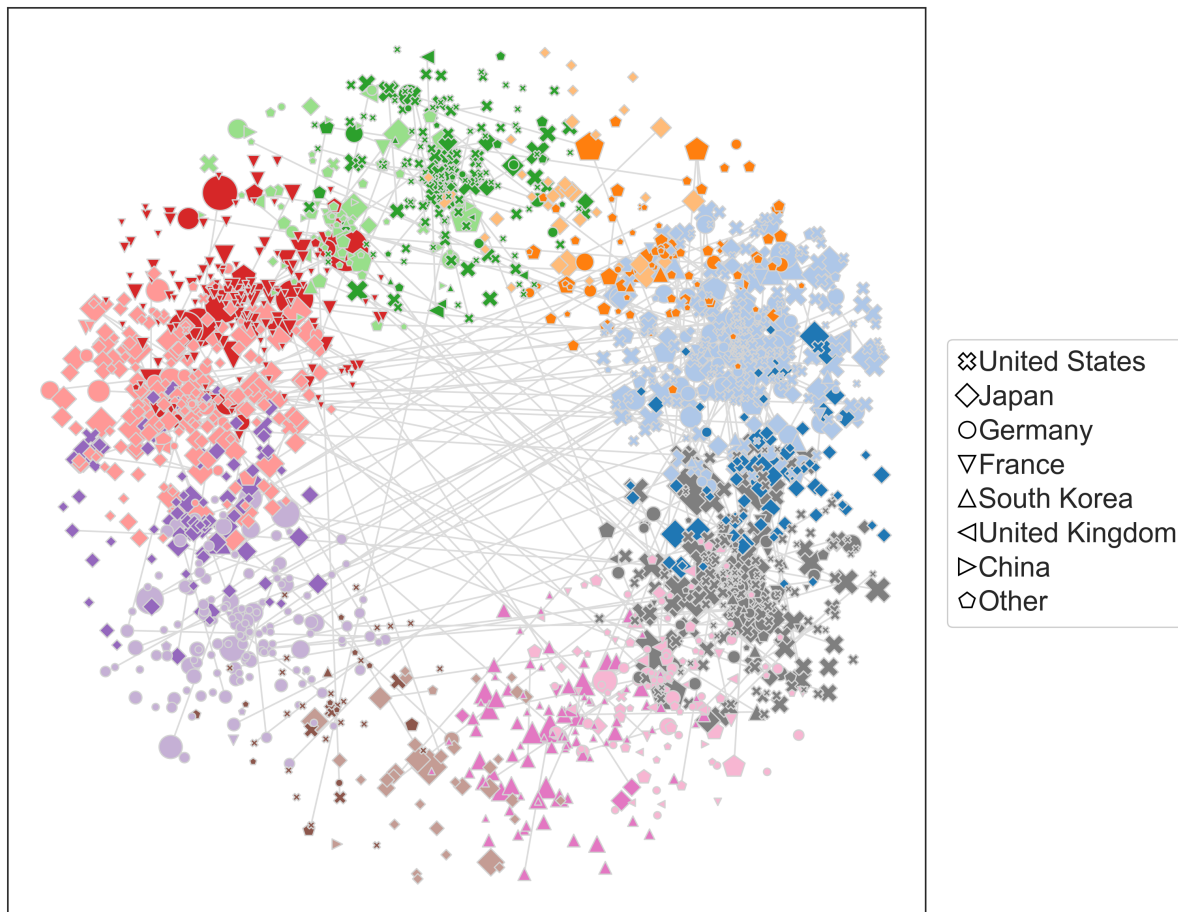
To perform the community detection task, we consider a probabilistic generative model that assigns a probability to a citation between two families that depends on the communities they belong to, as described in De Bacco et al. (2017). In our case we have access to relevant metadata about each triadic family, hence we consider the model of Contisciani et al. (2020), MTCOV, that is also able to incorporate the office at which priority was filed (which is often *not* a triadic office) as a node covariate to drive inference along with the network structural information. This covariate allows us to incorporate the home-bias of citations in early search reports (Bacchiocchi and Montobbio, 2010) and, to a lesser extent, industrial agglomeration patterns (Asheim and Gertler, 2005) (given the strong correlation between assignee location and priority office), to inform the inferred citation probability alongside explicit network structure. This model automatically balances the weight of the covariates’ contribution in determining the communities. In all our experiments we find that node covariates are indeed significant, in that they allow us to better quantify the probability of certain citation patterns. The optimal number of communities in each case is extracted through a cross-validation procedure, see Appendix A for details. In addition to being able to incorporate a covariate that may inform network structure at scales beyond individual links, MTCOV is scalable to large networks, allows overlapping communities, and is open-source,¹⁷ all of which are desirable features for the current work.

We chose ALL-AGG as a comparison because it contains all the same links as the multilayer network, and even accounts for link overlap among layers, but without context. As such, any differences in the extracted communities arise solely due to the addition of citation context, and the incorporation of this context into our network model. US-AGG is included in these comparisons as the most common citation network used in previous work. The USPTO also tends to make many more citations per patent, and so this single-jurisdiction layer is likely to be the most ‘complete’, with respect to the links in the full triadic network.

The communities extracted for MULTI are shown for a random subset of patent families in Figure 4. Analogous figures for the ALL-AGG and US-AGG networks can be found in Appendix B. While the model allows for overlapping communities (nodes can belong to multiple communities), in Figure 4 we colour nodes by their ‘hard’ communities, whereby each patent family is assigned to the community to which it displays the highest affinity. The optimal number of communities, calculated via the cross-validation exercise described in Appendix A.1, was found to be 15 for the multilayer citation network and 7 for ALL-AGG and US-AGG. Finally, the location of the assignee of each patent family

¹⁷<https://github.com/mcontisc/MTCOV>

Figure 4: **Community extraction.** This diagram shows the hard community membership partitions for MULTI. While inference was performed on the whole network, here we use a random sample of 2000 nodes and include any incidental links among these, for graphical clarity. The colouring shows the 15 communities found within the MULTI network. Node size is proportional to the number of outgoing and incoming citations, while node shapes denote the location of the assignee of each patent family.



(rather than the priority office, which is used as a covariate in the community detection procedure) is indicated by the shape of the node.

Between networks, several graphical observations can be made in the geographic composition of the extracted communities, despite the differing community sizes. First, country-based homophily is very clear. The most obvious example of this is that families filed by Japan-based assignees are primarily grouped with families that are also filed by Japan-based assignees, with the only observable difference between the networks being how many communities are found within this group of families (1 for ALL-AGG and US-AGG, and 5 for MULTI); however, this difference is expected as the optimal number of communities multilayer network is greater. The other consistently geographically-homogeneous communities include those families assigned to German firms and those assigned to South Korean firms. The existence of these groupings is somewhat expected

— geographic citation biases are a well-known phenomenon and have a wide range of drivers, including local industry agglomeration, shared language, prior-art search strategies, knowledge spillovers, and coordinated technological development strategies at the national level (Jaffe et al., 1993, Almeida and Kogut, 1999, MacGarvie, 2005, Bacchiocchi and Montobbio, 2010, Wada, 2016). Because priority office information is included in the community detection algorithm, the existence of the kind of geographic grouping we observe reflects that while technological similarity plays a big role in citation linkage at the micro-level, simple geographical metadata can be highly predictive of network structure at larger scales.

3.3 Network Communities and Technological Similarity

One would expect that the citations we consider in this work *should* link families with technological similarities and, therefore, the communities detected should group inventions with shared and legally relevant technological features. Indeed, the geographical biases in citation linkages that are observed above may be considered to be artifacts of the systems within which technological development occurs, and perhaps even hinder our understanding of the nature of innovation more generally. We assert that the multilayer framework is one way of mitigating some of these biases, as it integrates relevant technological relationships uncovered by several different, and geographically separated, patent agents and examiners working mostly independently. In aggregate, this information should give a more balanced view of technological similarity and down-weight those links that are heavily influenced by unwanted geographical and office-specific biases and conventions. However, the link weights in the ALL-AGG network may play a similar role. As such, we will now turn to the differences in the technological information contained in the three networks and examine the importance of citation context (i.e., source and justification) in assessments of technological similarity.

To do this, we directly compare the network of meso-level technological relationships that can be gleaned from extracted communities with externally-defined technological categories. First, we construct a weighted bipartite (two-mode) network of relationships between the extracted communities and the 3-digit Cooperative Patent Classification (CPC) codes that the families within each community were assigned upon application to the USPTO.¹⁸ CPC codes, henceforth referred to simply as *classes*, were chosen due to their status as the primary classification system at two of the triadic offices and widespread use in research, particularly in studies of technological evolution and forecasting technical change. The weight of each link in the bipartite network between communities and classes is proportional to the fraction of families in each community

¹⁸This choice was made for the sake of consistency. Different offices may make slightly different judgements regarding the particular set of classes assigned to an application. By using data from a single office, we do not have to be concerned with these systematic differences.

that were assigned to a given class. We then project onto the technology class nodes to obtain a network of classes wherein links exist between classes that were both found in the same community or communities. A higher link weight between two nodes in this projected network reflects a more similar distribution of those classes across the extracted communities (Vasques Filho and O’Neale, 2018).¹⁹

We then construct two basic comparison networks: co-classification (Engelsman and van Raan, 1994, Breschi et al., 2003) and inter-class citation linkage (Leten et al., 2007, Alstott et al., 2017). The former contains a link between (3-digit CPC) classes when a family is assigned to both, with weights proportional to the relative frequency of such occurrences. The latter network contains links between these classes with weights proportional to the number of citations made between families that were assigned to each class, normalised to the total made by each of the classes.²⁰ We keep self-loops in this network, as they are required for sensible link-weight normalisation. For example, if class A makes 10 citations (and receives none), one of which goes to a class B family but 9 return to other class A families, this is a very different situation from one in which all 10 go to class B families. Because we normalise link weights by total citations made, ignoring self-citations would give the link from A to B the same weight in both scenarios, rather differing by a factor of 10. Further description of the construction of all networks used in this section can be found in Appendix B.

Now that we have two externally defined, node-aligned class networks, we are able to directly compare their structure to those extracted from the community-class bipartite networks. Because the nodes in each network we wish to compare are labelled and the same for all networks, we are able to use known-node correspondence methods that allow for comparisons at the node-level in such a way that accounts for differences in relationships between specific node pairs and for higher-order relationships (Tantardini et al., 2019). For this exercise, we use two different methods of comparison: the Frobenius norm and DeltaCon (Koutra et al., 2013).

The Frobenius norm is applied to the raw differences in the adjacency matrices between two networks, and thus quantifies the entry-wise (link-level) differences in the matrices being compared. When the networks being compared are unweighted, this distance is simply the square root of the number of pair-wise differences between the networks. However, this method easily accommodates the weighted case, wherein each pair-wise difference can have a magnitude other than unity.²¹ The Frobenius norm is a crude com-

¹⁹Note that these classes are not directly used in the community detection process. However, the community detection process relies on citation linkages, and these citations are often found through searches within the technology classes to which the application under examination has been assigned (see, e.g., Demey and Golzio, 2020).

²⁰To compare this network to our (undirected) projected networks, we take the sum of the normalised weights of the directed links between classes to obtain an undirected link weight. This simplification is a necessary evil for the current purpose and may miss some nuance in certain technological relationships.

²¹Specifically, the Frobenius norm of a matrix $A_{m \times n}$ is defined as the square root of the sum of the

parison method that cannot account for higher-order relationships between nodes, such as the importance of a link in the overall structure of the network, but it is a good heuristic when making multiple comparisons as we do here. DeltaCon, on the other hand, is more sophisticated, and indirectly considers every possible path between two nodes. In this way, differences in the weights of links that are particularly important for the network structure at the macro-level are incorporated into the comparison. While the DeltaCon algorithm can be very computationally expensive on large networks, and an approximation is possible, our class network is small enough (535 nodes) that the exact form can be used (Koutra et al., 2013). Both the Frobenius norm and DeltaCon calculate a distance metric whereby smaller distances indicate more similar networks. These methods are implemented in Python using the `numpy` (Oliphant, 2006) and `netrd` packages (McCabe et al., 2021).

In addition to the network comparison methods, we are also able to quantify the diversity of technology classes within each community extracted. For this purpose, we make use of the Rao-Stirling diversity (RSD) (Rao, 1982, Stirling, 2007), which considers both the homogeneity of each community (with respect to the classes within it) and the level of ‘surprise’ that specific pairs of classes are found together. For the latter consideration, we operationalise class distance using the inter-class citation network described above, as citations are what we use to extract the communities in the first place.²² Calculating this index for all communities extracted from a particular network, we take the median index across these communities as a measure of their average diversity. The RSD is high for a particular community when classes co-occur in high proportions with other classes with which not many citations are exchanged. RSD is low when classes generally only co-occur in high proportions with other classes with which they exchange many citations. Specifics of the RSD can be found in Appendix B.2. This kind of analysis, when compared to the network comparison methods above, may be considered relatively myopic. It can only capture the internal composition of individual communities without accounting for the relationships between the pairs of technologies in other communities. However, this calculation may provide insight into the origin of differences we find for the network-level comparisons.

Lastly, we calculate the spread of technologies across the extracted communities. A priori, we do not know what the relationships between the communities are, so we cannot integrate a distance metric to account for the level of surprise that a family assigned a particular class is found in a given pair of communities (as we did for the previous diversity

absolute squares of its elements,

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}. \quad (1)$$

²²That is, it would be a ‘surprise’ to find a pair of classes that don’t cite each other, but are nonetheless found in the same community.

Table 2: **Network comparison and diversity measures.** Here we show the results of the network distance calculations as well as the diversity measures. DeltaCon (DC) and Frobenius (F) distances between the class-projected networks (leftmost block) and the externally defined co-classification (Co-class) and inter-class citation linkage (IC Cites) networks are displayed in the central block. The median Rao-Stirling class diversity (RSD) across communities and the median Herfindahl–Hirschman Index (HHI) of classes’ dispersion across communities are shown in the rightmost block. * indicates non-optimal partition. The lowest values within each comparison set are highlighted in bold.

		Comparison Network				Diversity	
		Co-class		IC Cites		RSD	HHI
	C	DC	F	DC	F		
MULTI*	7	32.43	115.98	30.40	115.01	6.81	0.185
ALL-AGG	7	34.00	149.40	31.96	148.31	7.02	0.263
US-AGG	7	34.54	150.63	32.49	149.59	7.05	0.271
MULTI	15	30.08	58.66	28.09	58.61	6.87	0.131
ALL-AGG*	15	31.47	78.87	29.49	78.56	6.87	0.169
US-AGG*	15	30.79	76.73	28.81	76.42	6.99	0.167

measure). As such, we implement the Herfindahl–Hirschman Index²³ (HHI) (Hirschman, 1945, Simpson, 1949, Herfindahl, 1950, Hirschman, 1964) to measure the extent to which classes are splintered across communities. The details of this calculation can be also found in Appendix B.2. The HHI is maximised when all families assigned a particular class are in the same community and minimised when there are the same number of these families in each community. Again, the median HHI across all communities is compared across the networks. Like the Rao-Stirling index above, this calculation may add additional colour to the more comprehensive network comparisons. It is important to note that while we believe that it is desirable that communities are able to capture, to some extent, the large-scale structure of the technology-level networks, neither the spread of technologies across communities, nor the internal diversity of communities, is a test of the performance of the community extraction exercise.

It is important to note that the optimal parameters for the community partitions for the three networks are different—the optimal number of communities found for the multilayer network is 15, while for the others it is 7. For this reason, we run the community-detection algorithm for each of the non-optimal partitions (7 for the multilayer network and 15 for the others) to obtain a complete set of networks with which we can make fair comparisons. In sum, we construct six bipartite (community-class) networks which we project onto the class nodes to compare with the co-classification and inter-class citation networks.

²³Sometimes referred to as the Simpson index.

The results of this analysis can be found in Table 2. First, we find that the communities in the multilayer network generate class networks that are more similar to the co-classification and inter-class citation networks than those generated by the other two networks. This finding holds for both individual-link-level comparisons (Frobenius) and when higher-order relationships are taken into account (DeltaCon), for both optimal and non-optimal partitions of the multilayer network. Further, we find that the average RSD of the individual communities is lowest, while classes are the most evenly distributed across the communities (low HHI), in the multilayer case.

These observations lend themselves to some interesting interpretations. When looking at all communities, in combination, those extracted from the multilayer network imply technological relationships that are closer to the explicit technology networks than the flattened or single-jurisdiction approaches. However, the diversity calculations suggest that this observation is not simply driven by the extraction of homogeneous communities that group technologies in a straightforward manner. In fact, technology classes are more thinly spread across communities in the multilayer case, while the average internal diversity of classes is generally lowest for this network once known technological similarities between classes are accounted for. This suggests that, on the micro-level, the multilayer (relative to the single-layer) network approach is more sensitive to citation linkages than co-classification, but is nonetheless better able to represent real technological relationships on the meso- and macro-levels.

Indeed, our results are consistent with the conclusion that the multifaceted nature of the technological relationships that are embedded in citation data may be partially lost when a multilayer network is flattened into a single-layer one. This view rests on an assumption that different technology types can be related to each other in different ways. For example, let’s assume that applicants filing a patent assigned to class A prefer to cite families assigned to class B, while examiners examining the same patent prefer to cite those assigned to class C. When, such as in this example, these different relationships are driven by different citing parties, the erasure of citation context will lead to the loss of this nuance. This problem may be exacerbated in the presence of higher-order effects, such as if the above citation behaviour only occurs when a fourth class D is also assigned to the patent application. In contrast, the multilayer network approach ensures these nuances and higher-order relationships remain accessible. The retention of this kind of technologically relevant information, particularly with respect to rare or subtle inter-class relationships, would be consistent with the findings displayed in Table 2.

4 Discussion and Conclusions

Historically, research informed by patent citation data has often ignored citation source and context. There can be a perfectly reasonable reason for this practice, such as when

one is only interested in citations made to and from patents in a single jurisdiction to study, for example, the effect of a local policy change. However, a truly comprehensive and global view of patented inventions and the relationships between them is only possible when data from multiple sources are integrated sensibly. It is in these contexts that the multilayer network is a natural framework for analysis.

In this work, we introduce the concept of multilayer patent citation networks as a natural way to present and analyse global patent information without loss of citation context. We conduct several empirical analyses to demonstrate the utility of the multilayer framework. All analyses are conducted on a subset of the full citation network, containing all triadic patent families classified into CPC class Y02 with US members granted from the year 2001. By design, this subset will give the most conservative estimates of the additional information that may be extracted from the multilayer network relative to its single-layer counterparts. Our results in this work suggest that not only is there, indeed, a considerable amount of additional information contained in the multilayer citation network relative to those single-layer counterparts, but this information is technologically relevant and captures nuanced aspects of the technological relationships between patented inventions.

First, an interdependence analysis shows that additional network layers, defined by citing office, contain complementary (rather than redundant) information that may be used to predict the link-level structure of other layers. To test whether this complementary information is important for characterising network structure more generally, we then conduct an exercise in community detection. This is carried out and compared across three different networks: the multilayer network, the flattened and weighted (single-layer) version of the multilayer network (containing all the links in the latter but without citation context), and the complete (flattened, single-layer) US citation network that is most commonly used in technological network analyses. While there is a notable similarity in the communities extracted from these networks, there is also significant disagreement, indicating that the information contained in the citation context may be important for characterising the mesoscopic structure of the global citation network.

To test whether the differences in community structure are technologically meaningful, we conduct direct comparisons between the technological relationships implied by the extracted communities and those of previously studied meso-scale networks of technological similarity: the co-classification and inter-class citations networks, at the CPC 3-digit level. These tests are conducted, in part, to show how the information content (i.e., citation context) contained in citation networks can be related to the meso-scale technological structures that are perhaps more established in the technology management community. To be able to draw a direct comparison, we construct the bipartite networks between communities and classes, then project onto the class nodes to obtain a class network wherein links reflect levels of co-occurrence in the communities. To add colour

to these comparisons, we also compute the Rao-Stirling diversities of these communities (across classes) and the Herfindahl–Hirschman Indices of class (across communities). Relative to the flattened networks, we find that while the communities extracted from the multilayer network are less diverse and the implied class network more similar to the co-classification and inter-class citation networks, classes are more evenly spread across communities. These results suggest that citation context is technologically relevant and a more realistic mesoscopic network structure can be inferred when we depart from the view that technological relationships are mono-faceted or driven by simple class-level technological similarity.

While we include the US citation network in our comparison exercises, this is only done as an acknowledgement of its position as the dominant data source in the extant literature. The flattened version of the multilayer network, on the other hand, contains all the links that are present in the multilayer network, but without the context that allows us to define the layers. As such, we consider this network the most appropriate comparison network, as any differences found must be driven by the absence of citation context. That the communities extracted from the multilayer network more closely replicate the established and explicit co-classification and inter-class citation networks indicates that citation context adds technologically relevant information in the aggregate, despite displaying higher within-community diversity of classes. This suggests that ignoring citation context results in a bias towards within-class citations (that are easier for all parties to search for and find), at the expense of the rarer inter-class citations and class combinations that play a larger role in both the network structure as a whole and, arguably, technological progress in the long-term (Castaldi et al., 2015, Verhoeven et al., 2016, Mewes, 2019, Kelly et al., 2021). Considering citation generation mechanisms, it is plausible that citation context provides important clues as to the relevance and nature of the technological relationship between citing and cited inventions (Criscuolo and Verspagen, 2008, Alcácer et al., 2009, Azagra-Caro et al., 2011, Li et al., 2014, Kuhn et al., 2020). As such, treating all these links as equal, with respect to their information content, is clearly not ideal for many use-cases.

4.1 Limitations

The main limitations of the empirical analyses conducted in this work are those restrictions we placed on the families we chose to include. As we describe in Section 2.3, these restrictions were put in place for a variety of reasons, including data availability, computational limitations, and a desire to demonstrate our approach in a conservative manner. Little can be done about data availability; however, this only affects our ability to examine citation context in the US case, and only for times earlier than the year 2001. In any case, we suggest that families granted after this time provide a sufficiently large sample for the purposes of this work.

The conservativeness of our approach is introduced with the decision to consider only those families with granted patents at all three triadic offices. This means that all offices had access to the same set of prior art, and had the opportunity to share information among themselves. In turn, this would introduce maximum redundancy between layers, and minimise the additional information that can be added by the inclusion of citation context. It is for this reason that we think of our approach as conservative. Extensions of the restrictive, special-case multilayer framework that we examine here are discussed below in Section 4.2, and highlight the potential of this framework going forward.

Lastly, to reduce the computational complexity of our analyses, we restrict the included families to those classified into CPC class Y02. While we maintain that this subset is an appropriate representation of the patent citation network as a whole, there may be arguments against its generalisability. However, in the case that this class contains a more homogeneous set of families than the set of all families (which is almost certainly true), then the inter-class structure that we are able to explore is likely to be *less* rich and *less* nuanced than that of the full network. Detecting higher-order nuances is precisely the domain in which we suggest the multilayer network excels, so following this logic would lead us to conclude that the current approach is, again, a very conservative one.

4.2 Future Work

This work aims to describe the construction of multilayer patent citation networks then conceptually and empirically justify their use. This framework may prove to be of particular interest to those who would prefer representations of technological relationships that are not as sensitive as extant frameworks to the idiosyncrasies of individual patent offices. However, both the layers that are selected to comprise the network and the appropriate empirical methods to extract information from this network will depend on the specific use-case. Here, we describe the myriad methodological doors that are opened with the introduction of patent-based multilayer networks into the broad field of science, technological, and innovation studies.

The obvious extension to the current work is to take a less conservative approach with respect to the subset of families and citations considered. This can take the form of additional layers, nodes, or links. The addition of layers corresponds to the addition of new citation contexts (such as in-text citations (Verluisse et al., 2020)) or the addition of new jurisdictions. The addition of nodes and links, on the other hand, would relax the condition that a family be triadic. Citations between triadic families only make up a tiny portion of all citations made and received by these families. For example, in Figure 1, we show the triadic ego network of the family with USPTO equivalent US-6819081-B2. In this restricted network, this family only receives 4 citations from other triadic families classified into class Y02. If we remove all restrictions on the patents we include in our

network, however, this family receives almost 50 citations; about 90% of these are from families that have a triadic member, and about 95% are from families that are also classified into Y02. As such, removing the triadic family requirement but keeping the network restricted to the triadic offices and the Y02 classification would dramatically increase the sample size.

Multilayer citation networks can also be flexibly aggregated. Just as one can analyse the inter-class citation network for a single jurisdiction or citation context (e.g., US applicant citations), it is also possible to include additional layers containing the equivalent information for other jurisdictions or contexts. In fact, in the same way that we use families to align layers in the current work, any metadata that connects groups of patents between network layers forms a natural multilayer configuration. Classes, firms, and inventors can all be linked across jurisdictions and citation contexts and their networks analysed in a multilayer framework. Even in the single-jurisdiction US case, for example, the relative positions of firms in the inter-firm citation network will depend on whether one uses examiner citations, applicant citations, third-party citations, or in-text citations. Because firms can be represented as nodes across all of these context-specific networks, multilayer network tools may be applied to obtain a comprehensive and integrative view of the network structure without abandoning citation context. Citations to non-patent literature such as scientific articles is challenging to incorporate into patent citations networks generally, but it is certainly possible to treat this information as family-level metadata — perhaps to construct a bipartite network similarly to how technology classification was used in Section 3.3. More complicated uses of this information could match institution and inventor data from patents onto scientific articles to extend recent work on the multilayered interplay between authorship and the broader dynamics of science and collaboration into the technological domain (Omodei et al., 2017, Nanumyan et al., 2020, Zingg et al., 2020).

In addition to these data extensions, the conceptual arguments against the omission of citation context lead to a strong case for the further application of novel tools designed for the study of multilayered systems. To return to the public transport analogy, it would be unwise to treat all modes of transport as equal if you are trying to find the fastest route between two places in the network. In the same way that the time and financial costs of using different modes affects the route choice between two points in a physical landscape (which will be moderated by the amount of time or money you had), citation networks are embedded in a technological landscape (Kauffman et al., 2000, Fleming and Sorenson, 2001) and different types of citation may traverse this landscape in different ways. This intuition has significant consequences for the analysis of citation networks. For example, any algorithm that ‘walks’ through the network, such as PageRank, should consider the ‘cost’ of each link in a similar way to one plotting a route through a multilayered transportation network. The application of multilayer network methods opens the

door to a menagerie of new analytical tools to develop more sophisticated and tailored metrics for studies of technical change and the nature of innovation systems. For example, the identification of patent thickets (Shapiro, 2000, Bessen, 2003) is often conducted through, or supported by, citation network analysis (Von Graevenitz et al., 2011, Zingg and Fischer, 2018, Yuan and Li, 2020). The multilayer framework may assist in these studies — thicket identification depends crucially on the citation context (blocking vs. non-blocking citations) and the jurisdiction (a thicket is necessarily a single-jurisdiction phenomenon). Adding citation context and linking families across jurisdictions for direct comparison may allow for thickets to be more easily distinguished from fields with dense, but non-overlapping, intellectual property rights. For example, when calculating clustering coefficients in multilayer networks, one can specify weights for different kinds of citation or penalise cycles that move between layers (De Domenico et al., 2013). This kind of flexibility can be used to operationalise the definition of thickets in a way that doesn't simply ignore applicant-provided citations or citations from other jurisdictions, which may not be entirely irrelevant, particularly at the firm level.

Network centrality is another important concept that is generalised in the multilayer case (Solá et al., 2013, Solé-Ribalta et al., 2014, De Domenico et al., 2015, Taylor et al., 2021), and can also be readily applied to citation networks. For example, without citation context, it is hard to know whether firms are central because they block the patents of competitors or are a source of knowledge from which other firms build. Further, firm centrality will likely depend on the jurisdiction one examines, so multilayer centrality may give a more holistic view of their centrality in global markets.

Both technology roadmaps (Lee et al., 2009) and technological trajectories (Verspagen, 2007) may be significantly altered by the incorporation of citation context, as different kinds of citation appear to hold different information, which may, in turn, be useful for forecasting or tracing different kinds of technical change (Acemoglu et al., 2016, Mariani et al., 2019). So-called 'main paths' in technological trajectory analysis (Hummon and Dereian, 1989, Verspagen, 2007) could be particularly sensitive to the weights that are placed on, or empirically determined for, different layers or citation contexts. The multilayer framework may also conceptually aid traditional economic analyses (Cai and Li, 2019), for which it is possible, for example, to allow layers to differ in importance when constructing proxy network variables that attempt to capture an abstract concept.

Lastly, pair-wise interactions may not be sufficient to describe the complex behaviour of interactions between the components of innovation systems that are accessible through citation networks. In particular, the interactions between firms or technology types that are visible in citation networks may be better represented through higher-order interactions (Lambiotte et al., 2019, Battiston et al., 2020, 2021). For example, the patenting and citing behaviour of firms may be described at several different scales. Higher-order representations allow us to differentiate changes in citation behaviour of a firm in re-

sponse to sector-wide changes from the pairwise interactions between a firm and every other firm in its sector. Higher-order interactions can exist within layers of multilayer networks and it is possible that different higher-order behaviours are observable in different patent systems. In any case, it is clear that applications of network frameworks beyond single-layer networks with dyadic links are very much in their infancy in the field of innovation studies, and hold huge potential as more realistic abstractions of innovation systems.

Acknowledgements

The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Martina Contisciani. MC and CDB were supported by the Cyber Valley Research Fund.

References

- D. Acemoglu, U. Akcigit, and W. R. Kerr. Innovation network. *Proceedings of the National Academy of Sciences*, 113(41):11483–11488, 2016.
- P. Aghion, A. Dechezleprêtre, D. Hemous, R. Martin, and J. Van Reenen. Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, 124(1):1–51, 2016.
- J. Alcácer, M. Gittelman, and B. Sampat. Applicant and examiner citations in US patents: An overview and analysis. *Research Policy*, 38(2):415–427, 2009.
- A. Aleta, S. Meloni, and Y. Moreno. A multilayer perspective for the analysis of urban transportation systems. *Scientific Reports*, 7(1):1–9, 2017.
- P. Almeida and B. Kogut. Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, 45(7):905–917, 1999.
- J. Alstott, G. Triulzi, B. Yan, and J. Luo. Mapping technology space by normalizing patent networks. *Scientometrics*, 110(1):443–479, 2017.
- B. T. Asheim and M. S. Gertler. The geography of innovation: Regional innovation systems. In *The Oxford Handbook of Innovation*. Oxford University Press, 2005.
- J. M. Azagra-Caro, P. Mattsson, and F. Perruchas. Smoothing the lies: The distinctive effects of patent characteristics on examiner and applicant citations. *Journal of the American Society for Information Science and Technology*, 62(9):1727–1740, 2011.
- E. Bacchiocchi and F. Montobbio. International knowledge diffusion and home-bias effect: Do USPTO and EPO patent citations tell the same story? *Scandinavian Journal of Economics*, 112(3):441–470, 2010.
- J. Bakker, D. Verhoeven, L. Zhang, and B. Van Looy. Patent citation indicators: One size fits all? *Scientometrics*, 106(1):187–211, 2016.
- P.-A. Balland and D. Rigby. The geography of complex knowledge. *Economic Geography*, 93(1):1–23, 2017.
- N. Barbieri. Fuel prices and the invention crowding out effect: Releasing the automotive industry from its dependence on fossil fuel. *Technological Forecasting and Social Change*, 111:222–234, 2016.
- F. Battiston, V. Nicosia, and V. Latora. Structural measures for multiplex networks. *Physical Review E*, 89(3):032804, 2014.
- F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92, 2020.
- F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, et al. The physics of higher-order interactions in complex systems. *Nature Physics*, 17(10):1093–1098, 2021.

- E. Berkes and R. Gaetani. The geography of unconventional innovation. *The Economic Journal*, 131(636):1466–1514, 2021.
- M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, and D. Pedreschi. Foundations of multidimensional network analysis. In *2011 International Conference on Advances in Social Networks Analysis and Mining*, pages 485–489. IEEE, 2011.
- J. E. Bessen. Patent thickets: Strategic patenting of complex technologies. *Available at SSRN 327760*, 2003.
- B. P. Biddinger. Limiting the business method patent: A comparison and proposed alignment of european, japanese and united states patent law. *Fordham L. Rev.*, 69: 2523, 2000.
- S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014.
- S. Breschi, F. Lissoni, and F. Malerba. Knowledge-relatedness in firm technological diversification. *Research policy*, 32(1):69–87, 2003.
- P. Bródka, P. Kazienko, K. Musiał, and K. Skibicki. Analysis of neighbourhoods in multi-layered dynamic social networks. *International Journal of Computational Intelligence Systems*, 5(3):582–596, 2012.
- J. Cai and N. Li. Growth through inter-sectoral knowledge linkages. *The Review of Economic Studies*, 86(5):1827–1866, 2019.
- C. Castaldi, K. Frenken, and B. Los. Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. *Regional Studies*, 49(5):767–781, 2015.
- C. Choi and Y. Park. Monitoring the organic structure of technology based on the patent development paths. *Technological Forecasting and Social Change*, 76(6):754–768, 2009.
- D. Chun. Patent law harmonization in the age of globalization: The necessity and strategy for a pragmatic outcome. *J. Pat. & Trademark Off. Soc’y*, 93:127, 2011.
- G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli. The statistical physics of real-world networks. *Nature Reviews Physics*, 1(1):58–71, 2019.
- J. R. Clough, J. Gollings, T. V. Loach, and T. S. Evans. Transitive reduction of citation networks. *Journal of Complex Networks*, 3(2):189–203, 2015.
- M. Contisciani, E. A. Power, and C. De Bacco. Community detection with node attributes in multilayer networks. *Scientific Reports*, 10(1):1–16, 2020.
- P. Criscuolo and B. Verspagen. Does it matter where patent citations come from? In-ventor vs. examiner citations in european patents. *Research Policy*, 37(10):1892–1908, 2008.
- J. Danguy. Globalization of innovation production: A patent-based industry analysis. *Science and Public Policy*, 44(1):75–94, 2017.

- C. De Bacco, E. A. Power, D. B. Larremore, and C. Moore. Community detection, link prediction, and layer interdependence in multilayer networks. *Physical Review E*, 95(4):042317, 2017.
- M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas. Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):041022, 2013.
- M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas. Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences*, 111(23):8351–8356, 2014.
- M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas. Ranking in interconnected multilayer networks reveals versatile nodes. *Nature Communications*, 6:6868, 2015.
- G. de Rassenfosse and B. van Pottelsberghe. A policy insight into the R&D–patent relationship. *Research Policy*, 38(5):779–792, 2009.
- A. Dechezleprêtre, S. Fankhauser, M. Glachant, J. Stoeber, and S. Touboul. Invention and global diffusion of technologies for climate change adaptation. Technical report, World Bank, Washington, DC, 2020.
- Y. T. Demey and D. Golzio. Search strategies at the European Patent Office. *World Patent Information*, 63:101989, 2020.
- H. Dernis and M. Khan. Triadic patent families methodology. Technical report, OECD, 2004.
- E. C. Engelsman and A. F. van Raan. A patent-based cartography of technology. *Research Policy*, 23(1):1–26, 1994.
- C. Fink, M. Khan, and H. Zhou. Exploring the worldwide patent surge. *Economics of Innovation and New Technology*, 25(2):114–142, 2016.
- L. Fleming. Recombinant uncertainty in technological search. *Management Science*, 47(1):117–132, 2001.
- L. Fleming and O. Sorenson. Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, 30(7):1019–1039, 2001.
- S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.
- M. D. Frakes and M. F. Wasserman. Is the time allocated to review patent applications inducing examiners to grant invalid patents? Evidence from microlevel application data. *Review of Economics and Statistics*, 99(3):550–563, 2017.
- R. J. Funk and J. Owen-Smith. A dynamic network measure of technological change. *Management Science*, 63(3):791–817, 2017.
- R. Gallotti, M. A. Porter, and M. Barthelemy. Lost in transportation: Information measures and cognitive limits in multilayer navigation. *Science Advances*, 2(2):e1500445, 2016.

- B. Hall. A note on the bias in Herfindahl-type measures based on count data. *Revue d'économie industrielle*, 110(1):149–156, 2005.
- E. Harvey, O. Maclaren, D. O’Neale, F. Patten-Elliott, S. Turnbull, and D. Wu. Network modelling of elimination strategy pillars: Prepare for it, stamp it out. Technical report, Te Pūnaha Matatini, 2021.
- I. Haščić and M. Migotto. Measuring environmental innovation using patent data. OECD Environment Working Papers 89, OECD, 2015.
- O. C. Herfindahl. *Concentration in the steel industry*. PhD thesis, Columbia University, 1950.
- K. Higham and T. Yoshioka-Kobayashi. Patent citation generation at the triadic offices: Mechanisms and implications for analysis. *Available at SSRN 4022851*, 2022.
- K. Higham, G. De Rassenfosse, and A. B. Jaffe. Patent quality: Towards a systematic framework for analysis and measurement. *Research Policy*, 50(4):104215, 2021.
- K. W. Higham, M. Governale, A. Jaffe, and U. Zülicke. Fame and obsolescence: Disentangling growth and aging dynamics of patent citations. *Physical Review E*, 95(4):042309, 2017.
- K. W. Higham, M. Governale, A. Jaffe, and U. Zülicke. Ex-ante measure of patent quality reveals intrinsic fitness for citation-network growth. *Physical Review E*, 99(6):060301, 2019.
- A. O. Hirschman. *National power and the structure of foreign trade*. Univ of California Press, 1945.
- A. O. Hirschman. The paternity of an index. *The American Economic Review*, 54(5):761–762, 1964.
- K. Hötte, S. J. Jee, and S. Srivastav. Knowledge for a warmer world: A patent analysis of climate change adaptation technologies. *arXiv preprint arXiv:2108.03722*, 2021.
- J. Huenteler, T. S. Schmidt, J. Ossenbrink, and V. H. Hoffmann. Technology life-cycles in the energy sector—technological characteristics and the role of deployment for innovation. *Technological Forecasting and Social Change*, 104:102–121, 2016.
- N. P. Hummon and P. Dereian. Connectivity in a citation network: The development of dna theory. *Social Networks*, 11(1):39–63, 1989.
- A. A. Ibrahim, A. Lonardi, and C. D. Bacco. Optimal transport in multilayer networks for traffic flow optimization. *Algorithms*, 14(7):189, 2021.
- A. B. Jaffe and G. de Rassenfosse. Patent citation data in social science research: Overview and best practices. *Journal of the Association for Information Science and Technology*, 68(6):1360–1374, 2017.
- A. B. Jaffe, M. Trajtenberg, and R. Henderson. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3):577–598, 1993.

- S. Kauffman, J. Lobo, and W. G. Macready. Optimal search on a technology landscape. *Journal of Economic Behavior & Organization*, 43(2):141–166, 2000.
- B. Kelly, D. Papanikolaou, A. Seru, and M. Taddy. Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3):303–20, 2021.
- M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014.
- D. Koutra, J. T. Vogelstein, and C. Faloutsos. Deltacon: A principled massive-graph similarity function. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 162–170. SIAM, 2013.
- J. Kuhn, K. Younge, and A. Marco. Patent citations reexamined. *The RAND Journal of Economics*, 51(1):109–132, 2020.
- F. Lafond and D. Kim. Long-run dynamics of the US patent classification system. *Journal of Evolutionary Economics*, 29(2):631–664, 2019.
- R. Lambiotte, M. Rosvall, and I. Scholtes. From networks to optimal higher-order models of complex systems. *Nature Physics*, 15(4):313–320, 2019.
- S. Lee, B. Yoon, C. Lee, and J. Park. Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping. *Technological Forecasting and Social Change*, 76(6):769–786, 2009.
- W. S. Lee, E. J. Han, and S. Y. Sohn. Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. *Technological Forecasting and Social Change*, 100:317–329, 2015.
- B. Leten, R. Belderbos, and B. Van Looy. Technological diversification, coherence, and performance of firms. *Journal of Product Innovation Management*, 24(6):567–579, 2007.
- R. Li, T. Chambers, Y. Ding, G. Zhang, and L. Meng. Patent citation analysis: Calculating science linkage based on citing motivation. *Journal of the Association for Information Science and Technology*, 65(5):1007–1017, 2014.
- M. MacGarvie. The determinants of international knowledge diffusion as measured by patent citations. *Economics Letters*, 87(1):121–126, 2005.
- M. S. Mariani, M. Medo, and F. Lafond. Early identification of important patents: Design and validation of citation network metrics. *Technological Forecasting and Social Change*, 146:644–654, 2019.
- C. Martinez. Insight into different types of patent families. Science, technology and industry working paper, OECD, 2010.
- C. Martínez. Patent families: When do different definitions really matter? *Scientometrics*, 86(1):39–63, 2011.
- S. McCabe, L. Torres, T. LaRock, S. A. Haque, C.-H. Yang, H. Hartle, and B. Klein. netrd: A library for network reconstruction and graph distances. *Journal of Open Source Software*, 6(62):2990, 2021.

- C. Mejia and Y. Kajikawa. Emerging topics in energy storage based on a large-scale analysis of academic articles and patents. *Applied Energy*, 263:114625, 2020.
- L. Mewes. Scaling of atypical knowledge combinations in american metropolitan areas from 1836 to 2010. *Economic Geography*, 95(4):341–361, 2019.
- R. G. Morris and M. Barthelemy. Transport on coupled spatial networks. *Physical Review Letters*, 109(12):128703, 2012.
- G. Morrison, E. Giovanis, F. Pammolli, and M. Riccaboni. Border sensitive centrality in global patent citation networks. *Journal of Complex Networks*, 2(4):518–536, 2014.
- H. Nakamura, S. Suzuki, Y. Kajikawa, and M. Osawa. The effect of patent family information in patent citation network analysis: A comparative case study in the drivetrain domain. *Scientometrics*, 104(2):437–452, 2015.
- K. Nakamura and A. Sasaki. 先行技術文献情報開示要件の実証分析：特許審査への影響 [Disclosure of information on prior art documents: Impacts on patent examination](in Japanese). *Kokumin Keizai Zasshi [Journal of Economics & Business Administration]*, 213(1):79–97, 2016.
- V. Nanumyan, C. Gote, and F. Schweitzer. Multilayer network approach to modeling authorship influence on citation dynamics in physics journals. *Physical Review E*, 102(3):032303, 2020.
- M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.
- V. Nicosia, G. Bianconi, V. Latora, and M. Barthelemy. Growing multiplex networks. *Physical Review Letters*, 111(5):058701, 2013.
- Y. Okada, Y. Naito, and S. Nagaoka. Making the patent scope consistent with the invention: Evidence from japan. *Journal of Economics & Management Strategy*, 27(3):607–625, 2018.
- T. E. Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.
- E. Omodei, M. De Domenico, and A. Arenas. Evaluating the impact of interdisciplinary research: A multilayer network approach. *Network Science*, 5(2):235–246, 2017.
- R. Parshani, C. Rozenblat, D. Ietri, C. Ducruet, and S. Havlin. Inter-similarity between coupled networks. *EPL (Europhysics Letters)*, 92(6):68002, 2011.
- P. G. Persoon, R. N. Bekkers, and F. Alkemade. The science base of renewables. *Technological Forecasting and Social Change*, 158:120121, 2020.
- E. Petit, B. Van Pottelsberghe, and L. Gimeno Fabra. Are patent offices substitutes? *ECARES Working Papers*, 2021, 2021.
- M. A. Porter. What is... a multilayer network. *Notices of the AMS*, 65(11), 2018.
- C. R. Rao. Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 1–22, 1982.
- C. Shapiro. Navigating the patent thicket: Cross licenses, patent pools, and standard setting. *Innovation Policy and the Economy*, 1:119–150, 2000.

- E. H. Simpson. Measurement of diversity. *nature*, 163(4148):688–688, 1949.
- L. Solá, M. Romance, R. Criado, J. Flores, A. García del Amo, and S. Boccaletti. Eigenvector centrality of nodes in multiplex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(3):033131, 2013.
- A. Solé-Ribalta, M. De Domenico, S. Gómez, and A. Arenas. Centrality rankings in multiplex networks. In *Proceedings of the 2014 ACM conference on Web science*, pages 149–155, 2014.
- O. Sorenson, J. W. Rivkin, and L. Fleming. Complexity, networks and knowledge flow. *Research Policy*, 35(7):994–1017, 2006.
- A. Stirling. A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15):707–719, 2007.
- B. Sun, S. Kolesnikov, A. Goldstein, and G. Chan. A dynamic approach for identifying technological breakthroughs with an application in solar photovoltaics. *Technological Forecasting and Social Change*, 165:120534, 2021.
- L. Tahmooresnejad and C. Beaudry. Capturing the economic value of triadic patents. *Scientometrics*, 118(1):127–157, 2019.
- M. Tantardini, F. Ieva, L. Tajoli, and C. Piccardi. Comparing methods for comparing networks. *Scientific Reports*, 9(1):1–19, 2019.
- D. Taylor, M. A. Porter, and P. J. Mucha. Tunable eigenvector-based centralities for multiplex and temporal networks. *Multiscale Modeling & Simulation*, 19(1):113–147, 2021.
- M. Vaiana and S. F. Muldoon. Multilayer brain networks. *Journal of Nonlinear Science*, 30(5):2147–2169, 2020.
- S. Valverde, R. V. Solé, M. A. Bedau, and N. Packard. Topology and evolution of technology innovation networks. *Physical Review E*, 76(5):056118, 2007.
- A. van der Marel, S. Prasher, C. Carminito, C. L. O’Connell, A. Phillips, B. M. Kluever, and E. A. Hobson. A framework to evaluate whether to pool or separate behaviors in a multilayer network. *Current Zoology*, 67(1):101–111, 2021.
- D. Vasques Filho and D. R. O’Neale. Degree distributions of bipartite networks and their projections. *Physical Review E*, 98(2):022307, 2018.
- V. Veefkind, J. Hurtado-Albir, S. Angelucci, K. Karachalios, and N. Thumm. A new EPO classification scheme for climate change mitigation technologies. *World Patent Information*, 34(2):106–111, 2012.
- D. Verhoeven, J. Bakker, and R. Veugelers. Measuring technological novelty with patent-based indicators. *Research Policy*, 45(3):707–723, 2016.
- C. Verluise, G. Cristelli, K. Higham, and G. de Rassenfosse. The missing 15 percent of patent citations. *Available at SSRN 3754772*, 2020.
- B. Verspagen. Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, 10(01):93–115, 2007.

- G. Von Graevenitz, S. Wagner, and D. Harhoff. How to measure patent thickets—a novel approach. *Economics Letters*, 111(1):6–9, 2011.
- I. Von Wartburg, T. Teichert, and K. Rost. Inventive progress measured by multi-stage patent citation analysis. *Research Policy*, 34(10):1591–1607, 2005.
- T. Wada. Obstacles to prior art searching by the trilateral patent offices: Empirical evidence from international search reports. *Scientometrics*, 107(2):701–722, 2016.
- T. Wada. When do the USPTO examiners cite as the EPO examiners? An analysis of examination spillovers through rejection citations at the international family-to-family level. *Scientometrics*, 125(2):1591–1615, 2020.
- S. Wasserman, K. Faust, et al. *Social network analysis: Methods and applications*. Cambridge University Press, 1994.
- L. Wu, D. Wang, and J. A. Evans. Large teams develop and small teams disrupt science and technology. *Nature*, 566(7744):378–382, 2019.
- B. Yan and J. Luo. Measuring technological distance for patent mapping. *Journal of the Association for Information Science and Technology*, 68(2):423–437, 2017.
- X. Yuan and X. Li. A network analytic method for measuring patent thickets: A case of fcev technology. *Technological Forecasting and Social Change*, 156:120038, 2020.
- M. Yuvaraj, A. K. Dey, V. Lyubchich, Y. R. Gel, and H. V. Poor. Topological clustering of multilayer networks. *Proceedings of the National Academy of Sciences*, 118(21), 2021.
- C. Zingg, V. Nanumyan, and F. Schweitzer. Citations driven by social connections? a multi-layer representation of coauthorship networks. *Quantitative Science Studies*, 1(4):1493–1509, 2020.
- R. Zingg and M. Fischer. The nanotechnology patent thicket revisited. *Journal of Nanoparticle Research*, 20(10):1–6, 2018.

Appendices

A Model description

For the layer interdependence and community detection analysis we use MTCOV,²⁴ the model developed by Contisciani et al. (2020). MTCOV is a probabilistic generative model that incorporates both the topology of interactions and node attributes to extract overlapping communities in directed and undirected multilayer networks. It works also with single-layer networks, since this is the special case for which there is only one layer in the ‘multilayer’ network. The model assumes conditional independence between the network and attribute data, given a set of latent variables (including the node community memberships). The likelihood function is a linear combination of the network and attribute information, adjusted by a scaling hyperparameter $\gamma \in [0, 1]$, which controls the relative contribution of the two terms: for $\gamma = 0$ the model only considers the network topology, while for $\gamma = 1$ it only considers the attribute information.

MTCOV has four parameters: two membership matrices accounting for outgoing and incoming links respectively, an affinity tensor that describes the density of links between each pair of groups among the different layers, and a parameter that matches communities and node attributes. The inference is performed with an Expectation-Maximization algorithm, and its implementation is efficient and scales to large datasets (such as the one studied here) because it exploits the sparsity of the dataset.

A.1 Cross-validation and hyperparameter settings

MTCOV has two hyperparameters, the scaling parameter γ and the number of communities C . For each network under analysis, we estimate the hyperparameters by using 5-fold cross-validation along with a grid-search to range across their possible values. For the current work, we choose to vary $C \in \{2, 3, 5, 7, 10, 12, 15\}$ and $\gamma \in \{0, 0.3, 0.5, 0.7, 1\}$. Specifically, we divide the dataset into five equal-size groups (folds), selected uniformly at random, and give the models access to four groups (training data) to learn the parameters; this contains 80% of the matrix entries and covariates. One then predicts both links and node attributes in the held-out group (test set). By varying which group we use as the test set, we get five trials per realization. For performance metrics, we measure the area under the receiver-operator characteristic curve (AUC) (for the link prediction) and the accuracy (for the node attribute prediction) on the test data, and the final results are averages over the five folds. The AUC is the probability that a random true positive is ranked above a random true negative; thus the AUC is 1 for perfect prediction, and 0.5 for random chance. The accuracy classification score is 1 for perfect recovery and 0 in the worst case of overfitting. In order to choose the best pair of hyperparameters $(\hat{C}, \hat{\gamma})$

²⁴<https://github.com/mcontisc/MTCOV>

Table 3: **Hyperparameters setting.** Values of the hyperparameters C and γ extracted by 5-fold cross-validation combined with grid-search.

	US-EXM	US-APP	EP-APP	EP-ISR	EP-SEA	JP-REJ	JP-BCK	US-AGG	ALL-AGG	MULTI
C	7	7	7	7	3	7	7	7	7	15
γ	0.3	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7

we look for the pair that performs best across both AUC and accuracy in the test set.

Since the networks are large, it is not always possible to compute the AUC on the whole training and test sets, hence we proceed with samples. In detail, we fix the number of comparisons we want to evaluate, here 10^5 , and for both the train and the test sets we sample 10^5 values from zeros entries (where there is no existing link) and we compute the link prediction on that sample (we save these values in a vector R_0); we do the same with the non-zeros entries (we save these values in a vector R_1). We then make element-wise comparisons and compute the AUC as:

$$AUC = \frac{\sum (R_1 > R_0) + 0.5 \sum (R_1 == R_0)}{|R_1|} \quad (2)$$

where $\sum (R_1 > R_0)$ stands for the number of times R_1 has a higher value than R_0 in the element-wise comparison; and $|R_1| = |R_0|$ is the length of the vector which is equal to the number of comparisons we fix. Moreover, when the network has a number of nodes bigger than 5000, we run the algorithm by computing the likelihood only on a batch of nodes (here a random subset with 5000 nodes) to speed up the computational time.

Table 3 shows the optimal hyperparameters obtained for all single-layer and multilayer networks used in the manuscript.

A.2 Layer interdependence analysis

The layer interdependence problem consists of identifying which sets of layers are structurally related, and quantifying the strengths of those relationships. To this end, we use the MTCOV model and we employ the method described in De Bacco et al. (2017). This method consists of performing link prediction in one layer with and without the information in another layer to quantify the extent to which these two layers are related. Thus, for our purposes, interdependence is based on the idea that two layers are interdependent if the structure of one layer provides meaningful knowledge about the structure of the other.

To test our ability to predict a set of target layers α , we perform experiments with 5-fold cross-validation following the same routine as above by using only the optimal pair of hyperparameters. The main difference from the community-detection procedure above is the way the training and test sets are built. In fact, for the layer interdependence task, we only split (5-fold) the links in the set of target layers α together with the attributes

for the nodes in this set, while giving full access to the set of layers β when they are added.

For this task, because we are mainly interested in link prediction, rather than in recovering covariates, we measure the AUC as in Equation (2). The final AUC is the average obtained over the five folds, each of which holds out a different subset of 20% of α . The value of the AUC depends both on the set of target layers α we are trying to predict, and on what set of other layers β we give the algorithm access to.

As described in Section 3.1, we restrict our analysis to the sublayers generated by the USPTO (separately) and the JPO and EPO layers (as sets of sublayers), without exploring all possible combinations of sublayers. In detail, we consider the following experiments:

- (a) $\alpha = [\text{US-APP}]$, $\beta_1 = [\text{US-EXM}]$, and $\beta_2 = [\text{US-EXM}, \text{EPO}, \text{JPO}]$.
- (b) $\alpha = [\text{US-EXM}]$, $\beta_1 = [\text{US-APP}]$, and $\beta_2 = [\text{US-APP}, \text{EPO}, \text{JPO}]$.
- (c) $\alpha = [\text{EPO}, \text{JPO}]$, $\beta_1 = [\text{US-APP}]$, $\beta_2 = [\text{US-EXM}]$, and $\beta_3 = [\text{US-APP}, \text{US-EXM}]$.
- (d) $\alpha = [\text{US-APP}, \text{EPO}, \text{JPO}]$, and $\beta_1 = [\text{US-EXM}]$.
- (e) $\alpha = [\text{US-EXM}, \text{EPO}, \text{JPO}]$, and $\beta_1 = [\text{US-APP}]$.

Note that for the JPO and EPO, we are using all sublayers of these two jurisdictions. Furthermore, when the set α contains only a sublayer of USPTO [(a), (b)], the hyperparameters used by the algorithm are $C = 7$ and $\gamma = 0.7$, which is the optimal choice for the US-AGG network. For [(c), (d), (e)] the algorithm uses $C = 15$ and $\gamma = 0.7$, which is the optimal choice for the multilayer network, for computational simplicity.²⁵

B Network comparison

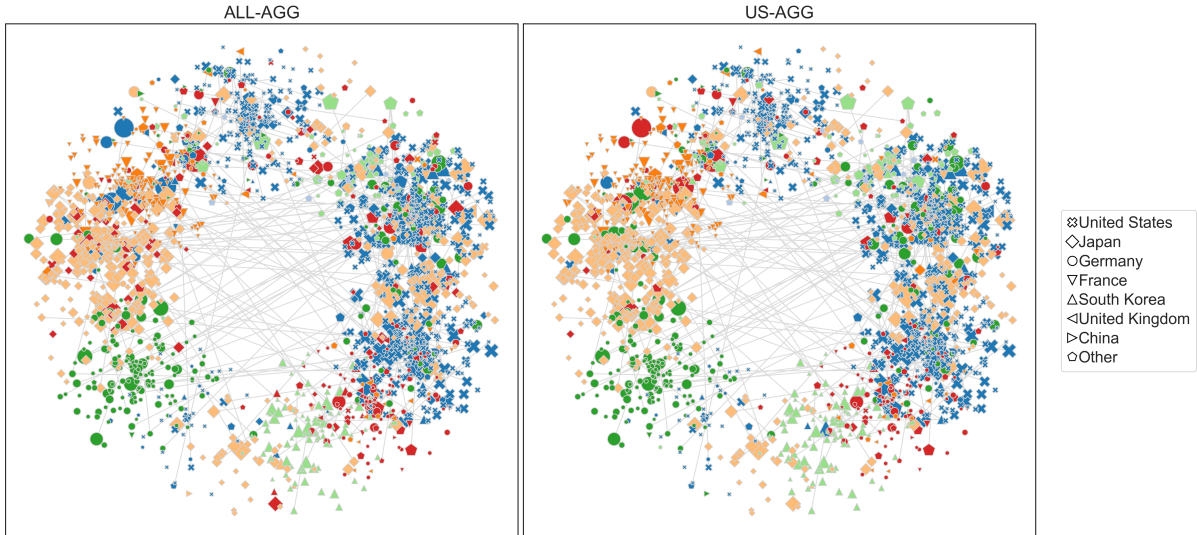
B.1 Class network construction

We use network comparison methods in order to quantify the differences in the technological information contained in the MULTI, ALL-AGG, and US-AGG networks. In particular, we directly compare a projection of the bipartite network of relationships between the extracted communities and the 3-digit Cooperative Patent Classification (CPC) classes with co-classification and inter-class citation networks. Figure 5 displays the communities extracted for a random subset of the nodes and edges in ALL-AGG and US-AGG.

To build the bipartite network between communities and classes, we first populate a matrix P whose dimensions are given by the number of families (22653) times the

²⁵A cross-validation procedure to detect the best pair for the different sets α was determined to be too computationally expensive.

Figure 5: **Community extraction for comparison networks.** This diagram shows the hard community membership partitions for the ALL-AGG and US-AGG networks. As for Figure 4, we use a random sample of 2000 nodes and include any incidental links. The layout is determined by the results for MULTI, for purposes of direct comparison, while the colouring shows the communities found for each network (7 communities for each of ALL-AGG and US-AGG). Node size is proportional to the number of outgoing and incoming citations, while node shapes denote the location of the assignee of each patent family.



number of classes (535). This is a binary matrix with non-zero entries when a family is assigned to a given class. We then normalize the matrix such that each column sums up to one. In this way, we can consider the matrix P to be the membership matrix of the classes among the patents. By multiplying the transpose of the membership matrix of the patents among the communities and the previous matrix P , we get the bipartite network $D = U^T P$ of relationships between the extracted communities and the classes. To ease the comparisons, we need to project this bipartite network onto the technology classes to obtain a network of classes. The projection onto the class nodes is computed through the matrix multiplication $D^T D$ between the bipartite matrix D and its transpose. This projection has non-zero entries when pairs of classes are both found in the same communities, with weights proportional to their relative frequencies within those communities. As baseline comparisons, we use the co-classification and the inter-class citation networks. The former is obtained by the matrix multiplication $P^T P$, while the latter is constructed as described in Section 3.3.

After running the community-detection algorithm for both the optimal and non-optimal partition of each of the three networks MULTI, ALL-AGG, and US-AGG, and only then can we obtain six projected networks among which we are able to make fair comparisons. Table 4 shows the performance of MTCOV on the citation networks (with

Table 4: **Results of link prediction and covariate prediction tasks.** We measure AUC (link prediction) and accuracy (covariate prediction) over 5-fold cross-validation for C equal to 7 (the optimal value for ALL-AGG and US-AGG) and 15 (the optimal value for the MULTI network); $\gamma = 0.7$ (the optimal value for all the networks).

	C	AUC	Accuracy
MULTI	7*	0.835	0.341
	15	0.852	0.422
ALL-AGG	7	0.730	0.402
	15*	0.739	0.393
US-AGG	7	0.736	0.426
	15*	0.749	0.406

non-optimal parameters identified with the symbol *) for the link prediction (AUC) and covariate prediction (accuracy) tasks, using 5-fold cross-validation.

After extracting communities, we construct the six bipartite (community-class) networks which we then project onto the class nodes to compare with the co-classification and inter-class citation networks.

B.2 Diversity measures

Two diversity measures are used in the main body of this work: Rao-Stirling diversity (RSD) and the Herfindahl–Hirschman Index (HHI). For each network, RSD is calculated at the *extracted-community level* and then a median is taken across communities. The RSD for community c is calculated as (Stirling, 2007):

$$RSD_c = \sum_{i,j,i \neq j} d_{ij} p_{i,c} p_{j,c}, \quad (3)$$

where d_{ij} is a known distance measure between 3-digit CPC technology classes i and j , while p_i and p_j are the proportion of families in the community that are assigned classes i and j , respectively. Two factors complicate this calculation. First, because each family can be assigned multiple categories, RSD can take on values greater than one. Because we are directly comparing the RSD for the same set of families (our networks have the same set of nodes), this is not a concern. In fact, we believe this is sensible for this data. That is, if a community consists of a set of families that are all assigned the same two classes i and j , our procedure here will treat these communities as consisting of 100% i and 100% j (minimal diversity) rather than 50% i and 50% j (maximum diversity), for a given d_{ij} . Second, because we allow overlapping communities (i.e., a node can be assigned multiple communities with different weights), p_i and p_j are the weighted sums

over patent families f in c :

$$p_{i,c} = \frac{\sum_{f \in i} w_{f,c}}{\sum_{\forall f} w_{f,c}}, \quad (4)$$

where $w_{f,c} \in [0, 1]$ is the weight of family f that is assigned to c .

For our purposes, d_{ij} is one minus the normalised link weight in the inter-class citation network constructed for our network comparison calculations. This metric is scaled such that distance zero corresponds to the strongest citation linkage for each class, and distance one corresponds to no citation linkage. These new weights act as proxies for the level of surprise, where a weight of zero indicates two classes that only ever cite each other, while a weight of unity indicates two classes that never cite each other. As such, the ‘level of surprise’ parameter d_{ij} down-weights combinations that we expect while exaggerating those that we don’t. This adjustment is important. For any given technology class, the number of classes with which it shares community membership depends crucially on both the classification system and the level of the hierarchy within this system that we choose to use. When a class starts to get too crowded, for example, it may be split to make technical search easier (Lafond and Kim, 2019) — after all, this is one of the primary goals of patent classification systems. For this reason, a distance measure like d_{ij} is crucial to incorporate into technological diversity measurements.

HHI, also called the Simpson diversity index, is calculated at the *technology level*, i , to measure the extent to which technology classes are split across extracted communities. A median across technology classes is then calculated. The HHI for class i is calculated as:

$$HHI_i = \frac{N \sum_c p_{i,c}^2 - 1}{N - 1}, \quad (5)$$

where $p_{i,c}$ is defined as in Equation (4) and N is the total number of communities into which families can be assigned (7 or 15, in our case). Equation (5) is the unbiased version of the HHI (Hall, 2005); this version corrects the $1/N$ offset that affects the standard version of the HHI (for which $1/N$ is the minimum value), which is the sum in the numerator of Equation (5). The HHI measures how much a technology class is splintered across communities, ranging from HHI=0 for maximally spread to HHI=1 for maximally concentrated. We note that the goal of the community detection process was *not* to replicate the CPC system as closely as possible. There are many valid reasons why a technology class may be split across communities, such as when a technology is particularly generalisable and is applied to (and cited by) many seemingly unrelated fields. Instead, the HHI gives us an idea of what is, or is not, driving the results we obtain for the direct network comparison.

C Ego network details

The diagram in Figure 1 shows the multilayer ego network of a triadic patent family, labelled **A**. Table 5 lists the seven families in this diagram, alongside their granted equivalents.

Table 5: **Example network subset details.** Details of each of the families displayed in Figure 1 are shown below. Priority indicates the month of first filing. All families consist of three triadic patents except for **C**, which includes multiple family members at the USPTO and EPO.

Node	DOCDB Family	Equivalent			Priority
		USPTO	EPO	JPO	
A	19192289	6819081	1333511	3671007	2002-01
B	17414436	6174618	0905803	3777748	1997-09
C	26411133	6211645, 6211646	0892450, 1030389, 1030390	4487967	1997-03
D	36242735	7615309	1695401	4527117	2003-12
E	37115311	7687192	1872418	4739405	2005-04
F	38522869	7967506	1994626	5133335	2005-03
G	37115315	7488201	1872421	4663781	2005-04