

## EMPIRICAL RESEARCH ARTICLE

# Exploring the Roles of Non-Commercial Entities in Open and Coopetitive Ecosystems

Anonymous author | The other one

<sup>1</sup>Not that department, At that University, In there

## Correspondence

Corresponding author: Super Mario

Email: SuperMario@ninten.do

## Present address

NY, Italy or Japan, not sure.

## Abstract

In this research, we address two contemporary management paradoxes in the high-tech sector: *competition versus cooperation* and *open-source versus proprietary* innovation within the high-tech sector. Focusing on the TensorFlow open-source platform led by Google, the research examines how competitors collaborate in the co-production of this widely-used machine learning platform. Through narrative and social network visualizations, the study explores the evolving roles of non-commercial entities within this ecosystem. The findings suggest that a permissive intellectual property regime can facilitate collaboration across industry, government, and non-commercial sectors, leading to innovation that benefits the commons. This case illustrates how aligned incentives and relaxed IP constraints can promote broad-based innovation, emphasizing the significant contributions of non-commercial entities in maintaining software commons and driving innovation. We outline motivations for high-tech giants to open-source platforms, and outline motivations for non-commercial organizations to contribute to open and coopetitive platform ecosystems. Furthermore, we propose an extension of the triple-helix model to better accommodate the important role that open-source foundations play in an open and competitive innovation model.

## KEYWORDS

Open-Innovation, Innovation Ecosystems, Coopetition, Open-Source, Open-coopetition, Triple-Helix, Non-Commercial Entities, Intellectual Property, Platform Ecosystems, Artificial Intelligence, TensorFlow.

## 1 | INTRODUCTION

Many see software as being produced by a single firm. Plenty of evidence shows however that software is often co-produced in networks (see ?????). Paradoxically, those networks can link rival and competing firms that cooperate with each other in an open-source way. For, instance it is known that Apple and Google cooperated in the development of open-source web browsing technologies while fighting expensive patent wars in the courts worldwide, or that Toyota, Ford, and Mitsubishi Motors cooperated in the co-production of open-source automotive software while fighting for car sales in overlapping geographical areas (??). Managing the production of software in an open and coopetitive modus operandi is increasingly popular (??) but remains challenging as value can erode via commoditization, free-riding and unintended spillover effects (???). As pointed out in a recent mapping study by (?), “since more and more companies contribute to open source software and/or develop their software products as open source, collaboration between developers of competing companies becomes an important issue. If developers from competing organizations contribute to the same project, this could lead to issues within a project” (?).

As research explaining coopetition in an open-source way (aka open-coopetition) remains scarce but gathers cross-disciplinary interest in Software Engineering, Strategic Management, Innovation Studies and Information Systems (e.g., ???), and as “companies use more and more open source in communities including their competitors it seems very important to know why, how and for which outcomes they follow this kind of strategy”(?), we conducted an exploratory case study guided by three broad and open research questions: “Why do tech giants like Google open-source advanced and complex technological platforms that started in-house?”, “Why are different organizations cooperating with their competitors in the co-production of those open-source platforms?” and “Why non-commercial organizations contribute to the co-production of those open-source platforms?”.

To do so, we take the case of TensorFlow, an advanced and complex technological platform for **ML!** (ML!) and **AI!** (AI!) that started in-house at Google but was released as an open-source project in 2015, and became a key and very popular building block for the ones embedding deep learning technology into thousands of products and services worldwide with impact on everyday life.

## 2 | THEORETICAL BACKGROUND

### 2.1 | Innovation networks and ecosystems

Even if many see firms as distinct and autonomous units of action, it has been both sighted and increasingly recognized that firms are embedded in networks comprised of close, robust and multidimensional connections that blur organizational boundaries (????). Organizational theory now stresses the need to understand how the relational context in which firms are embedded influences their behaviors (??).

Early identified drivers of inter-firm cooperation include: reducing costs through product rationalization and economies of scale (??), sharing total risks and total costs of large projects (??), accessing new knowledge and complementary assets (??), developing technology and accessing complementary markets (?), co-creation/co-production of value (??), solving market failures that emerge under conditions of bounded rationality (??), shaping competition with the aim of increasing or decreasing market competition (??), faster access to new markets (??), gaining legitimacy and reputation (??), securing investment (??) and reducing uncertainty from resource requirements (?).

The ecosystem construct, as a way of making the networked interdependence of the firm more explicit, has gained prominence in both research and practice (???). Theory underlying the ecosystem construct have focused on understanding coordination among partners in exchange networks that are characterized by simultaneous cooperation and competition (??). Studies in this arena have explored the challenges that arise when incentives across the ecosystem are not aligned (?), the role of established relationships with ecosystem partners in shaping firms motivations to compete for different market segments (?), the activities that focal firms undertake to induce exchange partners to favor their specific technology platforms (?), and the flow of activity among partner firms (?).

In our view, the ecosystems construct stresses the importance of actor-to-actor networked relationships. Therefore, our view on strategic networks (????) and ecosystems (???) approximates what others called "networked collaborations" (?), "lattices" (?), "webs" (?), "constellations" (?), "holonic organizations" (?), "interfaces" (?), "organizations networks" (?), "inter-organizational domain" (?) and "infrastructure" (?).

While attracting multi-disciplinary attention, the ecosystem construct also suffered from fragmentation in recent years. Many compound words surged such as "software ecosystem" (??), "digital ecosystem" (?), "socio-technical ecosystem" (?) and "platform ecosystem" (?) which are now growingly accepted across multiple disciplines ranging from Software Engineering and Information Systems to Marketing and Innovation Studies. Moreover, terms such as "business ecosystems" (???) and "innovation ecosystems" (????) have become widely employed compounds in organization, strategic management research and innovation studies. In our view, while the terms "software ecosystem", "digital ecosystem", and "socio-technical ecosystem" emphasize technology developed within a network of individuals and firms, the "business ecosystem" and "innovation ecosystem" words emphasize a network of the different partner organizations working together. Along the same line, we also point out that the "platform ecosystem" term often denotes that the ecosystem is centred around a central technology or service (i.e., core platform) controlled by a single provider.

Prior research has considered mainly the independent motivations and opportunities that guide ecosystem formation at the dyad-level (?). In this study, we explore the evolution of cooperation among competitors at multiple levels; we examine relational interactions at the inter-individual, inter-firm and inter-ecosystem level (i.e., zooming in and out). As pointed out by ?, distinctive issues concerned with the alignment of individual and collective networking interests should not be separated. Therefore, theory explaining the evolution of ecosystems should consider both inter-individual and inter-organizational relationships.

## 2.2 | The triple-helix model of innovation

### 2.2.1 | From linear to helix models of innovation

In the early twentieth Century, to explain innovation, the Linear model of innovation came into existence to explain the relationship between technology and development. However, due to its limitation of being unable to explain the growing relationship among academia, industry, and government. The growing network between a industry and academia created a mutual relationship that evolves into strategic alliance, allowing institutional development and technology-market evolution side by side. In this way, knowledge is becoming more integral to society's infrastructure. This growing co dependency propelled scholars to develop new ideas. Which also generated investigation on university's role into question.

Triple helix model is a framework to help explain the cooperation among University-Industry-Innovation relations into a model for studying knowledge-based societies in various developed and developing economies. The concept was opposite of the existing ideas focusing on industry-state or industry. The framework originated from and's work. Both Etzkowitz and Leydesdorff were inspired by sociological concept of triads (i.e., association of three) by Georg Simmel (see). The Triple Helix model is developed to aid in the explanation rather than at the level of the phenomenon that needs to be explained. It aids to explain the phenomenon of shifting knowledge from a traditional industry-government dual relationship to a growing triangular association among universities, industry, and the government in the knowledge society and innovation. Triple helix framework shows that to continue innovation and economic development in a knowledge society; the government, university, and industry must work together to develop various institutional & social mechanisms to produce, create, apply, and preserve knowledge. This framework implies continued innovation and development among the three mentioned spheres.

The triple helix activities start by collaboration among the mentioned three actors. For example, industries, universities, and local governments in a region can come together to participate in a discussion about the development of the region, the universities may train their students tailored for the region's needs, and companies building factories, and offices to employ those future graduated students. Companies also share their know-how with the universities, providing fund along with universities for projects and in return universities apply and mature those technologies. As new information and technology are produced, the triple helix shifts in orientation. Enhancing the performance of universities and other knowledge-producing institutions frequently emerges at this level of the triple helix as a crucial component of a plan to revitalize an aging economy or launch a new venture based on intellectual capital in one form or another, ranging from official research and development in government, university, and industrial laboratories to tacit knowledge emanating from already-existing industries.

In this framework, the university and other knowledge-producing organizations take the position of industry as the core spiral as new knowledge overlays and varied combinations of old and new knowledge become the foundation for information. Then, business and government may start assisting with academic advancement. A common tactic is to establish a research center to expedite the creation of scholarly research. The government and businesses provide the university with extra funding so it can better carry out one of its core missions, which is research. Due to this universities increasingly become a source of scientific knowledge economic development. Academic institutions are re oriented or founded for this purpose. The conventional role among University, Industry starts to blur as the university starts to license technologies, patenting intellectual property. This lead to university-government-industry falling into hybrid role where university becomes the main sphere for knowledge production. However each entities keep their independence from each other.

### 2.2.2 | Triple Helix Model and Software Industry

In the USA, tech cluster areas like Silicon Valley house companies like Alphabet, Tesla, Amazon, and Intel showcase the triple helix model in action. Recent studies show that Google was created thanks to the roles of Stanford University, UC Berkeley, funding from the US government via the National Science Foundation and Angel Investors. Modern innovations rarely came from the work of a sole individual or organization. Our research that looks at the co-production of the TensorFlow platform for AI and ML shows that innovations can be done in very large networks that embed thousands of actors and inter dependencies, over the internet without clear geographical boundaries that characterize tech clusters like Silicon Valley, Dublin, Seoul or the Shenzhen-Hong Kong-Guangzhou.

### 2.2.3 | Problem in Triple Helix Model

Many critics have shown their displeasure about Triple Helix Model. Some critics believe that triple helix model lacks a strong idealized foundation and it is not a neutral framework [1, 2, 3]. Others concur that the model has limited control in explaining many practical situations [4, 5, 6]. It is also said the framework has a limited capacity to explain when different conditions are compared with one another [7, 8]. The model is also criticized as a means for applying in every situation, attempting to use it as a universal system [9]. The Triple Helix Model also excludes civil society as [10] argues that "public" or "society" is hardly qualified to be the fourth helix as "university", "industry" and "university" are specialized spheres that communicate with each other in a formal, coded manner and evolves in a civil society.

Even if [11] and [12] argued that civil society is considered as the enabler for triple helix interaction. Criticism on the model, ultimately led to the proposal for developing a quadruple helix which includes a "media-based and culture-based public" [13] that favors quadruple helix over triple helix as the former conceptual framework clarifies the formal and informal role of approaches of supporting social innovation, emphasizing that it is really helpful to explain the role of community in innovation. [14] pointed out that triple helix model is insufficient for analyzing the various sectoral partnerships for entrepreneurship in modern society. Supporters of the triple helix model claim that due to its simplicity, triple helix model reduces the complexity of dynamics in innovation system of knowledge economy and it helps to elaborate the essential dynamics more clearly [15]. Leydesdorf proposed that empirical contexts and data are the key requirements to decide whether a triple helix or more than a triple helix framework is needed for analysis [16].

While not directly advocating for including NGOs in the Triple Helix Model, [17] highlighted NGOs role in bridging interactions between the three helices, particularly within the context of community empowerment projects. The work emphasized how NGOs can facilitate connections and foster a more integrated approach to innovation and development. Other research claims that an expanded model based on triple helix by keeping citizens in mind like what [18] proposed is a necessity of time and it is suitable for addressing new development in a contemporary society [19, 20]. Furthermore, it argues that by including NGOs in the triple-helix, the framework would better accommodate the important role that open-source foundations play in an open and competitive innovation model.

### 2.3 | Cooperation among competitors

As noted in the literature on innovation ecosystems [21], strategic cooperation among competitors (also known as cooptation) is not uncommon. This phenomenon can be observed in various industries such as automotive, pharmaceuticals, and airlines. For example, city car models like Toyota Aygo, Peugeot 108, and Citroën C1 are produced in the same factory through a joint venture between Toyota Motor Corporation and PSA Peugeot Citroën. In the pharmaceutical industry, leading companies establish consortia such as TransCelerate BioPharma to develop shared industry clinical-trial solutions. Similarly, alliances in the airline industry have blurred competition between individual firms by forming airline alliances against one another [22]. Additionally, companies like Apple, Google and Samsung have collaborated on open-source web-browsing technologies despite engaging in patent war across the courts worldwide [23].

From another perspective, economists have examined this phenomenon from both public and macroeconomic perspectives. For example [24] have studied cooperation among competitors in ATM and card-payment systems. Their work aligns with neoclassical economic theory, which assumes that competition leads to economic efficiency [25]. Antitrust and regulatory policies aim to prohibit agreements or practices that hinder free trading and competition between businesses [26]. While strategic management scholars (e.g., [27]) see cooptation as something desirable, economists call for regulation on collaboration that can hinder free trading and competition between companies [28].

Cooptation cases are not always success stories. As pointed out there is a risk of unwanted leakage of information and opportunistic behaviour [29]. This is especially critical in the open-source domain characterized by transparency, inclusiveness and a weak intellectual property regime [30]. As far as innovation is concerned, cooptation strategies have been revealed to be more conducive to than purely cooperative or competitive strategies [31]. There is a call for more research into the connection between innovation and cooptation [32]. Furthermore, and also regarding cooptation research, there have been multiple calls to analyze cooptation at multiple levels [33]. It was noted that the tensions and coping in cooperative relationships at the individual-level and the organisational level cannot be separated [34].

Although competition and collaboration have individually been extensively investigated, researchers have given limited attention to the fundamental issue of how these two concepts interact [35]. Additionally, existing literature on cooperation among

competitors is mostly based on joint ventures and R&D consortia where access is restricted to a few selected members (e.g., ??). In contrast, in open source software development, third-party actors generally do not require permission to contribute. Despite recognition of the importance of understanding open-source software from both competitive and cooperative perspectives in strategic management literature (see ?), there are very few empirical cases exploring cooperation among competitors in the open-source arena (see ?, for a recent review). This lack of empirical research on the topic is unfortunate as “companies use more and more open source in communities including their competitors it seems very important to know why, how and for which outcomes they follow this kind of strategy” (?).

The network-perspective is strongly embedded in the theories of coopetition as they explore relationships among individuals, firms and networks (?). For example, ? explored learning and knowledge sharing through networks. The importance of network characteristics and positions for the joint-development of competitive advantages was deeply investigated by ?, ? and ? as well. More recently, ? identified key determinants that constrain the emergence, morphology and evolutionary dynamics of inter-firm networks. Now that firms are changing from a former industrial logic focusing on internal resources to a new logic based on the ability to integrate external resources through networks (??), the network-perspective is gaining strong momentum.

## 2.4 | Open-source software

Much of the groundbreaking coding that drives software applications, operating systems, cloud servers, and the Internet is a result of “open-source” code - this means code that is freely distributed rather than being kept confidential or protected with a strong intellectual property regime. Numerous individuals contribute to open-source projects for various reasons such as aiding others, gaining recognition in their field, acquiring new skills and knowledge, finding enjoyment and fulfilment from their work, and out of altruism. Additionally, some are incentivized by employment or financial compensation to participate in open-source initiatives (?).

From an innovation studies perspective, ? and ? suggested that open-source software development shows that users program to solve their own as well as shared technical problems and freely reveal their innovations without appropriating private returns from selling the software. Such “free” user-to-user assistance has turned open-source into a remarkable example of user-innovation (?). It was also reported that the open-source trend has been so strong that previous, rather monolithic, organizations (e.g., SAP, Intel, Apple, Philips, Xerox, and IBM among others) decentralized research labs, opened up their proprietary technology, and increase their absorptive capacity for outside-in innovation processes within open-source ecosystems (?).

From an organizational perspective, ? also proposes that open-source software development is an example of a compound model of innovation that contains elements of both the private investment and the collective action models. They also argue that addressing open-source software may lead to a substantial rethinking of the concept of “organization for innovation” and a better understanding of innovation among distributed users who derive utility from freely revealing their information-based innovation to produce a collective good.

The open-source software phenomenon keeps evolving from the earliest purist views focusing on freedom (?) to newer perspectives considering open-source as an alternative and viable way of doing business (??). Moreover, it has expanded from open-source software to open-data (?), open-hardware (?), open-knowledge (?), and open-access (?), among other manifestations of increasing openness in the co-production of goods. Even if the open-source phenomenon started to attract early scholarly attention in computer science and software engineering, the phenomenon is more recently capturing the largest interest from business and management scholars (?). Therefore, as pointed out by ??, information systems as a discipline is well positioned to be at the centre of trans-disciplinary research addressing the phenomenon.

Given the rise of open-source outside the software domain and given the lack of consensus over the meaning of openness (?), ? proposed six dimensions of openness within high-networked socio-technical ecosystems. The six dimensions emphasize 1) architectural openness (hardware and software); 2) compliance with standards; 3) transparency and inclusiveness of governance; 4) free market policies rewarding innovation and entrepreneurship; 5) the absence of purposive lock-in mechanisms; and 6) an open regime of intellectual property. There is some agreement that, in the software world, open-source refers to products under a license approved by the Open Source Initiative (???). Outside the software world (e.g., services, platforms, ecosystems), what is meant by “open-source” and “openness” is rather blurry (?).

It is worth remarking that the research on open-source has attracted steady attention from multi-disciplinary scholars over the last decades (?????). To illustrate the growing academic relevance of the open-source phenomenon, we observed that many prominent academic outlets, including *Research Policy*, *IEEE Network*, *IEEE Software*, *Management Science*, *Criticism* and the

*Journal of the Association for Information Systems*, have published special issues on open-source software. Moreover, several comprehensive literature reviews are available (?????).

In an attempt to explain such trans disciplinary attention, ? proposed five characteristics for why open-source software is particularly attractive for examination in various fields and disciplines using a plethora of research methods, as follows: (1) impact: open-source software has an extensive impact on the economy and society; (2) theoretical tension: the phenomenon deviates sharply from the predictions and explanations of existing theory in different fields; (3) transparency: open-source software has offered researchers unprecedented access to data; (4) communal reflexivity: the community of open-source software developers frequently engage in a dialog on its functioning (it also has its own research community); and (5) proximity: the innovation process in open-source software resembles knowledge production in science (in many instances, open-source software is an output of research processes). Regarding point 2, (theoretical tension), open-source is particularly challenging to wide set of views on organizations and strategic alliances which emphasize the concepts of property, contract, and/or transaction (which are in turn de-emphasized in open-source software).

Many scholars have leveraged the network perspectives to study the open-source phenomenon from very different disciplines (?), who have based their network analysis on metadata from public source-code repositories [teixeira2015lessons], email data in bug-fixing contexts. (e.g., ??). In this research, we also conducted a network analysis; however, unlike most of the research mentioned above with cross-sectional analysis of static networks, we adopted a longitudinal view, as we are more interested in the evolution of cooperative network over time. Rather than extracting static quantitative indicators from the cooperative network by solely looking at digital artifacts, we also look at their surrounding industrial environment. All while considering inter-individual and inter-organizational behaviours at the same time (i.e., software developers and the organizations they are affiliated with).

Many scholars have used network perspectives to study the open-source phenomenon from various disciplines (?). For example, while ? conducted a network analysis based *who fixes bugs with who* on 120 projects hosted on SourceForge, while ? examined networks on *who reviews who* on the Linux kernel. In our research, we also performed a longitudinal view of the cooperative network's evolution over time. Unlike most previous studies that used cross-sectional analysis of static networks and extracted quantitative indicators solely from digital artefacts, we considered inter-individual and inter-organizational behaviours simultaneously (e.g., software developers and their affiliated organizations) while paying constant attention to the surrounding industrial environment in which social networks are embedded.

The motivations behind contributing to open-source software projects are multifaceted and have been extensively explored by researchers. Several studies have shed light on the key drivers for OSS participation. These include both intrinsic and extrinsic factors (see ?????). On the intrinsic side, the joy and intellectual challenge of creative problem-solving in OSS development are significant motivators, as highlighted by ?. The collaborative nature of OSS also enables knowledge sharing and collective learning, which are crucial for professional growth and innovation, as noted by Von Krogh's work. Complementing the intrinsic rewards are extrinsic incentives, such as enhanced reputation and career opportunities, which ? identify as significant drivers. Employing a social network perspective, ? further expand on the social and community aspects, where contributors gain social capital and peer recognition, fostering a sense of belonging and identity within the OSS community.

Additionally, the visibility and transparency of contributions in OSS projects can lead to career advancements and networking opportunities, as described by ?. Contributors often engage in OSS projects to solve personal or work-related problems, directly benefiting from the software they help develop, as explained in ? and ?. Overall, the confluence of intrinsic satisfaction, social rewards, professional advancement, and direct utility from the software itself creates a powerful motivation for individuals to contribute to the vibrant and dynamic OSS ecosystem. While extant theory provides a lot of explanation on why individuals contribute to open-source software projects, there is lack of knowledge about why organizations contribute especially from the point of view of non-commercial organizations.

### 3 | METHOD

Our case relied mostly on digital trace data publicly available on the Internet. To a larger extent, the collected data was naturally occurring, in the sense that it did not result from researchers' actions, but rather created and maintained by the TensorFlow community in their pursuits of developing an open-source platform for **ML!** and **AI!**. To make sense of coopetition within TensorFlow, we followed the methodological approach by ? that combines the qualitative analysis of archival data (QA), the mining software repositories (MSR), and Social Network Analysis (SNA) to reconstruct and visualize collaboration using (sociograms). The methodological approach by ? was previously explored by others (e.g., ?) and according to a recent systematic mapping study by ?, p. 12, it merits by not only modelling inter-developer collaboration but also considering "inter-company"

collaborations. The method and the tool<sup>1</sup> provided by ? associates software developers with each other by co-edits of the software source code, and attributes an organizational affiliation to each developer by the e-mail domain they use. Table ?? presents other established cross-disciplinary methodological notes that guided our research execution.

**TABLE 1** Guiding methodological notes.

Employed approach	Established within	Methodological notes
Case study rooted on archival data	Multidisciplinary	? ?
Mining software repositories	Software-Engineering	? ?
Social network analysis	Multidisciplinary	? ?
Use of digital trace data	Multidisciplinary	? ?
Network analysis of digital trace data	Multidisciplinary	? ?
Network analysis with emphasis on the visualization of cooperative activities	Biomedicine Innovation-Studies	? ?
Network analysis of open-source software projects	Information-Systems Software-Engineering	? ?
Semi-structured interviews	Multidisciplinary	? ?

As in ?, we aimed to integrate in a converging fashion information from multiple sources of evidence. Among them the software repository logs (retrieved using Git from the GitHub TensorFlow project), company announcements, financial reports, specialized press reports and actual information systems artefacts (e.g., the TensorFlow source code and its documentation) . All while encapsulating different qualitative, quantitative and relational social network data. We began our research qualitatively while learning about TensorFlow and its surrounding empirical background. This first stage, allowed us to review an immense amount of online information regarding the **ML!** and **AI!** industry. An internal web-based information system was used to keep and organize the retrieved qualitative materials. After attaining a better understanding of the industrial cooperative and competitive dynamics, we extracted and analyzed the social network of the TensorFlow open-source project by leveraging SNA (?).

At a later stage, we then retrieved the TensorFlow repository and its *changelog*. The commit logs were then scrapped, modelled and visualized as social networks using *Python NetworkX*(v3.3) (?), *Python Matplotlib*(v3.8.4) (?) and *Visone*(v2.27.1) (?). All guided by the methodology provided by ?. By mining digital traces of collaboration, and uncovering the evolution of social structure in the TensorFlow project, the computerized SNA also led to novel insights on open-source and competition. The combination of methods was not only fundamental for the retrieval of social structures but also for explaining them. We attempted to explain how the collaborative networks evolved year after year for the first 10 years of the history of the TensorFlow project. Our analysis started with the first code commit on Nov 7 2013 by Vijay Vasudevan (i.e. the birth of TensorFlow as an open-source project outside Google) and we closed our data analysis on 12 April 2024 for reporting the first results of this research<sup>2</sup>. Some of the networks were filtered to cover only developers affiliated with the top contributors that can be easily identified as affiliated with competing firms that market the same products and services on the same geographical areas (i.e., 'Google', 'microsoft', 'ibm', 'amazon', 'intel', 'amd', 'Nvidia', 'arm', 'meta', and 'bytedance').

After building many SNA visualizations capturing inter-individual and inter-organizational collaboration, we embraced the principles of engaged scholarship for (see ?) we started contacting developers to collect additional qualitative semi-structured interview data and triangulating our results obtained from online documentation and mining the TensorFlow repository with Social Network Analysis. We contacted first developers via e-mail without replies and we moved then to the LinkedIn online social network where three of the eight contacted software developers agreed for 20-minute interview about collaboration, competition and trust in the TensorFlow open-source project<sup>3</sup>. Their feedback led us to re-run the computerized SNA in a way

<sup>1</sup> See <https://github.com/jaateixeira/ScrapLogGit2Net>.

<sup>2</sup> Note we do not provide SNA visualization results for 2024 as the covered period is too small for drawing parallels with other years.

<sup>3</sup> Here we speculate that both prior research on the open-source phenomena by the first author and prior professional experience on the software domain by the first author might have helped in the data-collection

that would model affiliations using an **api!** (**api!**) provided by GitHub, and that hosted the TensorFlow code repository with minimal changes to the results. Both the visualizations capturing inter-individual and inter-organizational collaboration were validated by the developers, while we all agreed that the value of ?? is limited due to its density and complexity. All developers were promised anonymity, the video conference audio was not recorded but transcribed on a local secure machine with the whisper is an automatic speech recognition<sup>4</sup>.

As in prior multi-disciplinary studies (????), our analysis emphasizes the visualization of the cooperation network, which evolves over time, to reveal the dynamics among the OpenStack software developers. We then attempted to understand the visualized networks with our acquired understanding from the competitive cloud computing industry in general and OpenStack in particular. The visualization, together with a deeper understanding of the phenomenon under investigation, corresponds to the notion of figuration (???) as pointed out in several studies (??).

Studies that take an SNA approach are long established in organizational research in general (?????), and in innovation studies (??) and open-source research in particular (??). Even so, few studies have exploited the potential of social network visualizations for exploratory research, as recommended by ?. By capturing the evolution of cooperative (and some competitive behaviors) at the individual, firm and ecosystem levels, our multi-method approach aims to contribute to a broader range of methodologies capable of bridging the gap between causal processes at the macro-structural organization and those operating at the individual level (see ?).

## 4 | RESULTS

As in ??? our results aggregate qualitative, quantitative and social network evidence in a mix-methods design. Qualitative textual evidence collected from websites and semi-structured interviews are blended with quantitative simple counts and “pictures” of the collaborative social structure of the TensorFlow ecosystem. While the efforts of ?? that guided our research execution focused on the top 10 contributors to an open-source platform ecosystem (i.e., the OpenStack cloud computing infrastructure), our research focused on the top 10 commercial contributors as well as the top 10 non-commercial contributors (see ??? as we were interested in exploring the roles of non-commercial entities in open and cooperative ecosystems such as TensorFlow).

We started by analysing the inter-individual collaborative network of TensorFlow (i.e., a network of software developers that co-edit the same software source code), but as TensorFlow have more than 10 years of existence and attracted contributions from thousands of developers and hundreds of companies, the network is very dense to make sense of who works with the non-commercial organization in TensorFlow (see ??). We therefore transformed the inter-individual network where nodes are affiliated with an organization, into an inter-organizational network where nodes are the nodes and edges are the number of developers that participated in inter-organizational collaborative relationships - collaborative relationships traced from software source code repository. In other words, we transformed the inter-individual network (where nodes represent individuals affiliated with organizations) into an inter-organizational network (where nodes represent organizations), and for each pair of organizations, we counted the number of edges between individuals of the two organizations to assign the edge attribute to each edge inter-organizational network. That way, we lost information about who works with who at the individual level, but then we could easily visualize who works with who at the organizational level making the analysis much easier and presentable.

Our results are organized in different subsections aggregating information on (1) *what* are the top commercial and non-commercial contributors to TensorFlow?, (2) *how* they collaborate with each other and *who* works with whom?, and (3) *why* do they collaborate with each other in such an open and cooperative way?. We present ten inter-organizational networks, each focusing on the roles of the TOP 10 non-commercial contributors to TensorFlow (see ??) where two revealed to be open-source software communities legally represented as non-profit foundations, six universities and two research institutes from USA (n=5), Germany, Russia, China, Sweden and South Korea. The results are presented from the most active non-commercial contributor to TensorFlow (i.e., Seoul National University - SNU) to the 10th most active non-commercial contributor to TensorFlow (i.e., University of California, **Berkeley**). Given the selected theoretical background and the research question, the results are concerned with the role played by non-commercial organizations in open and cooperative platform ecosystems. The results aggregate quantitative (e.g., number of nodes), qualitative (e.g., web pages and semi-structured interview quotes) and relational evidence (e.g., networks visualizations), that contribute to further theorising the open-cooperation phenomena that according to strategic management scholars is increasingly popular (?). Furthermore, we also addressed a call by ? to integrate the open-cooperation phenomena with the triplex-helix innovation model.

<sup>4</sup> See <https://openai.com/index/whisper/>



## 4.1 | The *What* and *Who* in the co-production of TensorFlow

As a by-product of following the previously mentioned methodological approach by ? and the ScrapLogGit2Net scientific software tool that attempts to assign an organizational affiliation to each developer by using either via their e-mail or via the **api!** provided by GitHub<sup>5</sup>, we could get software-generated reports ranking the biggest contributors to the project. It was therefore easy to identify what are the commercial and non-commercial organizations that commit most code to the TensorFlow platform (see ????). By the time we finalized those rankings, we had already learned much about TensorFlow and its surrounding empirical background by reading plenty of online documentation. By then, it was not a surprise that Google and the chipset-makers (i.e., Intel, Nvidia, ARM, IBM and AMD) were the biggest contributors and more central nodes in the overall collaborative network. We were surprised however to see that Korean and Chinese organizations figured out among the top contributors even if very far from the Californian labs where Google founded the TensorFlow project. Having figured out who are the biggest commercial organizations and who are the biggest non-commercial organizations, it was time to figure out how they work with each other by analysing the social structure of their inter-individual and inter-organizational collaborative network.

**TABLE 2** TOP 10 organizational contributors that cooperate with others in the co-production of TensorFlow.

Order <sup>†</sup>	Organization	(n) contributors	Key markets
1	Google	979	cloud services, software tools, chipsets, training, and consulting
2	Intel	93	chipsets
3	Nvidia	61	chipsets
4	ARM	48	chipsets
5	IBM	41	hardware, cloud services, software tools, chipsets, training and consulting
6	AMD	27	chipsets
7	Microsoft	16	hardware, cloud services, software tools, operating systems, training and consulting
7	Huawei	12	hardware, cloud services
9	Amazon	11	cloud services, e-commerce, software tools, training and consulting
10	Naver	11	cloud services, e-commerce

<sup>†</sup> Order by number of contributors affiliated with the organization. If the number of contributors is the same, order is defined by the sum of the lines or code (LoC) submitted by each contributor.

## 4.2 | The *How* and *who* works with *whom*?

### 4.2.1 | Including them all - a very dense collaborative inter-individual network

After identifying and filtering for 'bots' that commit code to the project<sup>6</sup>, ?? captures the network of collaboration among developers in TensorFlow from Nov 2015 to April 2024. As it covers more than 10 years of digital trace data on the co-production of TensorFlow code (, the network is very dense. We could identify 4220 nodes/developers and 37831 edges/relationships. We note that many contributions were associated with email accounts for personal use (e.g., gmail, outlook, hotmail, ee and others) and therefore could not be associated with a specific organization. Google is visible in a central position, followed by chip-makers such as ARM, Intel, Nvidia, and AMD in the leaderboard of the collaborative software co-production efforts. As in ?? the inter-individual network have a legend for the TOP 10 organizations with most contributors. Besides providing a visual clue of a giant core component comprised of developers affiliated with Google, the figure is too complex to make sense of who works with the non-commercial organizations that can barely be seen in the very dense visualization. Therefore, the following figures take the TOP 10 non-commercial organizations as focal organizations into inter-organizational networks derived (i.e., transformed) from the same network visualized in ??.

Scrap log output for SNU

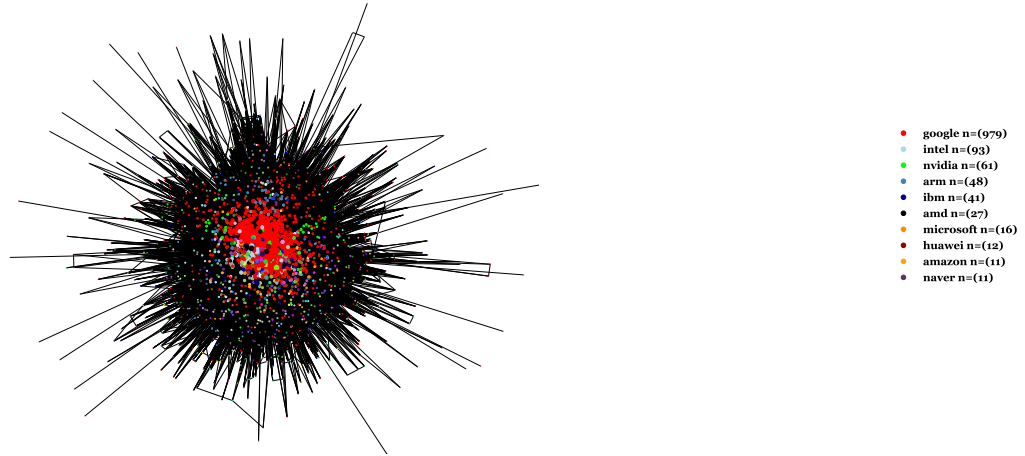
<sup>5</sup> Note that GitHub hosts the TensorFlow source code repository - our main source of digital trace data. See <https://docs.github.com/en/rest?apiVersion=2022-11-28> for the API details.

<sup>6</sup> Mostly the gardener@TensorFlow.org bot that commits code as part of continuous integration and quality assurance non-human automation within Google.

**TABLE 3** TOP 10 non-commercial contributors that cooperate with others in the co-production of TensorFlow.

Order <sup>†</sup>	Overall	Organization	(n) contributors	Type
1	30	Seoul National University - <b>SNU</b> (Korean: 서울대학교)	7	University
2	32	German Research Center for Artificial Intelligence (German: <i>Deutsches Forschungszentrum für Künstliche Intelligenz</i> - <b>DFKI</b> )	6	Research Institute
3	34	<b>Chromium</b> - a free and open-source web browser project	5	Community foundation
4	36	Institute for System Programming - <b>ISP</b> of the Russian Academy of Sciences (Russian: <i>Институт системного программирования</i> )	5	Research Institute
5	41	Peking University - <b>PKU</b> (Simplified Chinese :北京大)	5	University
6	43	Georgia Institute of Technology - <b>GeorgiaTech</b>	5	University
7	45	Massachusetts Institute of Technology - <b>MIT</b>	5	University
8	46	<b>Apache</b> Software Foundation - an American nonprofit corporation to support a number of open-source software projects	5	Community foundation
9	50	Royal Institute of Technology (Swedish: <i>Kungliga Tekniska Högskolan</i> - <b>KTH</b> )	4	University
10	59	University of California, <b>Berkeley</b>	3	University

<sup>†</sup> Order by number of contributors affiliated with the organization. If the number of contributors is the same, order is defined by the sum of the lines or code (LoC) submitted by each contributor.

**FIGURE 1** Sociogram capturing collaboration among developers during Nov 2014 - Apr 2024.

#### 4.2.2 | Pointing our lenses to SNU

Our first figure capturing inter-organizational network cooperation at an aggregated level (see ??), showcases relationships and interactions between SNU and the top ten contributors to the TensorFlow platform for **AI!** and **ML!** (i.e., Google, Intel, Nvidia, , ARM, IBM, AMD, Microsoft, Huawei, Amazon and Naver). Nodes in the network represent organizations, while edges (lines connecting the nodes) indicate the existence and strength of collaborative ties between them. The size of the nodes reflect the centrality and prominence of the organizations and the edges represent the intensity and frequency of collaborative interactions in the co-production of the TensorFlow software source code. Each node's label maps the organization name, and each edges's label represent the number of unique collaborations among software developers affiliated with the organizations the edge connects. In other words, the high the edge's label number, the more developers affiliated with the organizations cooperated

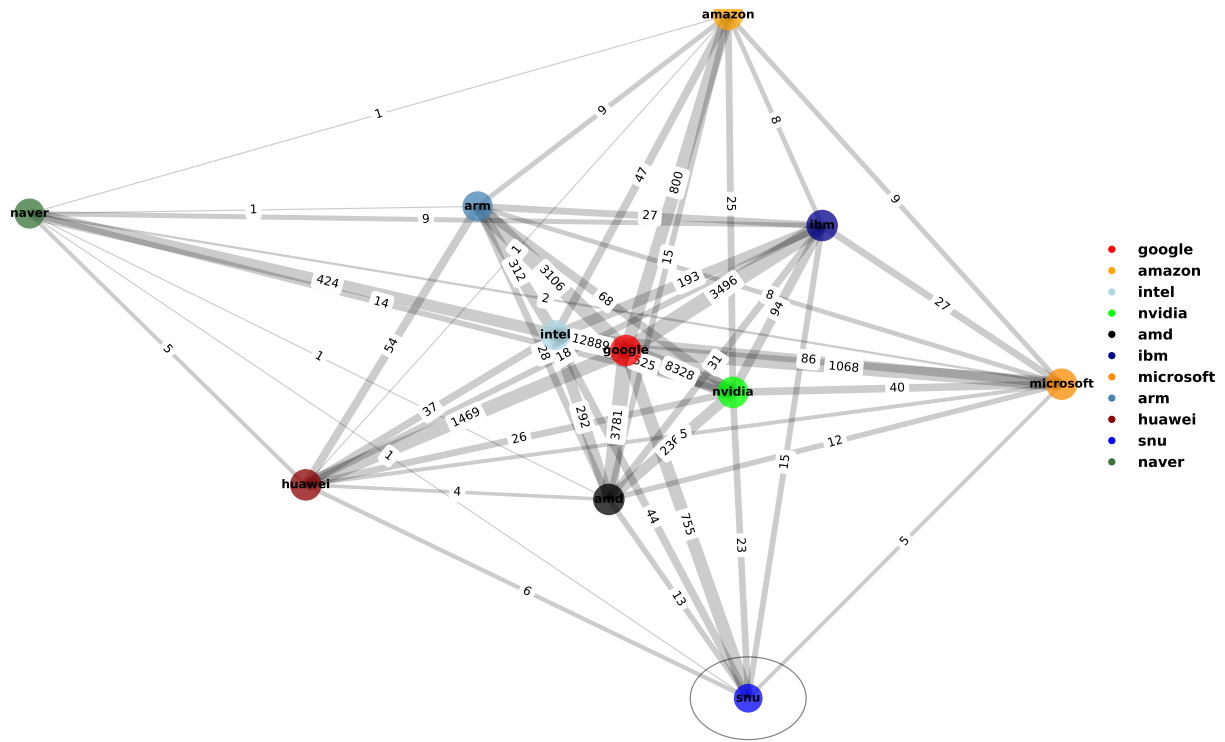


FIGURE 2 Scrap log output for SNU.

with each other. For example, Huawei and SNU had six unique inter-organizational collaborative relationships among software developers, while IBM and SNU had 16 developers<sup>7</sup>

As noted in ??, Google is the most central node as expected by being the founder, and biggest investor on TensorFlow. Following in terms of centrality, we note that chipset makers (i.e., Intel, ARM, Nvidia and Nvidia) and cloud vendors (Amazon, IBM and Microsoft) form central core of the TensorFlow collaborative network. SNU as University contributor remain in the periphery with strong ties with Google, Intel, AMD, IBM and Nvidia and weaker ties with Naver and Microsoft. Somehow surprising (as both Naver and SNU are Korean organizations), both organizations only had on inter-organizational collaborative dyad. The network also provides insights into strategic alliances where companies like AMD and Nvidia, despite being competitors, engage recurrent collaborative efforts. The fact that they sell competing cloud services or products or competing chipset do not reflect on the collaborative network. The collaborative network do not shows evidence that competition in the chipset and cloud computing market influences cooperation in the TensorFlow ecosystem.

On the one direction, and besides being in the network periphery (see ??), SNU plays a very important bridge role in the network. SNU acts as bridge to (1) researchers with high high-level expertise in developing and deploying **ML!** learning models, (2) as a bridge to talent that is recruited by the organizations that develop and deploy TensorFlow, and (3) as a bridge between theoretical research and practical real-world applications.

On the other hand, by contributing to TensorFlow, SNU researchers can stay at the forefront of AI and **ML!** research and access to the world-class cloud infrastructure maintained by Google. Furthermore, by contributing to TensorFlow, SNU can incorporate the latest **AI!** and **ML!** technologies into its curriculum, providing students with hands-on experience and preparing them for careers in **AI!** and **ML!**. As many researchers seek societal impact, by contributing to TensorFlow, researchers can see their model, their code and their algorithms applied to a vast cross-industrial domain that ranges from healthcare, telecommunications, automotive, social networks and robotics. By contributing to its development. While advances on TensorFlow immediately impact natural language processing, computer vision, and automated machine learning, those advancements can benefit society at large (better translations, self-driving cars, better hand-writing recognition, better diagnostics, automation of boring or dangerous tasks thus, leading to advancements that benefit society at large. Contributions to a widely-used and respected framework like

<sup>7</sup> Note that as SNU has 7 individual contributors and Huawei has 12 individual contributors, there is a maximum of  $7 \times 12 = 84$  possible inter-organizational relationship combinations involving both SNU and Huawei developers. However, as Google had 979 nodes, there is a maximum of 6853 possible combinations involving both SNU and Google developers.

TensorFlow and a strong partnership with Google and other big players in the ecosystem enhance SNU's reputation in the global research community, attracting funding, partnerships, and top-tier faculty and students. In summary, SNU's contributions to TensorFlow stem from a desire to advance research and access the required infrastructure, foster collaboration, enhance educational programs, create societal impact, and bolster institutional recognition. These efforts align with the university's goals of driving innovation and maintaining a leadership position in the fields of **AI!** and **ML!**.

### 4.3 | Pointing our lenses to DFKI

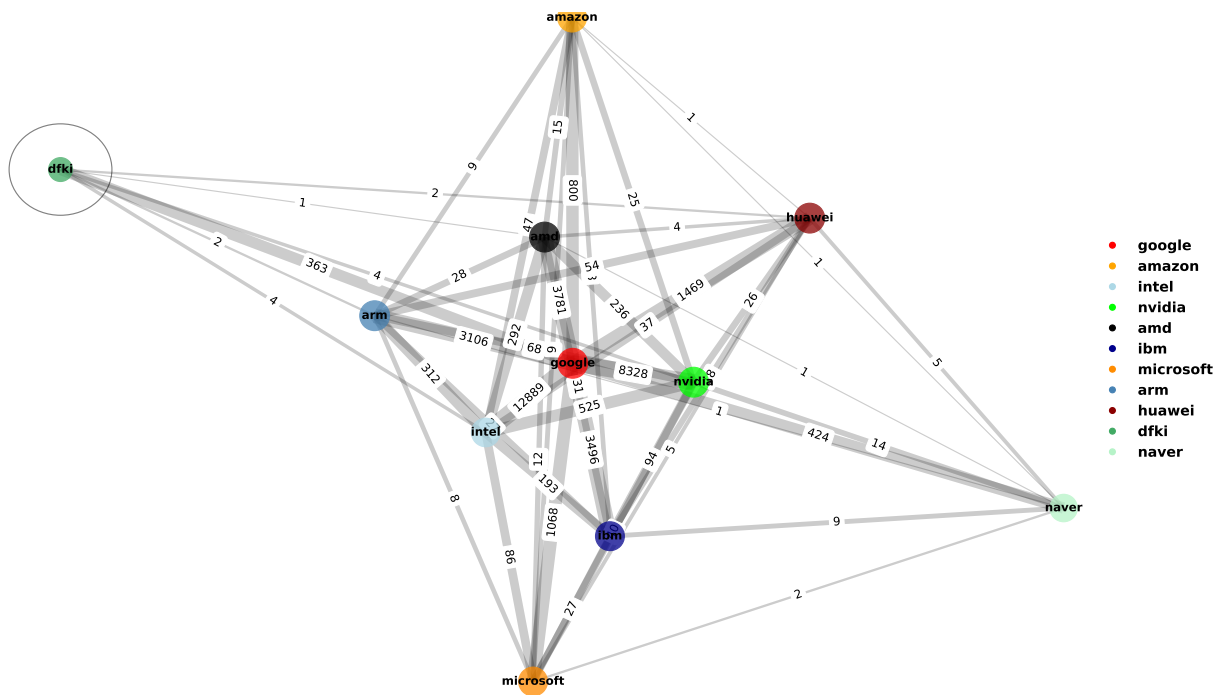


FIGURE 3 Scraplog output for dfki.

Our second figure capturing inter-organizational network cooperation at an aggregated level (see ??), showcases relationships and interactions between DFKI and the top ten contributors to the TensorFlow platform in the same fashion as in ??. Nodes represent organizations and edges represent collaboration as in the same way. Each node's is labeled with its organization name, and each edge is labeled with the number of unique collaborations among software developers affiliated with the organizations that the edge connects. In other words, the high the edge's label number, the more developers affiliated with the two nodes cooperated with each other.

DFKI as research institute contributor remain in the periphery with a unique strong tie with Google and weaker ties with Huawei, Nvidia, AMD, Intel and ARM. Somehow surprising (DFKI as a German research institute does not collaborate more with organizations that have large offices in Germany). The contributions of DFKI seem to be coupled with chipset design given that DFKI works with most of the chipset vendors but not with the most well know provider of cloud services (e.g., Amazon, IBM, and Microsoft).

On the one direction, and besides being in the network periphery (see ??), DFKI also plays a very important bridge role in the network. DFKI also acts as bridge to (1) researchers with high high-level expertise in developing and deploying machine learning models, (2) as a bridge to talent that is recruited by the organizations that develop and deploy TensorFlow, and (3) as a bridge between theoretical research and practical real-world applications. In the case of DFKI, our qualitative part of the investigation (i.e., interview with developers) highlight that DFKI cross-fertilized theoretical and practical competences in sector that are particularly important for the German economy (e.g., automotive, healthcare and manufacturing of precision machinery). This might explain why DFKI worker closer to chipset makers than to cloud vendors.

DFKI, along with its partners, secured funding from the European Union's Horizon 2020 program for projects that involve the development and application of AI technologies using TensorFlow. Funding from German national research agencies, such as the Federal Ministry of Education and Research (BMBF), has been utilized to support collaborative projects between DFKI and industry partners like Google. These funds supported research activities, including the hiring of researchers, purchasing of equipment, and running computational experiments. And were often done in collaboration with industrial partners that early jumped into testing and adopting **AI** and **ML** into their operations (e.g., OEM car-makers, hospital diagnostics centres). Without Google, it would be hard for DFKI and its partners to have ready to deploy high-performance computing (HPC) facilities for running large-scale machine learning experiments and training complex models using TensorFlow.

## 4.4 | Pointing our lenses to Chromium

The third non-commercial organization with the most contributors to TensorFlow is not a University nor a Research Institute. To our surprise, it was instead the Chromium open-source community which is an open-source project primarily maintained and managed by Google. The project started using the WebKit rendering engine<sup>8</sup>. When Google first launched Chromium and its commercial counterpart, Google Chrome, in 2008, they used WebKit as the browser's rendering engine. WebKit, initially developed by Apple for its Safari browser, was chosen due to its performance and compatibility.

However, in 2013, Google decided to create a separate fork of WebKit called Blink. This decision was driven by the desire to have more control over the development process and to innovate more rapidly without being tied to WebKit's development constraints and priorities. Blink allowed Google to make significant changes and optimizations specific to the Chromium project, leading to enhancements in performance, security, and compatibility.

The transition from WebKit to Blink marked a significant evolution in the Chromium project, allowing Google and other contributors to push the boundaries of what web browsers could achieve. Today, Blink is the rendering engine behind not only Google Chrome but also other Chromium-based browsers like Microsoft Edge, Opera, and Brave. Chromium serves as the foundational codebase for Google Chrome and several other web browsers, including Microsoft Edge, Brave, Opera, and Vivaldi. As an open-source project, Chromium invites contributions from developers worldwide, allowing for a diverse range of inputs and innovations. Google directs the overall development and ensures that contributions align with the project's goals and standards, but the open-source nature of Chromium means that it benefits from a wide array of perspectives and expertise, driving continuous improvement and adaptability.

The Chromium project operates under the permissive BSD license, which allows anyone to use, modify, and distribute the code. This licensing structure has enabled various companies and developers to build upon the Chromium codebase, integrating unique features and optimizations to cater to different user needs. For instance, Google Chrome includes proprietary enhancements and services on top of Chromium, while Microsoft Edge incorporates its features and integrations. The Chromium open-source community plays a crucial role in identifying bugs, proposing enhancements, and ensuring the project's security and stability.

It is important to notice that even if Google has significant control over the Chromium project as its primary maintainer, it cannot unilaterally close Chromium in the traditional sense of making it proprietary and removing access to the existing open-source code. Once code is released under the BSD license, those rights cannot be revoked for the code that has already been released. Furthermore, Chromium integrates large components of open-source code released under permissive copyleft licenses that would need to be rewritten at very high costs if Google chose to move the project from the open-source to the proprietary arena. Finally, if Google were to cease its open-source development of Chromium or attempt to close it, the community or other organizations could fork the project (see ??, for academic discussions on the implications of the forking mechanism). This means they could take the existing open-source code and continue development independently of Google under a different name. Closing Chromium could alienate the developer community and other stakeholders who contribute to the project, ultimately harming Google's interests.

By interviewing a developer that did not wish to reveal its identity, Google employs developers with @chromium.org email addresses to distinguish their contributions to the Chromium project from other Google-related activities. This practice serves several important purposes such as: (1) ensuring that contributions are correctly attributed to the open-source project rather than other Google initiatives. Clear Project Identification, (2) maintaining transparency within the open-source community, By having a distinct email domain, it is clear which contributions are made on behalf of Chromium even if by developers paid by Google, (3) create a sense of community and shared purpose among contributors, whether they are from Google or other organizations.

<sup>8</sup> See the work of ? on WebKit that coined the term open-coopetition.

Google developers may work on various projects simultaneously, however having a separate email domain for Chromium helps to differentiate communications and commitments specific to the Chromium project are managed distinctly from other Google projects.

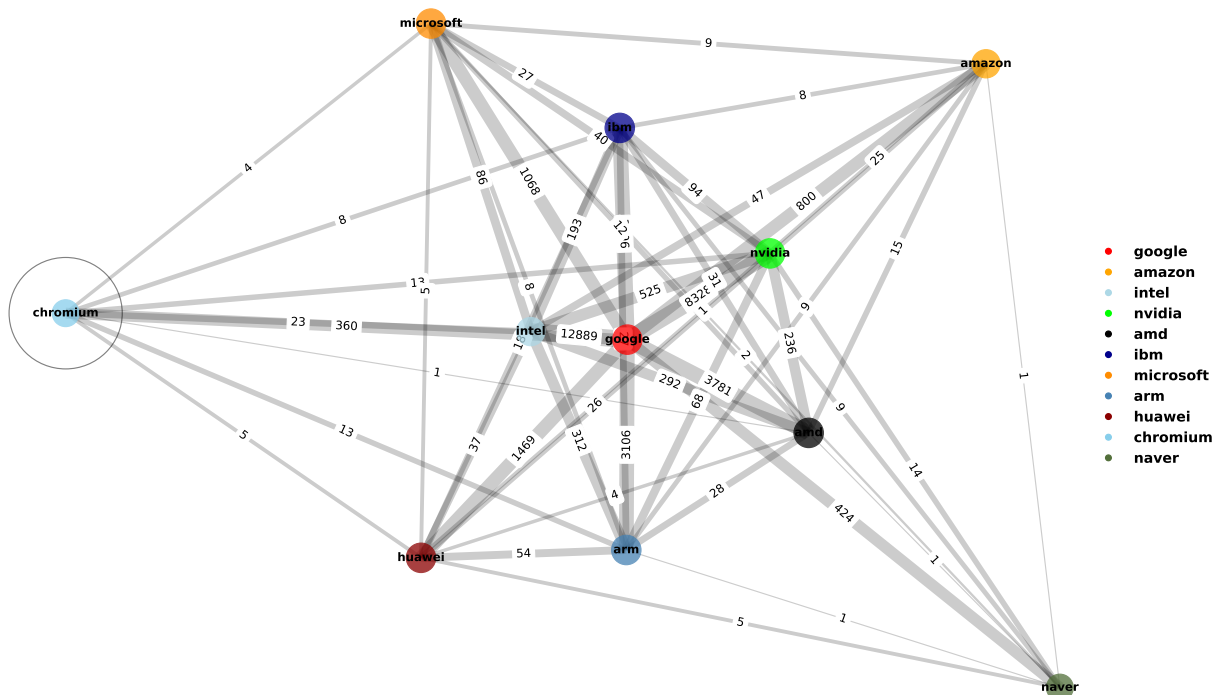


FIGURE 4 Scraplog output for chromium

The Chromium project, although primarily maintained by Google, aims to present itself as a neutral platform for browser development. This neutrality allows it to attract contributions from other major tech companies like Microsoft (for Edge), Opera, and Brave. By fostering an inclusive and collaborative environment, Chromium can benefit from a wide range of expertise and innovations, driving the project forward and ensuring its continued relevance and success. In line with other open-source communities like the ones represented by Linux Foundation, the Apache Foundation, and the Eclipse Foundation (see ?, for work on conflict of interest in open-source foundations), the Chromium project needs to be a neutral space where competitors can collaborate. Therefore organizations like the Chromium project play a very important dual role (1) maintain a software code base in an open-source way that for several reasons can not be turned into commercial software, and (2) be a neutral place that ensures openness and transparency in the orchestration of collective goals as evidenced by the following quote that pop-up during one semi-structured interview with a TensorFlow developer recruited via e-mail and LinkedIn.

“A neutral governance model ensures that everyone can contribute and benefit equally, which is essential for the health and growth of the project” ... “there needs to be a neutral space for competitors to push code” – Software developer and regular contributors to the TensorFlow during a semi-structured interview with the first author. Pseudonym Albert.

While SNU and DFKI roles on the network were more about bridging competence, the role of Chromium was more about bridging with an established open-source code base and bridging with a neutral space where organizations with competing interests work together.

## 4.5 | Pointing our lenses to ISP

Also to our surprise, the fourth non-commercial organization that most contributes to TensorFlow code was the Institute for System Programming of the Russian Academy of Sciences (ISP). Also very far away from California, ISP contributed to

TensorFlow, particularly in areas related to software verification and formal methods. Their contributions have primarily focused on enhancing the reliability and correctness of TensorFlow's codebase.

The ISP is a research institution focused on advanced software development and system programming. Founded in 1994, the ISP RAS engages in extensive research and development in areas such as formal methods, system programming, software verification, and programming languages. The institute collaborates with both domestic and international organizations, contributing significantly to the global software development community. The research institute has also a portfolio of open-source projects in the areas of software verification, physical dynamics simulators, functional verification of microprocessors and natural language processing. Their contributions to TensorFlow are another showcase of their advanced competencies on the international stage.

Our ?? captures inter-organizational network cooperation between ISP and the top ten organizational commercial companies that co-produce TensorFlow show that ISP as research institute organizational contributor remains in the periphery with strong ties with Google followed by ARM and Intel and weaker ties with Nvidia, Huawei, IBM and Naver. The contributions of ISP seem to be coupled with the chipset design makers. In the same way as DFKI, ISP works the most with Google and the chipset vendors but not so much with the most well-known providers of cloud services (e.g., Amazon, IBM, and Microsoft).

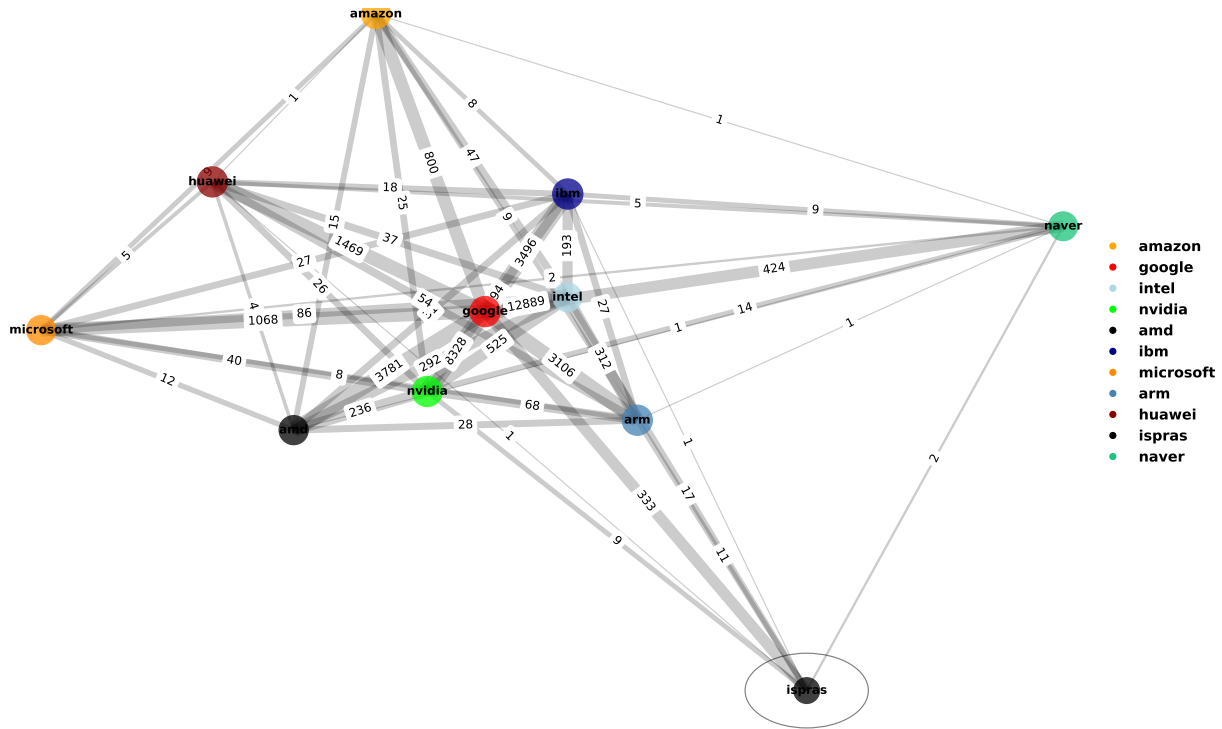


FIGURE 5 Scraplog output for ispras

From the one side, Google and its TensorFlow partners benefited from the ISP static analysis and verification tools for TensorFlow that were designed to identify and correct bugs, improve code safety, and ensure compliance with coding standards in the project. Furthermore, the ISP also provided tools and services for the formal verification of TensorFlow's algorithms. A difficult task in computer science that involves creating mathematical models of the algorithms and proving their correctness and robustness.

On the other side, the collaboration provided ISP RAS with access to Google's computational resources and development environments. This access enabled them to perform extensive testing and verification processes, which would be challenging to achieve independently. The joint efforts likely resulted in academic publications, conference presentations, and technical reports that contribute to the broader research community. These outputs disseminate knowledge, promote best practices in software verification, and showcase the ISP tools and methodologies in the growing AI industry.



## 4.6 | Pointing our lenses to PKU

The fifth non-commercial organization that most contributes to the TensorFlow code was Peking University (PKU). Like SNU, PKU is a University located quite far from California where the TensorFlow project was born. Both PKU and SNU are (1) top academic institutions in their own countries, (2) rank high in the international ranking organizations, (3) host many research institutes and centres, and (4) develop international partnerships with the industry on the global stage. In this sense, in terms of industry-university relations, we should not only see those universities as Korean or Chinese, but as “world” universities. The same applies to other universities and research institutes that contribute to TensorFlow listed in ??.

The PKU hosts the Institute for Artificial Intelligence that was established in April 2019 and hosts itself dozens AI labs with different expertise (e.g., healthcare, cognitive reasoning, and image recognition among others). PKU established multiple formal partnerships with Huawei and many Europeans and North American organizations when it comes to **AI!** and **ML!** (e.g., UCLA-PKU Joint Research Institute in Science and Engineering (JRI) or the ). By retrieving online documentation and speaking to TensorFlow developers we were surprised by the fact that Huawei, in collaboration with PKU, has developed its own **GPUs!** (**GPUs!**), called the Ascend series that is very competitive for **AI!** and **ML!** workloads making the Chinese AI industry less dependent on USA chipsets<sup>9</sup>. While Chinese tech giants like Alibaba, Tencent, and Baidu are reportedly turning to Huawei for **GPUs!**, Huawei has also been developing its **AI!** accelerator chips, which are similar to Google’s **TPUs!** (**TPUs!**).

In ??, we capture inter-organizational network cooperation between PKU and the top ten organizational commercial companies that co-produce TensorFlow show that ISP as a research institute organizational contributor remains in the periphery with a very strong tie with Google. Then PKU has weak ties with Huawei, Nvidia, and Naver. The multiple formal partnerships between Huawei and PKU are not be strongly visible in the TensorFlow code repository, as only two dyads of developers established an inter-organisational collaboration relationship between PKU and Huawei (see left side of ??). This suggests that the cooperation between PKU and Huawei in terms of advancing **AI!**, **ML!** happens somewhere else than in the TensorFlow core.

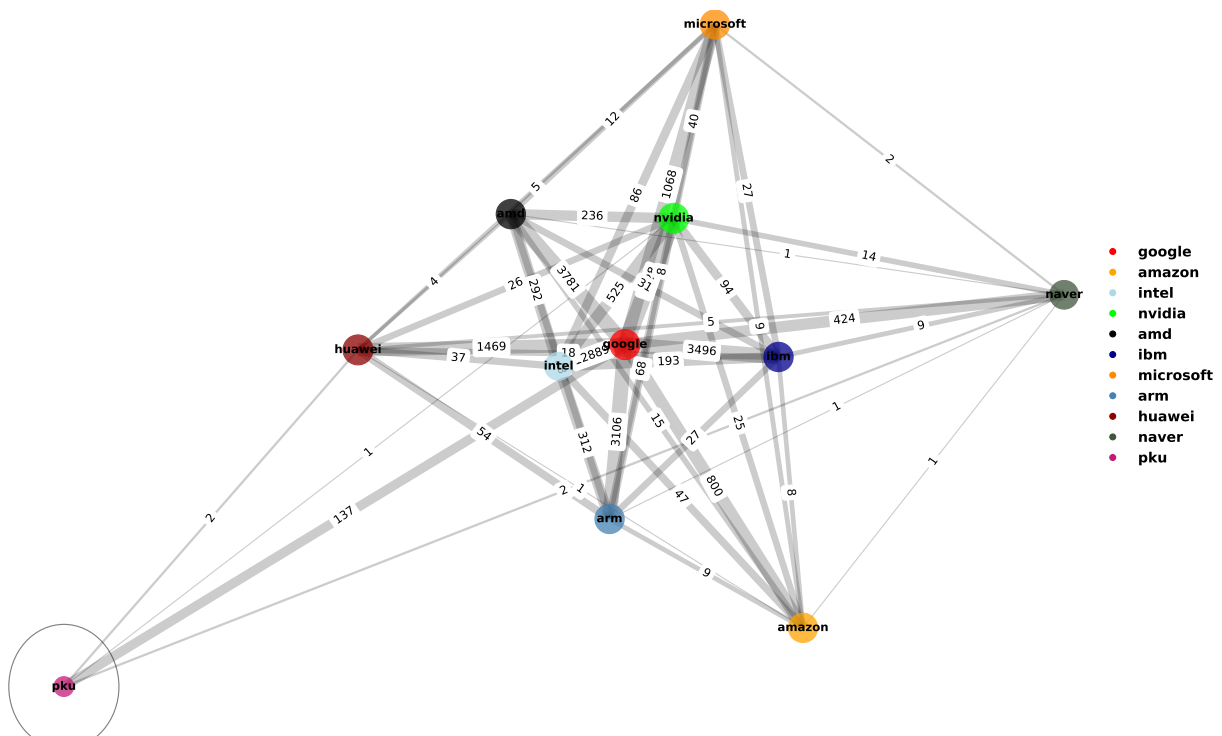


FIGURE 6 Scraplog output for pku

<sup>9</sup> See <https://www.tomshardware.com/news/huaweis-gpu-reportedly-matches-nvidias-a100-report>.



Significant contributions from PKU include research on Semantic Image Segmentation which involves detailed work on improving the accuracy and efficiency of semantic segmentation models. This project has seen contributions from both Google and academic researchers, including those from PKU, focusing on enhancing convolutional neural networks (CNNs) and developing new methods for object localization in images. Another example is the “Show and Tell” image captioning system, where researchers from both organizations worked on improving the computer vision components of TensorFlow to generate more detailed and accurate image descriptions. This collaboration not only resulted in advancements in image captioning techniques but also provided extensive open-source in the form of code to TensorFlow. PKU seems to act as a bridge to competence in computer image recognition that in the last years been improving a lot with the advance of **AI!** and **ML!**.

From the perspective of PKU, our analysis points out four motivations for its engagement in the co-production of TensorFlow. First, in terms of research and innovation, PKU’s participation in TensorFlow allows its researchers to pursue cutting-edge AI and machine learning research, enhancing the university’s academic and scientific capabilities. Collaborating with industry leaders like Google augments PKU’s research output and contributes to the advancement of state-of-the-art technologies. Second, and in terms of access to resources and expertise, collaborating on TensorFlow provides PKU with access to Google’s extensive resources, including computational infrastructure and expert knowledge in AI and machine learning. This partnership enables PKU researchers to work on large-scale projects and leverage Google’s experience to enhance their own research and development efforts. Thirds, and in terms of educational impact, PKU’s contributions to TensorFlow have significant educational benefits as PKU created educational resources, such as TensorFlow tutorials and courses, which help train the next generation of **AI!** and **ML!** researchers and practitioners. Finally, in terms of reputation, by participating in high-profile open-source projects like TensorFlow, PKU strengthens its global reputation and fosters international collaborations. Overall, PKU’s involvement in the co-production of TensorFlow is driven by its desire to advance research, gain access to valuable resources that would be harder to fulfil by itself, enhance educational offerings, and solidify its position as a leader in the **AI!** and **ML!** landscape.

## 4.7 | Pointing our lenses to GeorgiaTech

The sixth non-commercial organization that most contributes to the TensorFlow code is the North American Georgia Institute of Technology which is commonly known as GeorgiaTech. As visible ??, GeorgiaTech similar to all other non-commercial players listed ?? has a strong tie with Google. Besides the strong tie with Google, GeorgiaTech also has weaker ties with Nvidia (USA-based) and Huawei (China-based). For the later inter-organizational relationship, we could only find one pair of developers collaborating across the organizations.

Georgia Tech has been actively involved in projects that leverage the TensorFlow platform for **AI!** and **ML!** research. These projects often involve interdisciplinary collaborations, such as the work conducted by the Interactive Media Technology Center, which explores the usability and applications of wearable computing interfaces in assistive technologies. Georgia Tech’s extensive research infrastructure and computing resources have been instrumental in supporting TensorFlow-related projects. This includes utilizing high-performance computing clusters and specialized facilities like the GT-Bionics Lab, which focuses on developing assistive technologies to empower individuals with severe disabilities. In some cases, Georgia Tech has established industrial partnerships in the domains of AI and machine learning, which have helped secure research and development funding and bridge the gap between academic research and practical, real-world applications.

While conducting this research we learned that in the world of **AI!** deep learning, the infrastructure is not only about powerful computers with **GPUs!** or **TPUs!**. It is also about large amounts of memory and fast storage solutions, such as Solid-State Drives (SSDs) or distributed file systems, are crucial for managing vast datasets and model parameters that often deployed across multiple data center that require load balancing, parallelization efforts to work smoothly and consume fast amounts of electrical power and water for cooling. Many specialized frameworks, libraries and tools are also required for developing and optimizing deep learning models. Vast amounts of vast, clean, relevant, and readily accessible data are also important to deploy deep learning into production. The same applies to models, as the existence of reference pre-validated models is a key to developing newer and better models. So while some universities, research institutes and open-source communities might have access to powerful computing resources, Google can offer much more for deep learning that the hardware you can order from vendors it a very fast way to researchers interested in testing their latest models and algorithms for **AI!** or **ML!**.

As in prior covered universities and research institutes, by contributing to TensorFlow Georgia Tech can (1) advance research and education, (2) continue its strong tradition of supporting open-source initiatives, (3) bridge with the practice, industry and real-world applications, (4) nurture relationships with other leading institutions and industry players and, all while aligning with its strategy of leading in technology development and contributing to societal advancement through technological innovation.

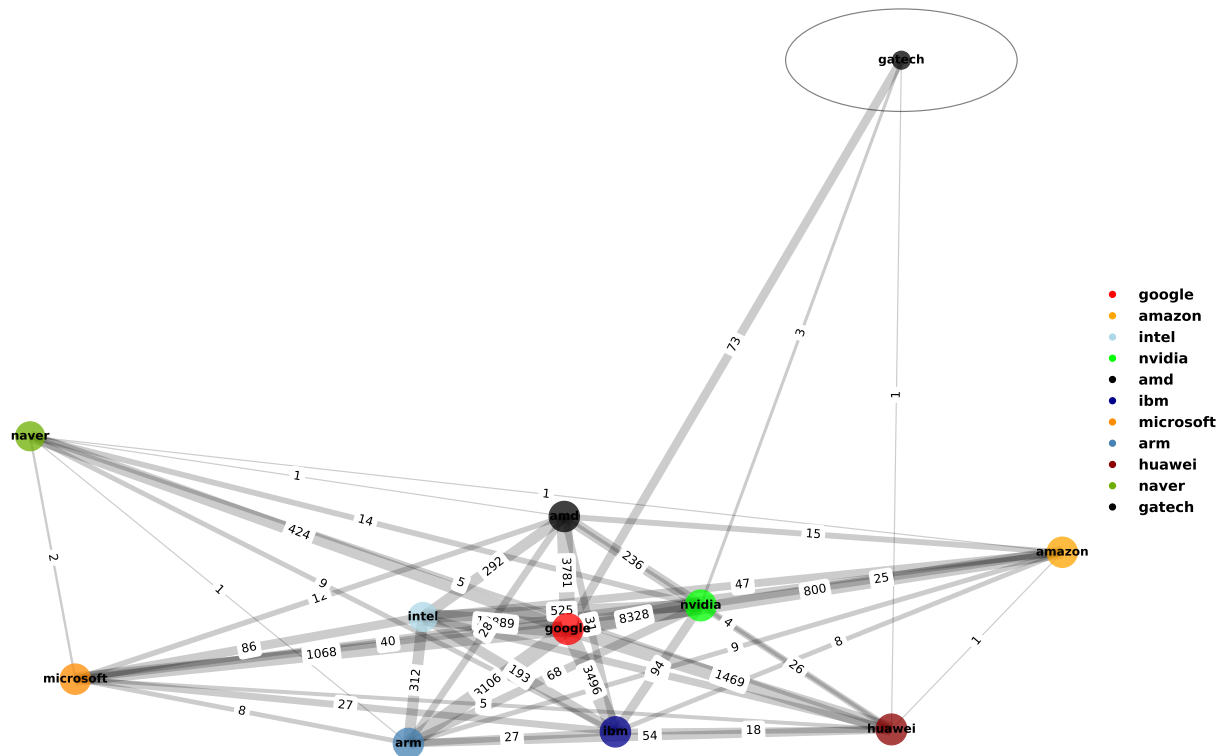


FIGURE 7 Scraplog output for gatech

## 4.8 | Pointing our lenses to MIT

The seventh non-commercial organization that most contributes to the TensorFlow code is the Massachusetts Institute of Technology which is commonly known as MIT. The famous Technological University had six developers coding with the top ten commercial contributors to TensorFlow (see ??) and the network shows five inter-organizational relationships. As visible ??, MIT similar to all other non-commercial players listed ?? has a strong tie with Google. Besides the strong tie with Google, MIT had a less strong tie with Intel. It also had weak ties with AMD, Naver (bases in South-Korea) and Nvidia and IBM. Here we must point out that we only consider developers as affiliated with MIT if the email they used to commit code had the *mit.edu* domain, we did not consider the *alum.mit.edu*. As the later email domain is provided by MIT Alumni Association we did not consider the association strong enough to associate developers with the MIT affiliation for this research purposes. We number of nodes in the collaborative inter-individual network affiliated with MIT would had three more nodes if considering the *alum.mit.edu* domain.

MIT have a strong association with the history of AI for many decades. It brought to the table, expertise in Neural Network Optimization where the work by MIT researchers on optimizing neural networks has influenced TensorFlow's implementation. Furthermore, MIT have been involved in the development of TensorFlow Federated, an extension of TensorFlow that facilitates federated learning. This library supports decentralized learning, where data remains on local devices and only model updates are shared, addressing privacy and data security concerns.

From a educational perspective, MIT has used TensorFlow extensively in its educational materials, such as the MIT OpenCourseWare (OCW) courses on ML and AI. These courses provide practical training on TensorFlow, helping students and practitioners learn how to build and deploy machine learning models. Many of the MIT courses expose students and future AI practitioner to TensorFlow with hands-on experience on building neural networks in TensorFlow<sup>10</sup>. The collaborative tie with IBM mirrors a formal partnership between MIT and IBM in the MIT-IBM Watson AI Lab, where TensorFlow is frequently used in joint research projects<sup>11</sup>.

<sup>10</sup> See <https://ocw.mit.edu/courses/6-s191-introduction-to-deep-learning-january-iap-2020/> for example

<sup>11</sup> See <https://mitibmwatsonailab.mit.edu/>.

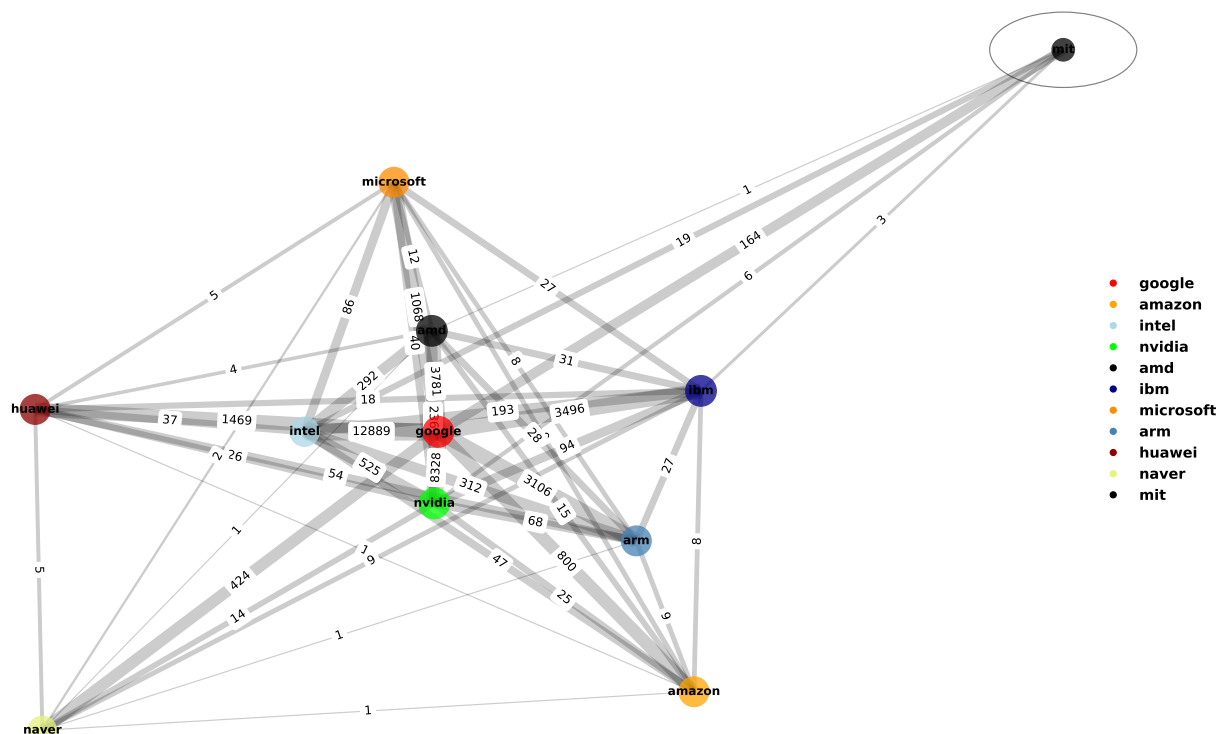


FIGURE 8 Scraplog output for MIT

## 4.9 | Pointing our lenses to Apache

The eighth non-commercial organization with the most contributors to TensorFlow is not a university or a research institute. It was the Apache open-source community primarily maintained and orchestrated by the Apache Software Foundation (ASF). The Apache Software Foundation (ASF) was established in 1999 as a non-profit organization to support the Apache HTTP Server Project and its community of open-source developers. The origins of Apache trace back to 1995, when a group of developers, including Rob McCool and Brian Behlendorf, collaborated to enhance the NCSA HTTPD web server software, leading to the creation of the Apache HTTP Server. This server quickly gained prominence for its reliability, flexibility, and open-source nature. The success of the Apache HTTP Server catalyzed the formation of the ASF to manage and oversee the growth of a wide range of open-source projects beyond the original HTTP server. Today, the ASF hosts numerous high-impact projects such as Apache Hadoop, Apache Spark, and Apache Kafka, continuing to promote open-source software development through its collaborative and community-driven model.

To our surprise, and as evidenced by ??, the Apache organization is connected to all the top 10 organizational contributors to TensorFlow (see ??) with 9 out of 9 possible inter-organizational edges. That connectedness beats even Chromium that was connected to 8 out of 9 possible inter-organizational edges (see ?? to notice that Chromium is not connected with Naver). Here, we must point out that open-source communities like Chromium and Apache establish more edges with the top 10 organisational contributors than universities and research institutes. This highlights their important role in the TensorFlow community as both TensorFlow, Apache and Chromium want to ensure compatibility and interoperability of their code-base.

Apache projects, such as Apache Hadoop and Apache Spark, provide frameworks for distributed data processing and storage, which are essential for handling the large-scale datasets typically used in machine learning tasks. TensorFlow, on the other hand, is a powerful machine-learning framework designed for building and deploying machine-learning models. They are therefore very complementary platforms. The integration of TensorFlow with Apache projects allows for better data processing pipelines. For example, TensorFlowOnSpark is an open-source framework that combines TensorFlow's machine learning capabilities with Apache Spark's efficient data processing and distributed computing power. This integration enables the training of machine learning models on large datasets stored in Hadoop's HDFS or processed using Spark, thus facilitating scalable and efficient machine learning workflows. Developers typically rely on formats, tools and libraries provided by the Apache foundation for processing data before loading them into machine learning workflows. Furthermore, Apache Beam, another project under the

Apache Software Foundation, offers a unified programming model to define and execute data processing pipelines, which can be integrated with TensorFlow Extended (TFX) for building production-scale machine learning pipelines. This collaboration enhances the ability to preprocess data, train models, and deploy them in a scalable and efficient manner, leveraging Apache's strengths in big data processing and TensorFlow's advanced machine learning algorithms.

On the one side, it is in the interest of the Apache Foundation to ensure that their solutions (e.g., Apache Hadoop and Apache Spark) integrate very well with TensorFlow. If Apache projects seamlessly integrate with TensorFlow they become more attractive and useful, they grow with each other. On the other hand, improving the interoperability between TensorFlow and Apache projects can result in more efficient data processing and model training workflows, which benefits both platforms and their users. When starting and evolving TensorFlow, Google and its partner did not need to reinvent the wheel. They could for example benefit from a number of ready-available open-source projects such as: (1) the Hadoop's Hadoop Distributed File System (HDFS) offers a highly scalable and fault-tolerant storage solution that can handle vast amounts of data across many distributed nodes making it ideal for storing the large datasets required for training deep learning models in TensorFlow, (2) the Hadoop's MapReduce and YARN (Yet Another Resource Negotiator) frameworks enable distributed data processing across a cluster of computers, (3) the Apache Spark analytics engine that offers in-memory data processing capabilities that can speed up data preprocessing and transformation tasks before feeding data into TensorFlow models, (4) the Apache Beam provides a unified programming model for batch and stream data processing, which can be integrated with TensorFlow to build end-to-end machine learning pipelines that handle data ingestion, preprocessing, model training, and serving, and (5) the Apache Kafka is used for real-time data streaming and ingestion. It would cost immense time and money for Google to re-write this well-established community-driven projects orchestrated by the Apache Foundation.

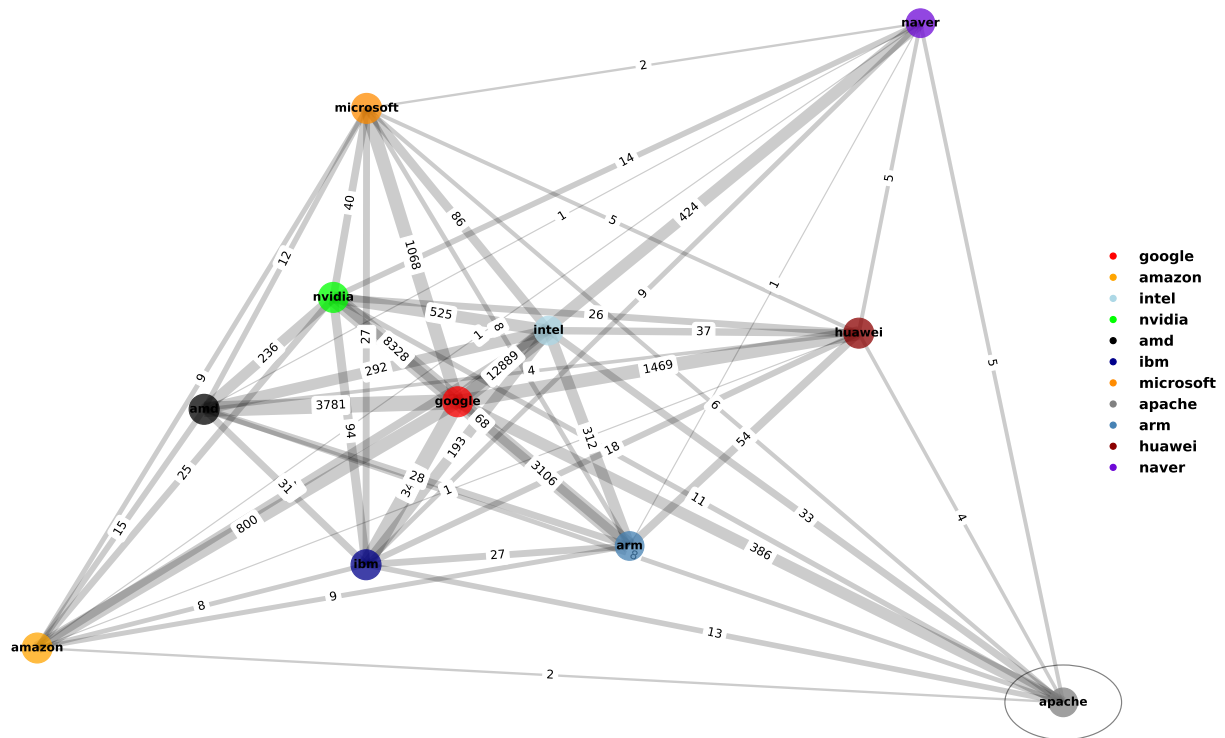


FIGURE 9 Scraplog output for Apache

As evidenced by the following quote that also emerged during a second semi-structured interview with a TensorFlow developer recruited via e-mail and LinkedIn, TensorFlow depends on multiple open-source software projects.

“We would not exist without many other open-source projects” ... “for example Python, NumPy, the Linux Kernel and the GNU tools for processing text.” Also Bazel, nGraph ’ ... “we also integrated Keras’ that started as an independent open-source

project’. – Software developer and regular contributors to the TensorFlow during a semi-structured interview with the first author. Pseudonym Berta.

Similarly to the Chromium project, but at a larger scale, the Apache Foundation also play a very important dual role (1) maintaining the software code base of multiple open-source projects that for several reasons can not be turned into commercial software, and (2) be a neutral place that ensures openness and transparency in the orchestration of collective goals of the different project contributors. In the case of Apache, we learned that “Required behaviours”, “Voting” and “Consensus” play a very important role in deciding the direction of the Apache projects under the foundation umbrella.

#### 4.10 | Pointing our lenses to KTH

The ninth non-commercial organization is the Royal Institute of Technology (Swedish: *Kungliga Tekniska Högskolan* commonly known as KTH. Based in Stockholm, Sweden, founded in 1827, is one of Europe’s leading technical and engineering universities. KTH hosts several institutes and labs dedicated to **AI!** and **ML!**. Among them is the KTH AI Research Centre, which focuses on advancing AI technologies and their applications across various domains. Additionally, KTH collaborates with the Swedish AI Society (SAIS) and participates in multiple AI and machine learning projects, both nationally and internationally having success in securing both EU and National R&D funds for AI research. The university’s School of Electrical Engineering and Computer Science is particularly active in AI research, with specialized labs such as the Robotics, Perception, and Learning Lab (RPL) and the Division of Speech, Music, and Hearing.

As visible in the computer-generated ?? obtained by mining the TensorFlow repository with Social Network Analysis, KTH has a strong tie with Google and several weaker ties with Naver, Nvidia, IBM (who have offices nearby KTH), AMD and Intel. Some **AI!** and **ML!** startups born at KTH. One example is Logical Clocks, a Swedish startup founded in 2016 by a team of researchers from KTH, which is one of the notable success stories emerging from KTH’s academic environment. The company was established by experts in distributed systems and big data, who were involved in significant research projects at KTH. Logical Clocks is best known for developing Hopsworks, a data platform that features the world’s first “Feature Store” for machine learning, which integrates seamlessly with TensorFlow. The collaboration between Logical Clocks, Google and KTH, which is to a large extent publicly funded, enhances TensorFlow’s functionality, particularly in handling big data and distributed machine learning tasks therefore bringing value from public money spent on research and education in the past back to the society as a whole.

Contributions from KTH to the TensorFlow ecosystem have included enhancements to its distributed training capabilities, improvements to data ingestion and pre-processing pipelines, and the development of new tools and libraries that extend TensorFlow’s functionality. Those are important as the training of complex models does not fit in one big machine. The work needs to be distributed across machines and process the aggregated results back - a very complex task as the allocation of computing resources is often limited, failures might occur across heterogeneous computing nodes and bottlenecks need to be minimized. From KTH side, the motivations to contribute go along the same lines as the Universities we covered before. Continue leading **ML!** research as one of the **ML!** research pioneers, showcasing unique technical competencies in **AI!** and **ML!**, curating industry-academia partnerships that are important for obtaining research grants in the EU and Swedish systems. Furthermore, researchers developing the latest **AI!** and **ML!** concepts, algorithms, models and tools can see their work applied in the real world via the TensorFlow ecosystem and the computing Infrastructure provided by Google and its partners.

#### 4.11 | And finally, pointing our lenses to Berkeley

The tenth non-commercial organization that most contributes to the TensorFlow code is The University of California, Berkeley also commonly known as UC Berkeley or simply Berkeley. The university is associated with many technological advancements that are in the public domain such as (1) the RISC-V chipset design instruction set released under a royalty-free open-source license released by Berkeley in the 2010s, (2) the Berkeley Software Unix Distribution (aka BSD) that played a very important role in powering servers, workstations and networking equipment during the 1980s, (3) the SPICE simulator that became the worldwide standard integrated circuit simulator, or (4) the TCL/TK programming language languages that remain very a popular choice for applications that need portability and graphical user interfaces that work across different operating systems. The

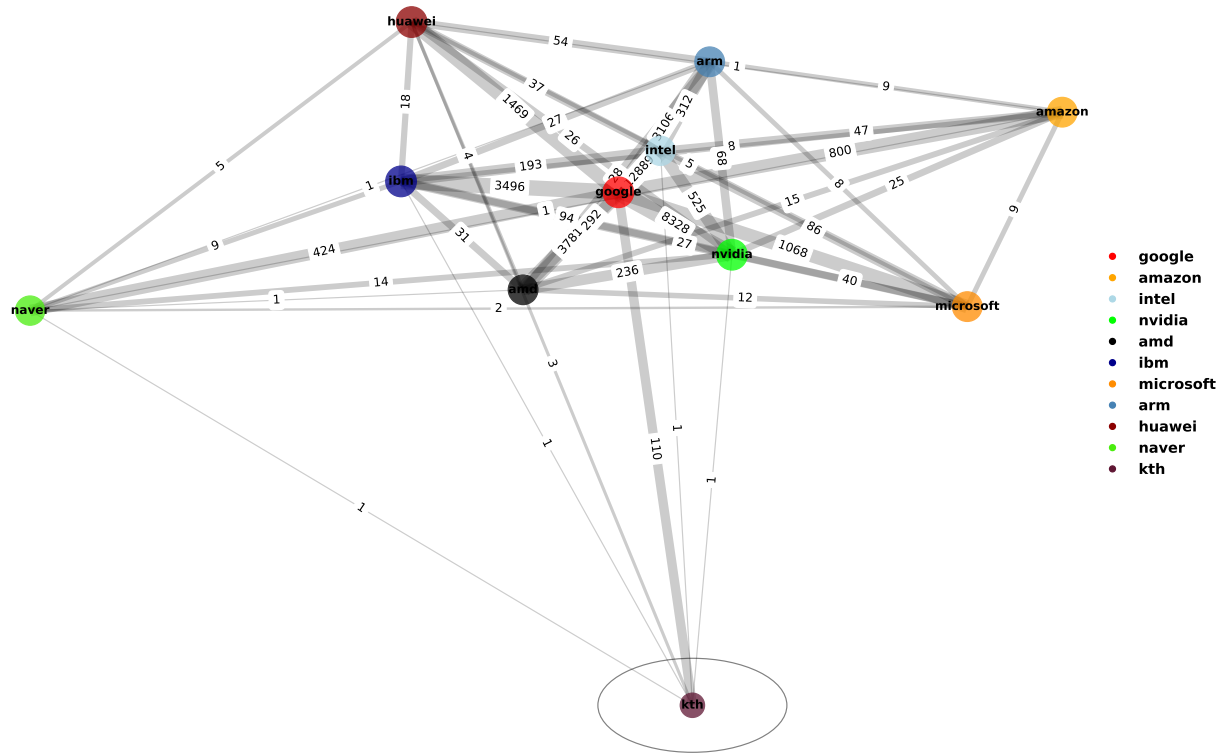


FIGURE 10 Scraplog output for KTH

Berkeley campus regularly hosts several open-source conferences for academics and practitioners (e.g., the Open Source Summit North America and the Scylla Summit) mounting evidence that Berkeley is not disconnected from the open-source world.

As visible in the computer-generated ??, Berkeley has a strong tie with Google and several weaker ties with Naver, AMD, Intel, Nvidia, IBM, and ARM. Here, it is important to note that even though Berkeley is geographically closer to Google, its position on the TensorFlow in the TensorFlow collaborative network is similar to other universities located somewhere else like MIT, SNU and the DFKU. The collaborative network in ?? also evidences that Berkeley is connected collaboratively to the big chipset makers (i.e., AMD, Intel, Nvidia, and ARM), something that is also visible in the RISC-V and SPICE open-source ecosystems. Contrary to many other non-commercial organisations covered by this study, Berkeley does not collaborate with the biggest vendors of cloud computing services (i.e., Microsoft and Amazon).

Berkeley University hosts two labs that play an important role in the TensorFlow ecosystem. The Berkeley Artificial Intelligence Research (BAIR), contributed to TensorFlow by developing new algorithms and techniques for **ai!** (ai!) and **ml!** (ml!). Another lab hosted by Berkeley, the Algorithms, Machines, and People Lab (AMPLab), with expertise in large-scale data processing and distributed computing, has helped improve TensorFlow's capabilities in these areas. Berkeley's expertise in the realm of chipset design and operating systems contributed to Google's and Intel's efforts to improve TensorFlow's performance on Intel processors and accelerators. Even if Nvidia currently designs the most desired chips for **ai!** and **ml!** applications, due to the number of cores available for matrix multiplication within a single chip (see ?), the processors of Intel are still the most deployed one across servers and clouds worldwide. Therefore, it from the interest of both Intel and Google to improve TensorFlow's performance on Intel processors and accelerators. Otherwise, consumers might move to other platforms. Berkeley expertise offered not only expert consulting, but also code contributions from faculty scholars and students who co-developed new features, bug fixes, and performance improvements, which were shared with the broader TensorFlow community.

By engaging in the co-production efforts of TensorFlow, Berkeley advanced research and innovation, enhanced the educational resources, boosted their students and staff with competencies that are applicable in the real-world, cultivated relations with strong industrial leaders, and showcased their abilities to improve the performance and scalability of **ai!** and **ml!** training environments by tuning the software layers of code that work closer to the hardware and interface with the operating systems, CPUs and GPUs. Furthermore, UC Berkeley continued its strong tradition of supporting open-source projects, which aligns with the collaborative

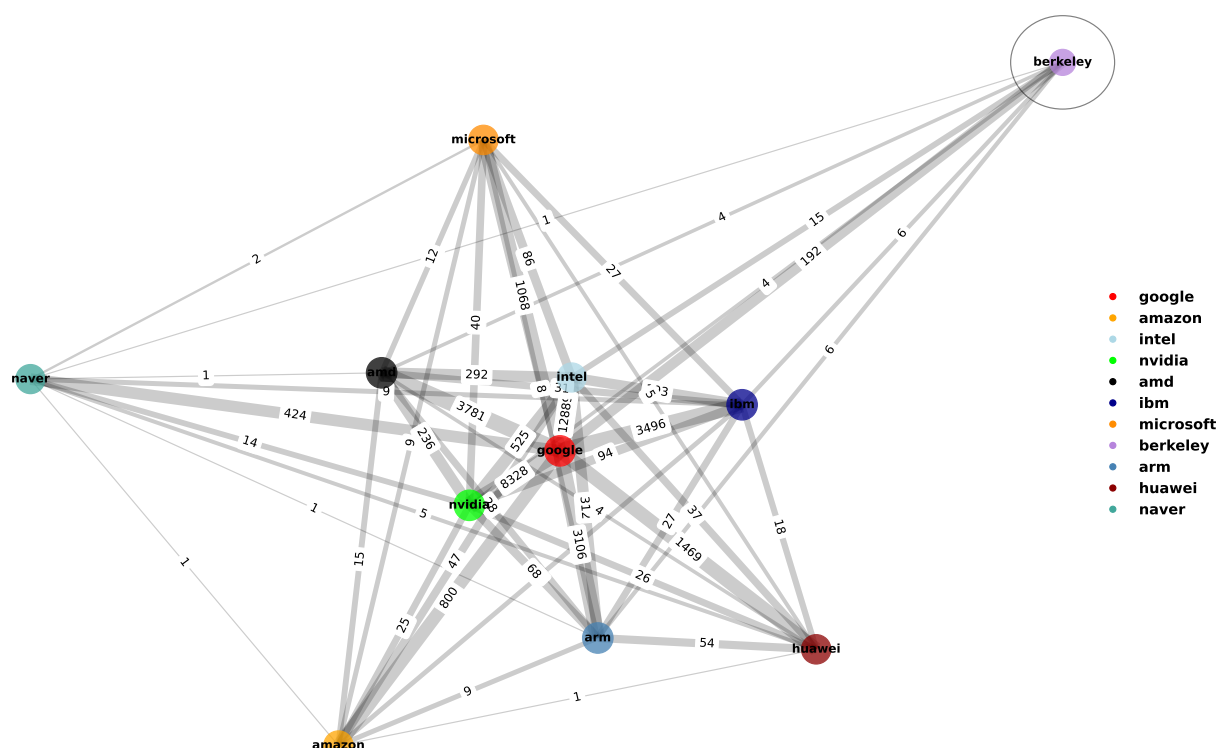


FIGURE 11 Scraplog output for berkeley

and accessible nature of academic research. In our view, this makes sense as much of the excellence comes from projects that are publicly funded by state and federal budgets.

#### 4.12 | The Why of co-producing TensorFlow in an open and cooperative way

After reporting our results centred around the top ten non-commercial organizational contributors to TensorFlow and how they work with the top commercial contributors to TensorFlow, it is now time to report on why co-producing TensorFlow in an open and cooperative way. Or in some other words, **Why not co-producing in the traditional proprietary way that emphasises centralized control and intellectual property rights ?**. First, we take the perspective of the Google high-tech giant that started TensorFlow as an open-source project, and then the perspective of its non-commercial partners in the TensorFlow platform ecosystem.

From the perspective of Google, we could identify four reasons to open up TensorFlow from the collected evidence. First, by doing it in an open-source way, Google could integrate and build on top of established open-source technologies such as the ones provided by the Python, NumPy, GNU, Linux, Apache and Chromium communities. It would be expensive and time-consuming for Google to ignore those open and ready-available technologies under a permissive intellectual property regime<sup>12</sup>. A second reason is scientific prestige, as noted in the following quote by a Google employee, Google researchers had mixed feelings about external entities developing tools or technologies directly derived from Google's own research efforts. On the other way, Google was criticized in the scientific community for publishing new ideas, new models, and new results, but not sharing the blueprints (e.g., the model definition, or the software implementation). By providing the source code of the software related to new **AI!** or **ML!** advances presented at prestigious conferences, Google boosted its reputation and gained trust from leading researchers **AI!**, **ML!**, computer vision and sound recognition from academia and industry experts. They stopped doing tools using Google's latest research, instead, they started implementing their ideas, models and tools on top of TensorFlow. Both sides benefited from the flow of expertise, ideas, models, and tools in the the platform ecosystem.

<sup>12</sup> Note here that while the distribution license terms of those technologies allow Google to integrate them, they force Google to keep TensorFlow open as well, including to its competitors.



“After we published a paper, we could see that new open-source tools would come up in just a few weeks based on our latest research. Then we publish another paper, another model, or another algorithm. And guess what, somebody would come with a new tool or a new module just doing that.” – Machine Learning researcher with expertise in computer vision, Google employee and regular contributor to TensorFlow during a semi-structured interview with the first author. Pseudonym Charles.

Third, and another valid motivation for doing it in an open-source way are the practical benefits of producing software in the public domain. The open-source approach fosters a collaborative environment where diverse developers, researchers, and organizations can contribute to TensorFlow’s evolution. No individual or organization needs formal or approval to contribute to work on their innovation. All without much time negotiating IP licensing agreements, software end-user agreements, implementing gate-keeping and access-control procedures and mechanisms, and knowledge-sharing tensions that characterise the software business as usual often in supplier-client relationships. And all in the public domain over the Internet, where being far from Google does not seem to be a big problem to contribute to TensorFlow. Furthermore, by making TensorFlow freely available with models and curated data with it, Google established it as a standard machine learning platform. Widespread adoption by academic, research, and industry organizations promotes its use and strengthens Google’s position in the AI ecosystem and the AI hype of the 2010s. Also importantly, it removed and reduced barriers for others to play with **AI!** and **ML!** increasing the overall size of the market and creating demand for complementary products and services such as applications with **AI!** features, rentable cloud computing power and TPUs that are specially designed to optimize the deep matrix multiplication operations for training **AI!** and **ML!** models.

From the point of view of non-commercial organizations, our research results data aggregated and evidenced ten motivations for contributing to the TensorFlow. To our surprise, the analysis centered around ten non-commercial organization unveiled to ten motivation to contribute to a open-source and cooperative platform ecosystem such as TensorFlow as we following enumerate:

**Advancing research and innovation** Non-commercial entities, such as universities and research institutes, engage in TensorFlow’s development to stay at the forefront of AI and machine learning research. Participation allows these entities to pursue cutting-edge projects, pushing the boundaries of existing knowledge and technology in AI and ML.

**Showcasing specialized competence** Contributing to a high-profile open-source project like TensorFlow provides an opportunity for non-commercial entities to demonstrate their technical expertise and specialized competencies in AI and ML. This can enhance their reputation and attract further research funding and collaborations. This is also important for open-source communities and their respective foundation so that they can also showcase their technical solutions and competences.

**Accessing world-class infrastructure** Collaborating on TensorFlow provides non-commercial entities access to Google’s extensive computational resources and infrastructure, which might otherwise be unattainable. This access enables them to conduct large-scale experiments and research that require significant computational power.

**Facilitating partnerships** By contributing to TensorFlow, non-commercial entities can “expose” themselves form partnerships with industry leaders, such as Google, and other academic and industrial institutions. These partnerships can lead to collaborative research projects, increased funding opportunities, and shared expertise that can be immediate or in the long run.

**Setting up communities of knowledge** Engaging with TensorFlow allows non-commercial entities to participate in and help build a global community of AI and ML researchers and developers. This community fosters the exchange of ideas, best practices, and innovations, accelerating the overall progress in the field.

**Enhancing educational programs** Contributions to TensorFlow can have significant educational benefits. Non-commercial entities can develop educational resources, such as tutorials and courses, based on their hands-on experience with TensorFlow, which helps train the next generation of AI and ML researchers and practitioners.

**Societal impact** Non-commercial entities are often motivated by the potential societal impact of their contributions. By advancing AI and ML technologies, they can address real-world problems and contribute to societal well-being, such as through advancements in healthcare, environmental monitoring, and more. To our view, there is natural desire of seeing their theoretical advancements in practice in the real world.

**Institutional recognition** Participation in TensorFlow enhances the global reputation of non-commercial entities, positioning them as leaders in AI and ML research. This recognition can attract top talent, students, and additional funding to the institution.

**Influencing the development agenda** By contributing to TensorFlow, non-commercial entities can influence the direction and priorities of AI and ML development. This ensures that their research, business, social or technological interests and goals



are considered in the evolution of the platform. This is especially important for organization that want TensorFlow to evolve according their own interests.

**Ensuring interoperability** Contributions from diverse organizations help ensure that TensorFlow remains interoperable with various systems and technologies. This compatibility encourages widespread adoption and integration of TensorFlow with different tools, libraries, solutions and services. This is especially important for open-source projects and respective foundations that want TensorFlow to integrate well with their code-base.

We finalize our results section with ?? that maps motivation to each non-commercial organization before discussing how our results integrate with extant theory.

**TABLE 4** Motivations for non-commercial entities to contribute to the TensorFlow open-source platform for AI and ML.

	SNU	DFKI	Chromium	ISP	PKU	GTech	MIT	Apache	KTH	Berkeley
Advancing research & innovation	×	×		×	×	×	×		×	×
Showcase specialized competence	×	×	×	×	×	×	×	×	×	×
Accessing unique infrastructure	×	×		×	×			×	×	
Facilitating partnerships	×	×		×	×			×	×	×
Setting up knowledge communities	×	×		×	×	×	×	×	×	×
Enhancing educational programs	×	×		×	×	×	×		×	×
Societal impact	×	×		×	×	×	×		×	×
Institutional recognition	×	×	×	×	×	×	×	×	×	×
Influencing development agenda	×	×	×	×	×	×	×	×	×	×
Ensuring interoperability			×	×			×	×	×	×

## 5 | DISCUSSION

Our research addressed a direct call by ? that stated the importance of knowing “why and for which outcomes companies use more and more open source in communities including their competitors” (?). By addressing the research questions “*Why do tech giants like Google open-source advanced and complex technological platforms that started in-house?*”, “*Why are different organizations cooperating with their competitors in the co-production of those open-source platforms?*” and “*Why non-commercial organizations contribute to the co-production of those open-source platforms?*”, we aim to increase our understanding of the open-coopetition phenomena. As a consequence of our theoretical interests, we also end up narrating the first circa ten years of TensorFlow with complementary visualizations of its social structure. Furthermore, to be able to explain the ‘computerized’ social network visualizations over time we also came across plenty of naturally occurring qualitative material that explained why Google open-sourced TensorFlow. Furthermore, we also came across material explaining the emergence of “communities of competitors” in the co-production of TensorFlow. All within the triple-helix framework.

The first surprising empirical results are captured in ?. Many of the code contributions to TensorFlow were identified with email accounts commonly associated with personal use (i.e., gmail, hotmail, yahoo, outlook and Chinese equivalent QQMail). We noted that many researchers associated with universities, research institutions and startup founders contributed using personal email accounts. In this sense, Google was successful at opening innovation processes via the open-source community. Like in other cases of open-coopetition in the automotive industry (see ?), they open up to get third-party contributions from enthusiasts, students, hackers and academics among others. In the words of Teixeira, “no one should need a permit to innovate on top of their product platforms” (p. 6 ?). In this case, we noted that Google got many contributions from researchers in Germany, Russia, Cyprus, Portugal, France, Japan, Korea, South Africa, Australia, and many other geographies miles away from the Google office where TensorFlow was first developed. Innovation was not confined to a regional cluster.

The second surprising empirical result is visible in many of the figures capturing collaboration among the top 10 contributors without counting Google. We can note that chipset makers (i.e., amd, Nvidia, arm, and intel), and cloud computing vendors (microsoft, amazon, and ibm ) have developers collaborating directly with each other. This is in the sense that they co-edit the

same source-code files of the TensorFlow code repository. They are not only in competition solely with Google in the ecosystem, they are also in competition with many other direct competitors while fighting for the same revenue in globalized markets. In this sense, these results align with prior studies that found low levels of homophily by company affiliation in the production of complex software ecosystems (see ??).

Even if our research questions were theoretically driven, in order to explain the evolution of TensorFlow year after year, we came across naturally occurring qualitative material that directly addressed *Why do tech giants like Google open-source advanced and complex technological platforms that started in-house?*. That qualitative material often taking the form of interviews<sup>13</sup>, televised expert commentaries<sup>14</sup>, and Internet forums<sup>15</sup>, was rich and explanatory even if not provoked by us. While many wonder why giving up so much intellectual property already on the table, many outlined other explanations such as crowd sourcing innovation, fomenting open-innovation, resource complementary, sharing risk, increased cooperation with researchers, enhancing collaborative efficiency in product development, practical benefits that can lead to faster product development and maintenance, reuse of software artefacts, exposition of the technology to thousands of developers in the open-source community as well as fearing the threat of new entrants to the market.

While the set explanations from the previously mentioned qualitative material were already covered by literature in open-source, open-innovation, user-innovation, competition, and open-competition (see e.g., ????), we did find another novel explanation (i.e., expected outcome) for releasing software in an open-source way - that is **increasing the market size for the technology**. As Google released TensorFlow under an open-source license, it became much easier and more efficient to deploy deep learning. As time passed, barriers to entry decreased, technology evolved, and people with skills to train and deploy efficient deep learning models passed from a few hundred (often researchers with advanced doctoral education) to many thousands. The range of products and services embedding deep learning increased exponentially and all that created additional demand for the computing and data services commercialized by Google. Furthermore, as a bonus, Google started also selling TPUs (i.e., integrated circuits for neural network ML!) that had been specifically designed for TensorFlow and started deploying them on their Google Pixel Phones and many other products and services at Google (e.g., speech and image recognition). As pointed out by Rajat Monga, deep learning used to be done by researchers, but he has been now seeing high-schoolers building, training and deploying deep learning models.

Based on those observations we contribute to the further theorize the phenomenon of open-competition by laying out the following proposition:

**Theoretical Proposition 1.** – Within a high tech context, releasing a complex technological platform in an open-source way will lead to a loss of intellectual property, but it can also lead to an increased size of the market that in turn can increase demand for complementary products and services.

Regarding the second research question *Why are different organizations cooperating with their competitors in the co-production of those open-source platforms*, we found it easier to explain *what would happen if they fail to do so*. First, let's take the example of Microsoft. If Microsoft would ignore TensorFlow, we would likely see the demand for its Azure cloud computing services and for its Windows operating system decrease to other players that put the effort in steering, customizing and optimizing TensorFlow according to their services and product needs (e.g., Amazon cloud services and Linux). As a second example, if Nvidia would ignore TensorFlow, we would likely see the demand for its GPUs chipset sales decreasing to other chipset vendors performing better at running TensorFlow (e.g., ARM, Google, Intel, AMD and Samsung). Furthermore, if ByteDance would also ignore TensorFlow, it would be in a much harder position to maintain its distributed deep learning infrastructure that supports several apps such as TikTok, CapCut and Douyin. Finally, companies like IBM that help their customers run customized deep learning models, could also lose businesses by not contributing to TensorFlow. Contributing to TensorFlow can bring good reputation to the firm to win projects that deploy TensorFlow deep learning in production and can also help them to meet their customer requirements.

While we suggest that open sourcing might increase the size of the market, failing to engage in open-competition can negatively impact the market share. Based on our observations we further theorize the phenomenon of open-competition by laying out the second and last proposition of this study:

<sup>13</sup> See for example the interview with Rajat Monga that leads the R&D of TensorFlow at Google, conducted by Lex Fridman, a YouTuber and research scientist at the MIT Laboratory for Information and Decision Systems. Transcript publicly available at <https://transcript.lol/read/youtube/@lexfridman/652310c5033150beacd17e4e>.

<sup>14</sup> See *Why Google wants everyone to have access to TensorFlow* via Fox News <https://www.foxnews.com/video/4611174773001>.

<sup>15</sup> See <https://www.quora.com/Why-did-Google-open-source-TensorFlow-Whats-in-it-for-them>.

**Theoretical Proposition 2.** – Within a high tech context, firms might be forced to engage in coopetition in the co-production of complex technological platforms to protect the market share of their complementary products and services.

Besides the two theoretical propositions that theorize the open-coopetition phenomenon, we also contribute research in open-source software by outlining motivations for high-tech giants to open-source platforms, and outline motivations for non-commercial organizations to contribute to open and cooperative platform ecosystems. Furthermore, we propose an extension of the triple-helix model to better accommodate the important role that open-source foundations play in an open and competitive innovation model. Given that the open-source communities Chromium and Apache were the most connected to all the top ten contributors to TensorFlow, and as open-source foundations play an important dual role in (1) maintaining a software code base in an open-source way that for several reasons can not be turned into commercial software, and (2) being a neutral place that ensures openness and transparency in the orchestration of collective goals, we call for the NGOs like open-source foundations to be included in the triple helix framework.

From a practical point of view, we warn managers to not fall into the trap of being overfocused on protecting intellectual property and market share<sup>16</sup>. By being so concentrated on growing intellectual property assets, licensing, sales and market share, they can neglect the potential size of the market that a technology can bring. The open-sourcing of TensorFlow is now perceived as a success story, the value that TensorFlow brought to Google as an open-source project in terms of innovation and complementary demand is much higher than the value that TensorFlow could bring if kept locked and protected within the organization. In certain conditions, increasing the market size can offset the loss of intellectual property, licensing revenues, sales and market share. Managers must conciliate advice and interests from the legal and marketing teams with the advice and interests from the R&D and engineering teams, to not miss up opportunities to scale the market.

## 6 | CONCLUSION

In this research, we followed the TensorFlow open-source software ecosystem where competitors often cooperate in the co-production of a widely-used platform for **ML!** and **AI!**. We contributed to an enhanced understanding of (1) why open-sourcing technology that was made in-house, and (2) why cooperate with competitions in an open-source way, and (3) the role and motivation of non-commercial organization in the co-production of open and cooperative platforms.

We found out that on the one hand, by open-sourcing TensorFlow, Google benefited upstream by continuously developing the state of the art of deep learning. This as it counted with contributions from researchers from all over the world in the form of ideas, models, methods, patches, pull requests, bug fixes, security patches, etc. On the other hand, Google passed some of those benefits downstream by making it more easy and more efficient to deploy deep learning into products and services that impact real people. That in turn increased the size of the market and created complementary demand for the computing and data services that Google commercializes. After all, the best-performing deep learning models depend on large amounts of quality data and computational power that Google have at hand.

The main message of the paper is that Google increased the size of the market by open-sourcing TensorFlow and created complementary demand for its computing and data services. Also, Google brought innovation from outside (researchers, specialized startups, users, competitors, deployers, etc). Many competitors also need to cooperate with the development of TensorFlow otherwise the demand for their complementary products and services (e.g., chipsets, computing services, and operating systems among others) would decrease to other players that steered, customized, and optimized TensorFlow according to their own needs.

Our work also illustrates how aligned incentives and relaxed IP constraints can promote broad-based innovation, emphasizing the significant contributions of non-commercial entities in maintaining software commons and driving innovation. We contribute to extant theory on open-source software by outlining motivations for high-tech giants to open-source platforms, and outline motivations for non-commercial organizations to contribute to open and cooperative platform ecosystems. Furthermore, we call for the triple-helix model to better accommodate the important role that open-source foundations play in an open and competitive innovation model along with the lines of prior work by (?) that highlighted the importance of NGOs in innovation and development.

<sup>16</sup> Note that many organizations have incentives that reward employees and managers based on intellectual property and market share.