



Integrating the gender dimension to disclose the degree of businesses' *articulation of innovation*

Giacomo di Tollo¹ · Joseph Andria² · Stoyan Tanev³ · Sara Ghilardi⁴

Received: 17 July 2023 / Accepted: 23 October 2023
© The Author(s) 2023

Abstract

In this contribution, we examine the relationship between the presence of women in companies' Boards and innovation communication claims: we propose a framework to quantitatively assess the presence of women and the online articulation of innovation, in order to understand whether some correlations hold between these two variables. We also introduce a neural network approach to predict the innovation metric that uses, amongst the predictors, the gender component, and we compare it with a linear regression analysis. Results indicate that neural networks may be used to predict the articulation of innovation by using a predictor set that includes the gender component of the Board of Directors, and also that the use of the gender metric improves previous predictions about the articulation of innovation model's output.

Keywords Gender diversity · Board of directors · Innovation metrics · Online communication · Neural networks

✉ Joseph Andria
joseph.andria@unipa.it

Giacomo di Tollo
giacomo.ditollo@unisannio.it

Stoyan Tanev
stoyantanev@cunet.carleton.ca

Sara Ghilardi
sara.ghilardi97@gmail.com

¹ Dipartimento di Diritto, Economia, Management e Metodi Quantitativi, Università degli Studi Del Sannio, Benevento, Italy

² Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università degli Studi di Palermo, Palermo, Italy

³ Sprott School of Business, Carleton University, Ottawa, Canada

⁴ Sales Department, Regalgrid Europe Srl, Treviso, Italy

Introduction

Innovation in corporate organizations represents a driver of competitiveness and productivity growth, which in turn have a positive impact on both firms' value and reputation. In the last decades, we have witnessed fundamental changes in the innovators' workplaces, that had a significant influence on both the business collaboration practices and the management of (bulky) information [83, 99, 104]. On one side, the figure of the innovator is changing, and though innovation has been perceived for years in a predominantly masculine context, a strong evidence of gender differences in the decision-making process which impact firms' innovation strategies has been reported in literature [30]. On the other side, the increasing intensive use of web-based technologies is changing the way firms communicate all the activities related to their business mission and objectives. The interdependence of these activities has opened debates about who actually drives the innovation process, and the effect of the people driving innovation on *how* the innovation is articulated and communicated to potential stakeholders.

In this context, it is important to understand whether the diversity of people behind innovation is an asset for companies, and whether this asset has an impact on the ways creativity and innovation are communicated. This task is difficult since, as we shall see in what follows, it is difficult to assess the human aspect of firm's contribution to innovation, probably due to the invisibility of people in innovation.

In our paper, we want to study whether the gender composition of the corporate Boards of Directors could be related to how the information about innovation is *articulated* on the company's website, and to how it is *communicated* to potential customers and other stakeholders. Corporate boards are becoming increasingly engaged with their company's innovation journey and can provide important oversight and support along the way [12, 88, 103]. Our research design presupposes a relation between corporate website content and board's innovation strategy. Such presupposition does not assume a fine-grained control of the board on the content of company websites. We do however assume that the communication practices and the content of company websites are aligned with the key messages and decisions of the board with respect to the corporate innovation strategy.

This paper is thus an attempt to address the ongoing debate about the relationship between leadership and innovation, by examining:

- whether the presence of women as part of the leadership of complex organizations may have an influence on the communication of innovation-related topics;
- the extent to which human attributes (in our case, the gender attribute) may be used to derive useful information about company innovation communication practices.

Following the ideas of [107], we want to quantify the ability of businesses to articulate and communicate the innovative aspects of their products, services and

processes opportunities. Any innovation assessment procedure should be devised w.r.t. the type of innovation that one wants to take into account. In our study, we focus on the products, processes and services innovation. Such focus allows us to relate our results to previous studies that have investigated the relationship between the Board's gender diversity and product innovation [84]. In particular, gender diversity may lead to a broader range of ideas, which could trigger innovation, since more diverse ideas, in their number and in their diversity, may increase the likelihood to introduce new products or new services by the company [78, 89]. Although gender aspects have been already studied w.r.t. product innovation, results are still far from being unique: for example, [46] bring evidence of significant associations between Board diversity and product innovation, also stating that in some cases gender diversity revealed a negative impact on product innovation. In our contribution, we will go beyond traditional products innovation metrics by focusing on the companies' degree of articulation of the innovative aspects of their new products, processes and services.

After addressing the gender-aspects of innovation, we aim to use these insights to help modeling the extent to which businesses articulate the innovativeness of their products and services, based on the straightforward assumption that a company's online communication practices have a potential impact on customers and other stakeholders perceptions of its innovativeness, sustainability, and commitment [21, 26]. Starting from the consideration by [66], that stresses out the importance of considering how innovativeness is perceived as a whole, we implement a metric that measures the extent to which a company articulates the innovativeness of its products, processes and services. This metric (referred to as *articulation of innovation*) measures the frequency of use of keywords associated to innovation, and it is related to company's communication practices, based on its claims about the innovative aspects of its products, processes or services. In order to determine the values of this metric, we have used a web search tool that measures the frequency of a preliminary designed regular expression on the websites of firms from specific business sectors. Please notice that this procedure has been devised by [105], in a contribution aimed to express the articulation of innovation by using different types of *co-creation* activities as predictors [36, 37]. In our work, we want to extend these approaches, by adding the *gender metrics* in the predictors set.

In a nutshell, in our paper, we aim:

- To examine the relation between the *articulation of innovation* and the presence of women in companies' Boards of Directors. To this end, we will perform a correlation analysis between the gender metric and the articulation of innovation, and we will interpret the results to test the hypothesis that the specific expertise and knowledge brought by women could be as associated with a broader range of online claims about the innovativeness of company's products and services and, thus, with a better articulation of firm's innovativeness [69].
- To predict the *articulation of innovation* by adding the gender component to the predictors set.

Since some authors state that female Board representation is positively associated with performance only with respect to firms for which innovation and creativity play a particularly important role [27], the purpose of our exploration is to understand whether there is a difference between the relationship between *gender* and *articulation of innovation*, and between *innovation-driven* and *more traditional* businesses: we investigate firms for which innovation and creativity play a particularly important role to determine whether there is a correlation between gender and innovation articulation; then, we compare our findings to firms that are representative of a whole economy. This will provide (or not) evidence to confirm that female Board representation is positively associated with innovation performances and that is tightly intertwined with the most innovative and creative firms.

Our paper is organized as follows: “[Motivation](#)” illustrates the motivations that led to the present work; “[Co-creation, gender aspects and innovation](#)” outlines the ongoing trends of the innovation and gender related literature; “[Our data](#)” describes the data we have used in our experiments; “[Correlation analysis](#)” performs a first correlation analysis of the variables taken into account. Then, our Neural Network approach is outlined in “[Neural network approach](#)”, and is compared to a linear regression approach in “[Comparison with linear regression](#)”, before concluding and outlining future directions of research in “[Conclusions and future works](#)”.

Motivation

The motivation for our study is twofold. First, the relationship between the gender composition of the Board of Directors and the degree of articulation of the innovative aspects of companies’ new products and services has not been studied in detail before: there are no previous studies that have suggested a specific tool and model that could be used to examine the practical aspects of companies’ online communications, and in this sense, our study addresses a gap in the existing innovation management literature. Second, there is not any preliminary assumption about the nature of the above relationship, and this lack of preliminary knowledge justifies the adoption of a generic approach like neural networks in parallel to traditional linear regression methods, which pre-assumes a simple linear relationship. In this sense, a key feature of our study consists in the generic nature of the exploration of the relationship between gender and the articulation of innovation: we will show that the neural network approach appears to be better at detecting the potential role of gender on the articulation of innovation on companies’ websites, when used jointly with other indicators. We want to stress out that the only contribution to this regard has been made by [35], that concluded that “it is not possible to detect a universal relationship between innovation perception and gender diversity on the Board of Directors, since traditional business show a low magnitude of such correlation and, conversely, an higher positive relationship is found on businesses that are referred to as innovative”: in both cases, [35] has found an increasing relationships between the two aggregates over time (data have been collected over the years 2013 and 2019), and in the current contribution, we will check whether this trend is confirmed over new observations.

The novelty of the study is the focus on the relationship between the presence of women in the Boards of Directors and the articulation of innovation on companies websites: to the best of our knowledge this relationship has not been studied in-depth so far, and the results of our study are expected to provide preliminary insights that could potentially inform and support executive management decisions related to the online communication practices of their companies, as well as to explore ways of enhancing them by ensuring a stronger engagement with the female representatives of their Boards of Directors. More importantly, the study provides a basis for future studies that could address the above relation in more explicit ways. Furthermore, our analysis can be used by policy makers interested in the relation between innovation management, the communication of companies' innovativeness and gender balance. Accordingly, we can assess the potential impact of gender balance on firms performance and innovation management: our data are important for understanding whether gender policies (promoted by the country in which the firm is based) are successfully implemented, as well as for comparing gender policies promoted in the different countries, in order to set up an experimental comparison over time and across different countries. In the second part of our work, we perform a computational prediction task in order to assess whether gender aspects may be useful to predict the online articulation of innovation by businesses coming from different sectors.

Our work falls within the strand of recent research highlighting the impact of gender on corporate innovation [47, 115, 117]. However, our approach differ in several respects, mainly in the way we measure the innovativeness of a business. Instead of using traditional measures, such as patent counts and patent citations (which capture only a partial aspect of corporate innovation), we go beyond by using the value co-creation dimension as a result of all those practises which challenge the traditional ways of innovation management [38, 105].

Gender quotas regulations for company boards is nowadays widely debated and, although their implementation has contributed to boost the promotion of gender diversity policy in today's labor market, they represent a very controversial subject. While the political arguments for gender quotas are intrinsically motivated by principles of fairness and equality, this paper, by examining to what extent gender diversity at the board level affects innovation, looks more in-depth as it discloses relevant policy guidelines and practical implications.

Co-creation, gender aspects and innovation

In this section, we are providing a literature review about the main concepts upon which our study is based: co-creation and gender, and their implications in the innovation debate. As stated in "[Introduction](#)", there are some contributions that assess the relationship between co-creation and articulation of innovation, and we want to understand whether the gender component can be used to improve this assessment task.

Value co-creation is becoming more and more important in marketing and innovation: it can be defined as a marketing paradigm able to satisfy the needs of

heterogeneous groups of customers [109], through the involvement of end users and other stakeholders in the creation of the final product or service [90]: it fosters a transformation of the customer role, that becomes an *active* participant in the creation of its value [95]. In this paradigm, the business and the customer build a dialogue that should be interpreted as a process of co-learning to achieve a shared goal for both parties [13]. For this reason, the involvement of customers in the design of the product or service has to be seen as an interactive process, often undertaken unconsciously by the customer [64, 90]. Customers' preferences are fundamental elements which shape the value co-creation process [108]: in order to satisfy these preferences, the business has to provide its customers with information, knowledge, skills and resources to be used during the processes; in addition, the business must be able to influence the value co-creation process in order to enable customers to make the most efficient use of the resources [87].

Due to the advancement in Information and Communication Technology (ICT), more and more efficient technological tools allow individuals to receive news and information in real time and at a global level. As a consequence, people become more and more aware of their needs to be satisfied by a product or a service [38]. Furthermore, it should be noted that the interaction between the business and the customer during the value co-creation process can be guaranteed by a channel of communication between them, and this channel often consists of an online platform [101]. The development of new technologies has improved the efficiency of these online platforms, making the collaboration between businesses and customers easier and faster.

In this contribution, we adopt the definition of co-creation introduced by [36–39] in which *co-creation* has been identified as a single concept. Please notice that in some contributions [44, 97] the concept of co-creation has been partitioned into *value co-creation* and *co-production*, in which the former refers to the customer's involvement in the phase of use and consumption of the product or service, and the latter refers to the involvement in the phase of design and production.

The adoption of value co-creation practices leads to benefits for both the customer and the supplier: with regards to customers, they can get the product or service that actually meets their preferences [116], and they feel actively involved in the production process and this stimulates trust and loyalty towards the supplier [90]; from a supplier's perspective, value co-creation clearly highlights customer preferences and provides the opportunity to create the products and services able to meet their interests [109]. In addition, this learning may lead to create new products (and services), leading the business to gain a competitive advantage and a greater degree of innovation [116]. In this regard, it has been stated by several authors [65, 82, 94] that the involvement of customers in value co-creation activities has a positive influence on the results of innovation; in particular, this activity is able to reduce innovation costs and time-to-market, and to increase the quality of the new product or service and the business' development skills. For this reason, the development of new platforms that allow collaboration in the creation of value is considered more important and it is placed among the fundamental pillars of a business strategic plan [82, 96].

Given the increasing importance of co-creation activities [109] and that a possible positive relationship between the concepts of value co-creation and

innovation has been found [36–39], one of the main purposes of this study is to investigate this relationship and contribute to its prediction and assessment.

Over the last decades, the innovation debate paid more and more attention to the gender aspect, especially to its relations to top management [52]: leading academics and policy makers have started to investigate the diverse implications of the presence of women in (and more generally, the gender diversity of) corporate Board of Directors [72]. According to [32], the presence of women in the Board of Directors is apt to diversify perspectives, experiences, working styles, knowledge, and expertise with respect to their male counterpart [51, 56], motivating an increasing attention towards the role of women, especially in innovative (and creative) businesses [27, 76].

While the sex of a person is a feature defined by biology, gender is something that refers to a social construction: it is a cultural aspect [70] and it is assimilated through the interaction with other people. Therefore, individuals learn how to behave according to their gender, and they also learn the attitudes to avoid in order not to be inappropriate [114]. As a consequence, there is a high probability that people expect women to behave according to their “femininity”, while men according to their “masculinity” [112]. Therefore, there are stereotypes of the role attributed to both genders [43] and these can be classified as 1) *descriptive gender stereotypes*, that refer to what a man or woman is like, and 2) *prescriptive gender stereotypes*, that refer to the socially required behavior of woman and a man. The elements just described can shape the behavior of individuals and have an influence on their decisions: men and women are partly influenced by what is *normal to do* for the society according to “masculinity” or “femininity”. With this regard, a relevant example is given by the segregation of the labor market, i.e., the lack of males or females in a certain sector or profession [3].

The behavior of individuals is also affected by the context in which they find themselves because some contexts seem to require specific attitudes and roles. The reaction of people who find themselves in this circumstance is usually to adapt themselves to the context and behave as it requires, assuming traits that recall masculinity or femininity [24]. A relevant context in which individuals’ behavior is influenced by the context itself is represented by the *entrepreneurial world*: from a psychological and competence point of view, it requires leadership skills [80, 118], which are often identified in typical masculine traits such as aggressiveness, risk-taking and autonomy [4]. On the contrary, femininity seems to be attributed to a warm, calm and communal attitude and therefore, if a woman intends to undertake an entrepreneurial activity, she *should* assume the afore mentioned masculine traits [2]. Furthermore, according to the Global Report 2020/2021 [22] compiled by the Global Entrepreneurship Monitor (GEM), men involved in entrepreneurial activities are significantly more than women. In particular, the percentage of men entrepreneurs exceeds that of women in almost the entire world in the Total Early-stage Entrepreneurial Activity (TEA), which accounts for all those new entrepreneurial activities undertaken during the year; the exceptions are represented by only six countries (Kazakhstan, Indonesia, Oman, Saudi Arabia, Togo and Angola).

In the context of entrepreneurship, there are significant differences between the two genders in innovation-related issues [85], due to the fact that the innovation concept is characterized by a risk-taking component (women are more risk-adverse) and by the implementation of new technologies. This last aspect may be explained by the different fields of study chosen by men versus women: it has been shown [15, 23, 57] that the presence of women in STEM curricula (Science, Technology, Engineering, and Mathematics) is significantly lower compared to men. As a result, more men acquire training and familiarity with technology and are therefore more likely to adopt innovative management practises that require the application of new technologies.

We emphasize that the definition of innovation is gender-neutral, but it has been argued that the way innovation is operationalized and measured is strongly gendered with masculine connotations [81, 92]: a strong association is found between innovation (and technology) with masculinity [111], and women are seen as less innovative than men, leading to a perception of women's underperformance [31, 75] in innovation practices [20, 74, 81].

More recently, theories asserting the masculinity character of innovation have been partly discarded: there is a growing focus on the role of women in innovation, particularly in companies that are strongly innovation-driven [76]. In particular, a positive influence of the women's presence on Boards in innovation results has been demonstrated [84]. The presence of women on Boards leads to a greater heterogeneity in the group, and, consequently, to a greater diversity of perspectives, experiences, work styles, knowledge and skills [32]. With regard to skills and knowledge, it is found that women are more adept at recognizing customer behaviors and expectations; this leads to the identification of innovative products, services and processes that more closely reflect customer needs [58, 107]. Furthermore, women on Boards usually have better understanding of customer behaviors and expectations, and in this regard, the presence of women on Boards is more suitable to widen the range of ideas and perspectives to innovate the product and to identify market opportunities [51, 107]. In a nutshell, as conjectured by [86] the specific expertise and knowledge brought by women may contribute to broaden the range of new products and services, and one may argue that in some cases a correlation exists between the presence of women on Boards and innovation [107]. Nevertheless, the studies focused on identifying the contribution of the presence of women (or gender diversity) to firm innovation are still limited, probably due to the invisibility of people in innovation [106], as opposed to the limelight nature of entrepreneurship [7]. In our contribution, also in view of the particular focus by policy makers (i.e., European institutions) on assessing and closing the gender gap in the economic and financial sector [77], we are going to investigate the relationship between the gender aspect and the companies' articulation of innovation, which is strictly related to the customers' perception: this latter point is a key concern of innovation, since it is right through this one that companies influence their customers' innovation perception. In this regard, there exist in literature several studies that propose a qualitative assessment approach, e.g. by performing surveys [68] that are also used to assess the perception of innovation barriers amongst firms [16, 53]. A substantial amount of literature exists on the customer perception of innovation [67]; the effects of innovation

perception have been explored in different sectors, e.g. banking [41], educational [60], hospitality [59]. A unified framework for consumer perceived innovativeness, embedding both qualitative and quantitative aspects has been proposed by [71], and a further abstraction is proposed by [66], that define the perceived firms' innovativeness as a consumer-centric view of innovation, based on the customer assessment of the business' capability to endure, whose indicator is the capability of creating and commercializing novel, creative, and impactful ideas and solutions. Please notice that the procedure we have used to assess the articulation of innovation on company web-sites is gender neutral, to avoid masculine connotations in the way it is measured.

In what follows we outline the research ideas developed to investigate the relationship between *co-creation* and *innovation*, that will be used as the main *building blocks* of our approach. In particular, we will present the origin and development of the metrics used to measure the two aspects (i.e., innovation and co-creation) and the approaches and tools used to assess their relationships. In order to assess these two aspects, one approach suggests to compute the frequency of keywords related to the concept of co-creation and innovation within the websites of a large sample of businesses. The approach based on using the frequency and co-occurrence of words to examine the relation between specific organizational activities was introduced by [45] where it has been analyzed the sequence of competitive actions of a sample of firms to define their own business strategies. The same method has then applied to innovation-related topics by [50] to analyze the business models of innovative businesses, by defining a set of keywords related to innovation and by counting (and normalizing according to the size of the websites) their occurrences: thanks to this approach, the authors were able to classify businesses based on the degree and the type of innovation communicated on their websites.

The same approach was applied to analyze value co-creation by [6], that introduced a set of keywords (detailed in Table 1) related to value co-creation in order to classify the different co-creation activities practiced by a large sample of businesses.

A further development was provided by [105], that looks for the occurrence of keywords related to co-creation and innovation on businesses' websites, and implements a linear regression to investigate the relationship between the two aspects, in order to assess the influence of the practice of co-creation activities on the degree of innovation communicated by businesses on their websites. This idea was further investigated by [38], that introduced a Neural Network approach to investigate the relationship between value co-creation and innovation among a large sample of businesses, by exploiting the neural networks' skills to analyze different kinds of relationships amongst variables. In this latter paper the authors tested the hypothesis that businesses with a higher degree of involvement in co-creation activity had a greater number of possibilities, occasions and contexts in which to apply innovation in their products, services and processes [44, 96], and a positive association between these concepts has been found. Later contributions led to comparable results on different sets of data [36, 37, 73], on top of implementing Self-Organizing Maps to classify businesses according to their degree of involvement in co-creation and innovation activities.

Table 1 Co-creation keywords in [5]. Keywords are represented as regular expressions in which concepts are identified by plain text and connected by logical connectors (\vee denotes *OR* and \wedge indicates *AND*)

Co-creation keywords defined by Allen et al. 2009

(customer \vee *user)* \wedge (*learn* \vee *learning*)
(customer \vee *user)* \wedge (*communities* \vee *community* \vee *network* \vee *networking* \vee *forum*)
(customer \vee *user)* \wedge (*suggest* \vee *suggestion* \vee *input* \vee *request* \vee *demand*)
(customer \vee *user)* \wedge (*dialog* \vee *dialogue* \vee *communicate* \vee *communication* \vee *conversation* \vee *contact* \vee *feedback* \vee *call* \vee *interact* \vee $\}$ *information sharing* \vee *engage*)
internal \wedge (*expertise* \vee *resource*)
cost \wedge (*reduce* \vee *reduction* \vee *saving*)
customer \wedge (*partnerships* \vee *interaction* \vee *relationship* \vee *participate* \vee *participation* \vee *activity* \vee *action*)
(design \vee *process)* \wedge (*flexibility* \vee *flexible* \vee *adaptable*)
(customer \vee *user)* \wedge (*cooperate* \vee *cooperation* \vee *collaboration* \vee *partnership*)
(customer \vee *user)* \wedge (*riskmanage* \vee *management* \vee *control* \vee *assess* \vee *reduce* \vee *reduction* \vee *potential* \vee *exposure*)
trust \vee *honesty* \vee *integrity*
(customer \vee *user)* \wedge (*options* \vee *choice* \vee *choose*)
integrated \wedge *online* \wedge *services*
customization \vee *customize* \vee *customized* \vee *personalize* \vee *individualize* \vee $\}$ *add feature* \vee $\}$ *added feature* \vee $\}$
(product \vee *process)* \wedge (*modularity* \vee *modular* \vee *module*)
ecosystem \vee $\}$ *value network* \vee $\}$ *value constellation* \vee $\}$ *multiple partners* \vee $\}$ *external contributor* \vee $\}$ *external source* \vee $\}$
(customer \vee *user)* \wedge (*disclose* \vee *inform* \vee *disseminate* \vee *reveal*)
(customer \vee *user)* \wedge (*produce* \vee *assemble* \vee *manufacture*)
(customer \vee *user)* \wedge (*IP* \vee $\}$ *intellectual property* \vee $\}$)
(customer \vee *user)* \wedge (*test* \vee *trial* \vee *beta*)

Last, [39] examined the relationship between the degree of involvement in co-creation activities by businesses, the degree of articulation of their service value attributes and their innovativeness, by using different methods: Principal Component Analysis was used in order to identify the components of co-creation activities; neural networks and correlation analysis were used to analyze the aforementioned researched relationship; K-means cluster analysis¹ and Self-Organizing Maps (SOMs) approach allowed to classify the businesses according to their degree of involvement in different co-creation activities, articulation of their service value attributes and their innovativeness. The results of this contribution show the presence of a statistically significant relationship between the degree of involvement in co-creation activities by businesses, the degree of articulation of their service value attributes and their innovativeness.

All the above studies and findings suggest that it may exist a relationship between the articulation of innovation and co-creation activities. In this contribution,

¹ K-means cluster analysis is an algorithm which allows the user to classify a group of objects on the basis of their characteristics.

we attempt to deepen this line of research by trying to understand whether adding the gender component may shed more lights to detect (or not) this relationship.

Our data

In the past decades, there has been a significant growth in the amount of unstructured data: the elaboration of texts proves to be more and more complex and time-consuming, so the adoption of text mining techniques for the analysis of textual data has proved to be very useful in order to simplify and improve the activity of researchers [11] and other stakeholders. Text mining refers to techniques capable of analyzing an unstructured text, processing its information, and creating a structured content [55]. *Text mining* has been used in the field of innovation research [40, 83], and its use over the Internet (Web scraping) can be useful to detect specific business' information: according to [26], businesses' online communication of innovation-related content influences customers' and other stakeholders' perceptions of innovation. Starting from this assertion, we want to quantify the amount of online content about innovation, which from now on we will call *articulation of innovation*. Articulation of innovation will be used in what follows as a metric to evaluate the degree of innovation of a business, and instead of using a pre-defined dictionary based on text mining procedures [11], we use the frequency of the companies' online comments about their own new products, processes and services to assess their articulation of innovation: we do not aim to apply an automated content analysis, but rather to assess the *degree* of the articulation, i.e. to which extent companies' news announcements concern the innovative aspects of their products, services and processes on their websites. Actually, the meaning of this measure is twofold: first, it could be seen as an expression of the firm's perception of its own innovativeness; second, it represents the communication of the innovative image the company wants to offer to customers and other stakeholders.

This metric has been introduced by [105], and it is computed by detecting any online statement containing the combination of the words *new* and *product*, or the words *new* and *service*, or *new* and *process* etc..²

Other research studies have focussed on the relationship between gender and innovation, applying more traditional and tangible metrics (number of new products, new services, new processes, new patents, etc.) such as the ones described in the OSLO manual.³ The present study does not use these metrics, and our innovation metric provides a quantitative evaluation of how often firms articulate the innovative aspects of their market offers on their websites: it embeds the advantage of emphasising the ability of a firm to differentiate itself by articulating the innovative

² Formally, this metric counts the occurrences of the following regular expression: $new \wedge (product \vee service \vee process \vee application \vee solution \vee feature \vee release \vee version \vee launch \vee introduction \vee introduce \vee (new\ product) \vee (new\ service) \vee (new\ process))$ and then normalises the value found w.r.t. the number of pages taken into account.

³ Available at https://www.oecd-ilibrary.org/science-and-technology/oslo-manual-2018/introduction-to-innovation-statistics-and-the-oslo-manual_9789264304604-4-en

Table 2 The co-creation keywords used in our approach

Co-creation	
Keyword	
C1	<i>customer</i> ∨ (<i>user</i> ∧ <i>dialog</i>) ∨ <i>dialogue</i> ∨ <i>conversation</i> ∨ <i>feedback</i> ∨ <i>call</i> ∨ <i>interact</i> ∨ (<i>information</i> ∧ <i>exchange</i>) ∨ (<i>information</i> ∧ <i>sharing</i>) ∨ (<i>information</i> ∧ <i>access</i>) ∨ <i>engage</i>
C2	<i>customer</i> ∨ <i>user</i> ∨ <i>forum</i> ∨ <i>connect</i> ∨ <i>network</i> ∨ <i>networking</i>
C3	<i>lease</i> ∨ <i>rent</i> ∨ <i>license</i> ∨ (<i>self</i> ∧ <i>serve</i>) ∨ (<i>self</i> ∧ <i>service</i>)
C4	<i>customer</i> ∨ (<i>user</i> ∧ <i>cooperate</i>) ∨ <i>cooperation</i> ∨ <i>collaboration</i> ∨ <i>partnership</i>
C5	<i>customer</i> ∨ (<i>user</i> ∧ <i>suggest</i>) ∨ <i>suggestion</i> ∨ <i>input</i> ∨ <i>request</i> ∨ <i>demand</i>
C6	(<i>internal</i> ∧ <i>expertise</i>) ∨ <i>resource</i>
C7	<i>customer</i> ∨ (<i>user</i> ∧ (<i>risk</i> ∧ <i>manage</i>)) ∨ <i>management</i> ∨ <i>control</i> ∨ <i>assess</i> ∨ <i>reduce</i> ∨ <i>reduction</i> ∨ <i>potential</i> ∨ <i>exposure</i>
C8	<i>customer</i> ∨ (<i>user</i> ∧ <i>IP</i>) ∨ (<i>intellectual</i> ∧ <i>property</i>)
C9	<i>customer</i> ∨ (<i>user</i> ∧ <i>learn</i>) ∨ <i>learning</i>
C10	<i>product</i> ∨ <i>process</i> ∨ (<i>service</i> ∧ <i>evolution</i>) ∨ <i>evolve</i>
C11	<i>customer</i> ∨ (<i>user</i> ∧ <i>experience</i>)
C12	<i>customer</i> ∨ (<i>user</i> ∧ <i>test</i>) ∨ <i>trial</i> ∨ <i>beta</i>
C13	(<i>integrated</i> ∧ (<i>online</i> ∧ <i>services</i>))
C14	<i>simulation</i> ∨ <i>simulate</i> ∨ <i>model</i> ∨ <i>modelling</i> ∨ (<i>virtual</i> ∧ <i>world</i>) ∨ (<i>reference</i> ∧ <i>design</i>) ∨ (<i>reference</i> ∧ <i>flow</i>) ∨ (<i>demo</i> ∧ <i>application</i>) ∨ <i>toolkit</i> ∨ <i>tutorial</i> ∨ <i>sdk</i> ∨ (<i>software</i> ∧ (<i>development</i> ∧ <i>kit</i>))
C15	<i>product</i> ∨ (<i>process</i> ∧ <i>modularity</i>) ∨ <i>modular</i> ∨ <i>module</i>
C16	<i>customer</i> ∨ (<i>user</i> ∧ <i>produce</i>) ∨ <i>assemble</i> ∨ <i>manufacture</i>
C17	<i>customer</i> ∨ (<i>user</i> ∧ <i>options</i>) ∨ <i>choice</i> ∨ <i>choose</i>
C18	<i>design</i> ∨ (<i>process</i> ∧ <i>flexibility</i>) ∨ <i>flexible</i> ∨ <i>adaptable</i>
C19	(<i>customer</i> ∧ <i>partnership</i>) ∨ <i>interaction</i> ∨ <i>relationship</i> ∨ <i>participate</i> ∨ <i>participation</i> ∨ <i>activity</i> ∨ <i>action</i>
C20	(<i>cost</i> ∧ <i>reduce</i>) ∨ <i>reduction</i> ∨ <i>saving</i>
C21	<i>customer</i> ∨ (<i>user</i> ∧ <i>survey</i>) ∨ <i>review</i> ∨ <i>voting</i> ∨ <i>vote</i> ∨ <i>rate</i> ∨ <i>rating</i>
C22	<i>trust</i> ∨ <i>honesty</i> ∨ <i>integrity</i> ∨ <i>transparency</i>
C23	<i>customer</i> ∨ (<i>user</i> ∧ <i>disclose</i>) ∨ <i>inform</i> ∨ <i>disseminate</i> ∨ <i>reveal</i>
C24	<i>customer</i> ∨ (<i>user</i> ∧ <i>dashboard</i>) ∨ <i>statistics</i>

aspects of its products and services [36], and it accounts for the claims of innovativeness of a company about its own products and services.

The *value co-creation* refers to the degree of customers' involvement in a business in the creation of new products and services. According to the related literature outlined in “**Co-creation, gender aspects and innovation**”, this degree of customers' co-participation has been computed by searching for co-creation-related keywords (detailed in Table 2) on businesses' websites.

In order to assess the *articulation of innovation* and the *co-creation* components, we have implemented a *Web Scraper for Regular Expressions*, that

is specifically tailored to our data needs: it allows the user to calculate the frequency of a predefined set of regular expressions (that has to be provided as input by the user), related to the innovation and co-creation activities of a business. For each website, the scraper identifies all available webpages belonging to the web domain; for each page, the source code is translated into text, and the conditions imposed by the user are translated into regular expressions: the occurrences of these regular expressions are computed for each page, aggregated for the whole website, and rescaled with respect to the number of pages.

In order to carry out our analysis, we have collected data about three different business' attributes: the afore mentioned *articulation of innovation*, the *co-creation* components, and the *gender* component. As for the *gender component*, [1, 48] introduced the idea of using gender (i.e., sex) as a variable, often to analyse differences and similarities between men and women. Starting from this, we have decided to compute the percentage of female directors on the companies' Boards, that reflects either the extent of female directors appointments or Boards homogeneity/heterogeneity (with values, respectively, equal to 0 or 1), as suggested by [72]. By using this metric (referred to as *gender metric*), we want to test the assertion in [33, 100], according to which many firms portrayed as leaders in innovation activity are managed by teams that include both men and women. Moreover, such approach appears to be in agreement with the concepts stated by [25], which remarks that research should focus less on gender differences and similarities, and pay more attention into understanding how gender is embedded in processes, for instance by the counting of women and men involved in innovation processes.

We have computed the *articulation of innovation*, the *co-creation* components, and the *gender* component with respect to the year 2021 over five different sets of data:

- 287 Open Source (OS) businesses associated with the Eclipse OS Foundation which embrace those businesses that are more apt to adopt open innovation and co-creation practices. OS companies have been commonly considered as a unique example of open innovation and creativity [42, 91, 110]. This set of data will be referred to as *Eclipse*;
- the other four sets of data consist of businesses listed on the NASDAQ (98 businesses), FTSE100 (95 businesses), DAX30 (30 businesses), and CAC40 (40 businesses), which refer to stock markets of different countries (respectively, USA, UK, Germany, and France) and are considered to be representative of the country's overall economic and financial condition (details can be found at https://docs.google.com/spreadsheets/d/1gaZepe_m_kqoQLtBM3dBpiBvpBppIAZj/edit?usp=sharing&ouid=100079799523990491093&rtpof=true&sd=true).

Table 3 shows the main statistics about the *gender metric* which is defined, as said, by the ratio between the number of female Board members and the total number of Board members along with the *articulation of innovation* over the examined sets of data for the year 2021. The following tables also contain the

Table 3 Main statistics over the year 2021 of: the ratio of women to total members of the Board; the articulation of innovation

	Women to total Board				Art. of innovation			
	Mean	STD	Min	Max	Mean	STD	Min	Max
Eclipse	0.66	0.32	0	1	0.17	0.32	0	2
NASDAQ	0.28	0.10	0	0.54	0.10	0.18	0	1.00
FTSE	0.35	0.11	0	0.64	0.08	0.17	0	1.04
DAX	0.16	0.10	0	0.38	0.11	0.16	0	0.89
CAC	0.38	0.14	0	0.63	0.11	0.23	0	1.00
Overall	0.26	0.16	0	1	0.14	0.29	0	2

statistics and results for the overall set of data containing aggregated data from the five instances pooled together, that will be referred to as *Overall*.

Nevertheless, we want to emphasize that by applying the above described gender metric we overcome the criticism about the reliability of the innovation-related measurements in literature as considered to be gender-biased [81]. Please notice that our analysis is not meant to define a model to explain innovation, that is why we are not building a model by identifying the most suitable control variables: the purpose of our experimental study is to verify whether:

- the relation between gender and innovation is different between innovation-related and more traditional businesses;
- the gender metric can better explain the innovation metric.

We bring to the reader's notice that here we are not addressing the role that time plays in the innovation process or in making claims on company websites, for example by examining the time lag between the joining to the Board of a woman and her fully realized ability to influence the business innovation claims [17]. This latter aspect would require a separate investigation by means of a dedicated script able to autonomously detect changes in the articulation of innovation and gender component, which goes beyond the scope of our current implementation and goals.

Looking at Fig. 2, we can draw some insights from the graphical representation of the relationship between innovation and the percentage of women on the Board of Directors. For all sets of data, we plotted on the x-axis the percentage of women on the Board of Directors while the frequencies of the articulation of innovation are shown on the y-axis. At a first glance, we do not observe a straight correlation between the variables, and the same conclusion holds also by looking at Fig. 1, where the articulation of innovation frequencies appear on the y-axis, and the number of men on the Board of Directors are displayed on the x-axis.

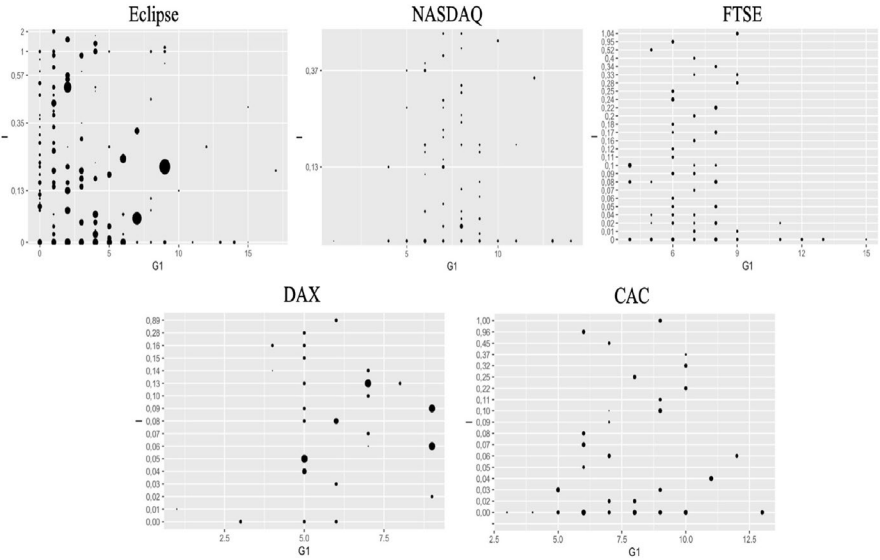


Fig. 1 Scatter plot of the articulation of innovation (referred to as I on the y-axis) vs number of men in the Board of Directors (referred to as G1 on the x-axis) in the five sets of data taken into account. The point size is proportional to the number of women in the Board of Directors

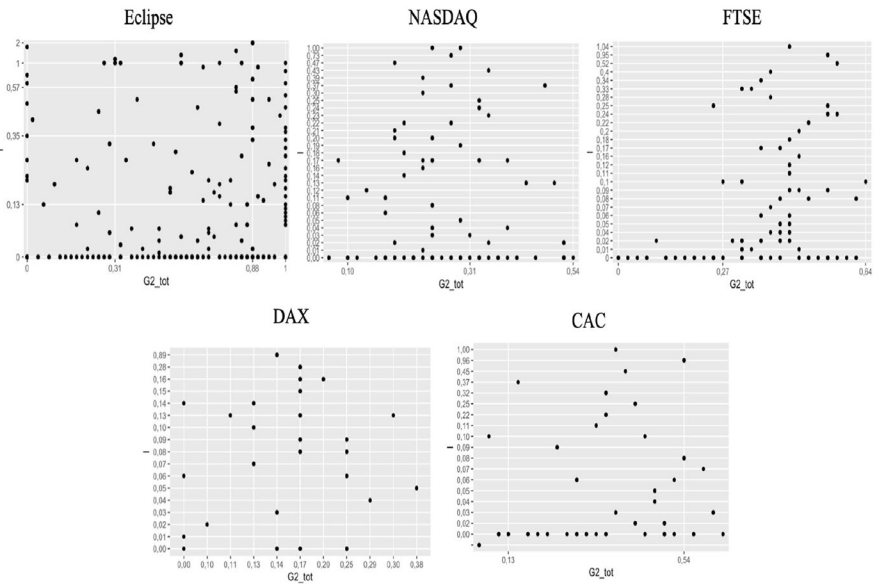


Fig. 2 Scatter plot of the articulation of innovation (referred to as I on the y-axis) vs percentage of women in the Board of Directors (referred to as G2_tot on the x-axis) in the five sets of data taken into account

Correlation analysis

In this section, we outline a correlation analysis between the *gender component* and the *articulation of innovation* with respect to the investigated sets of data. The purpose of this analysis is to understand whether the gender ratio on the Board of Directors and businesses' innovation activity are somewhat related to each other. Moreover, with respect to this aspect, we want to investigate whether there are some differences between businesses that are known to be innovative and technology-driven (i.e., *Eclipse*) and those more traditional. We start by remarking that the average ratio of women on the Board of Directors is highest in the *Eclipse* set of data: this seems to confirm what has been observed in literature on gender and innovation, i.e., a higher presence of women involved in the decision-making process in innovation-related businesses than traditional ones. In addition, we report that 71% of the businesses in the *Eclipse* instance have an absolute majority of women on the Board and that the second highest absolute majority of women is shown by CAC (15%), while the other sets of data show small values, down to 0% in DAX. We also remark that the maximum female presence on the Board is still found in the *Eclipse* set of data (100%). The other sets of data contains also businesses in which there is an absolute majority of women, except for the businesses listed in the German stock market, where the maximum female presence is 38%. Last, we want to stress that the ratio *female to Board* show a minimum equal to 0 on all sets of data, and that the percentage of businesses without women in the Board is between 1% and 2% in all sets of data but DAX, that show a higher number. These results seem to confirm that on a large scale entrepreneurship is still largely gendered [63], but, compared to the past, there is a greater presence of women in both businesses where innovation is more intense and more traditional businesses [35].

In what follows we draw some comments on the articulation of innovation metric. By looking at Table 3, the highest average is found in *Eclipse*, and this seems to confirm what was expected, that is, these businesses are exactly those known to be more technology-driven and innovative. The same holds for the maximum. When analysing the minimum instead, it is possible to observe that all sets of data show the value zero: in all sets of data there are therefore businesses that seem to have no interest in communicating their innovation to potential customers and stakeholders through their website. In particular, we report that in all sets of data the percentage of businesses that have zero frequency related to the articulation of innovation is about 50%, except for DAX, that shows the 20%. In table 4 we report, respectively, a Pearson, Rank-based and Mutual Information statistical dependence analysis aimed at determining the relationships between the two variables. In particular, Mutual Information (MI) dates back to the beginnings of information theory [102]. It was initially applied to calculate the capacity performance of communication systems, and thereafter it served as a fundamental tool of investigation for many other research areas, such as mathematics, computer science and economics [8, 9]. MI has been successfully adopted in feature-selection problems to assess both the relevancy of a subset of features in predicting

Table 4 Pearson, Rank Based and Mutual Information (MI) dependence coefficients between the *gender component* and the *articulation of innovation*

	Pearson	Rank-based	MI
Eclipse	0.02	0.10	0.53
NASDAQ	0.01	0.01	0.44
FTSE	0.23	0.03	0.41
DAX	- 0.06	- 0.10	0.24
CAC	0.09	0.07	0.33
Overall	0.03	- 0.003	0.37

the target variable and the redundancy with respect to other variables. Although the correlation coefficient is a long-standing measure of the strength of statistical dependence, MI has advantages over it: it is able to detect non-linear relationship between features and it is model neutral.

The results show a very low Pearson correlation coefficient and almost close to zero (except for FTSE) with a p -value significantly greater than 0.05, suggesting that there is no correlation between the two variables over all sets of data. Similar remarks can be drawn by looking at the ranked-based correlation analysis.

As a last takeaway, we draw a comparison with the results about the pairwise mutual information between the articulation of innovation and the gender component over the years 2013 and 2019 reported in [35]. Results therein presented show an increase of the mutual information over the investigated years, and this trend appears to hold true for 2021 as well. In particular, Eclipse shows the highest Mutual Information value (0.53 in 2021, while it was 0.32 in 2013 and 0.34 in 2019), while the highest incremental change over years is observed for FTSE which goes from 0.05 in 2013 to 0.41 in 2021.

We conclude this analysis by remarking that the gender component and the articulation of innovation do not show any linear or rank-based correlation. On the other side, the Mutual Information between the two variables is significant (≥ 0.5) over three sets of data out of five: since the Mutual Information helps in reducing the uncertainty for a given random variable (in our case, the *articulation of innovation*) when another variable (in our case, the *ratio women to total Board*) is known, this could be a good signal suggesting us that the use of the gender component may be used to predict the articulation of innovation. The next Sections will aim to investigate if this holds true.

Neural network approach

Artificial Neural Networks (ANN) are high-level algorithms inspired by the behavior of the brain, of which they represent a simplification [49]: they have been used to perform a wide range of computational tasks in many fields in which there is no assumption about the relations amongst input and output variables. They are composed of elementary units (i.e., neurons), which are connected through weighted edges (i.e., synapses) in a given topology.

Synapses are associated with a value, which determines the magnitude of transmission. A neuron receives inputs from the neurons that are connected to it, which are aggregated as the weighted sum of inputs and synapses' values; then an activation value is computed by applying a given activation function to the weighted sum of inputs. This value is sent, through synapses, to the connected neurons. We can devise three main categories of neurons:

- Input neurons, whose activations represent the inputs of the task to be performed by the network;
- Output neurons, whose activations represent the output of the network;
- Hidden neurons, the remaining neurons, so called because they are not visible from the external environment.

As for the topologies, the most common are the layered one (in which neurons are subdivided in layers) and the completely connected (in which all neurons are all connected with each other). Each neuron in the layered topology is connected to all neurons in adjacent layers, whilst there are no connections between neurons in the same layer: in this topology the information flow of the network is unidirectional, and the resulting network is referred to as a *feed-forward* architecture.

To use the neural networks, the user needs first to set the values of synapses: this is done by the learning algorithm, which can be seen as the ability of the network to modify its behavior in order to obtain the right output given some input through the modification of synaptic connections (weights). This happens by modifying these weights until a stopping criterion is met (usually, when the difference between the desired and the actual output produced by the network is smaller than a given threshold).

Several algorithms have been proposed to train neural networks, e.g., genetic algorithms [79], simulated annealing [62], and backpropagation [113]. This last algorithm is still the most used, and it is based on the back-propagation of the error from the output units to the input ones: in a first phase, the activation of the input units is propagated through the activation functions; then, the weights of the synapses are modified through the technique of the descendent gradient.

We have performed three different sets of experiment (models):

1. A model to identify a relationship between the articulation of innovation and the gender component only, in which the input is given by the gender metric and the output by the articulation of innovation. In what follows this model is referred to as *GE model*;
2. A model in which the input set is composed of all co-creation components and the output by the articulation of innovation. This model is referred to as *CO model*;
3. A model in which the input set is composed by all the co-creation components plus the ratio female to cardinality of Board of Directors of the business taken into account: this model combines the co-creation components with the gender aspect, and is referred to as *CO-GE model*.

In what follows, we will outline the main features of our neural network approach: “Data pre-processing” will outline the data pre-processing operations; “Training and test set” will detail how we have partitioned data into *training* and *test* sets; “Network topology” will detail the Network topology, while the learning parameter and the algorithm’s performances will be outlined in “Learning parameters” and “Algorithm performances”, respectively.

Data pre-processing

The analysis of data at hand represents an important phase in the experimental settings for a neural network approach: this operation aims to explore data features, to detect eventual anomalies, and to preserve the most information as possible, avoiding at the same time over-fitting shortcomings. In our approach, we have used the pre-processing operations defined by [10, 29], that we outline in what follows:

Removal and replacement

The issue of incurring in missing or wrong values is a key-point in real-world applications, and all neural networks approaches resort to procedures to take into account this aspect. We have used the approach introduced by [10], that suggests to remove the indicators containing more than 30% of missing or wrong values. All sets of data introduced in “Our data” do not reach this portion of missing or wrong data, hence we are using, in the experimental phase, the whole set of variables. In detail, no wrong values have been detected in our sets of data; as for missing data, we want to remark that for some businesses it was not possible to collect all data referring to the gender components, so we have some missing data referring to the number of male and female in the Board of Directors: in this last case, we have followed the guidelines indicated by [10, 29] and replaced the missing values by the variable’s average over all businesses.

Normalization

When applying neural networks, a widely used rule-of-thumb suggests to perform data-normalization in order to feed the neural network with data belonging to the same range. Many mathematical formulations have been suggested to this aim [61], and we are using the logarithmic transformation used by [36, 37, 39], which is defined as follows: let x_i be the value before normalisation of input x for business i , and \bar{x}_i be its normalised value. The relation between normalised and pre-normalised data can be defined as follows:

$$\bar{x}_i = \log_u (x_i + 1), \quad (1)$$

where $u = x_{\max} + 1$, in order to have $\bar{x}_i \in [0, 1]$. We want to remark that the original formulation proposed by [10] was the following:

$$\bar{x}_i = \log_u (|\min(0, x_{\min})| + x_i + 1), \quad (2)$$

and this was due to the fact that the authors have tackled a problem arising from the use of different variables sets, in which there were observations whose values were negative. All values belonging to our sets of data represent weighted occurrences,

and they cannot be negative by definition, hence we are not considering the possibility of encountering negative values, that would hinder the application of a logarithmic transformation. Tables 5, 6 and 7 show the main statistics of data at hand after the pre-processing operations.

Training and test set

As for the experimental framework, in all experiments, we have sampled two disjoint sets of observations out of the total number of businesses belonging to a set of data: the training set (used to estimate the networks' parameters) and the test set (used for assessing their performances). Businesses have been randomly allocated to these two sets to have the training set consisting of 75% of the total businesses, and the test set consisting of 25% of the total businesses. This sampling has been repeated 30 times, each time leading to a different definition of training and test set, which was needed to compute a robust weighted error function over the diverse train-test partition. The procedure has been performed over the five different sets of data at hand.

Network topology

As for the network topology, we have used a feed-forward architecture trained by backpropagation: in this topology, the most important parameter to be set is the number of hidden neurons and number of hidden layers. The literature on the topic has introduced many approaches to enhance these issues, ranging from recognised rules of thumbs to adaptive procedures to change the topology of the neural network over time. In this direction, we have performed preliminary experiments by using the adaptive procedure proposed by [28], in which hidden neurons are added to the network until no improvements on the performances of the system are detected. This approach requires a computational time that is about 10 up to 50 times higher than a basic feed-forward one, and a comparison of the results based on a Wilcoxon signed rank test (tested against a 0.05 significance level) led us to reject the null hypothesis of difference between the two approaches. Hence, we have resorted to well-established rules of thumb in order to set the number of hidden neurons, as reported by [34], and we have used one hidden layer, setting the number of hidden neurons equal to the number of parameters (variables). The reason for choosing one layer only is due to the small cardinality of some sets of data at hand, that could hinder the learning algorithm's ability to learn the higher number of parameters associated to the synapses. We want to stress out that two-layers neural networks are formally recognised as universal function approximators [54], but the comparison with the adaptive procedure performed in the preliminary experiments suggests that one hidden layer is enough for our purposes.

Table 5 Main statistics of the response and explanatory variables after pre-processing operations, for the sets of data *Eclipse* and *Nasdaq*. Variables *C1* – *C24* refer to co-creations components; variable *I* denotes the articulation of innovation; variable *G* denotes the ratio women to total members of the Board

Variable name	Average	STD	Min	Max
<i>Eclipse</i>				
Number of businesses = 287				
<i>C1</i>	0.30	0.24	0	1
<i>C2</i>	0.27	0.25	0	1
<i>C3</i>	0.18	0.18	0	1
<i>C4</i>	0.25	0.25	0	1
<i>C5</i>	0.31	0.27	0	1
<i>C6</i>	0.16	0.25	0	1
<i>C7</i>	0.32	0.24	0	1
<i>C8</i>	0.23	0.24	0	1
<i>C9</i>	0.31	0.31	0	1
<i>C10</i>	0.21	0.20	0	1
<i>C11</i>	0.26	0.26	0	1
<i>C12</i>	0.26	0.27	0	1
<i>C13</i>	0.22	0.20	0	1
<i>C14</i>	0.29	0.24	0	1
<i>C15</i>	0.22	0.24	0	1
<i>C16</i>	0.23	0.24	0	1
<i>C17</i>	0.34	0.28	0	1
<i>C18</i>	0.15	0.19	0	1
<i>C19</i>	0.20	0.24	0	1
<i>C20</i>	0.11	0.18	0	1
<i>C21</i>	0.25	0.25	0	1
<i>C22</i>	0.05	0.11	0	1
<i>C23</i>	0.23	0.24	0	1
<i>C24</i>	0.23	0.25	0	1
<i>I</i>	0.12	0.20	0	1
<i>G</i>	0.23	0.20	0	1
<i>NASDAQ</i>				
Number of businesses = 98				
<i>C1</i>	0.35	0.29	0	1
<i>C2</i>	0.27	0.26	0	1
<i>C3</i>	0.28	0.26	0	1
<i>C4</i>	0.23	0.24	0	1
<i>C5</i>	0.33	0.25	0	1
<i>C6</i>	0.16	0.23	0	1
<i>C7</i>	0.38	0.25	0	1
<i>C8</i>	0.23	0.24	0	1
<i>C9</i>	0.30	0.27	0	1
<i>C10</i>	0.30	0.28	0	1
<i>C11</i>	0.25	0.26	0	1
<i>C12</i>	0.27	0.27	0	1
<i>C13</i>	0.31	0.32	0	1

Table 5 (continued)

Variable name	Average	STD	Min	Max
<i>C14</i>	0.21	0.19	0	1
<i>C15</i>	0.19	0.22	0	1
<i>C16</i>	0.22	0.24	0	1
<i>C17</i>	0.37	0.26	0	1
<i>C18</i>	0.12	0.17	0	1
<i>C19</i>	0.23	0.25	0	1
<i>C20</i>	0.18	0.26	0	1
<i>C21</i>	0.25	0.27	0	1
<i>C22</i>	0.25	0.27	0	1
<i>C23</i>	0.12	0.19	0	1
<i>C24</i>	0.23	0.26	0	1
<i>I</i>	0.12	0.20	0	1
<i>G</i>	0.55	0.21	0	1

Learning parameters

As for the neural network learning parameter, the literature on the topic reports that the typical values for a neural network with standardized inputs (or inputs mapped to the $[0, 1]$ interval, which is our case) has to be smaller than 1 and greater than 10^{-6} [18]; furthermore, there is no way of determining the learning rate a-priori [98]. For these reasons, we have resorted to a parameter tuning procedure to set the learning rate: F-Race [19], by whose execution we obtained a learning parameter equal to 0.12 for the *CO* model and 0.10 for the *CO-GE* model. Please notice that for both models the momentum value as determined by the F-Race procedure was close to 0, hence we have not introduced any momentum parameter into our analysis.

Algorithm performances

As for the neural network learning performances, the basic rule consists of using the supervised learning technique on the test set, while, in order to avoid overfitting, the goodness of the neural network has to be assessed by computing an error metric over the validation set, that can be also used in order to set the termination criterion for the algorithm. Several error measures can be used, such as the mean absolute error (MAE), the mean absolute percentage error (MAPE), the root mean square error (RMSE), etc. In our experiments, we have used the mean square error (MSE), defined as:

$$\frac{1}{n} \sum_{i=1}^n (e_i - a_i)^2, \quad (3)$$

where n is the number of observations, e is the expected output and a is the actual neural network output.

Table 6 Main statistics of the response and explanatory variables after pre-processing operations, for the sets of data *FTSE* and *DAX*. Variables *C1* – *C24* refer to co-creations components; variable *I* denotes the articulation of innovation; variable *G* denotes the ratio women to total members of the Board

Variable name	Average	STD	Min	Max
<i>FTSE</i>				
Number of businesses = 95				
<i>C1</i>	0.48	0.30	0	1
<i>C2</i>	0.30	0.30	0	1
<i>C3</i>	0.24	0.25	0	1
<i>C4</i>	0.24	0.26	0	1
<i>C5</i>	0.33	0.28	0	1
<i>C6</i>	0.13	0.22	0	1
<i>C7</i>	0.48	0.28	0	1
<i>C8</i>	0.19	0.25	0	1
<i>C9</i>	0.24	0.24	0	1
<i>C10</i>	0.33	0.30	0	1
<i>C11</i>	0.26	0.26	0	1
<i>C12</i>	0.19	0.25	0	1
<i>C13</i>	0.30	0.26	0	1
<i>C14</i>	0.32	0.25	0	1
<i>C15</i>	0.24	0.27	0	1
<i>C16</i>	0.19	0.24	0	1
<i>C17</i>	0.34	0.26	0	1
<i>C18</i>	0.18	0.21	0	1
<i>C19</i>	0.25	0.27	0	1
<i>C20</i>	0.14	0.18	0	1
<i>C21</i>	0.20	0.21	0	1
<i>C22</i>	0.13	0.19	0	1
<i>C23</i>	0.18	0.24	0	1
<i>C24</i>	0.18	0.25	0	1
<i>I</i>	0.10	0.18	0	1
<i>G</i>	0.59	0.19	0	1
<i>DAX</i>				
Number of businesses = 30				
<i>C1</i>	0.53	0.23	0	1
<i>C2</i>	0.41	0.24	0.04	1
<i>C3</i>	0.42	0.25	0	1
<i>C4</i>	0.45	0.28	0	1
<i>C5</i>	0.43	0.26	0.01	1
<i>C6</i>	0.33	0.30	0	1
<i>C7</i>	0.64	0.22	0	1
<i>C8</i>	0.36	0.27	0	1
<i>C9</i>	0.35	0.26	0	1
<i>C10</i>	0.56	0.28	0	1
<i>C11</i>	0.42	0.29	0	1
<i>C12</i>	0.38	0.29	0	1
<i>C13</i>	0.51	0.25	0	1

Table 6 (continued)

Variable name	Average	STD	Min	Max
<i>C14</i>	0.51	0.30	0	1
<i>C15</i>	0.47	0.32	0	1
<i>C16</i>	0.37	0.28	0	1
<i>C17</i>	0.48	0.20	0.02	1
<i>C18</i>	0.37	0.30	0	1
<i>C19</i>	0.40	0.30	0	1
<i>C20</i>	0.37	0.29	0	1
<i>C21</i>	0.47	0.25	0	1
<i>C22</i>	0.22	0.28	0	1
<i>C23</i>	0.37	0.27	0	1
<i>C24</i>	0.38	0.29	0	1
<i>I</i>	0.15	0.19	0	1
<i>G</i>	0.46	0.26	0	1

Neural Networks have been implemented in Python. Experiments have been run on a cluster with AMD Opteron 2216 dual core CPUs running at 2.4 GHz with 2x1 MB L2 cache and 4 GB of RAM under Cluster Rocks distribution built on top of CentOS 5.3 Linux. We have run the neural network on all the obtained *train–test* partitions. For each partition, the algorithm has been run 100 times, and the best run (with respect to the MSE) has been recorded. In order to test the neural network performances and the robustness of the approach, we report in Table 8 the statistics of the best runs obtained over the 30 partitions and with respect to the *GE*, *CO*, and *CO-GE* models.

By looking at the results, we can say that the neural network is not able to generalize the *GE* model, hence we conclude that a general relationship between the gender component alone and the articulation of innovation cannot be explained. On the contrary, good results were obtained from the application of the neural network to the *CO* and the *CO-GE* models: the average of the total error on the single test sets varies between 0.035 and 0.120 in with respect to the *CO-GE* model and between 0.041 and 0.171 for the *CO* model; furthermore, the overall instances show the average of the total error between the minimum and maximum provided by the single test sets, but features a lower standard variation, due to better performances obtained on bigger sets of data. As for the single sets of data, the lowest MSE (on both models) is found for the FTSE set of data and the highest error is found for the CAC set of data, which may be due to the low number of observations. The standard deviation of both models is comparable, meaning that the better performances on the *CO-GE* model do not come at the cost of an higher variability of the results. In all cases, we can say that the introduction of the gender component improves the skills of the neural approach to predict the articulation of innovation.

Table 7 Main statistics of the response and explanatory variables after pre-processing operations, for the sets of data *CAC* and *Overall*. Variables *C1* – *C24* refer to co-creations components; variable *I* denotes the articulation of innovation; variable *G* denotes the ratio women to total members of the Board

Variable name	Average	STD	Min	Max
<i>CAC</i>				
Number of businesses = 40				
<i>C1</i>	0.54	0.26	0	1
<i>C2</i>	0.36	0.32	0	1
<i>C3</i>	0.31	0.32	0	1
<i>C4</i>	0.28	0.28	0	1
<i>C5</i>	0.35	0.33	0	1
<i>C6</i>	0.21	0.24	0	1
<i>C7</i>	0.46	0.23	0	1
<i>C8</i>	0.26	0.30	0	1
<i>C9</i>	0.31	0.31	0	1
<i>C10</i>	0.26	0.25	0	1
<i>C11</i>	0.28	0.30	0	1
<i>C12</i>	0.27	0.32	0	1
<i>C13</i>	0.35	0.30	0	1
<i>C14</i>	0.37	0.31	0	1
<i>C15</i>	0.17	0.24	0	1
<i>C16</i>	0.26	0.31	0	1
<i>C17</i>	0.44	0.25	0	1
<i>C18</i>	0.24	0.29	0	1
<i>C19</i>	0.37	0.32	0	1
<i>C20</i>	0.23	0.28	0	1
<i>C21</i>	0.34	0.31	0	1
<i>C22</i>	0.19	0.28	0	1
<i>C23</i>	0.26	0.31	0	1
<i>C24</i>	0.26	0.30	0	1
<i>I</i>	0.13	0.24	0	1
<i>G</i>	0.66	0.22	0	1
<i>Overall</i>				
Number of businesses = 550				
<i>C1</i>	0.45	0.28	0	1
<i>C2</i>	0.35	0.25	0	1
<i>C3</i>	0.34	0.25	0	1
<i>C4</i>	0.35	0.27	0	1
<i>C5</i>	0.39	0.27	0	1
<i>C6</i>	0.25	0.26	0	1
<i>C7</i>	0.43	0.25	0	1
<i>C8</i>	0.33	0.27	0	1
<i>C9</i>	0.36	0.26	0	1
<i>C10</i>	0.35	0.25	0	1
<i>C11</i>	0.36	0.27	0	1
<i>C12</i>	0.34	0.26	0	1
<i>C13</i>	0.36	0.25	0	1

Table 7 (continued)

Variable name	Average	STD	Min	Max
<i>C14</i>	0.40	0.26	0	1
<i>C15</i>	0.30	0.26	0	1
<i>C16</i>	0.33	0.27	0	1
<i>C17</i>	0.42	0.26	0	1
<i>C18</i>	0.28	0.25	0	1
<i>C19</i>	0.32	0.27	0	1
<i>C20</i>	0.25	0.26	0	1
<i>C21</i>	0.32	0.25	0	1
<i>C22</i>	0.18	0.23	0	1
<i>C23</i>	0.33	0.27	0	1
<i>C24</i>	0.33	0.27	0	1
<i>I</i>	0.17	0.22	0	1
<i>G</i>	0.28	0.12	0	1

Table 8 Neural Network's overall errors (MSE) on the *GE*, *CO-GE*, and *CO* models. We report the statistics on the best runs obtained on the 30 different partitions of overall error

	Training set				Validation set			
	Average	STD	Min	Max	Average	STD	Min	Max
<i>GE model</i>								
Eclipse	0.073	0.034	0.012	0.085	0.081	0.054	0.023	0.099
Nasdaq	0.052	0.012	0.040	0.065	0.058	0.012	0.048	0.075
Ftse	0.052	0.012	0.038	0.069	0.042	0.013	0.035	0.049
Dax	0.054	0.019	0.034	0.079	0.066	0.018	0.043	0.076
Cac	0.127	0.013	0.112	0.152	0.076	0.019	0.053	0.098
Overall	0.084	0.041	0.021	0.137	0.095	0.052	0.032	0.215
<i>CO-GE model</i>								
Eclipse	0.042	0.005	0.037	0.058	0.045	0.011	0.022	0.084
Nasdaq	0.044	0.015	0.024	0.068	0.043	0.017	0.015	0.088
Ftse	0.032	0.014	0.013	0.044	0.035	0.011	0.014	0.081
Dax	0.047	0.022	0.021	0.096	0.058	0.036	0.012	0.114
Cac	0.112	0.031	0.042	0.148	0.120	0.056	0.025	0.237
Overall	0.043	0.007	0.022	0.153	0.053	0.009	0.031	0.173
<i>CO model</i>								
Eclipse	0.047	0.006	0.028	0.060	0.048	0.010	0.026	0.079
Nasdaq	0.042	0.014	0.015	0.068	0.048	0.021	0.014	0.086
Ftse	0.037	0.012	0.019	0.049	0.041	0.028	0.011	0.084
Dax	0.044	0.022	0.017	0.089	0.046	0.034	0.010	0.119
Cac	0.171	0.120	0.128	0.190	0.171	0.123	0.118	0.201
Overall	0.045	0.013	0.025	0.170	0.059	0.016	0.039	0.201

Comparison with linear regression

From the observation of the Table 8, it is possible to detect the strength and the ability of generalization of the neural network: results have been validated with respect to the test set, hence on different sets of data than those used for the training of the network. We want to stress the importance of this conclusion by performing a linear regression on the same sets of data, in order to test the over-fitting of the resulting model. We have performed linear regressions for the three models, but we are reporting only results of the *CO* and *CO-GE* models, since the *GE* model do not lead to satisfactory performances. Please notice that in literature there are contributions that introduce linear regression to predict the articulation of innovation, in which the predictor set is given by the same co-creation components used in our analysis (see “[Co-creation, gender aspects and innovation](#)”): we want to stress out that these contributions also introduce a pre-processing phase in which Principal Component Analysis [37] is used to reduce the input space. In our contribution instead, we do not apply such data reduction techniques, since they may lead to biased or misleading conclusions. We used all the indicators shown in Table 2 to form the predictors set, resulting in a procedure that is highly reliable and that does not require the intervention of an external user.

Please notice that we are not interested in assessing the goodness of the predictors, so that we do not report their estimates along with the corresponding *p*-values, and we just report the following performance measures:

- The coefficient of determination R^2 , along with its variants: the *Adjusted R^2* , that is used to adjust this performance measure to take into account the number of predictors (i.e., it decreases when the cardinality of predictors does not improve the goodness of the fit by a sufficient amount), and the *Predicted R^2* , that takes into account a sampling procedure to assess overfitting (i.e., lower values are an indicator of overfit);
- The *p*-value of the F-test, that represents the probability to obtain an F-statistic value greater than the F-value of the model, under the null hypothesis that the regression is not significant [14];
- The Akaike’s Information Criteria (AIC), that measures the goodness of a statistical model by assessing the relative amount of information loss (lower values identify better models with regards to the specific set of data).

We have performed two linear-regression analyses based on two different sets of experiments: first, we have performed the analysis on the whole set of data, without partitioning into training and test sets; then, we have adopted the same subsampling procedure to identify thirty different partitions of training and test sets as defined in “[Training and test set](#)”, and then we have computed the aggregated performance indicators.

As for the experiments performed without partitioning the data into training and test set, we obtain the results listed in Table 9. We can notice that adding the gender component (to the predictor set composed by the co-creation indicators) does not

Table 9 Estimates of the linear models *CO* and *CO-GE*: estimated parameters, *p* – values and significance statistics of the linear models. Columns highlighted in bold mean that the corresponding regression is significant according to the p-value of the F-test

Statistics	CO model					CO-GE model					Overall	
	Eclipse	NASDAQ	FTSE	DAX	CAC	Overall	Eclipse	NASDAQ	FTSE	DAX		CAC
R^2	0.460	0.858	0.722	0.958	0.903	0.660	0.673	0.588	0.779	0.396	0.310	0.661
<i>Predicted</i> R^2	0.283	0.323	0.388	0.739	0.473	0.602	0.568	0.154	0.508	0.011	0.013	0.601
<i>Adjusted</i> R^2	0.408	0.813	0.632	0.796	0.762	0.642	0.640	0.437	0.695	0.125	0.185	0.642
AIC	2999	709.1	772.5	215.7	335.4	-571.7	-184.3	44.221	-39.634	30.750	30.243	-570.6
p-value (F-test)	< 0.001	< 0.001	< 0.001	0.0174	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.084	< 0.001	< 0.001

Table 10 Linear regression errors (MSE) on both *CO* and *CO-GE* models. We report the statistics on the best runs obtained on the 30 different partitions of training-validation set

	Training set				Validation set			
	Average	STD	Min	Max	Average	STD	Min	Max
<i>CO model</i>								
Eclipse	0.029	0.012	0	0.52	0.115	0.093	0	0.719
Nasdaq	0.025	0.019	0.001	0.57	0.061	0.052	0.014	0.312
Ftse	0.041	0.017	0.001	0.072	0.081	0.061	0.035	0.102
Dax	0.002	0.002	0.001	0.004	0.022	0.014	0.001	0.059
Cac	0.004	0.003	0.001	0.007	0.149	0.192	0.046	0.512
Overall	0.015	0.001	0.014	0.017	0.013	0.002	0.008	0.018
<i>CO-GE model</i>								
Eclipse	0.026	0.0138	0.0005	0.066	0.109	0.142	0	0.827
Nasdaq	0.0312	0.016	0.0009	0.101	0.073	0.081	0.015	0.351
Ftse	0.038	0.013	0.0007	0.065	0.078	0.038	0.046	0.172
Dax	0.021	0.012	0.009	0.036	0.075	0.057	0.009	0.137
Cac	0.082	0.058	0.007	0.157	0.274	0.251	0.241	0.325
Overall	0.016	0.001	0.014	0.017	0.012	0.002	0.008	0.015

always lead to an improvement in the goodness of the fit when the regression performance is assessed by the R^2 . On the contrary, it worsens the performances over three (NASDAQ, DAX, CAC) out of five sets of data. Anyhow, the regression is always significant, as confirmed by the p -value of the F-test (values smaller than 0.025 indicate that the regression is significant, and they are highlighted in bold in the table).

Once more, we could conclude that the operation of adding the gender component not only does not lead to a significant adjuvant effect but it worsens the regression performance in the case of the DAX set of data. Anyhow, the goodness of the fit assessed by the R^2 seems to indicate that linear regression is appropriate, with the exception of the sets of data *DAX* and *CAC* for the *CO-GE* model, and of the set of data *Eclipse* with respect to the *CO* model. We also notice that adding the gender component improves the linear regression performance over the *Eclipse* set of data.

Although this result may sound promising, this is a situation affected by overfitting, as witnessed by low values of the predicted R^2 over the five sets of data. In order to demonstrate this, we have introduced the same sampling procedure defined in “[Training and test set](#)” and computed the MSE, by using Eq. 3, as for the neural networks. Results are reported in Table 10. We can see that linear regression achieves better performances over the train set with respect to four sets of data out of five for both *CO-GE* and *CO* models, whilst when dealing with the test set, neural networks show the best performances over the *CO-GE model* for all sets of data, and on three out of five of them with respect to the *CO* model. In this case, we see that the addition of the gender component plays a significant role when comparing these two different approaches.

In a nutshell, the introduction of the gender component into the predictor set leads to better neural network's performances in the validation set in three sets of data out of five (Eclipse, Nasdaq, Ftse); on the remaining two (Dax, Cac), its introduction worsen the neural network performances. Overall, we can state that the introduction of the gender component may help to improve the performances of a neural network approach. As for the linear models, the introduction of the gender component may help to improve the performances, but it is highly affected by over-fitting on small-sized instances, which makes the approach not reliable. On bigger sized instances (i.e., the overall set of data) instead, the linear regression analysis show performances that are better than the neural network approach, suggesting that the two approaches have to be used jointly.

Conclusions and future works

In this contribution, we have proposed a framework to predict the articulation of innovation, and to understand whether the gender component of businesses (coming from different scenarios) can be used to this extent: since innovation activities are managed by Board of Directors that include both men and women, we have tried to understand whether the gender component can be embedded in (and whether it can be used to explain) the innovation processes, such as, in our case, the articulation of innovation. Our results confirm previous findings [35] and reveal that it is not possible to detect a universal relationship between innovation perception and gender diversity on the Board of Directors, and demonstrate that neural networks may be used to approximate this relationship. The increasing relationships between *articulation of innovation* and *gender component* over time is also confirmed by more recent observations.

Our findings corroborate the idea that female Board representation is positively associated with more innovative and creative firms, but we have seen that this statement can also be applied to some more *traditional* business. We have introduced a novel support tool based on a neural network approach that is able to grasp the relationship between *gender* and *communication of innovation* on both *innovation-related* and *more traditional* companies. To this end, we have taken into account other features related to the innovation attitude of a business (such as the value co-creation) since we have shown that it is not possible to grasp such relationship when the gender component is used as a stand-alone predictor: this suggests that female Board representation is positively associated with innovation performances only when it is integrated into the business culture and vision, and our approach has shown that it is computationally pointless to advocate a direct relation between these two aggregates taken as *stand-alone*. At the current stage, we can assert that the gender component may be useful to predict the innovation articulation of a business under several scenarios.

Our findings seem to suggest that the relation between gender and innovation is strong in innovation-driven businesses, but it does not mean that it does not hold also for more traditional businesses: our approach offers satisfactory performances to this regard over traditional sets of data, but it fails on two benchmarks: this could

due to some features of the benchmarks themselves that lead to over-fitting, and further research has to be devoted to investigate this aspect. For the same reason, we can state that the gender metric might be helpful in predicting the articulation of innovation, but this statement has to be read in the light of what previously remarked about the evidence that the gender metrics cannot be used alone.

It has been argued that women are either absent or made absent in innovation processes, even if they work in a R &D role and are responsible of innovation activities [93]. In this contribution, we have shown the opposite, by devising an approach based on the ideas from [25]: we focussed on how gender is embedded in business processes, by the counting of women and men involved in innovation processes, and we have defined a metric able to quantify the level of openness of a company to innovation.

Our contribution spreads light on the ongoing discussion about the conditions under which gender diversity is an asset and a stimuli for creativity and innovation: we have investigated businesses with different gender composition and different attitude to creativity and innovation. In a way, this is a first step in the definition of a gender neutral concepts to be used when examining innovation and related issues, in order to develop methods to examine what people do, rather than how they talk about it. At the current step, it seems that the presence of women has a relation with the business' innovation claims: we cannot identify a causal link, but this indicates a direction for future research on what makes women more influential in the Boards, and on the extent to which women ideas about innovation are considered and implemented in companies (and comparing these findings with those relative to the male counterpart). Furthermore, we want to examine the extent to which women and men intervene in the innovation discourses: for this goal, we have to define different keywords and regular expressions to be associated to innovation and gender analysis.

Further research will be also devoted to replicating this approach to other geographical contexts: the first step will consist in analysing emerging markets, in which gender and innovations appear to be more intertwined. The tools we have introduced are apt to work in English, hence the first attempts will be concerned with English-speaking countries. A further analysis will be carried out after having defined the new keywords to be used in other languages.

Furthermore, we want to extend our approach by considering the role that time plays in the innovation process and in making claims on companies websites. In this direction, it would be interesting to investigate on the effects induced by time lags between the moment when a woman joins the Board of Directors and the time she can exert an influence on firms' innovativeness. Such analysis would need high frequency input data to continuously monitor changes in the Boards of Directors and in the regular expressions occurrences on the firms' websites. It will be will valuable for future studies to include more traditional ways of measuring innovation (number of new products, number of new product features, number of new services, number of patents, etc.) as covariates in models similar to the one presented here. This analysis was left out of the scope of the present study since it requires a completely different data collection approach.

Funding Open access funding provided by Università degli Studi di Palermo within the CRUI-CARE Agreement.

Data availability The set of data analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Åsberg, C., & Lykke, N. (2010). Feminist technoscience studies. *European Journal of Women's Studies*, 17(4), 299–305.
2. Abele, A. E. (2003). The dynamics of masculine-agentic and feminine-communal traits: findings from a prospective study. *Journal of personality and social psychology*, 85(4), 768.
3. Akerlof, G. A., & Kranton, R. E. (2010). *Identity economics: How our identities shape our work, wages, and well-being*. USA: Princeton University Press.
4. Alexander, D., & Andersen, K. (1993). Gender as a factor in the attribution of leadership traits. *Political Research Quarterly*, 46(3), 527–545.
5. Allen, S., Tanev, S., & Bailetti, T. (2009). *Components of co-creation*. Special issue on value co-creation: Open Source Business Review Online Journal.
6. Allen, S., Tanev, S., & Bailetti, T. (2009). Towards the development of a research methodology for studying the nature of value co-creation in internet driven businesses.
7. Alsos, G. A., Ljunggren, E., & Hytti, U. (2013). Gender and innovation: state of the art and a research agenda. *International Journal of Gender and Entrepreneurship*, 5(3), 236–256.
8. Andria, J., di Tollo, G., & Pesenti, R. (2020). A heuristic fuzzy algorithm for assessing and managing tourism sustainability. *Soft Computing*, 24, 4027–4040.
9. Andria, J., di Tollo, G., & Pesenti, R. (2021). Fuzzy multi-criteria decision-making: An entropy-based approach to assess tourism sustainability. *Tourism Economics*, 27(1), 168–186.
10. Angelini, E., di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *The quarterly review of economics and finance*, 48(4), 733–755.
11. Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R & D Management*, 50(3), 329–351.
12. Asensio-López, D., Cabeza-García, L., & González-Álvarez, N. (2019). Corporate governance and innovation: a theoretical review. *European Journal of Management and Business Economics*, 28(3), 266–284.
13. Ballantyne, D. (2004). Dialogue and its role in the development of relationship specific knowledge. *Journal of Business & Industrial Marketing*.
14. Basso, A., & di Tollo, G. (2021). Prediction of UK research excellence framework assessment by the departmental h-index. *European Journal of Operational Research*.

15. Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011). Women in stem: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, (04–11).
16. Belitz, H., & Lejpras, A. (2015). Financing patterns of R &D in small and medium-sized enterprises and the perception of innovation barriers in Germany. *Science and Public Policy*, 43(2), 245–261. 07.
17. Bendell, B. L., & Nesij, H. M. (2020). Does stakeholder engagement through corporate social and environmental behaviors affect innovation? *Journal of Business Research*, 119(C), 685–696.
18. Bengio, Y. (2012). *Practical Recommendations for Gradient-Based Training of Deep Architectures* (pp. 437–478). Berlin, Heidelberg: Springer.
19. Birattari, M., Yuan, Z., Balaprakash, P., & Stützle, T. (2010). *F-Race and Iterated F-Race: An Overview* (pp. 311–336). Berlin, Heidelberg: Springer.
20. Blake, M. K., & Hanson, S. (2005). Rethinking innovation: Context and gender. *Environment and Planning A: Economy and Space*, 37(4), 681–701.
21. Boateng, H., & Abdul-Hamid, I. K. (2017). An evaluation of corporate social responsibility communication on the websites of telecommunication companies operating in ghana: Impression management perspectives. *Journal of Information, Communication and Ethics in Society*, 15(1), 17–31.
22. Bosma, N., Hill, S., Ionescu-Somers, A., Kelley, D., Guerrero, M., & Schott, T. (2021). Global entrepreneurship monitor: 2020/2021 global report. Technical report, Global entrepreneurship monitor. Accessed 15 May 2021.
23. Card, D., & Payne, A. A. (2021). High school choices and the gender gap in stem. *Economic Inquiry*, 59(1), 9–28.
24. Carr, P. B., & Steele, C. M. (2010). Stereotype threat affects financial decision making. *Psychological Science*, 21(10), 1411–1416.
25. Carter, S. L., & Shaw, E. (2006). Women's business ownership : recent research and policy developments.
26. Cebi, S. (2013). Determining importance degrees of website design parameters based on interactions and types of websites. *Decision Support Systems*, 54(2), 1030–1043.
27. Chen, J., Leung, W. S., & Evans, K. P. (2018). Female board representation, corporate innovation and firm performance. *Journal of Empirical Finance*, 48, 236–254.
28. Cooper, R. (2012). The gender gap in union leadership in australia: A qualitative study. *Journal of Industrial Relations*, 54(2), 131–146.
29. Corazza, M., De March, D., & di Tollo, G. (2021). Design of adaptive elman networks for credit risk assessment. *Quantitative Finance*, 21(2), 323–340.
30. Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448–74.
31. Dahlstrand, Å. L., & Politis, D. (2013). Women business ventures in swedish university incubators. *International Journal of Gender and Entrepreneurship*, 5(1), 78–96.
32. Daily, C. M., & Dalton, D. R. (2003). Women in the boardroom: a business imperative. *Journal of Business Strategy*, 24(5), 747–763.
33. Dautzenberg, K. (2006). Gender differences of business owners in technology-based firms. *International Journal of Gender and Entrepreneurship*, 4(1), 79–98.
34. di Tollo, G. (2005). *Reti neurali e rischio di credito: stato dell'arte e analisi sperimentale*. Technical Report R-2005-003, Dipartimento di Scienze, Università "G. D'Annunzio" Chieti–Pescara.
35. di Tollo, G., Andria, J., & Tanev, S. (2021). *Neural Networks to Determine the Relationships Between Business Innovation and Gender Aspects*. Cham: Springer.
36. di Tollo, G., Stoyan, T., De March, D., & Zheng, M. (2012). Neural networks to model the innovativeness perception of co-creative firms. *Expert Systems with Applications*, 39(16), 12719–12726.
37. di Tollo, G., Stoyan, T., Slim, K. M., & De March, D. (2014). Determining the relationship between co-creation and innovation by neural networks. In M. Faggini & A. Parziale (Eds.), *Complexity in Economics: Cutting Edge Research* (pp. 49–62). Cham: Springer International Publishing.
38. di Tollo, G., & Tanev, S. (2010). *A neural network approach for the evaluation of the innovation outcomes of value co-creation practices in technology-driven firms*. Working paper, University of Southern Denmark.

39. di Tollo, G., Tanev, S., Liotta, G., & De March, D. (2015). Using online textual data, principal component analysis and artificial neural networks to study business and innovation practices in technology-driven firms. *Computers in Industry*, 74, 16–28.
40. Dobrev, D. (2012). A definition of artificial intelligence. arXiv.
41. Doran, J., & Ryan, G. (2012). Regulation and firm perception, eco-innovation and firm performance. *European Journal of Innovation Management*, 15(4), 421–441.
42. Ebert, C. (2007). Open source drives innovation. *IEEE Softw.*, 24(3), 105–109.
43. Ellemers, N. (2018). Gender stereotypes. *Annual review of psychology*, 69, 275–298.
44. Etgar, M. (2008). A descriptive model of the consumer co-production process. *Journal of the academy of marketing science*, 36(1), 97–108.
45. Ferrier, W. J. (2001). Navigating the competitive landscape: The drivers and consequences of competitive aggressiveness. *The Academy of Management Journal*, 44(4), 858–877.
46. Galia, F., & Zenou, E. (2012). Board composition and forms of innovation: does diversity make a difference? *European Journal of International Management*, 6(6), 630–650. PMID: 50425.
47. Griffin, D., Li, K., & Xu, T. (2021). Board gender diversity and corporate innovation: International evidence. *Journal of Financial and Quantitative Analysis*, 56(1), 123–154.
48. Harding, S. (1986). The science question in feminism. *Bulletin of Science, Technology & Society*, 6(4), 400–400.
49. Haykin, S. (2009). *Neural networks and learning machines*. Pearson Education, Upper Saddle River: NJ, third edition.
50. Hicks, D., Libaers, D., & Porter, A. L. (December, 2006). Identification of the technology commercialization strategies of high-tech small firms. *Small Business Research Summary*.
51. Hillman, A. J., Cannella, A. A., Jr., & Harris, I. C. (2002). Women and racial minorities in the boardroom: How do directors differ? *Journal of Management*, 28(6), 747–763.
52. Højgaard, L. (2002). Tracing differentiation in gendered leadership: An analysis of differences in gender composition in top management in business, politics and the civil service. *Gender, Work & Organization*, 9(1), 15–38.
53. Hölzl, W., & Janger, J. (2014). Distance to the frontier and the perception of innovation barriers across European countries. *Research Policy*, 43(4), 707–725.
54. Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.
55. Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. In: *Ldv Forum*, (vol. 20, pp. 19–62). Citeseer.
56. Huse, M. (2007). *Boards, Governance and Value Creation: The Human Side of Corporate Governance*. New Zealand: Cambridge University Press.
57. Kahn, S., & Ginther, D. (June 2017). Women and stem. Working Paper 23525, National Bureau of Economic Research.
58. Kang, H., Cheng, M., & Gray, S. J. (2007). Corporate governance and board composition: diversity and independence of australian boards. *Corporate Governance: An International Review*, 15(2), 194–207.
59. Karakaş, A., Bilgin, Y., & Raşit Yildiz, M. (October 2019) Impact of Managers' Innovation Perception on Innovation Activities and Innovation Strategies in Hotel Businesses. In: Mehmet Huseyin Bilgin, Hakan Danis, Ender Demir, and Ugur Can, editors, *Eurasian Economic Perspectives*, Eurasian Studies in Business and Economics, (pp. 55–70). Springer.
60. Kartal, S., & Bektas, Ø. (2020). Being an innovative teacher in 21th century: Studying innovative perception. *International Journal of Education Technology and Scientific Researches*, 5(11), 423–504.
61. Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications*, 37(9), 6233–6239.
62. Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. (1998). *Optimization by Simulated Annealing*. Cambridge, MA, USA: MIT Press.
63. Koellinger, P., Minniti, M., & Schade, C. (2013). Gender differences in entrepreneurial propensity. *Oxford Bulletin of Economics and Statistics*, 75(2), 213–234.
64. Korkman, O. (2006). *Customer value formation in practice: a practice-theoretical approach*. Svenska handelshögskolan.
65. Kristensson, P., Matthing, J., & Johansson, N. (2008). Key strategies for the successful involvement of customers in the co-creation of new technology-based services. *International journal of service industry management*.

66. Kunz, W., Schmitt, B., & Meyer, A. (2011). How does perceived firm innovativeness affect the consumer? *Journal of Business Research*, 64(8), 816–822.
67. Lee, E. M., Temel, S., & Uz Kurt, C. (2016). The effect of consumers' innovation perception on internet usage behaviors. *International Journal of Innovation Science*, 8(2), 100–112.
68. Linäker, J., Munir, H., Regnell, B., & Schrewelius, C. A survey on the perception of innovation in a large product-focused software organization. In J. Fernandes, R. Machado, and K. Wnuk, editors, *LSoftware Business. ICSOB 2015*, volume 210 of *Lecture Notes in Business Information Processing*. Springer, Cham.
69. Liu, Y., Wei, Z., & Xie, F. (2014). Do women directors improve firm performance in china? *Journal of Corporate Finance*, 28, 169–184. Inside the Board Room.
70. Lorber, J. (2018). The social construction of gender. In: *Inequality in the 21st Century*, (pp. 347–352). Routledge.
71. Lowe, B., & Alpert, F. (2015). Forecasting consumer perception of innovativeness. *Technovation*, 45–46, 1–14.
72. Mahadeo, J. D., Soobaroyen, T., & Hanuman, V. O. (2012). Board composition and financial performance: Uncovering the effects of diversity in an emerging economy. *Journal of Business Ethics*, 105(3), 375–388.
73. De March, D., di Tollo, G., & Tanev, S. (2012). Using online innovation metrics to cluster co-creative firms by soms. In: *ISPIM Conference Proceedings, page 1. The International Society for Professional Innovation Management (ISPIM)*.
74. Marlow, S., & McAdam, M. (2012). Analyzing the influence of gender upon high-technology venturing within the context of business incubation. *Entrepreneurship Theory and Practice*, 36(4), 655–676.
75. Marlow, S., & McAdam, M. (2013). Gender and entrepreneurship: Advancing debate and challenging myths; exploring the mystery of the under-performing female entrepreneur. *International Journal of Entrepreneurial Behavior & Research*, 19(1), 114–124.
76. Martin, L., Jerrard, B., & Wright, L. (2020). Identity work in female-led creative businesses. *Gender, Work & Organization*, 27(3), 310–326.
77. Martino, R., Ventre, V., & di Tollo, G. (2023). Analytic hierarchy process for classes of economic behavior in the context of intertemporal choices. *Ratio Mathematica*, 47.
78. Milliken, F. J., & Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *The Academy of Management Review*, 21(2), 402–433.
79. Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press.
80. Mumford, M. D., Zaccaro, S. J., Connelly, M. S., & Marks, M. A. (2000). Leadership skills: Conclusions and future directions. *The Leadership Quarterly*, 11(1), 155–170.
81. Nählinder, J., Tillmar, M., & Wigren, C. (2012). Are female and male entrepreneurs equally innovative? reducing the genderbias of operationalizations and industries studied. In S. Andersson, K. Berglund, J. G. Torslund, E. Gunnarsson, & E. Sundin (Eds.), *Promoting Innovation - Policies, Practices and Procedures*. Stockholm: VINNOVA.
82. Nambisan, S., & Baron, R. A. (2009). Virtual customer environments: testing a model of voluntary participation in value co-creation activities. *Journal of product innovation management*, 26(4), 388–406.
83. Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management: Reinventing innovation management research in a digital world. *MIS Quarterly*, 41(1), 223–238.
84. Nielsen, S., & Huse, M. (2010). Women directors' contribution to board decision-making and strategic involvement: The role of equality perception. *European Management Review*, 7, 03.
85. Nissan, E., Carrasco, I., & Castaño, M. S. (2012). Women entrepreneurship, innovation, and internationalization. In: *Women's entrepreneurship and economics*, (pp. 125–142). Springer.
86. Nissan, Edward, Carrasco, Inmaculada, & Castaño, María-Soledad. (2012). *Women Entrepreneurship, Innovation, and Internationalization* (pp. 125–142). New York: Springer.
87. Normann, R. (2001). *Reframing business: When the map changes the landscape*. USA: John Wiley & Sons.
88. Oktem, C., Pederson, K., & Kanazawa, M. Four ways boards are driving innovation in a volatile environment.
89. Østergaard, C. R., Timmermans, B., & Kristinsson, K. (2011). Does a different view create something new? the effect of employee diversity on innovation. *Research Policy*, 40(3), 500–509.

90. Payne, A. F., Storbacka, K., & Frow, P. (2008). Managing the co-creation of value. *Journal of the academy of marketing science*, 36(1), 83–96.
91. Pénin, J. (2011). Open source innovation: Towards a generalization of the open source model beyond software. *Revue d'économie industrielle*, 136, 65–88.
92. Pettersson, K., & Lindberg, M. (2013). Paradoxical spaces of feminist resistance: Mapping the margin to the masculinist innovation discourse. *International Journal of Gender and Entrepreneurship*, 5(3), 323–341.
93. Poutanen, S., & Kovalainen, A. (2017). *Gender and Innovation in the New Economy: Women. Identity and Creative Work*: Palgrave Macmillan.
94. Prahalad, C. K., & Krishnan, M. S. (2008). *The new age of innovation*. Tata McGraw-Hill Education.
95. Prahalad, C. K., & Ramaswamy, V. (2000). Co-opting customer competence. *Harvard business review*, 78(1), 79–90.
96. Prahalad, C. K., & Ramaswamy, V. (2003). The new frontier of experience innovation. *MIT Sloan management review*, 44(4), 12.
97. Ranjan, K. R., & Read, S. (2016). Value co-creation: concept and measurement. *Journal of the Academy of Marketing Science*, 44(3), 290–315.
98. Reed, R. D., & Marks, R. J. (1998). *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. Cambridge, MA, USA: MIT Press.
99. Rindfleisch, A., O'Hern, M., & Sachdev, V. (2017). The digital revolution, 3d printing, and innovation as data. *Journal of Product Innovation Management*, 34(5), 681–690.
100. Rosa, P., & Dawson, A. (2006). Gender and the commercialization of university science: academic founders of spinout companies. *Entrepreneurship & Regional Development*, 18(4), 341–366.
101. Sawhney, M., Verona, G., & Prandelli, E. (2005). Collaborating to create: The internet as a platform for customer engagement in product innovation. *Journal of interactive marketing*, 19(4), 4–17.
102. Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27, 623–656.
103. Sierra-Morán, J., Cabeza-García, L., González-Álvarez, N., & Botella, J. The board of directors and firm innovation: A meta-analytical review. *BRQ Business Research Quarterly*, 0(0), 23409444211039856, 0.
104. Sorescu, A. (2017). Data-driven business model innovation. *Journal of Product Innovation Management*, 34(5), 691–696.
105. Tanev, S., Bailetti, T., Allen, S., Milyakov, H., Durchev, P., & Ruskov, P. (2011). How do value co-creation activities relate to the perception of firms' innovativeness? *Journal of Innovation Economics*, 1(7), 131–159.
106. Torchia, M., Calabró, A., & Huse, M. (2011). Women directors on corporate boards: From tokenism to critical mass. *Journal of Business Ethics*, 102(2), 299–317.
107. Toyah, M., & Del Carmen Triana, M. (2009). Demographic diversity in the boardroom: Mediators of the board diversity-firm performance relationship. *Journal of Management Studies*, 46(5), 755–786.
108. Vargo, S. L., & Lusch, R. F. (2004). The four service marketing myths: remnants of a goods-based, manufacturing model. *Journal of service research*, 6(4), 324–335.
109. Vargo, S. L., & Lusch, R. F. (2006). *Service-dominant logic*. The service-dominant logic of marketing: Dialog, debate, and directions.
110. von Hippel, E. (February 2003). *Open Source Projects as Horizontal Innovation Networks - By and for users*. Working papers 4366-02, Massachusetts Institute of Technology (MIT), Sloan School of Management.
111. Wajcman, J. (2009). Feminist theories of technology. *Cambridge Journal of Economics*, 34(1), 143–152. 01.
112. Watson, J., & Newby, R. (2005). Biological sex, stereotypical sex-roles, and sme owner characteristics. *International Journal of Entrepreneurial Behavior & Research*.
113. Werbos, P. (1994). *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*. USA: Wiley-Interscience.
114. Witt, M. G., & Wood, W. (2010). Self-regulation of gendered behavior in everyday life. *Sex Roles*, 62(9–10), 635–646.
115. Wu, Q., Dbouk, W., Hasan, I., Kobeissi, N., & Zheng, L. (2021). Does gender affect innovation? evidence from female chief technology officers. *Research Policy*, 50(9), 104327.

116. Zhang, X., & Chen, R. (2008). Examining the mechanism of the value co-creation with customers. *International Journal of Production Economics*, 116(2), 242–250.
117. Xie, Luqun, Zhou, Jieyu, Zong, Qingqing, & Qian, Lu. (2020). Gender diversity in r & d teams and innovation efficiency: Role of the innovation context. *Research Policy*, 49(1), 103885.
118. Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2001). Team leadership. *The leadership quarterly*, 12(4), 451–483.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.