Rijksdienst voor Ondernemend Nederland

dialogic
innovatie • interactie

Orchestrating public-private R&D networks

government-affiliated intermediary organizations as a policy intervention

by Maya J. Bogers

M.J.Bogers@students.uu.nl
student no: 3516903

Supervised by
Dr. ir. Matthijs J. Janssen
M.J.Janssen@uu.nl

Second read by
Dr. Iris Wanzenböck
I.Wanzenbock@uu.nl

Master thesis, MSc Innovation Sciences
Copernicus Institute of Sustainable Development
Department Innovation, Environment and Energy Sciences
Faculty of Geosciences, Utrecht University

Utrecht, November 1st, 2017

Universiteit Utrecht

This paper was written to obtain the degree Master of Science in Innovation Sciences from Utrecht University.

Summary

We investigate the role of government-affiliated intermediary organizations (GAIO) as a policy intervention for 'orchestrating' the Dutch public-private research collaboration network in the years 2013-2016. GAIO are type of innovation intermediaries of which very little is known in literature, despite their assumed relevance in facilitating innovation. Our key hypothesis is that GAIO do not only directly stimulate new partnership formation, but also alter the natural collaboration tendencies of firms as described by the proximities theory (Boschma, 2005). We test this by analysing an unexplored database of public-private R&D collaborations in the Netherlands, where in 2013 the Topconsortia for Knowledge and Innovation (TKI) have been implemented as industry-specific GAIO to facilitate the formation of public-private research consortia. The TKI were introduced as part of the national innovation policy, the "Topsector approach".

Results indicate that firms who are both members of the same GAIO at t_0 have an up to three times higher probability to form a new partnership together at t_1 than firms who are members of different GAIO at t_0. Cognitive proximity is a significant influencer of new partnership formation, but only when firms do not share membership of a GAIO. This indicates that GAIO are able to overcome the hurdle that cognitive distance poses for new collaborations to arise, allowing for knowledge recombination to occur over larger cognitive distances. Contrary to previous empirical evidence, social proximity has shown to be more relevant within subgroups of GAIO members than between them. This indicates that GAIO exacerbate the natural tendencies of firms to act locally in their embedded networks and preferentially collaborate with their partners' partners. Lastly, an important finding is that GAIO mostly facilitate large firms to form new partnerships, preferentially with each other and to a lesser extent also with SME. Collaboration amongst SME is also enhanced, but to a lesser extent.

This study provides the first quantitative empirical evidence on GAIO, providing a stronger foundation for their role. From a policy perspective, it appears that GAIO can play an important role in connecting firms over relatively large cognitive distance, that were otherwise unlikely to collaborate. At the same time however, the GAIO are also constrained to some extent by the borders of the specific industrial sector they were assigned to.

Acknowledgements

First and foremost I would like to thank Matthijs Janssen for his supervision. Matthijs, thank you for your time and input, your fast feedback, your high level of involvement (thinking along before I even officially started!) and for enabling me to make the most of this thesis.

I'd also like to thank Iris Wänzenbock for being my second reader. Iris, thank you for providing me with critical feedback to my research proposal, and thank you for taking the time to read these 80 pages and attending my presentation.

A special thanks goes to Jarno Hoekman, for taking out the time to provide me with some feedback on my methodology.

A big, big thank you to Arjan Wolters and Youri van Leeuwen, who helped me unravel a database, gave me the specifics on the TKI policy, introduced me to many people at RVO, and provided me the opportunity to present my results to a larger audience. Arjan and Youri, thank you.

Last but not least, my sincerest gratitude goes out to the two organizations that enabled me to do this research: the Dutch Enterprise Agency (Rijksdienst voor Ondernemend Nederland) and Dialogic Innovatie & Interactie. Thank you both for coming together and providing me with not only all the resources I needed, but also a very pleasant work environment with wonderful colleagues over the course of eight months.

Table of contents

Summary 2
Acknowledgements 3
Table of contents 4
1. Introduction 6
2. Theory 9
2.1. The case for policy to steer R&D collaboration 9
2.2. Government-affiliated intermediary organizations 10
2.2.1. Definition 10
2.2.2. Functions 11
2.3. Determinants of R&D collaboration and the role of GAIO 12
2.3.1. Cognitive proximity 12
2.3.2. Social proximity 13
2.3.3. Geographical proximity 13
2.3.4. Organizational proximity 14
2.3.5. Institutional proximity 14
2.3.6. Government-affiliated intermediary organizations 14
2.4. Conceptual model 17
3. Data and Methodology 18
3.1. Research design 18
3.2. Case description 18
3.3. Data description 19
3.4. Constructing variables 23
3.4.1. Dependent variable 23
3.4.2. Independent variables 23
3.4.3. Control variables 26
3.5. Statistical model 26
3.5.1. Logit function 26
3.5.2. Quality of the model 28
3.5.3. Rare events data 28
3.5.4. Interaction effects in logit models 29
3.5.5. Background analyses 30

4. Results 32
4.1. Descriptives 32
4.2. Stylized facts 34
4.3. Split models 35
4.4. AME models 38
4.4.1. How to interpret AMEs 40
4.4.2. AME models 2 and 3 41
4.4.3. AME Results Interaction Models 43
4.4.4. Effect Plot 47
5. Conclusion 49
6. Discussion 53
6.1. GAIO as a policy tool 53
6.2. Academic contribution 57
6.3. Shortcomings and future research 58
References 60
Appendices 65
Appendix A 66
Appendix B 70
Appendix C 72
Appendix D 74
Appendix E 76
Appendix F 80

1. Introduction

As early as 1934, Schumpeter posed the term “Neue Kombinationen”, recognizing the importance of combining existing knowledge and resources to develop new technologies and products (Schumpeter, 1934). Increasingly, firms rely on external partners to access new knowledge bases, forming large interorganizational knowledge networks. External knowledge has become essential for innovators. Firms profit not only from their direct partners, but from the knowledge available throughout the network as a whole (Powell et al., 1996; Ahuja, 2000; Rigby and Zook, 2002; Rycroft and Kash, 2004; Schilling and Phelps, 2007).

From a societal perspective, collaboration between firms has several benefits. Firstly, collaborations are desirable as especially breakthrough innovations, those of which we expect high economic and societal value, require the recombination of specialized knowledge (Ahuja and Lampert, 2001; Fleming, 2001). Secondly, the knowledge networks that form as a result of interfirm collaborations are expected to generate positive externalities, or knowledge-spillovers, that have a positive effect on innovation and economic growth in the entire region or country (Cohen, 2006).

However, not every collaboration is equally valuable. To create ‘recombinant growth’, firms must access a knowledge base that complements their own (Weitzman, 1998; Cassiman and Veugelers, 2006). Knowledge is complementary when it is both understandable and provides novelty value. In other words, organizations that access each other’s knowledge bases must have the absorptive capacity to effectively recognize, absorb and use each other’s knowledge, whilst still have enough cognitive distance to learn something new (Cohen and Levinthal, 1990; Nooteboom et al., 2007). This concept of combining different, but related, knowledge as a driver of (breakthrough) innovation and economic value has been proven to hold empirically, on the level of individual inventors (Kaplan and Vakili, 2015), technologies (Arts and Veugelers, 2014), firms (Tödtling and Grillitsch, 2015) and even regions (Frenken et al., 2007).

The search for partners who provide such complementary knowledge is not easy. It has been posed that due to limited cognitive capabilities, it may be difficult for decision-makers to identify potentially valuable knowledge (Nooteboom, 2000). Indeed firms tend to collaborate with partners that have a similar knowledge base (Scherngell and Barber, 2009). Moreover, they tend to rely on familiarity and trust in choosing their partners, leading them to connect “locally” within their embedded networks (Gulati, 1995; Rycroft and Kash, 2004; Baum et al., 2010). Due to these constraints, recombination of knowledge often occurs with concepts that were already familiar (Fleming, 2001).

From the perspective of the national innovation systems literature, this myopic behaviour of firms can be considered a form of system failure resulting from interaction failure (a suboptimal number of interactions between actors in the network) or infrastructure failure (actors unaware of external sources of knowledge) (Klein Woolthuis et al., 2005; Russo et al., 2016)[1] Such a system failure is a legitimization for policy intervention.

One particular way in which governments may aim to address this system failure is by steering the knowledge network(s) in their country to achieve the 'best' collaborations from a societal and economic perspective. Recently, such conscious 'orchestration' of knowledge networks has gained increased attention in literature (Dhanaraj and Parkhe, 2006; Hurmelinna-Laukkanen et al., 2012). An important role is ascribed to innovation intermediaries, which have been defined as *"an organization or body that acts as an agent or broker in any aspect of the innovation process between two or more parties"* (Howells, 2006). More specifically, there are intermediaries that operate at a high systemic level, e.g. sectors or countries, who deal with complex networks and problems and are important in facilitating and coordinating efforts for long-term change (van Lente et al., 2003; Kilelu et al., 2011; Hannon et al., 2014). Such systemic innovation intermediaries (SII) may aid in connecting firms in networks that were otherwise 'unlikely' to collaborate, spanning structural holes within networks (Howells, 2006; Kilelu et al., 2011; Abbate et al., 2013; Hannon et al., 2014).

Despite the widespread recognition that innovation on a national level requires broad systemic support, research on SSI at the national or sectoral level has received little attention, both in academics and in policy. Empirical evidence for their role is minimal (Dalziel, 2010; Abbate et al., 2013; Levén et al., 2014), though important qualitative contributions have been made on the role of SII in agriculture (Klerkx and Leeuwis, 2009/7, 2008; Kilelu et al., 2011), energy (Hannon et al., 2014; Kivimaa, 2014) and eco-innovation (Kanda et al., 2015). All of these studies endorse the importance of the 'network building' and/or brokerage' function of SII to connect actors in their network. However, as far as the author is aware, no study exists that assesses quantitatively the extent to which SII influence firms' collaboration choices in knowledge networks.

Against this backdrop, this paper aims to contribute to the literature by conducting a quantitative study of SII. The focus will be specifically on a set of government-owned SII, which have been termed "government-affiliated intermediary organizations" (GAIO) in previous literature (Kivimaa, 2014). We assess to what extent these GAIO influence the formation of new partnerships between firms, and to what extent GAIO influence the natural collaboration tendencies of firms as described by the proximities theory (Boschma, 2005). We focus specifically on interfirm collaborations. This brings us to the following research question:

To what extent do government-affiliated intermediary organizations influence the formation of new partnerships between firms, and to what extent do they alter the influence of different types of proximity on new partnership formation?

In 2012, the Dutch government implemented specialized entities, the Topconsortia for Knowledge and Innovation (TKI), to orchestrate the national public private R&D network.

[1] From the perspective of neoclassical economics, this may also be considered a market failure, resulting from information asymmetry or coordination failure.

These TKI are part of a larger innovation policy, the “Topsector approach”, which is seen as a leading example of modern industrial policy (Warwick and Nolan, 2014). The TKI are concerned with the formation of public private R&D consortia and fulfil a variety of functions that are characteristic of SII. As such, TKI can be considered GAIO. Using data from the TKI, this paper will assess the influence of Boschma’s proximities on new interfirm partnerships in public-private R&D consortia, and whether these collaboration choices are altered when firms have the possibility to connect through a GAIO. One could argue that GAIO only fulfill their function properly when they at least to some extent influence the natural tendencies of firms to collaborate, be it reinforcing existing powers at work, or reducing the influence of determinants, thereby allowing firms to ‘broaden their horizon’ with regards to potential collaboration partners.

The contribution of this study is two-fold. Firstly, a contribution will be made to the literature on SII, as GAIO are in essence a sub-type of SII. As mentioned, empirical evidence for the role of SII is lacking and this study aims to shed more light on one important function of SII: network formation. Specifically, this study draws from two theoretical fields, namely that of R&D collaboration choices based on proximities theory, and the SII literature, to provide insights on the role of GAIO in R&D partnership formation. Secondly, by assessing a case where GAIO are implemented as a policy intervention, this study may result in empirical evidence with regards to the effectiveness of GAIO in steering a national public-private R&D network. This can provide important insights for policy-makers who aim to stimulate networked forms of R&D.

The remainder of this paper is organized as follows. First a theory section will describe relevant background literature on government-affiliated intermediary organizations and propose hypotheses on interfirm collaboration within this context, and propose a conceptual model by which to test these hypotheses. The next section will be devoted to describing the data and some background to the TKI, the GAIO under investigation. Subsequently, the empirical methodology will be explained. Then follows a results section, and finally a conclusion and discussion.

2. Theory

This section will first focus on why governments create policy for R&D collaborations. Secondly, we focus on government-affiliated intermediary organizations: what are they, and what do they do? Thereafter, we look at what drives firms to collaborate, following the proximities theory by Boschma (2005). In this section, we also examine how GAIO may influence these determinants of collaboration.

2.1. The case for policy to steer R&D collaboration

The main rationale for governments to stimulate R&D collaboration is based on two key elements. First, there is a market failure that prevents firms from investing in R&D in general, resulting from the high risk and costs, and low appropriability associated with R&D. This is especially true for basic research, or R&D that is committed to the 'public good' (Knockaert et al., 2014/1; Cohen, 2006; Feldman and Kelley, 2006). Moreover, increasing pressure on firms to deliver measurable results, makes the reluctance to invest in long-term knowledge development even greater (Dalziel, 2010). Thus, investment in R&D is suboptimal when left to the market, both from an economic and societal perspective. Governments can help overcome this market failure by providing incentives to invest in R&D, amongst others through pooling of public and private resources, and sharing risk, in public private R&D partnerships.

Secondly, knowledge production, diffusion and innovation is increasingly the result of R&D networks and the knowledge available throughout that network as a whole, rather than individual firms (Powell et al., 1996; Ahuja, 2000; Phelps et al., 2012). This is true especially in knowledge-based economies, where development of technology is complex and high-risk, and thus requires the exchange and recombination of resources and (complementary) knowledge (Rycroft and Kash, 2004; Cohen, 2006). Stimulation of R&D collaboration between firms can assist in creating or shaping such knowledge networks, which are then expected to generate knowledge-spillovers that have a positive effect on economic growth of the country as a whole (Cohen, 2006).

Unfortunately, governments too have limited resources. As such, the question arises where to focus (the gross of) the investments and efforts. Recently, several countries with knowledge-based economies, including The Netherlands, have adopted 'narrow' innovation policies, aimed at specific sectors or industries that are current strongholds (Cohen, 2006; Warwick and Nolan, 2014). The rationale of focussing on a select number of industries is that to sustain a competitive advantage, a hard-to-imitate, deep and specialized knowledge base is required (Porter, 1986). Such knowledge only builds up through a process of knowledge and experience accumulation (Asheim et al., 2011). Vertical policies thus build on local strengths, of which most economic growth is expected.

The question remains then, where the greatest potential for collaboration lies. Returning to the notion that breakthrough innovation is most likely to arise from recombination of different knowledge bases, it seems only smart to aspire to collaborations that bridge those. More specifically, a case is to be made to focus government efforts on "cross-over collaborations" between a country's stronghold industries, as in these industries, the country's most specialized, deep knowledge resides (Janssen, 2015; Frenken, 2017).

Thus, from a policy perspective, it is sensible to stimulate R&D collaborations between firms, and to orchestrate the resulting networks in such a way that new ties are formed to bring together firms with different knowledge bases, those that were otherwise unlikely to meet. This is where government-affiliated intermediary organizations come in.

2.2. Government-affiliated intermediary organizations

2.2.1. Definition

The literature on innovation intermediaries has been quite dispersed, and has emerged from different research fields. Several different terms are used in these fields that refer to the same entities. This makes a comprehensive definition complicated (Howells, 2006; Abbate et al., 2013). Howells defined innovation intermediaries as *"an organization or body that acts as an agent or broker in any aspect of the innovation process between two or more parties"* (Howells, 2006). In earlier literature, the focus was primarily on firms as central 'hubs' who shape and manage their own (R&D) network of partners, as a side-activity to their core business (Doz et al., 2000; Dhanaraj and Parkhe, 2006; Howells, 2006; Gassmann et al., 2011). In recent years however, the term innovation intermediary is often attributed to those organizations or entities that operate more independently than firms and exclusively focus on enabling other organizations to innovate, rather than being involved in the development and implementation of innovations themselves (Winch and Courtney, 2007; Batterink et al., 2010; Abbate et al., 2013).

From a systems of innovation perspective, the term "systemic innovation intermediary" (SII) has been posed to refer to those entities that operate at a higher systemic level, e.g. local clusters, regions or countries, and who deal with more complex networks and problems and/or transitions at the systemic level. SII often work as a nonprofit or public organization (Kilelu et al., 2011; van Lente et al., 2011; Hannon et al., 2014). As the latter is not always the case per se, Kivimaa (2014) added more clarity to the ownership of SII by describing government-affiliated intermediary organizations (GAIO) as a sub-type of SII that are *"quasi-autonomous government agencies, government-owned companies or government-initiated foundations, as they fall between traditional public and private sector actors" (Kivimaa, 2014)*.

Combining definitions from previous work, we define GAIO as: *"Quasi-autonomous government agencies, government-owned companies or government-initiated foundations that operate as entities at the interface between multiple innovation actors, working to facilitate and coordinate innovation activities at the system-level." (Howells, 2006; Kilelu et al., 2011; Hannon et al., 2014; Kivimaa, 2014)*.

It is these GAIO that this paper focuses on. However, as there is no literature on government-affiliated intermediary organizations yet, with the exception of the study by

Kivimaa (2014), this paper will focus on SII literature to build hypotheses and a conceptual framework.

2.2.2. Functions

To facilitate innovation between parties, SII perform a range of functions. These include, but are not limited to (van Lente et al., 2003; Dalziel, 2010; Kilelu et al., 2011):

- demand articulation and strategy development;
- network building and sustainment;
- knowledge brokerage;
- trust building amongst actors in the network;
- process management of long-term and/or complex innovation projects;
- organizing discourse, alignment and consensus;
- institutional support;
- creating conditions for learning by doing, using, interacting and searching;
- provision of tailor-made (strategic) information;
- R&D funding.

As SII are often active within a specific industry or sector, industry promotion may also be one of the functions that SII perform (van Lente et al., 2003; Winch and Courtney, 2007; Dalziel, 2010). SII do not perform R&D or innovate themselves, but enable others to innovate. They have often been founded especially to undertake the role of facilitator as their core business rather than as a byproduct of other activities (Winch and Courtney, 2007; Kilelu et al., 2011).

To perform their functions, SII typically position themselves at 'handover points' in the innovation system and they act as broker between the various parties (van Lente et al., 2003; Hannon et al., 2014). What these handover points are, may depend on the specific function of the SII. In this study, the focus is on GAIO that are involved with connecting firms within and across sectors with each other, and with public research organizations and knowledge institutes, with the goal of forming public-private R&D consortia. A visualization of their representative position is illustrated in Figure 1.

Networking is often considered an important function of SII (van Lente et al., 2011; Hannon et al., 2014). Indeed, in their case study on GAIO, Kivimaa et al. (2014) find that stakeholders assign high value to the role of GAIO in terms of network formation (Kivimaa, 2014). This study will further focus on this networking function of GAIO, more specifically how they establish new partnerships between firms.

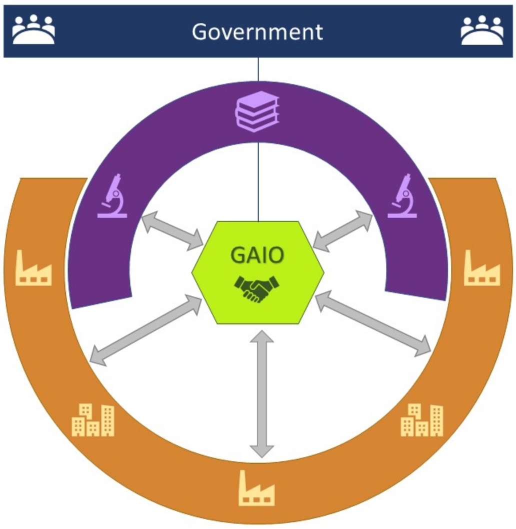


Figure 1: Position of government-affiliated intermediary organizations (GAIO) amongst other organisations or bodies. GAIO position themselves between and within government (blue), industry (orange) and public research organizations or knowledge institutes (purple). Note that they also play an explicit role in bringing firms together, and as such are also an intermediary within and between industry sectors.

2.3. Determinants of R&D collaboration and the role of GAIO

It becomes of interest then, what drives firms to collaborate. In essence, firms choose to collaborate based on the expected utility of collaboration. This utility can be derived from the direct partner by sharing of resources and risks (Williamson, 1981), access to a partner's unique resources, including indirect partners and a knowledge network, (Wernerfelt, 1984; Schilling and Phelps, 2007), or the opportunity to engage in organizational learning and joint knowledge production (Powell et al., 1996; Nooteboom, 2000; Graf, 2006). Regarding the latter, it has been posed that five proximities between agents facilitate the effective transfer of knowledge between agents: cognitive proximity, social proximity, geographical proximity, organizational proximity and institutional proximity (Boschma, 2005). Considering the notion that we are looking here at R&D collaborations, where knowledge generation is arguably the most important goal, we use these proximities, plus concepts from the SII literature, as a base for the conceptual model.

2.3.1. Cognitive proximity

In assessing a collaboration choice, the firm determines whether they are able to learn from the other party. At the same time, the firm must also have the absorptive capacity to be able to learn and to even recognize that there is an opportunity to collaborate (Cohen and Levinthal, 1990; Nooteboom, 2000; Nooteboom et al., 2007). Indeed, it has been shown empirically that firms are more likely to collaborate when cognitive proximity is higher (Autant-Bernard et al., 2007; Cantner and Meder, 2007). So despite the notion that high cognitive proximity may lead to a lesser learning experience, this seems to be ignored in making the initial partner choice. Following theory and empirical evidence, the following hypothesis is posed:

Hypothesis 1

> *The higher the cognitive proximity between any two firms, the more likely they are to form a partnership in a public private research consortium.*

2.3.2. Social proximity

In order to determine whether to collaborate with another party, a firm must first be aware of the existence of the other party, and be able to assess whether the partner is a good match. In other words, to determine the utility of a potential collaboration, firms must have a "window" on other actors' capabilities and assets. Information on potential partners may diffuse through 'prior acquaintances' (Barabasi and Albert, 1999), so this window of information tends to be rather local in the network (Rycroft and Kash, 2004). Moreover, a certain level of familiarity and trust is required for learning to occur (Boschma, 2005). Thus, firms may choose to repeat collaborations, or cooperate with indirect partners rather than "strangers" (Gulati, 1995; Baum et al., 2010). As a consequence, in order to assess the benefits of collaboration, not only individual determinants, but also the firm's network position must be taken into account as a determinant of cooperation (Bala and Goyal, 2000). This leads us to the following hypothesis:

Hypothesis 2

> *The higher the social proximity between any two firms, the more likely they are to form a partnership in a public private research consortium.*

2.3.3. Geographical proximity

It is generally agreed upon that knowledge production and spillovers are, at least to a large extent, geographically localised (Audretsch and Feldman, 1996; Feldman, 1999; Ponds et al., 2007). Short geographic distances between collaborating partners may increase knowledge transfer by frequent interaction and face-to-face contact, which in turn can facilitate trust (Boschma, 2005). Indeed, empirically, it has been shown that large geographical distances decrease the likeliness (Paier and Scherngell, 2011) or intensity of collaboration (Ponds et al., 2007; Hoekman et al., 2009). On the other hand, it has been found that geographical effects may not exist within certain contexts, where other proximities do play a role (Autant-Bernard et al., 2007). This is in line with theory posed by Boschma, posing that geographical proximity "is neither necessary nor sufficient" for interactive learning to take place, as other forms of proximity may compensate for a lack of geographical proximity, and the other way around: geographical proximity may compensate for a lack of other proximities (Boschma, 2005).

Despite the notion that this paper assesses a national knowledge network, influenced by national innovation policy, and the relatively small size of The Netherlands as a country, we pose that geographical proximity may play a role due to the localization of knowledge clusters. As such, we hypothesize the following:

Hypothesis 3

> *The higher the geographical proximity between any two firms, the more likely they are to form a partnership in a public private research consortium.*

2.3.4. Organizational proximity

There is some ambiguity as to what the concept of organizational proximity exactly entails (Knoben and Oerlemans, 2006). According to Boschma (2005), organizational proximity refers to the rate of autonomy and control that can be exerted in organizational arrangements, which has to do with hierarchy in the governance structure or arrangement (Boschma, 2005). Within the context of one policy measure though, the governance structure is not expected to differ. At the dyadic level, it has been posed that organizational proximity refers to whether two firms share a similarity in 'organizational context' in which they operate (Torre and Rallet, 2005; Knoben and Oerlemans, 2006). This context is created by explicit and implicit rules and routines within and surrounding the organization. These rules and routines influence the interactions between actors. Similarity in these rules and routines, or business reality in which they operate, can either facilitate or hamper interaction and thereby the ability to collaborate (Torre and Rallet, 2005; Knoben and Oerlemans, 2006). We pose the following hypothesis:

Hypothesis 4

> *The higher the organizational proximity between any two firms, the more likely they are to form a partnership in a public private research consortium.*

2.3.5. Institutional proximity

Institutional proximity can be defined as the extent to which "actors [share] the same institutional rules of the game, as well as a set of cultural habits and values" (Boschma, 2005). Sometimes this is referred to in practical sense as agents being in the same country, and other times it is referred to as agents being in the same institutional form, e.g. public or private organization, university, etc (Ponds et al., 2007; Balland, 2012). Considering the specific focus of this paper is on collaborations between firms within a single country only, all firms share the same legislation, cultural habits and values, and institutional form. As such, institutional proximity does not differ for the actors in our set and therefore it is disregarded as a determinant of collaboration in this study.

2.3.6. Government-affiliated intermediary organizations

Now we turn our attention to GAIO as a determinant of R&D collaboration. As far as the author is aware, the influence of GAIO, or SII, on collaboration choices or knowledge networks has not been tested empirically so far (Dalziel, 2010; Abbate et al., 2013; Levén et al., 2014). As such, we base our hypotheses on concepts from the SII literature, rather than empirical evidence.

Based on what we know of GAIO so far regarding the functions such as networking and goal alignment, it can be assumed that GAIO promote collaboration between the firms that are somehow affiliated to them, hereafter referred to as their ‘members’. Moreover, as SII are often active within a specific industry or sector, it can be expected that their members are also concerned more or less with the same topics and interests. Thus, it can be expected that firms that share membership of the same GAIO are more likely to collaborate than those who do not share a membership, both through a ‘selection’ effect and the broker effect of the intermediaries. This leads to the first hypothesis:

Hypothesis 5

> *When two firms are both members of the same GAIO, they are more likely to form a partnership in a public private research consortium than when they are not.*

In section 2.3.1. it was discussed that firms must have a certain level of absorptive capacity to even recognize that there is an opportunity to collaborate (Cohen and Levinthal, 1990; Nooteboom, 2000; Nooteboom et al., 2007). It has been posed that intermediaries are able to establish communication between parties, making them aware of their matching goals (Backhaus, 2010). Mahnke et al. (2008) perform a single case-study on a for-profit intermediary and conclude that, based on their preliminary evidence, intermediaries are indeed able to bridge cognitive distance between agents (Mahnke et al., 2008). As such, we hypothesize that GAIO may help overcome the hurdle that high cognitive distance poses for collaboration.

Hypothesis 1a

> *When two firms are both members of the same GAIO, the positive effect of cognitive proximity on partnership formation is reduced (negative moderation of hypothesis 1).*

As mentioned in section 2.2.2., network building is often considered a key task of SII (van Lente et al., 2003; Abbate et al., 2013). It has been noted that intermediaries spur innovation by accessing and brokering direct linkages across disciplinary boundaries. They promote knowledge sharing amongst sets of actors that would normally not interact or collaborate (Abbate et al., 2013). Indeed, they play a role similar to that of a knowledge broker in a network, spanning structural holes between local clusters (Burt, 2004). This function of SII can help firms to gain a “window” on potential partners’ capabilities and assets, and can help build trust between firms, increasing the likeliness of a collaboration occurring. This brings us to the following hypothesis:

Hypothesis 2a

> *When two firms are both members of the same GAIO, the positive effect of social proximity on partnership formation is reduced (negative moderation of hypothesis 2).*

However, an opposite effect may also exist. As SII are also involved in demand articulation (van Lente et al., 2003; Abbate et al., 2013), it may occur that the SII create a positive feedback-loop of articulating goals and interests of a select group of firms and subsequently, bringing these exact same firms together in collaborations. It has been noted that in demand-driven policy initiatives, a relatively small range of actors is included based on pre-existing relations. This leads to a select group of organizations benefitting from policy (Fromhold-Eisebith and Eisebith, 2005). As such, the following hypothesis is also investigated.

Hypothesis 2b

When two firms are both members of the same GAIO, the positive effect of social proximity on partnership formation is increased (positive moderation of hypothesis 2).

Considering the notion that GAIO operate their networking functions at a high systemic level, in this case at the national sectoral level, we propose that GAIO are able to overcome the negative influence of geographical distance on collaboration. As such, we state the following hypothesis:

Hypothesis 3a

When two firms are both members of the same GAIO, the positive effect of geographical proximity on partnership formation is reduced (negative moderation of hypothesis 3).

Lastly, we look at organizational proximity. As described, this refers to whether two firms share a similarity in 'organizational context' in which they operate. Though it is unlikely that a GAIO may alter the way in which their member firms operate, or the context in which they do so, GAIO have been said to organize discourse, alignment and consensus (van Lente et al., 2003). As such, it can be posed that GAIO can play a role in providing two firms, holding different sets of rules and routines, with a platform for discourse. As such, we pose the last hypothesis:

Hypothesis 4a

When two firms are both members of the same GAIO, the positive effect of organizational proximity on partnership formation is reduced (negative moderation of hypothesis 4).

2.4. Conceptual model

Integrating theory and hypotheses, brings us to the conceptual model as illustrated in Figure 2 below.

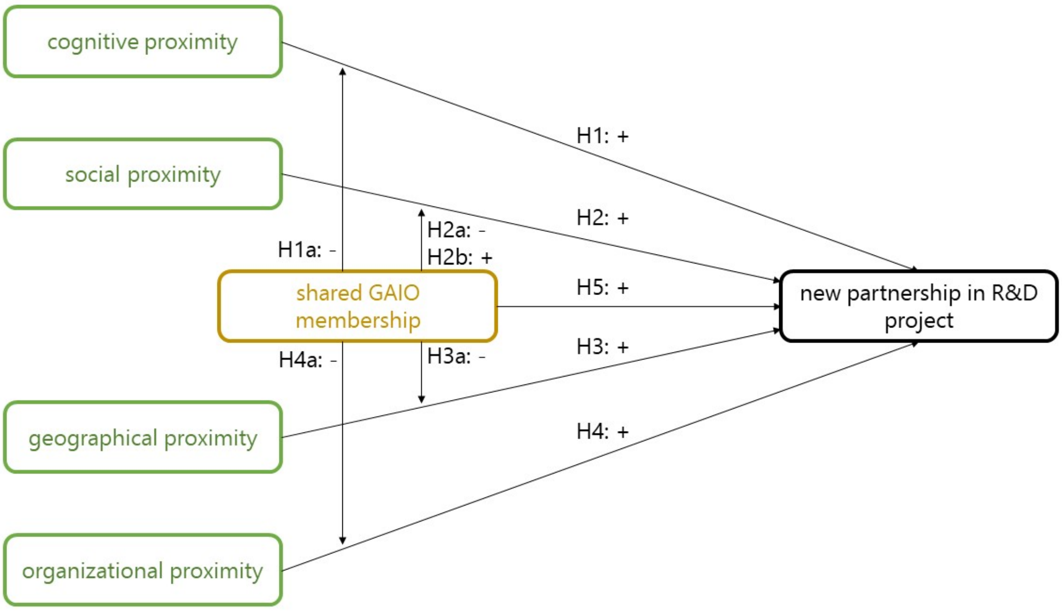


Figure 2: Conceptual model to be assessed. We expect four proximities to positively influence new partnership formation. Shared GAIO membership is hypothesized to reduce the influence of all four proximities, with the exception of social proximities: here we hypothesize both a negative and positive moderation.

3. Data and Methodology

This chapter describes the data and methodology by means we test the hypothesized effects of the proximities and GAIO on new partnership formation between firms.

3.1. Research design

The focus of this research is on government-affiliated innovation intermediaries and how they affect the determinants of interfirm collaboration. The aim is to investigate the influence of several independent variables, based on the proximities theory by Boschma (2005), and influence of GAIO, on the likeliness of new partnerships being formed between firms in a knowledge network. The unit of analysis is thus the dyad between each firm pair i and j.

This study will take a quantitative approach, as this is a suitable method to test hypotheses in deductive research (Bryman, 2015). Also, the use of quantitative research methods in studying the determinants of collaboration is well-established (Autant-Bernard et al., 2007; Scherngell and Barber, 2009; Paier and Scherngell, 2011).

The goal of this research design is to make causal inferences. As such, a time lag between the independent variables and the dependent variable is required (Bryman, 2015). As all variables are observed at one point in time only, this research uses a cross-sectional research design (Bryman, 2015).

3.2. Case description

As mentioned in the introduction, the Dutch government implemented the ‘Topsector’ policy approach in 2012, a narrow innovation policy focussed on nine specific sectors. In these nine sectors, the Dutch aim to excel scientifically and technologically. One of the means to achieve this excellence is the use of multilateral public-private partnerships for R&D, or public-private R&D consortia. These are facilitated by twelve Topconsortia for Knowledge and Innovation (TKI).

The TKI are semi-autonomous, but government-owned entities that operate at the national level. They perform a range of functions that are characteristic of SII, including: shared goal-setting by establishing research agendas in a bottom-up style; demand articulation, e.g. by spreading the word about research calls; network building activities, e.g. by organizing networking events; industry lobbying; and R&D funding. (“Topconsortia voor Kennis & Innovatie,” 2016; Janssen et al., 2016). The TKI operate at the intersection of private organizations, such as small and medium-sized enterprises and large firms, and public organizations, such as PROs and universities. Importantly, TKI also play an explicit role in connecting firms with each other, be it within the same industry or between industries. Their networking role is thus explicitly not limited to academia to firms or vice-versa.

Taking these functions, their level of operation, and ownership into account, we state that it is fair to consider the TKI as government-affiliated intermediary organizations.

The twelve TKI are each connected to one of the nine Topsector industries. They each have their own research and innovation agendas, which they establish with input from industry, academia and government. The TKI “earn” money by registering existing public-private R&D partnerships (PPP) that fit their research agendas to the Dutch Enterprise Agency (Rijksdienst voor Ondernemend Nederland, hereafter referred to as RVO). For each € 1,- of privately invested funds in these PPP, the TKI receive € 0,25 of allowance from the Dutch Ministry of Economic Affairs. This allowance can be used by the TKI to fund new PPP-projects or, to a lesser extent, activities such as networking events. The TKI thus provide stimulation for new PPP to be formed, by initiating new projects that fit their research agenda, connecting actors to join in these projects, and providing additional funding for projects.

Importantly, when the Topsector policy started, the then current PPP projects were registered by the newly-founded TKI, to generate their first ‘allowance’. These projects however, were clearly not set-up with assistance of the TKI, as at the start of these projects (before 2013), the TKI didn’t exist yet, nor did a similar organization. Hence, data on these projects can be used to re-create a network of R&D collaborations as it existed without interference of the TKI. It should be noted that it is mostly the universities and public research organizations that register the PPP projects to the TKI. When a project is registered to a TKI, all participating firms are also coupled to a TKI.

3.3. Data description

Data is made available by RVO, part of the Dutch Ministry of Economic Affairs. This data denotes all Dutch public private R&D consortia, as part of the Topsector policy described above, that were ongoing or started in 2013 or thereafter. The data contains systematic information on projects, including their participants and under which GAIO their projects were registered. This data thus allows us to construct a one-mode actor-network that is a representation of the Dutch public-private research network in several recent years (2013-2016), and it provides information as to which firms are ‘members’ of the different GAIO.

Public-private R&D consortia are defined here as R&D collaborations that involve at least three and maximum twenty-five actors, of which at least one actor is a public institute, and at least one is a for-profit organization. Any projects in the data that do not meet this requirement are excluded from the analysis.

The actor set is determined by taking all firms that participated in at least one public-private R&D project in 2013. These projects were established before the GAIO existed, and hence their formation has not been influenced by the GAIO. In 2013, there was a total of 381 ongoing projects. Within those projects, there were 674 unique Dutch participants, both for-profits and not-for-profits.

A baseline knowledge network (t_0) of these 674 actors is constructed to obtain two independent variables for social proximity, see section 3.4.2. Two actors are connected by a tie when they participate in at least one project together. This results in a non-weighted, one-mode participant network. Note that the variables for social proximity are determined with not-for-profits included in the network. These not-for-profits include universities and public research organizations. Though the analysis focuses on interfirm collaboration, it is recognized that not-for-profits may have an important brokerage role and hence their presence is taken into account in determining the social proximity between firms.

The comparison knowledge network (t_1) is constructed using the same actor set as the t_0 network, new actors are not taken into account. It is assessed whether new ties are formed between actors that were not previously linked (not linked in t_0). These ties are based on projects that were newly started in 2014-2016 (three years aggregated) and received funding from the GAIO. As such, it is assumed that GAIO played at least some role in setting up these new ties. Only ties between actors that did not previously collaborate ($0 \rightarrow 1$) are considered, as from an innovation intermediary perspective bringing together actors that already knew each other is not as relevant as stimulating new ties.

A visualization and some descriptives of the knowledge network(s) are given in Table 1 and Figure 3, respectively. Note that the maximum degree of both networks seems quite high (175 and 218, for t_0 and t_1, respectively). This is a result of including the not-for-profit actors in the network as described above. The high numbers can be declared by one outlier, namely the national Netherlands Organisation for Applied Scientific Research (TNO), which is involved in many projects in all different industrial sectors and thus has a lot of unique partners (degree). In Appendix A, a network visualization for each year (2013-2016) separately is given to provide some idea of how the connections between the partners have grown have the years.

As mentioned, the regression analysis focuses on interfirm collaborations only. In the actor set of 674 participants that were used to construct the knowledge network at t_0 and t_1, there were 589 firms. This results in a total of 589*(589-1) / 2 = 173 166 possible interfirm ties. As the dyad is the unit of analysis, these are our observations. However, considering that we are only interested in modelling newly formed ties, we exclude all dyads that already collaborated (had a tie) at t_0. These are 1968 ties, see Table 1. As such, we are left with a total of 173 166 - 1968 = 171 198 observations at the dyad level.

Table 1: Descriptives of the Dutch public-private knowledge network in t_0 (projects on-going in 2013) and t_1 (projects started in 2014-2016 under the influence of GAIO).

Indicator	2013 (t_0)	2014-2016 (t_1)
Projects	381	531
Actors (Dutch only)	674	674
-- of which firms	589	589
Possible ties	173166	173166
Collaborations (ties)	3172	3813
-- of which interfirm ties	1968	2213
New ties at t_1	-	641
-- of which interfirm ties	-	245
Total network density	0,0140	0,0168
Mean degree (number of unique partners)	9,41	11,32
Min degree	1	1
Max degree	175	218
Mean tie weight (collaboration intensity)	1,28	1,30
Min tie weight	1	1
Max tie weight	15	15
Mean shortest path length	3,30	2,96
Longest shortest path	7	6

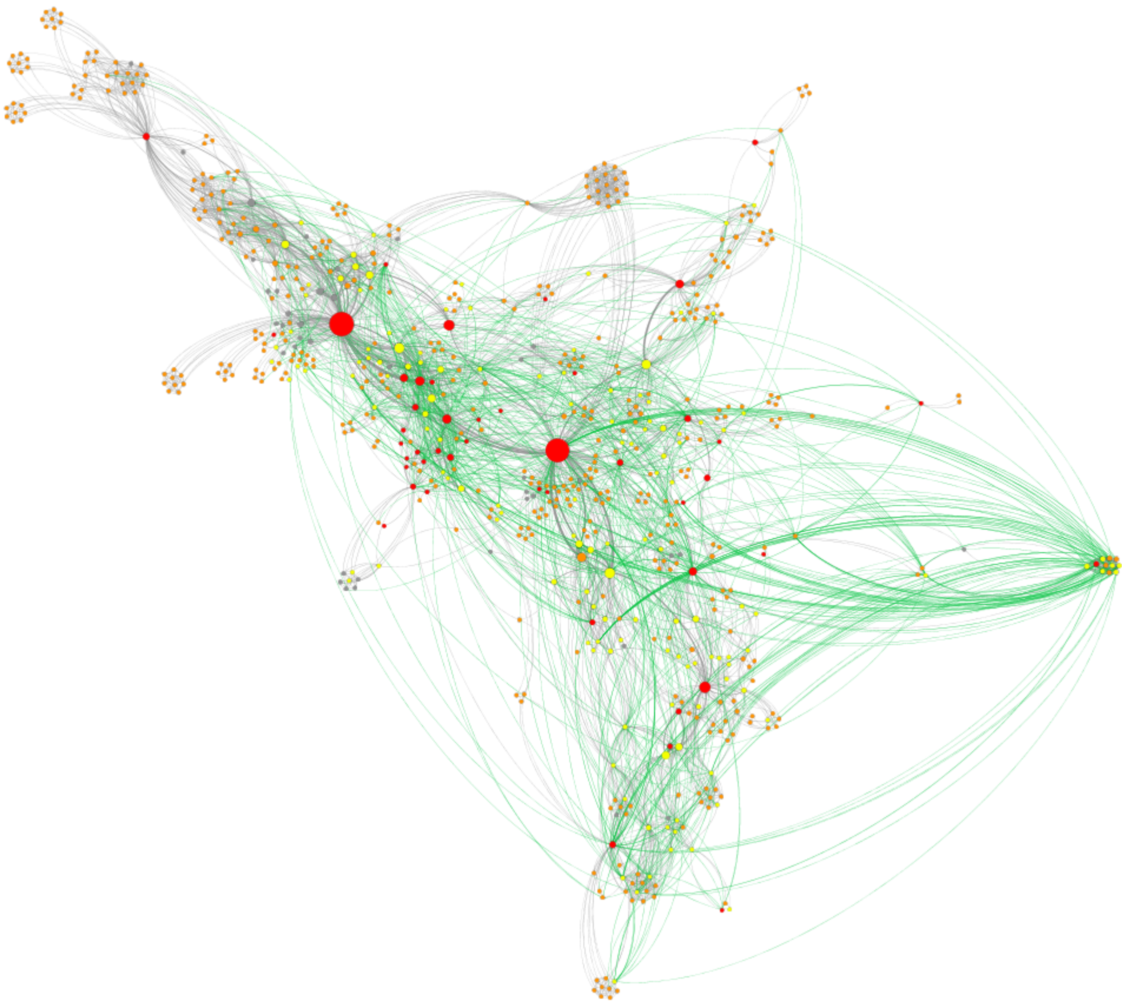


Figure 3: Visualization of the the Dutch public-private knowledge network. Nodes represent public research organizations/knowledge institutes (red), SME (orange) or large firms (yellow). Node size indicates the number of project participations per actor at t_0. Ties represent joint collaboration in a project. Grey ties represent collaboration in t_0, when no GAIO were present. Green ties represent new ties (0→1) formed in t_1, when GAIO were present.

3.4. Constructing variables

The section below describes the operationalization of the concepts used in the hypotheses.

3.4.1. Dependent variable

The dependent variable is binary and indicates whether a new tie is formed between firms i and j at t_1, *Y*. This is the case for 245 observations. As mentioned in the data description, there is a total of 171 198 interfirm ties considered (the total number of possible ties minus the already existing ties at t_0). Thus, the overall probability of a new tie forming Pr(Y=1) is 245/171 198 = 0,00143 or 0,143%.

3.4.2. Independent variables

The **influence of GAIO** is assessed by a binary variable that indicates whether firms i and j in the dyad have a shared membership of at least 1 GAIO, *GAIO*. Shared membership is defined here as both firms i and j having participated in at least one project that was registered under the same GAIO at t_0. In this case, the GAIO variable takes on value 1, otherwise it is 0. This is represented schematically in Figure 4. When the GAIO were first established in 2013, all then ongoing projects were registered at a GAIO, which gave the GAIO their first 'members' immediately. In Figure 4, actors a and b have membership of GAIO X only; actor c has membership of GAIO X and Y; actors d to h have membership of GAIO Y only; and actors i, j and k have membership of GAIO Z only. Note that projects can only be registered under one GAIO, but as actors can participate in multiple projects, actors can be members of multiple GAIOs. At the dyad level, it can then be determined whether each firm pair i and j are both members of at least one same GAIO, yes or no (GAIO = 1 or GAIO = 0).

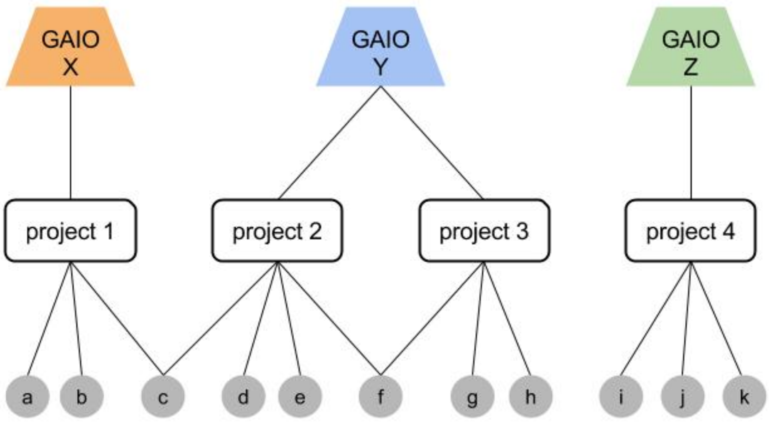


Figure 4: Hierarchy of GAIO (coloured trapezes), projects (transparent squares) and project participating actors (grey circles).

Cognitive proximity between firms i and j is approximated by a revealed skill relatedness measure. This concept refers to revealed similarities between the skills and knowledge required by workers in different industries. Following the method by Neffke, Otto and Weyh (2016), skill relatedness is determined based on labour mobility between each pair of 4-digit NACE-codes in The Netherlands (Neffke et al., 2016). Labour mobility data was obtained from a previous study by Dialogic Innovatie & Interactie and the Netherlands Environmental Assessment Agency (Planbureau voor de Leefomgeving, PBL), who derived it from the Dutch Central Bureau for Statistics (CBS). The data contains the number of labour market transfers in 2009 and 2010 from each 4-digit NACE code to the other. The skill relatedness between an origin (o) and destination (d) is calculated using the following equation:

$$Skill\,Rel_{o\rightarrow d} = F_{o\rightarrow d} * F_{total} / F_{o\rightarrow all} * F_{all\rightarrow d} \tag{1}$$

Where $F_{o\rightarrow d}$ represents the number of labour transfers from origin to destination industry; F_{total} represents the total numbers of labour transfers from all origins to all destinations; $F_{o\rightarrow all}$ represents the total outflow from the origin to all other industries; and $F_{all\rightarrow d}$ represents the total inflow into the destination industry from all other industries. The result is a weighed relatedness measure taking on values [0, ∞]. As this measure is strongly right-skewed, it is normalized:

$$SRNorm_{o\rightarrow d} = (Skill\,Rel_{o\rightarrow d} - 1) / (Skill\,Rel_{o\rightarrow d} + 1) \tag{2}$$

This normalized SR measure ranges from [-1,1] with 1 indicating the highest possible, and -1 indicating the lowest possible skill relatedness.

For the firms in our dataset then, we have the 4-digit NACE codes available and hence for each firm pair i and j we can determine skill relatedness. Note that the (normalized) skill relatedness is directed, whereas partnership formation is 'undirected'. Hence for each firm pair i and j, we have two measures: i → j and j → i. To approach the situation where one firm reaches out to the other if cognitive proximity is high enough, we use the maximum of the directed normalized skill relatedness between the NACE-code of firms i and j as an undirected measure: *SR*. As a higher skill relatedness indicates higher cognitive proximity, we expect a positive effect of SR on Pr(Y=1).

Social proximity is approximated by two measures that represent the extent to which firms i and j are able to 'discover' each other as possible new partners. As mentioned in theory, this knowledge about potential partners and the creation of required trust often flows through previous partners. Hence, first we measure social proximity by the number of direct partners firms i and j share in t_0, *Shared_partn*. As a higher number of shared partners would hypothetically lead to more chance of 'discovering' each other, we expect a positive effect of Shared_partn on Pr(Y=1). The second measure for social proximity is the geodesic distance, or shortest path length, between firms i and j in the collaboration network at t_0, *ND* (Network Distance). As firm pairs that previously collaborated are excluded, the network distance possibly has a range of [2, Inf]. Unconnected nodes pose a problem in statistical analysis, however inspection of the data shows that ND in the used dataset has a range of [2, 7], so there is no issue in this case. In this case, a higher value for ND indicates a lower social proximity and less opportunity for firms to 'discover' each other. Hence, we expect a negative effect of ND on Pr(Y=1). Note that these two social proximity measures are determined with not-for-profits included in the network at t_0, as described in section 3.3.

Geographical proximity between firms is measured as geographical distance in kilometers 'as the crow flies', divided by 100, *Geog*. The address of each firm was made available by RVO. Using Google's geocoding API and the "RGoogleMaps" package in R, addresses were turned into latitude and longitude coordinates (Loecher and Ropkins, 2015; "GoogleMaps Geocoding API," 2017). Then, the distance between coordinates could be calculated using the "Imap" package in R (Wallace, 2012). As geographical proximity is expected to facilitate collaboration, we expect a negative sign of Geog on Pr(Y=1).

Organizational proximity between firms is approximated by two measures. Firstly, as it is recognized that SME (<250 employees) and large firms (≥250) differ in their take on knowledge management and new product development (McAdam and Reid, 2001; Nicholas et al., 2011), we determine whether firms i and j are both SME, both large firms or one SME and one large firm. This create a categorical variable, *Firm_types,* at the dyad level with three levels: SME_SME, SME_Large and Large_Large.

The second approximation of organizational proximity is made by assessing whether firms i and j are involved in fundamental or applied research, *Research.* This is determined based on their projects at t_0. For each project, at least one public organization is involved, as these are public-private research projects. For each public organization then, it is determined whether it engages mostly in fundamental or applied research. This was done by three individuals separately, to ensure reliability. If a project involved at least one public organization that is considered 'fundamental', the project is classified as a fundamental research project, otherwise it is classified as an applied project. Now we must translate this to the firm-dyad level. We use the same approach: if a firm participates in at least one fundamental research project, we consider that firm as being a "fundamental research firm". If not, we consider it an "applied research firm". This means that firms that participate only in either fundamental or applied research projects are classified accordingly, and firms that participate in both fundamental and applied research projects are classified as "fundamental research firms". A categorical variable is created at the dyad level with three levels: Fund_Fund, Appl_Appl, Fund_Appl.

We choose this approach of giving more weight to fundamental than applied research as it is assumed that all firms engaged in research are willing to conduct applied research, but only a rather small number will be interested in engaging in fundamental research, due to the high costs and risks involved (Cohen, 2006). At the same time, it seems unlikely that firms are willing to invest only in fundamental research, without also engaging in applied research, as that goes against the profit-making goal of the firm (Rosenberg, 1990).

For both variables for organizational proximity we expect that the 'same' actors are more likely to collaborate than 'different' actors, following the argument of proximity and homophily. Hence, if we use SME_Large and Fund_Appl as reference levels, we expect a positive effect of the categorical variable taking on the other possible levels, in both cases.

3.4.3. Control variables

To account for alternative factors that explain the likeliness to form a new tie, several control variables are included. Firstly, to account for firms R&D potential, the total number of R&D projects that firms participated in at t_0 is taken into account. At the dyad level, this is calculated as the mean of the number of projects of i and j: *Mean_proj* = (projects firm i + projects firm j) / 2.

Secondly, to account for firms' collaboration potential, the number of unique partners of each firm at t_0 is taken into account. At the dyad level, this is calculated as the mean number of unique partners of firms i and j at t_0: *Mean_partn* = (partners firm i + partners firm j) / 2.

Lastly, to control for sectoral differences, a dummy is included that indicates whether two firms are in the same sector, based on their 2-digit NACE code. If this 2-digit NACE code is the same, then the dummy, *Dummy_same_industry*, takes on value of 1, and 0 otherwise.

An overview of all the variables is given in Table 2 on page 27.

3.5. Statistical model

3.5.1. Logit function

Following Autant-Bernard et al. (2007) and Paier and Scherngell (2008), a latent variable model is considered, that is observed as a binary logit model (Autant-Bernard et al., 2007; Paier and Scherngell, 2008). The decision to form a new partnership is in essence a binary choice model: either you do or you don't. This decision reflects an expected payoff from the new collaboration by firms i and j, this is our latent variable. This payoff, Y^*_{ij}, is a continuous variable and can be considered a linear function of all characteristics of firms i and j, as described in the hypotheses:

$$Y^*ij \;=\; \beta_0 \;+\; \beta_1{}^*X_1 \;+\; \beta_2{}^*X_2 \;\; + \ldots \;\; + \beta_n{}^*X_n + \varepsilon_{ij} \tag{3}$$

where parameter β_0 indicates the intercept, β_1 to β_n are the parameters to be estimated, X_1 to X_n are the independent and control variables, and ε_{ij} is a random error.

As Y^*_{ij} is not observable, it is assumed that when payoffs are positive, collaboration will occur, and otherwise it will not. This results in a binary dependent variable, Y_{ij}. Y_{ij} follows a Bernoulli distribution, taking on values (1) or (0) with probabilities *p* and *1-p*, respectively. Thus, the probability, p, of a new link forming between i and j, is given by:

$$p \;=\; Pr(Y_{ij} \;=\; 1) \;=\; Pr(Y^*ij \;>\; 0) \tag{4}$$

As probabilities are bound between 0 and 1, but the linear function for the latent variable Y^*_{ij} can take on all values, p is logit transformed:

$$logit(p) \;=\; ln(odds) \;=\; ln\left(p \;/\; (1-p)\right) \;=\; Y^*_{ij} \tag{5}$$

Table 2: Overview of dependent, independent and control variables. Exp. sign = Expected sign.

Type	Concept	Name (exp. sign)	Description	Calculation of score
Dependent	New partnership formation	Y	New collaboration between firms i and j.	(1) if firms i and j form a new partnership by participating in at least one project together at t_1; (0) otherwise. Binary variable, 0 or 1.
Independent	Cognitive proximity	SR (+)	Skill-relatedness between firms i and j.	Following the method by Neffke et al. (2011). When skill-relatedness is high, ~1. If no relatedness, 0. Most unrelated, ~ -1 Interval variable [-1, 1]
	Social proximity	ND (-)	Shortest path between firms i and j.	Geodesic distance between firms i and j in baseline knowledge network (t_0). Count variable [2, ∞]
		Shared_partn (+)	Direct partners that i and j have in common.	Count of the number of partners that i and j have in common in baseline knowledge network (t_0). Count variable [0, ∞]
	Geographical proximity	Geog (-)	Geographical distance between firms i and j.	Distance in 100km ‘as the crow flies’ between addresses of firms i and j. Continuous variable [0, ∞]
	Organizational proximity	Firm_types (+) for both levels compared to baseline	Whether firms i and j are SME, Large or SME and Large.	Categorical variable with three levels: SME_SME; Large_Large; SME_Large. The latter is the baseline level. Categorical variable, 3 levels.
		Research (+) for both levels compared to baseline	Whether firms i and j are involved in fundamental or applied research.	Categorical variable with three levels: Fund_Fund; Appl_Appl; Fund_Appl. The latter is the baseline level. Categorical variable, 3 levels.
	Systemic innovation intermediary (SII)	GAIO (+/-)	Whether firms i and j are members of the same GAIO	(1) if firms i and j have both participated in at least one project registered under the same GAIO, (0) otherwise. Binary variable, 0 or 1.
Control	R&D potential	Mean_proj (+)	Number of projects that i and j participated in.	Mean number of project participations for i and j at t_0. Mean_proj = ($Proj_i$ + $Proj_j$) / 2 Continuous variable, 1 - ∞
	Collaboration potential	Mean_partn (+)	Mean number of project partners for i and j	Mean number of unique partners (degree) for i and j in the network at t_0. Mean_partn = ($Partn_i$ + $Partn_j$) / 2 Continuous variable, 1 - ∞
	Collaboration propensity	Dummy_same_industry	Dummy indicating whether i and j are in same industry	(1) if 2-digit NACE code of i and j is identical, (0) otherwise. Binary variable, 0 or 1.

Which leads to the binary logistic regression model to be estimated given by:

$$p = e^{Y*ij} / (1 + e^{Y*ij}) = e^{\beta_0 + \beta_1 * X_1 + ..+ \beta_n * X_n + \varepsilon_{ij}} / (1 + e^{\beta_0 + \beta_1 * X_1 + ..+ \beta_n * X_n + \varepsilon_{ij}}) \quad (6)$$

The logit model is estimated using the statistical program R, and the interface program Rstudio (RStudio Team, 2015; R Core Team, 2017). The R base function 'glm' is suitable to estimate binary logit models, and uses maximum likelihood estimation (MLE) to estimate the model.

3.5.2. Quality of the model

Models will be compared using the likelihood ratio test, implemented in R in the "lmtest" package (Zeileis and Hothorn, 2002).

With regards to model fit, there is no R2 statistic that explains the proportion of variance in the dependent variable that is explained by the independent variables. Some pseudo-R2 measures have been proposed, though their interpretation is not very straightforward, but can be used to simply assess a better or lesser fit of one model over the other. We report McFadden's pseudo-R2, a popular measure for binary logistic regression models (Hoetker, 2007). This value can be obtained from R using the "pscl" package (Jackman, 2015). Note that though like in OLS regression, a higher value for R2 indicates a better fit, overall McFadden's pseudo-R2 values tend to be lower than OLS R2 values. McFadden's pseudo-R2 values of 0.2 to 0.4 represent excellent fit (Hensher and Stopher, 1979).

Some argue that binary logit models should be evaluated based on the number of correct predictions they make on a test dataset, as the pseudo-R2 is not interpretable in practice (Hoetker, 2007). However, this requires the separation of data in a model and test set, and hence a loss of data for creating the model. In our case this can be problematic due to the rarity of 1 values on the outcome variable. Hence, we stick to McFadden's pseudo-R2 only.

3.5.3. Rare events data

As the overall probability of the Y-variable taking on the value of 1 is only 0,143% (see section 3.4.1.) we are dealing with rare events data. This may bias results of the model, specifically the estimated probability of Y=1 and the estimated coefficients (King and Zeng, 2001). To reduce the bias, King and Zeng recommend using an alternative estimation method, which is implemented in the relogit-model in the 'Zelig' package for R (Imai et al., 2008; Choirat et al., 2017). Another method to reduce bias is penalized likelihood estimation, also called the Firth method (Firth, 1993). This is implemented in the 'logistf' package for R (Heinze and Ploner, 2016).

To assess for bias due to rare events data, results of the three different methods are compared for a model with all independent and relevant control variables, but no interactions. The results show that there are minimal differences in the estimated coefficients between the three methods, see Appendix B. Thus, there seems to be no reason to assume the normal logit model gives biased results with regards to rare events.

3.5.4. Interaction effects in logit models

A second issue in the analysis may result from estimating interaction effects in a non-linear model. There is a long-standing discussion as to how interaction effects in nonlinear models should be modelled and interpreted (Wolfinger and Rosenstone, 1980; Nagler, 1991; Ai and Norton, 2003; Brambor et al., 2005; Hoetker, 2007; Berry et al., 2010; Greene, 2010; Karaca-Mandic et al., 2012; Onukwugha et al., 2015; Rainey, 2016).

The problem is with the functional form of the logit model, namely its nonlinearity. This causes the effect of each independent variable on the outcome Pr(Y=1) to *always* be conditional on the values of all other independent variables. In other words, each independent variable *always* interacts with all others (Huang and Shields, 2000; Hoetker, 2007; Berry et al., 2010). The questions that arise then, are whether or not to include a product term in a binary logit model, and how to assess the influence of this product term.

With regards to the first, literature is reasonably unanimous: when one hypothesizes an interaction effect to be present, based on theoretical foundations, then one should include a product term in the model (Nagler, 1991; Berry et al., 2010; Williams, 2012; Rainey, 2016).

Trouble arises, however, with the second question: how to assess the effect of a product term? As mentioned, all variables interact with one another in a nonlinear model. Thus, there is always interaction between all variables present, which influences the outcome variable Pr(Y=1). Some researchers consider this 'built-in interaction effect' or 'compression' theoretically irrelevant, simply an artifact of using a nonlinear model. They argue that a significant product term in the model is a requirement to show that any 'substantive interaction', e.g. besides compression, is present (Nagler, 1991). Others however, argue that compression ís substantially meaningful, as long as one poses hypotheses about the probability of an event occurring, rather than about the latent variable. Compression clearly influences the Pr(Y=1), and thus, when hypothesizing interactive effects specifically on Pr(Y=1), compression is as theoretically relevant as the effect of a product term (Berry et al., 2010). They also argue that compression is very much a real-world phenomenon. As Huang and Shields (2000) posed it: *"Those whose fitted probabilities are located somewhere in the middle will be more sensitive to changes in variables. Everyone knows that it takes much more effort [...] to raise a person's probability, say, from .8 to .9 than it takes to raise the probability from .4 to .5 [...] if we ignore the transformation function of the [nonlinear] model, we also ignore the built-in ceiling and floor effects"* (Huang and Shields, 2000; Berry et al., 2010).

Considering the more recent view that compression is indeed relevant, and the notion that we indeed pose hypotheses about the probability that firms collaborate, and not the latent variable of utility from collaboration, it seems logical to follow the latter stream of thought. Hence, we consider compression as much a relevant effect as any other interaction effect.

When it comes to assessing model results then, a statistically significant product term is *neither necessary nor sufficient* for variables to meaningfully interact in influencing Pr(Y=1). There may be a significant interaction effect, even when the product term is not significant, and vice-versa (Ai and Norton, 2003; Brambor et al., 2005; Powers, 2005; Hoetker, 2007; Berry et al., 2010; Karaca-Mandic et al., 2012). Thus, an alternative method must be used to assess the sign, significance and effect size of interacting variables. Considering the ambiguity surrounding interactions in binary logit models, we choose to apply two separate methods for robustness.

Firstly, we circumvent the entire problem by conducting a split model analysis. One model for all observations where GAIO = 0, and one for all observations where GAIO = 1. This will give insight into the sign and significance of the independent variables on Pr(Y=1) for the subgroups. This, in essence, models the interaction of being in a subgroup 0 or 1, with all of the other independent (and control) variables.

Secondly, we follow a method proposed Karaca-Mandic, Norton and Dowd (2012) which is also underscribed by several other authors (Ai and Norton, 2003; Powers, 2005; Greene, 2010). This method proposes to look at the Average Marginal Effects (AME) from a model with a product term to interpret interactions. Very short, the AME indicates how much, in absolute percentage points, Pr(Y=1) changes for a one unit increase in an independent variable, on average over all the observations in the sample. More discussion with regards to the calculation and interpretation of AME is given in section 4.4.1. The AME can be obtained from R by using the "margins" package (Leeper, 2017). Following this method, we build several models on the full dataset, so including both observations where GAIO = 0 and GAIO = 1. Subsequently, to assess interaction effects, we include a product term for GAIO*[variable of interest] in the models, as it was described that inclusion of a product term is necessary when this is hypothesized (Berry et al., 2010; Karaca-Mandic et al., 2012; Williams, 2012). Then, we calculate the AME for the variable of interest, while setting GAIO to 0 or 1, and we assess whether there is a significant effect of the variable of interest in each situation.

A benefit over the second approach is that while the split model provides insight into the sign and significance of predictors, it is not easy to compare the extent of the effect of a predictor in each of the models, e.g. the extent to which an x-unit increase in the variable changes Pr(Y=1) for GAIO = 0 or GAIO = 1. In other words, if GAIO abolishes or reverses the effect of a predictor, we will see that in the split models. However, if GAIO reduces or exacerbates the influence of a predictor, its sign and significance in the GAIO = 1 split model will be the same sign as in the GAIO = 0 split model. From the AME in the second model, we can look at differences in Pr(Y=1) between the subgroups, caused by the independent variables, which will enable us to see an exacerbated or reduced effect as well.

3.5.5. Background analyses

In addition to the main analyses described above, we perform two background analyses to provide some stylized facts. As this study takes a cross-sectional design, we are looking how values of predictors at t_0 influence the formation of new ties at t_1, with the latter being while the GAIO were present and t_0 being before the GAIO were implemented. However, as we only analyse the influence of predictors on ties at t_1, any results that we derive do not provide information as to the effect of the presence of GAIO in the knowledge landscape overall. Rather we assess how firms form new partnerships either when sharing a GAIO or not sharing a GAIO.

To get some additional insight into how the firms' collaboration choices were influenced before the GAIO were implemented, and how this changed when the GAIO were implemented, we need to assess predictors on the collaborations at t_0 and compare these with predictors on the collaborations at t_1. Such an analysis is not intended to assess the entirety of effects that occurred from implementation of the GAIO, and clearly there can be many other predictors that influence changes in collaboration determinants over time. However, for the sake of context, we provide some basic insights with regards to the situation before and after the GAIO were implemented. For that purpose, we conduct two background analyses to complement the main analyses.

The first is a binary logit regression of the baseline ties (at t_0) with some of the independent variables. These are all independent variables that have not been derived from the project data at t_0: *SR*, *Geog, Firm_types, Dummy_same_industry.* The Y-variable of this regression is whether a baseline tie (at t_0) exists between firms i and j or not. The second is an analysis with the same independent variables, but now performed on the new ties at t_1, so the Y-variable is the same as in the main analysis. However, in this background analysis we only include the variables *SR*, *Geog, Firm_types, Dummy_same_industry.* In this way, we can check whether the influence of these predictors has changed from t_0 to t_1, whereby we assume that the values of these predictors are relatively stable over time.

In the results section, we will first discuss these background analyses, before continuing with the results of the main analyses.

4. Results

This section will first look at the descriptives, both for the full dataset and the subsets for GAIO = 0 and GAIO = 1. Then, we'll turn our attention to the background-analyses, to provide some first insights. Thereafter we'll discuss the main analyses, first the split model and then the AME model.

4.1. Descriptives

As mentioned in the data description (section 3.3), we consider a total of 171 198 observations for the main analysis, we consider this the full dataset. Descriptives of the variables for the full dataset are given in Table 3. Correlations of the numeric variables of the full dataset are given in Table C.1 in Appendix C.

Table 3: Descriptives of the variables of the full dataset (n = 171198). DV = Dependent variable, IV = Independent variable, CV = Control variable, Min = minimum, Max = maximum, SD = Standard deviation.

Numerical variables	Mean	Min	Max	SD
IV: SR	-0,037	-1	1	0,583
IV: ND	3,374	2	7	0,906
IV: Shared_partn	0,204	0	11	0,520
IV: Geog	0,896	0	3,19	0,508
CV: Mean_proj	1,829	1	23	1,805
CV: Mean_partn	8,357	1	54	4,872
Categorical variables	**Count of 0**	**Count of 1**	**Perc. 0 (%)**	**Perc. 1 (%)**
DV: Y	170 953	245	99,86	0,14
IV: GAIO	150 494	20 704	87,91	12,09
IV Firm_types: SME_Large	-	60 357	-	35,26
IV: Firm_types: SME_SME	-	102 214	-	59,72
IV: Firm_types: Large_Large	-	8627	-	5,04
IV: Research: Fund_Appl	-	55 888	-	32,65
IV: Research: Appl_Appl	-	108 412	-	63,33
IV: Research: Fund_Fund	-	6898	-	4,03
CV: Dummy_same_industry	159 865	113 33	93,38	6,62

For the split model analyses, we split the observations into two groups. One where GAIO = 0 and the other where GAIO = 1. As can be seen from Table 3, this results in two subsets of the data with n = 150 494 for the first, and n = 20 704 for the latter. We will further refer to these subsets as the ‘GAIO-0 subset’ and ‘GAIO-1 subset’. Descriptives of the GAIO-0 and GAIO-1 subsets are given in Table 4. Correlations can be found in Appendix C, in Tables C.2. and C.3., respectively.

Table 4: Descriptives of the variables of the GAIO-0 (left four columns) and GAIO-1 data subsets (right four columns). DV = Dependent variable, IV = Independent variable, CV = Control variable, Min = minimum, Max = maximum, SD = Standard deviation.

Dataset:	GAIO-0 subset (n = 150494)				GAIO-1 subset (n = 20704)			
Numerical variables	Mean	Min	Max	SD	Mean	Min	Max	SD
IV: SR	-0,068	-1	1	0,572	0,183	-1	1	0,615
IV: ND	3,486	2	7	0,868	2,560	2	6	0,742
IV: Shared_partn	0,121	0	6	0,348	0,802	0	11	0,973
IV: Geog	0,897	0	3,19	0,505	0,886	0	3,17	0,529
CV: Mean_proj	1,780	1	22	1,699	2,183	1	23	2,415
CV: Mean_partn	8,132	1	44	4,573	9,995	1	54	6,421
Dataset:	**GAIO-0 subset (n = 150494)**				**GAIO-1 subset (n = 20704)**			
Categorical variables	**Count of 0**	**Count of 1**	**Perc. 0 (%)**	**Perc. 1 (%)**	**Count of 0**	**Count of 1**	**Perc. 0 (%)**	**Perc. 1 (%)**
DV: Y	150 370	124	0,08	99,92	20583	121	99,42	0,58
IV Firm_types: SME_Large	-	53690	-	35,68	-	6667	-	32,2
IV: Firm_types: SME_SME	-	89445	-	59,43	-	12 769	-	61,67
IV: Firm_types: Large_Large	-	7359	-	4,89	-	1268	-	6,12
IV: Research: Fund_Appl	-	50540	-	33,58	-	5348	-	25,83
IV: Research: Appl_Appl	-	94296	-	62,66	-	14 116	-	68,18
IV: Research: Fund_Fund	-	5658	-	3,76	-	1240	-	5,99
CV: Dummy_same _industry	142 375	8119	94,61	5,39	17 490	3214	84,48	15,52

4.2. Stylized facts

Before moving to the main analyses, we take a short look at the background analyses performed, as described in section 3.5.5. The results are given in Appendix D. These results provide a first indication as to what predictors influenced the formation of the t_0 ties, so without the influence of GAIO, and the formation of new ties at t_1, so with the influence of GAIO. We will call these the t_0-bg-model and t_1-bg-model, respectively, in the remainder of this section ('bg' for 'background').

Firstly, it should be noted that the overall probability of outcome Y taking on the value 1, Pr(Y=1) , differs for the two models, as they use a different Y-variable. For the t_0-bg-model, the overall Pr(Y=1) is 1,14%, see Table D.1 in Appendix D. For the t_1-bg-model, it is the same as in the main analysis, namely 0,14%, see Table 3.

In both models, Skill Relatedness has a positive and significant effect, indicating that in both cases firms who have a higher cognitive proximity are more likely to form a partnership. Looking at the coefficients and the p-values, it appears that SR has a slightly stronger effect in the t_0-bg-model than in the t_1-bg-model. Thus, it appears that, when comparing t_1 to t_0, SR has become somewhat less important as a predictor of collaboration when the GAIO were present than when they were not.

In both models, Geographical distance has a negative and significant effect. Again, this effect is stronger (more negative coefficient and lower p-value) in the t_0-bg-model than in the t_1-bg-model. This would indicate that, while Geographical distance has a negative effect on all collaborations, the negative effect is less strong for collaborations that were formed after the GAIO were implemented than for those that were formed before.

Looking at the categorical variable Firm_types, we see that, in the t_0-bg-model, two SME firms do not differ significantly in their likeliness to collaborate compared to an SME and a large firm. For two large firms, the effect on the outcome variable is positive and significant, indicating that two large firms are more likely to collaborate compared to one large and one SME. In the t_1-bg-model, we see the same for two large firms: a positive and significant effect. However, now we also see a negative and significant effect for two SME, meaning that two SME are less likely to collaborate than an SME and large firm. So before introduction of the GAIO, there were relatively more collaborations in large-large configuration compared to both other configurations. After introduction of the GAIO, there were relatively more collaborations in large-large configuration compared to sme-large, and this sme-large configuration occurred relatively more than the sme-sme configuration. Note that this says nothing about true probabilities of collaboration, but merely something about how often the different firm configurations occur relatively to each other.

The last item to note about the background analyses is the *Dummy_same_industry.* This predictor is positive and significant in the t_0-bg-model, but not significant in the t_1-bg-model. This indicates that before the introduction of the GAIO, firms were more likely to collaborate with firms in the same industry (based on 2-digit NACE), rather than with firms from other industries. After introduction of the GAIO, this is no longer true, firms appear to have no preference for collaboration with another firm from their own or from another industry.

As mentioned in section 3.5.5. the stylized facts we describe above should be considered as merely that, stylized facts. No conclusions with regards to causal effects of the implementation of GAIO can be drawn from these results, simply that we see different patterns of collaboration before and after the GAIO were implemented. Equally, these changes in patterns may be attributable to many other factors that undoubtedly changed over time from t_0 to t_1.

4.3. Split models

Now we turn our attention to the split models, where we separately estimate the effects of the predictors on the subset of observations where GAIO=0 and the subset where GAIO=1.

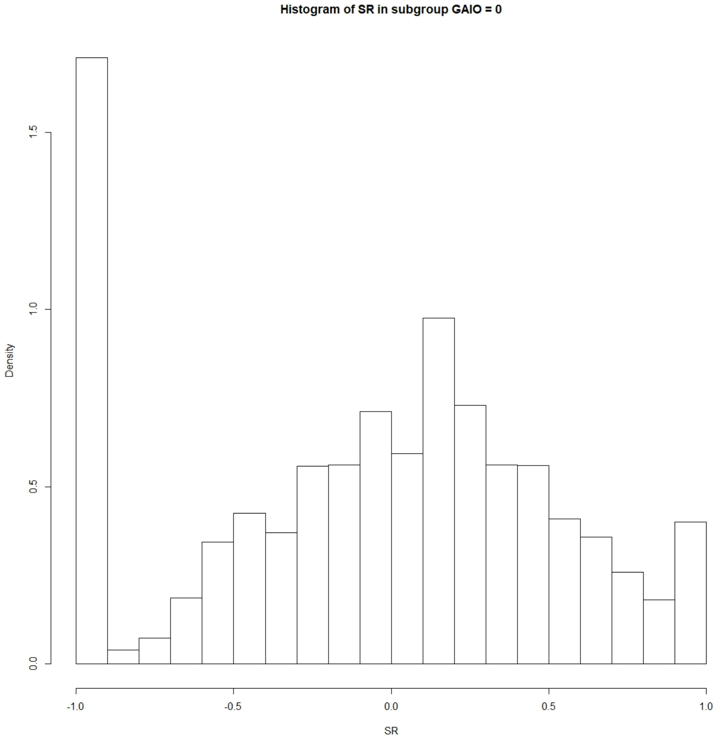
Firstly, it should be noted from the descriptives in Table 3 (page 32), that the overall probability of collaboration, Pr(Y=1) amongst the subset of GAIO=0 is 0,08% and amongst the subset of GAIO=1 it is 0,58%. This gives a first indication that firms are more likely to collaborate when they share a GAIO than when they are members of different GAIO.

Secondly, it should be noted that the subgroups differ in terms of their mean values of the numerical independent variables and the proportional distribution over levels of the categorical variables. Student's t-tests reveal that these differences are all significant at the 1% level. This indicates that there are differences between the subgroup of dyads where firms share membership of a GAIO and the subgroup of dyads where firms do not share a GAIO. Given that the mean or proportions differ between the groups, is not necessarily problematic, as there is still a reasonable distribution of the variables in both groups. As an illustrative example, the histograms of Skill Relatedness for both subgroups are given in Figure 5, on the next page. As another indication for the distribution, note also from Table 4 (page 33) that for most variables, including SR, the standard deviation in the GAIO = 1 subset is higher than in the GAIO = 0 subset (with the exception of ND).

The results of the logistic regression on the subgroups are given in Table 5. Due to multicollinearity between *Mean_proj* and *Mean_partn*[2], the latter was removed.

The results in Table 5 indicate that when GAIO = 0, Skill Relatedness has a positive and significant effect on new tie formation. However, when GAIO = 1, Skill Relatedness does not have a significant effect. This indicates that when firms are members of different GAIO, a higher cognitive proximity facilitates innovation. When firms are members of the same GAIO, this predictor no longer matters, which indicates that GAIO may be able to help firms overcome cognitive distance in forming new partnerships.

[2] Variance Inflation Factors for Mean_proj and Mean_partn in the split models were both over 2,5. This can be cause for concern in logistic regression (Midi et al., 2010).



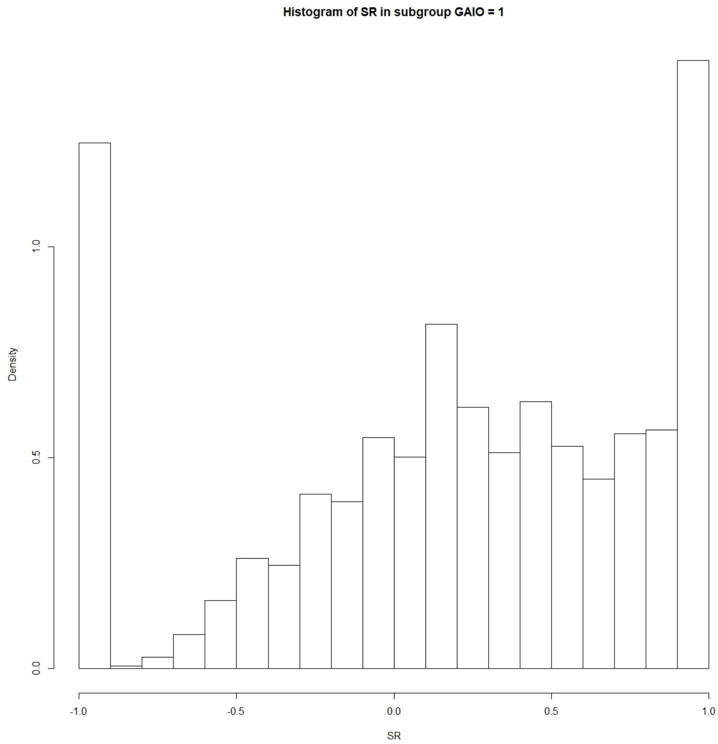


Figure 5: Histograms of Skill Relatedness for subgroups GAIO = 0 (top left) and GAIO = 1 (bottom right). Skill Relatedness is given on the X-axis, ranging from -1 to 1. On the Y-axis the relative occurrence of each value in the dataset is given.

Table 5: Results of split model logistic regression on subgroups of the dataset where GAIO = 0 (left) and GAIO = 1 (right).

	Subgroup GAIO = 0				Subgroup GAIO = 1			
	Coeff.	**SE**	**p-value**	**sign.**	**Coeff.**	**SE**	**p-value**	**sign.**
SR	0,468	0,170	0,006	**	0,260	0,180	0,148	n.s
ND	0,255	0,128	0,046	*	-1,033	0,308	0,001	***
Shared_partn	0,190	0,195	0,332	n.s.	0,281	0,065	1,37e-05	***
Geog	-0,337	0,196	0,086	n.s.	-0,193	0,193	0,319	n.s.
Firm_typesSME_SME	-1,419	0,293	1,24e-06	***	-1,022	0,289	4,02e-04	***
Firm_typesLarge_Large	1,845	0,203	< 2e-16	***	1,473	0,211	3,20e-12	***
ResearchAppl_Appl	0,009	0,222	0,967	n.s.	-0,693	0,240	0,004	**
ResearchFund_Fund	0,805	0,261	0,002	**	0,239	0,251	0,341	n.s.
Mean_proj	0,156	0,028	3,68e-08	***	-0,008	0,028	0,765	n.s.
Dummy_same_industry	-0,326	0,439	0,457	n.s.	-0,246	0,333	0,459	n.s.
Constant	-8,095	0,516	< 2e-16	***	-2,600	0,728	3,56e-04	***
AIC	1743,5				1247,9			
McFadden's pseudo-R2	0,143				0,175			

Coeff. = Coefficient; SE = Standard Error; sign. = Significance. Significance levels: 0,000 *** 0,001 ** 0,01 * 0,05 ‘n.s.’ 1

Also, for Network Distance, we see an interesting result when GAIO = 0. In that model, ND has a positive and significant effect on new tie formation, indicating that firms at a larger distance in the initial collaboration network are more likely to form a new partnership. When GAIO = 1, however, ND has a strong significant and negative effect. This indicates that within the GAIO member groups, there is a tendency of firms to form new ties with actors that are relatively close to them in the network: the firms act locally within their embedded networks. A similar pattern can be observed from *Shared_partn*, which has a positive and significant effect when GAIO = 1. This indicates that amongst members of a GAIO, firms tend to form new ties with their partners' partners: 'triadic closure' occurs.

The next interesting result is that for *Firm_types.* In both subgroups, we observe a similar pattern: negative and significant for SME-SME firm pairs compared to SME-Large firm pairs; and positive and significant for Large-Large dyads as compared to SME-Large. Remember from the background models that at t_0, there was only a positive significant effect for Large-Large firm pairs. Now looking at these results, it appears that the negative significant effect that appears at t_1 for SME-SME pairs holds for both subgroups of GAIO = 0 and GAIO = 1. To further investigate these results, and assess whether these effect are stronger in one of the subgroups, we resort to estimating the AME from a full model containing a product term for *GAIO*Firm_types*. More on this in section 4.4.3.

Lastly, we look at the categorical variable *Research*. Noticeably, in subgroup GAIO = 0, two fundamental research firms are more likely to form new ties (positive and significant effect) than a fundamental and an applied research firm. For two applied research firms, there is no significant effect. In subgroup GAIO = 1, there is a negative and significant effect for two applied research firms, indicating that they are less likely to form a new partnership than a fundamental and applied firm. Here, there is no significant effect for two fundamental firms. Again, we further investigate these results by estimating the AME from a full model with a product term for *GAIO*Research*, in section 4.4.3.

Lastly, we see that the control variable *Mean_proj* has a positive and significant effect in subgroup GAIO = 0, but no effect in subgroup GAIO = 1. As *Mean_proj* is intended as a proxy for R&D collaboration potential, this would indicate that a higher joint collaboration potential at t_0 leads to a higher probability of collaboration at t_1, but only if two firms do not share a GAIO.

The variables *Geog* and *Dummy_same_industry* are not significant in the split models.

4.4. AME models

Now we turn our attention to the main analysis, where we follow the method proposed by Karaca-Mandic et al. (2012) to fit a logistic regression model on the full dataset, incorporate product terms for the interactions of interest, and look at the Average Marginal Effects (AME) to assess the significance and extent of interaction effects. Hereafter we'll refer to this set of models on the full dataset as "AME models". The AME models will give insight into the change in Pr(Y=1) that each predictor variable can make when GAIO is either 0 or 1.

A total of 9 AME models were estimated on the complete dataset, an overview of which variables they include is given in Table 6 in below. As the focus of this method is on interpreting the Average Marginal Effects, these are provided in text, while the AME model estimates are provided in Appendix E. Again, note that a significant coefficient on the product term is not required for significant interaction to be present (Ai and Norton, 2003; Powers, 2005; Greene, 2010; Karaca-Mandic et al., 2012)

AME model 1 includes only the control variables. Both *Mean_proj* and *Mean_partn* have a significant influence on the outcome variable, *dummy_same_industry* has no significant effect, see Appendix E, hence it is not included in further models. Due to multicollinearity between *Mean_proj* and *Mean_partn*[3], only the predictor with the strongest effect was included in further AME models: *Mean_proj*.

Table 6: Overview of the variables included in each of the AME models. An X indicates that the variable or product term was included in the model.

	AME model								
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
GAIO			X	X	X	X	X	X	X
SR		X	X	X	X	X	X	X	X
ND		X	X	X	X	X	X	X	X
Shared_partn		X	X	X	X	X	X	X	X
Geog		X	X	X	X	X	X	X	X
Firm_types		X	X	X	X	X	X	X	X
Research		X	X	X	X	X	X	X	X
Mean_proj	X	X	X	X	X	X	X	X	X
Mean_partn	X								
Dummy_same_variable	X								
*GAIO*SR*				X					
*GAIO*ND*					X				
*GAIO*Shared_partn*						X			
*GAIO*Geog*							X		
*GAIO*Firm_types*								X	
*GAIO*Research*									X

[3] Variance Inflation Factors for Mean_proj and Mean_partn in the full model were assessed and were both ~4,18, where VIF values over 2,5 can be cause for concern in logistic regression (Midi et al., 2010).

AME model 2 includes the control variable and all independent variables except GAIO, AME model 3 includes all independent variables.

AME models 4-9 each add a single product term to AME model 3. AIC values and McFadden's pseudo-R2 reported in Appendix E indicate that adding product terms (Models 4-9) does not increase the model's performance as compared to AME model 3. A notable exception is Model 5, in which a product term for *Network distance * GAIO* is added, which leads to an improvement in model fit compared to AME model 3 (lower AIC, higher pseudo-R2). This improvement is significant at the 0.1% level, as measured with a likelihood ratio test, see Appendix F. Though other models with product terms also have somewhat lower AIC or higher pseudo-R2 values than AME model 3, none of these differences are significant at the 5% level, as shown by likelihood ratio tests.

Variance Inflation Factors were tested for all AME models, and were all under 2,5 except when product terms were included and except in AME model 1, which was solved as described above. As such, there was no cause for concern with regards to multicollinearity (Midi et al., 2010).

4.4.1. How to interpret AMEs

Interpretation of AMEs requires some knowledge with regards to what they mean, which is most easily obtained by explaining how they are calculated. For categorical variables, their calculation, and hence interpretation, is relatively straightforward. For a binary variable, e.g. in our set GAIO, the Marginal Effect (ME) for a single observation is obtained by first setting GAIO to the value 0 for that observation and calculating the predicted probability of Y = 1 using the logit model. Then, GAIO is set to 1, and again Pr(Y=1) is calculated. The difference between the predicted probabilities is the ME of GAIO for that observation. The AME is then the average of the ME over all observations in the dataset. Hence, the AME of a binary variable indicates how much Pr(Y=1) changes, on average, when the binary variable is changed from 0 to 1. For a categorical variable, the AME indicates how much Pr(Y=1) changes, on average, when that categorical variable takes on the level of interest compared to the baseline level.

For continuous variables, the interpretation is somewhat less straightforward. Similarly, the AME is the average of the ME over all observations. However, the Marginal Effect (ME) for a continuous variable is calculated as the first derivative of the probability function relating the independent variable, X, to Pr(Y=1) (StataCorp, 2013; Leeper, 2014). In other words: the slope or the *instantaneous rate of change* for that function. The function is not linear per se, thus the slope can change for different values of the independent variable. For interpretation purposes, this means that, if X increases by some very small unit, e.g. 0.0001, then Pr(Y=1) would change by about 0.0001*AME (Richard Williams, 2017), which is the slope at that 'moment'. A common interpretation then, one also followed by Karaca-Mandic et al. (2012), is that if X increases by 1 unit, Pr(Y=1) changes by 1*AME (Karaca-Mandic et al., 2012). However, this is not exactly true, as the slope of the function changes when the value of X changes (R. Williams, 2017). As a metaphor: stating that Pr(Y=1) will change by the AME if X increases by 1 unit is no more valid than stating that someone driving a car at 80 kilometers per hour will have travelled 80 kilometers, if the time driving is increased by 1 hour. It is probably a good approximation, but there really is no guarantee, as the driving speed can change, as can the AME.

Moreover, a 1-unit increase in X may not always provide a realistic value. E.g. the continuous variable SR in this analysis is bound [-1, 1]. Hence, a 1-unit increase only makes sense for observations with SR ≤ 0.

That said, even the criticizers have indicated that the common interpretation of “a 1-unit increase in the predictor variable leads to an x increase in Pr(Y=1)” provides a good approximation, or at least one that is understandable in practice (R. Williams, 2017; Schechter, 2017). So, we shall continue the interpretation of results by assuming that the AME indicates the change in Pr(Y=1) for a 1-unit increase in X, while keeping in mind the above-mentioned objections against using such an interpretation.

Continuing, the interpretation of interactions also becomes easier in this way. Following Karaca-Mandic et al. (2012), we can now state that the interaction effect of a continuous and a binary variable is the change in Pr(Y=1) for a 1-unit change in the continuous variable, as the binary variable changes from 0 to 1. In our case, the interaction effect of e.g. *SR*GAIO* is the difference between the AME of Skill Relatedness when GAIO = 0 and the AME of Skill Relatedness when GAIO = 1.

These AMEs are calculated as follows: First, GAIO is set to 0 for all observations, leaving all other independent variables as they are. Then, the ME for each observation is calculated for the variable of interest as described above (e.g. the first derivative for continuous variables, the absolute difference in Pr(Y=1) for a change in categorical variables). Again, the average of these ME when GAIO = 0 is the AME for GAIO = 0 for the variable of interest. Then, GAIO is set to 1 for all observations, and the same calculation is repeated, giving the AME for GAIO = 1 for the variable of interest. The difference between these two is, in essence, the interaction effect (Karaca-Mandic et al., 2012; Williams, 2012).

4.4.2. AME models 2 and 3

The Average Marginal Effects of **AME models 2 and 3** are given Table 7 on the next page. Comparing the AME for AME models 2 and 3, we overall see a similar pattern of sign and significance of most predictor variables. The variables *SR, Shared_partn, Firm_typesLarge_Large, ResearchFund_Fund* and *Mean_proj* all have a significant and positive effect on new tie formation in both models. The variables *Geog* and *Firm_typesSME_SME* have a significant and negative effect in both models. The variable *ND* has a negative significant effect in Model 2, but when GAIO is added, it loses its significance. Also, when GAIO is added as a variable, the influence of SR as a predictor is slightly decreased: the AME is slightly lower and the p-value becomes higher. GAIO itself has a positive and significant effect in Model 3.

Table 7: Average Marginal Effects of AME Models 2 and 3 predictors on Pr(Y=1).

AME Model 2					
Variable	**AME**	**SE**	**p-value**	**CI**	**sign.**
SR	0,0006	0,0002	0,0003	[0,0003; 0,0009]	***
ND	-0,0004	0,0001	0,0074	[-0,0006; -0,0001]	**
Shared_partn	0,0006	0,0001	0,0000	[0,0004; 0,0008]	***
Geog	-0,0004	0,0002	0,0348	[-0,0008; -0,0000]	*
Firm_typesSME_SME	-0,0010	0,0002	0,0000	[-0,0014; -0,0007]	***
Firm_typesLarge_Large	0,0060	0,0008	0,0000	[0,0043; 0,0076]	***
ResearchAppl_Appl	-0,0003	0,0002	0,2021	[-0,0007; 0,0001]	n.s.
ResearchFund_Fund	0,0008	0,0004	0,0259	[0,0001; 0,0016]	*
Mean_proj	0,0001	0,0000	0,0018	[0,0000; 0,0001]	**
AME Model 3					
Variable	**AME**	**SE**	**p-value**	**CI**	**sign.**
GAIOyes	0,0026	0,0004	0,0000	[0,0017; 0,0034]	***
SR	0,0004	0,0002	0,0095	[0,0001, 0,0007]	**
ND	0,0000	0,0001	0,7437	[-0,0003, 0,0002]	n.s.
Shared_partn	0,0004	0,0001	0,0000	[0,0002, 0,0005]	***
Geog	-0,0004	0,0002	0,0410	[-0,0008, 0,0000]	*
Firm_typesSME_SME	-0,0011	0,0002	0,0000	[-0,0014, -0,0007]	***
Firm_typesLarge_Large	0,0062	0,0009	0,0000	[0,0045, 0,0079]	***
ResearchAppl_Appl	-0,0003	0,0002	0,1237	[-0,0007, 0,0001]	n.s.
ResearchFund_Fund	0,0009	0,0004	0,0199	[0,0001, 0,0017]	*
Mean_proj	0,0001	0,0000	0,0012	[0,0000; 0,0001]	**

CI = 95% Confidence Interval. Sign. = Significance level, codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 'n.s.' 1. SE calculated using delta-method (Leeper, 2017).

If we take a closer look at the results of AME model 3, and interpret the AME, we can state that when two firms are in the same GAIO, their probability to form a new tie is, on average, 0,26 percentage points (p.p.) higher than two firms that are not in the same GAIO, holding all other variables constant. As mentioned, for continuous variables such as SR we interpret that, for a 1-unit increase in *SR*, two firms are 0,04p.p. more likely to form a new tie on average, holding all other variables constant. A one-unit increase in *Shared_partn* leads to a 0,06 p.p. increase in Pr(Y=1), and a one-unit increase in ND leads to a 0,04p.p. decrease in Pr(Y=1). For *Geog*, a 1-unit increase in *Geog* (meaning a 100km increase in distance), reduces the probability of a new tie forming by 0,04 p.p., on average. For the *Firm_types* categorical variable, we interpret that, on average, two SME are 0,11p.p. less likely to collaborate when compared to an SME and large firm and two large firms are 0,62p.p. more likely to collaborate than an SME and large firm, holding all other variables constant. For the *Research* categorical variable, we can see that two firms involved in fundamental research are 0,09p.p. more likely to collaborate compared to pairs where one firm is involved in fundamental and the other in applied research. For firm pairs where both are involved in applied research, there is no significant difference in the probability of new tie formation compared to when one firm is involved in fundamental and the other in applied research.

4.4.3. AME Results Interaction Models

Now we look at the AME models with product terms included, AME models 4-9. As mentioned, the coefficient for the product terms in the models is not a reliable source to assess the interaction effect (Hoetker, 2007; Berry et al., 2010). Hence, following Karaca-Mandic et al. (2012), we assess the effect of the interactions by looking at the AMEs for predictor variables when GAIO = 0 and when GAIO = 1.

The AME for AME models 4-9 are given in Table 8. For the sake of brevity, we only report the AMEs for the product term variables, rather than the AMEs for all variables.

First **AME model 4**, where a product term for *SR*GAIO* is included. From Table 8, we can see that the AME for SR when GAIO = 0, is 0,0004 (95% CI: [0,0001; 0,0007]). This result is significant at the 5% level. This means that when the observations are treated as having no shared GAIO, a 1-unit increase in SR leads to a 0,04p.p. increase in Pr(Y=1), on average. The AME for SR when GAIO = 1 is not significant at the 5% level. These results show the same patterns as the split models: when firms share a GAIO, their likeliness to collaborate is not affected by cognitive distance, but when they don't share a GAIO, cognitive proximity facilitates collaboration.

Looking at the AME for **AME model 5**, where *ND*GAIO* is included as a product term, ND has an AME of -0,0021 (95% CI: [-0,0031; -0,0010]) that is significant at the 0,1% level when GAIO = 1, but no significant effect when GAIO = 0. Thus, when firms share a GAIO, a one unit increase in ND leads to a 0,21p.p. lower probability of collaboration. If firms don't share a GAIO, there is no effect of ND on the probability to collaborate.

When we compare this with the split models, we see the same pattern when GAIO = 1: the negative effect of ND indicates that firms tend to form new ties locally in their embedded networks. However, when GAIO = 0, we (surprisingly) saw a positive effect from ND in the split models. This effect is not visible in the AME of AME model 5, thus there is a discrepancy in what the two model types show. We'll get back to this in the conclusion and discussion.

In **AME model 6** a product term for *Shared_partn*GAIO* is included. We see a significant positive AME of 0,0009 (95% CI: [0,0005; 0,0013]) when GAIO = 1, meaning that on average, a 1-unit increase in *Shared_partn* leads to a 0,09p.p. higher probability of a new tie forming, but only if firms share a GAIO. When firms do not share a GAIO, there is no significant AME of Shared_partn. These results of the AME are in line with the results from the split models and indicate that firms who share a GAIO tend to form new partnerships with their partners' partners.

In **AME model 7** a product term for *Geog*GAIO* is included. Recall that in AME model 3, the AME for Geog was negative and significant. In AME model 7, the AME for *Geog* is no longer significant, either for the group GAIO = 0 or GAIO = 1. Hence, inclusion of a product term *Geog*GAIO* appears to render the effect of Geographical distance on new tie formation non significant in terms of the AME. The latter is in line with results from the split models, where there was also no significant effect of Geog in either subgroup model.

In **AME model 8** a product term for *Firm_types*GAIO* is included. From Model 3, it can be seen that two SME firms, on average, have a 0,11p.p. lower probability of new tie formation than an SME and large firm. Two large firms, on average, have a 0,62p.p. higher probability of new tie formation than an SME and large firm.

Recall that from the split models, the effect of this categorical variable was the same for the two subgroups GAIO = 0 and GAIO = 1, namely in both cases positive and significant for two large firms and negative and significant for two SME, both compared to an SME and large firm. Now when we look at the AME from AME model 8, we see the same pattern: the AME are significant when GAIO = 0 and when GAIO = 1, for both SME-SME and Large-Large firm combinations, when compared to SME-Large. If we look closer at the AME, we see that amongst firms who do not share a GAIO (GAIO = 0), the probability that an SME-SME firm duo forms a new tie, is 0,07p.p. lower, on average, than the probability that an SME-Large duo forms a new tie. The probability that a Large-Large duo forms a new tie, is 0,52p.p. higher than that an SME-Large duo forms a new tie. When firms do share a GAIO (GAIO = 1), these patterns stay the same, but the differences between the duos in terms of Pr(Y=1) becomes larger. Now, the probability that an SME-SME firm duo forms a new tie is 0,27p.p. lower, on average, than the probability that an SME-Large duo forms a tie. Large-Large firm duos now have a 1,25p.p. higher probability, on average, than SME-Large duos.

What becomes more obvious now, and was not directly clear from the split models, is that the effects are much stronger, in absolute percentage points change in Pr(Y=1), when firms share a GAIO than when they don't. Thus, it appears that GAIO exacerbate the existing tendencies of Large firms to preferentially collaborate with each other, and to a lesser extent with SME. A further exploration of this effect is done through an effect plot, see section 4.4.4.

In **AME model 9** a product term for *Research*GAIO* is included. When GAIO = 0, the AME for Fund.-Fund. firm pairs (both involved in fundamental research) is significant and positive compared to Appl.-Fund. pairs. For Appl.-Appl. pairs, the AME is not significant. This is in line with the results from the split models. From the AME, we can interpret that, when firms do not share a GAIO, two firms involved in fundamental research are, on average, 0,13p.p. more likely to form a new partnership between them than firm pairs where there is only one, or zero firms involved in fundamental research. Thus, when firms do not share a GAIO, it appears that firms involved in fundamental research preferentially collaborate with each other.

When GAIO = 1, there is no significant AME from AME model 9. This indicates that the type of research that firms are involved in does not influence their probability to collaborate when they share a GAIO. Thus, it seems that sharing a GAIO takes away the tendency of firms involved in fundamental research to search each other out.

With regards to two applied research firms forming partnerships, we see a discrepancy between the split model results and the AME model 9 results. In the split model, there was a negative effect on collaboration for Appl.-Appl. firm pairs when they shared a GAIO. This effect is not visible in the AME. We get back to the discrepancy in the conclusion and discussion.

Table 8: AME of interacted variables in Models 4-9, split over GAIO.

Model	Variable	GAIO	AME	SE	p-value	CI	sign.
Model 4 (GAIO*SR)	*GAIOyes*	-	0,0026	0,0005	0,0000	[0,0017; 0,0035]	***
	SR	no	0,0004	0,0002	0,0202	[0,0001; 0,0007]	*
		yes	0,0007	0,0006	0,1893	[-0,0004; 0,0018]	n.s.
Model 5 (ND*SR)	*GAIOyes*	-	0,0014	0,0003	0,0000	[0,0008; 0.0021]	***
	ND	no	0,0001	0,0001	0,2960	[-0,0001; 0,0003]	n.s.
		yes	-0,0021	0,0005	0,0001	[-0,0031; -0,0010]	***
Model 6 (Shared_partn*SR)	*GAIOyes*	-	0,0025	0,0004	0,0000	[0,0016; 0,0034]	***
	Shared_partn	no	0,0002	0,0002	0,3729	[-0,0002; 0,0005]	n.s.
		yes	0,0009	0,0002	0,0000	[0,0005; 0,0013]	***
Model 7 (Geog*SR)	*GAIOyes*	-	0,0026	0,0004	0,0000	[0,0017; 0,0034]	***
	Geog	no	-0,0003	0,0002	0,0892	[-0,0007; 0,0000]	n.s.
		yes	-0,0008	0,0007	0,2330	[-0,0021; 0,0005]	n.s.
Model 8 (Firm_types*GAIO)	*GAIOyes*	-	0,0026	0,0005	0,0000	[0,0018; 0,0035]	***
	SME_SME	no	-0,0007	0,0001	0,0000	[-0,0010; -0,0004]	***
		yes	-0,0027	0,0007	0,0002	[-0,0042; -0,0013]	***
	Large_Large	no	0,0052	0,0009	0,0000	[0,0035; 0,0069]	***
		yes	0,0125	0,0029	0,0000	[0,0067; 0,0183]	***
Model 9 (Research*GAIO)	*GAIOyes*	-	0,0026	0,0005	0,0000	[0,0018; 0,0035]	***
	Appl_Appl	no	-0,0001	0,0002	0,4509	[0,0018; 0,0035]	n.s.
		yes	-0,0012	0,0008	0,1298	[-0,0027; 0,0003]	n.s.
	Fund_Fund	no	0,0013	0,0005	0,0058	[0,0004; 0,0021]	**
		yes	0,0005	0,0011	0,6397	[-0,0017; 0,0027]	n.s.

CI = 95% Confidence Interval. Sign. = Significance level, codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 'n.s.' 1. SE calculated using delta-method (Leeper, 2017).

4.4.4. Effect Plot

Though the AME provide insight, some interactions are best understood with effect plots (Huang and Shields, 2000; Hoetker, 2007). As an addition to the AME results, a further exploration of the interaction effect of *Firm_types* and *GAIO* is done through such an effect plot.

We plot the curves of how Pr(Y=1) changes depending on Firm_types for both GAIO = 0 and GAIO = 1, based on AME model 8. The levels of Firm_types become the values on the x-axis, and Pr(Y=1) is plotted on the y-axis. All other variables must be set at certain values, most commonly one uses the mean of the sample for each variable (Huang and Shields, 2000; Williams, 2012). For numerical variables (*SR, ND, Shared_partn, Geog, Mean_proj)*, we follow this method. However, for categorical variables *(Research*), the mean does not make much sense, hence a sensible approach would be to set these at their baseline levels (*Fund_Appl*). If we use the coefficients from AME model 8 to create the function, the link function for Pr(Y=1 | Firm_types, SR) based on AME model 8, $\boldsymbol{\phi}$, then becomes:

$$\Phi = -6.850 + 1.516*GAIO - 1.474*Firm\_typesSME\_SME + 1.913*Firm\_typesLarge\_Large + 0.369*GAIO*Firm\_typesSME\_SME - 0.497*GAIO*Firm\_typesLarge\_Large + 0.299*-0.037 - 0.006*3.374 + 0.293*0.204 - 0.291*0.896 + 0.067*1.829$$

Subsequently, we plot the full function for Pr(Y=1 | Firm_types, SR):

$$Pr(Y=1) = e^{(\Phi)} / (1 + e^{(\Phi)})$$

The result of this function plot can be seen in Figure 6 on the next page. As discussed in section 4.4.3., the trend for firm types influence Pr(Y=1) is the same for GAIO = 0 and GAIO = 1, with the difference being that within GAIO = 1 the effects are stronger in terms of absolute change in Pr(Y=1). This becomes immediately clear from the visual representation.

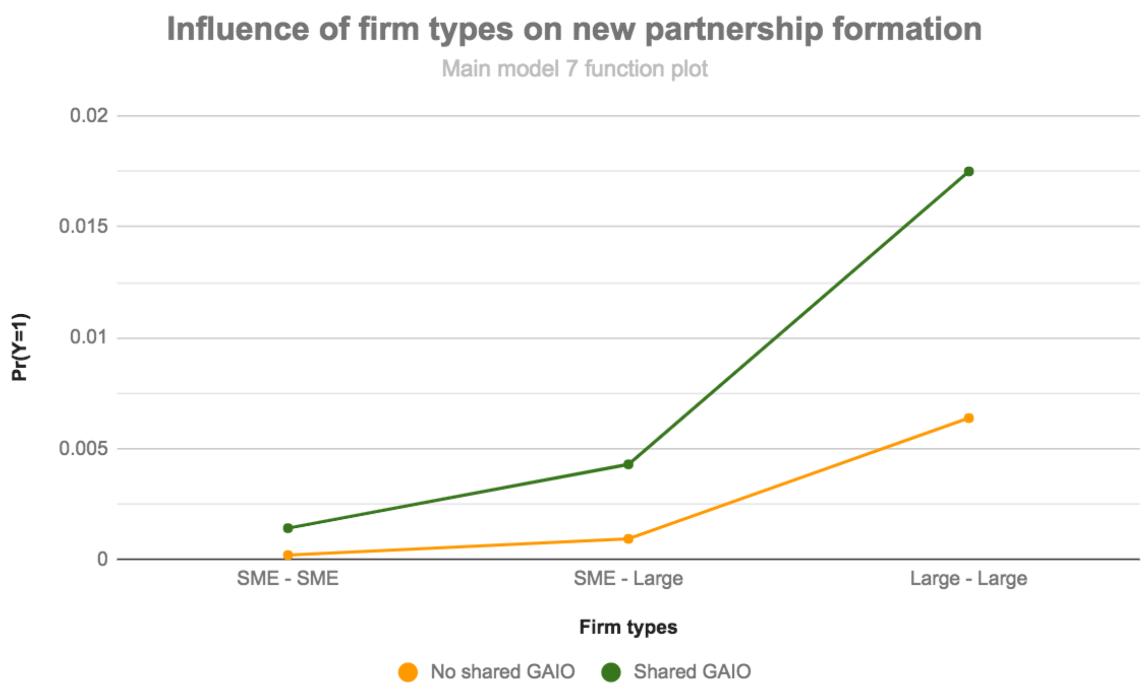


Figure 6: Function plot for Pr(Y=1) as a function of Firm_types, for different values of GAIO. All other variables are set at their mean value (numerical variables) or baseline level (categorical variables).

5. Conclusion

Using a binary logit regression model, it was assessed which determinants influence the formation of new partnerships between firms in public-private R&D partnerships in The Netherlands. The determinants were derived from Boschma's five proximities, which are well-established in literature to influence interfirm collaboration. Indeed, their influence was present in this analysis too. The influence of government-affiliated intermediary organizations was also assessed by examining their direct influence on new partnership formation, and on the proximity determinants. Results show that GAIO affect collaboration both directly and alter the natural tendencies of firms to collaborate.

Firstly, shared membership of a GAIO was found to be a direct predictor of the probability of a new partnership forming between firms. On average, two firms that were members of the same GAIO had a 0,26p.p. higher probability of forming a new partnership than two firms that preivously did not share a GAIO. This result is significant at the 0,1% level, and has a 95% confidence interval of [0,17 - 0,34]. Though this increase in probability seems small, keep in mind that the overall probability of forming a new partnership in the dataset was 0,14%. A first conclusion can thus be that sharing a GAIO can almost triple the probability of forming a new tie between two firms, on average, providing evidence to **accept Hypothesis 5**.

Regarding cognitive proximity, it was showed that revealed skill relatedness, as a proxy for cognitive proximity, has a positive and significant effect on the probability of a new partnership forming between two firms. The extent of this effect is hard to express numerically, as there are many shortcomings in the use of AMEs (discussed extensively in section 4.4.1) and because skill relatedness as a measure itself is quite abstract: a 'one unit increase in skill relatedness' does not mean much in practice. What can be said however, is that firms who share a high cognitive proximity are more likely to collaborate than those who don't, confirming theory and previous empirical evidence (Nooteboom, 2000; Frenken et al., 2007; Paier and Scherngell, 2011). These results provide evidence to **accept Hypothesis 1**.

Looking at the interaction of skill relatedness and GAIO, we see that when two firms do not share a GAIO, skill relatedness has a positive significant effect on collaboration. When firms do share a GAIO, skill relatedness has no significant effect on collaboration. Thus, it appears that GAIO are able to overcome the hurdle to collaborate that cognitive distance poses. This provides evidence to **accept Hypothesis 1a**. This is the first quantitative empirical evidence to support case study evidence and theoretical frameworks that GAIO, or SSI, are able to create awareness of matching goals amongst actors, thereby facilitating collaboration between actors of which, given their cognitive distance, we would have not expected a collaboration (van Lente et al., 2003; Backhaus, 2010; Kilelu et al., 2011; Hannon et al., 2014).

With regards to social proximity, two proxies were used: network distance in the collaboration network at t_0 and the number of mutual direct partners between firms i and j at t_0. Network distance had no significant effect in the model on the full dataset without product terms (AME model 3). This indicates that the network distance between two firms in the R&D collaboration network did not, at least significantly, affect the chance of new tie formation at a later point. The second proxy used for social proximity is the number of mutual partners that firms i and j have. This determinant is positive and significant in model 3, which is in line with previous findings and the concept of 'triadic closure' (Gulati, 1995; Hanaki et al., 2010). On average, for each one unit increase in the number of mutual partners (e.g. one more mutual partners), two firms are 0,04p.p. more likely to form a new partnership. Thus, with regards to Boschma's proximities theory, it seems that the concept of triadic closure holds true in this analysis, but network distance is not a significant predictor. As such, there is **support for Hypothesis 2**.

Looking at the role that GAIO play with regards to social distance, we see some interesting results. Firstly, there is a discrepancy between the models with regards to *Network distance* when GAIO = 0: in the AME model, we see a non-significant effect, but in the split models, we see a slightly positive effect (p = 0,046). The latter indicates that when firms do not share a GAIO, they preferentially collaborate with firms that are farther from them in their embedded networks. Clearly, this effect requires further research as it is counterintuitive, and the models do not show the same result.

With regards to firms that do share a membership of a GAIO (GAIO = 1), both the models on the full dataset and the split models indicate the same: Network Distance has a negative influence on the probability of a new tie forming. This indicates that within the 'subgroups' of GAIO members, there is an effect of firms 'acting locally' within their embedded networks'. Firms mostly form new partnerships with firms that are indirectly familiar to them, which is in line with previous findings (Rycroft and Kash, 2004; Baum et al., 2010).

A similar pattern becomes visible for the number of shared partners: this predictor is positive and significant when GAIO = 1, and not significant when GAIO = 0, in both the main and split models. This indicates that the triadic closure concept does hold true in the subgroups of GAIO members, but not amongst firms that do not share a GAIO. Taking together all results, it appears that social proximity does not influence firms who do not share a GAIO, but it has a positive and significant effect on new partnership formation amongst those firms who do share a GAIO already. This provides **support to accept Hypothesis 2b (and reject Hypothesis 2a)**.

Regarding geographical proximity, we see a negative significant effect of geographical distance on the probability of collaboration in the AME model without product terms (AME model 3). On average, for each 100km increase in distance between two firms, they are 0,04p.p. less likely to collaborate. This provides **support for Hypothesis 3** and is in line with previous empirical findings of the effect of geographical distance on collaboration (Autant-Bernard et al., 2007; Paier and Scherngell, 2011).

The combined effect of GAIO and geographical distance is not entirely clear. In both the AME model and split models, there is no significant effect of geographical distance, either when GAIO = 0 or GAIO = 1. Thus, it appears that geographical distance is simply obsolete overall as a predictor of new tie formation when we let GAIO and Geog interact. Importantly, it should be noted that the effects of geographical distance are quite weak in AME model 3 as well, with a p-value only just under the 'significant threshold' of 0.05. Based on this data with these p-values, it seems quite bold to draw conclusions on the effect of GAIO on geographical distance as a determinant of collaboration, as there is no unequivocal evidence pointing in a single direction. Hence, further research is warranted and **Hypothesis 3a is neither rejected nor supported**.

Regarding organizational proximity, two proxies were used. The first is whether the firm dyad concerns two SME, two large firms or one large firm and one SME. From the basic model without product terms, it became clear that large firms are more likely to form new partnerships than SME. On average, two SME are 0,11p.p. less likely to form a new partnership, and two large firms are 0,62p.p. more likely to form a new partnership than an SME and large firm. Hence, SME-SME dyads have the lowest probability to collaborate, then large-SME dyads, and large-large dyads have the highest probability to form new partnerships, holding all other variables constant. The second proxy for organizational proximity is whether firms are involved in fundamental research, or not (in which case they are involved in applied research). In the basic model, without product terms, we see a positive significant effect for two fundamental firms forming a new partnership: on average, two firms involved in fundamental research are 0,09p.p. more likely to form a new partnership than a fundamental and applied research firm. There was no significant effect for two applied firms. Thus, firms who engage in fundamental research appear to specifically search each other out, but firms who do not engage in fundamental research appear to have no preference with regards to the type of research their collaboration partner does. Though the results from the SME and Large collaboration preferences are interesting, if we look strictly at the hypothesis with regards to organizational proximity, there does not appear to be a 'homophily' effect in the collaboration choices. The exception being the fundamental research firms. Hence, there is **little evidence to support Hypothesis 4.**

With regards to the influence of GAIO on the two proxies for organizational proximity, there are two notable findings. First, we see that shared membership of a GAIO increases the probability to collaborate for all firm size combinations, due to the positive effect of GAIO. However, the absolute increase in probability for two large firms collaborating when GAIO = 1, is much higher than the absolute increases for an SME and large firm, or two SME collaborating. In essence then, sharing a GAIO leads all firms to have a higher probability to form new partnerships, but this effect is stronger for large firms. GAIO thus appear to enhance the existing effects of firm sizes on the likeliness to form new partnerships.

With regards to the second proxy, we do see some of the hypothesized effect. When firms have a shared membership of a GAIO, the tendency of fundamental research firms to search other out is no longer present, compared to when firms do not share a GAIO. Thus, it appears that GAIO can overcome this homophily tendency of fundamental research firms, which provides **some support for Hypothesis 4a**. Notably, there is a discrepancy of the model results when it comes to two applied research firms collaborating when the share a GAIO: it is not clear whether there is no effect or a negative effect. This, and the notion that we see the moderating effect of GAIO on only one level of one proxy for organizational proximity, leads us to say that **Hypothesis 4a cannot be confirmed nor rejected without further research**. It should be noted that, within the context of public-private R&D partnerships, the results with regards to organizational proximity are not necessarily surprising. We will further discuss this in section 6.1.

This paper started with the following research question:

To what extent do government-affiliated intermediary organizations influence the formation of new partnerships between firms, and to what extent do they alter the influence of different types of proximity on new partnership formation?

Now we can state that sharing a GAIO can, on average, almost triple the probability that any two firms form a new partnership. Sharing a GAIO also reduces the influence of cognitive proximity on new partnership formation, and thereby GAIO act as facilitators of collaborations where knowledge recombination can take place. With regards to social proximity, there is quite strong evidence that amongst members of the same GAIO, existing tendencies of firms to engage in 'triadic closure' and act locally within embedded networks, are enhanced. GAIO do not act as network-wide brokers, but rather cluster the network amongst their members. With regards to organizational proximity, GAIO appear to reduce the tendency of firms involved in fundamental research to preferentially collaborate with one another. Lastly, GAIO exacerbate the natural tendencies of mostly large firms to collaborate with each other, and, to a lesser extent, with SME.

6. Discussion

This study has provided some first quantitative insights into the influence of GAIO on new partnership formation, and their role in influencing the natural tendencies of firms to collaborate. The goal of this paper was twofold: to provide insight for policymakers who aim to implement GAIO as a means to stimulate collaboration; and to contribute to academic literature on GAIO, and thereby also to some extent on SII. Both contributions will be discussed below, after which we finish the paper with the most important shortcomings and future research avenues that are worth exploring.

6.1. GAIO as a policy tool

Some of the results were contrary to the hypothesized effects, but are not necessarily surprising when considering the context of the case, where GAIO are a policy tool to set up public-private R&D consortia.

The findings with regards to SME and large firms collaborating were contrary to the hypothesized homophily effect of organizational proximity. However, the patterns are not inexplicable at all. Firstly, as large firms generally have more resources available than SME, and R&D is a high cost and risky business, it is not surprising that they are more likely to engage in R&D collaborations. Previous research on R&D collaborations in public-private context shows this as well (Autant-Bernard et al., 2007). Secondly, one of the specific aims of the GAIO in this study case was to involve more SME in public-private research. At the same time however, large firms are often needed in these projects for their (financial) resources and extensive knowledge. Thus, it makes sense that the combination of SME and large firms collaborating is also increased amongst members of the same GAIO. Looking at the background analyses, or stylized facts, we also see that the probability that an SME and Large firm collaborate was not significantly different from the probability that two SME collaborate before the GAIO were implemented, but it was significant after. Thus, in the period after the GAIO were implemented, relatively more SME-Large combinations were forged as compared to SME-SME combinations. Though, as mentioned, we cannot talk of causal effects, these patterns are interesting and may point to the notion that the presence of GAIO leads to the involvement of relatively more SME in public-private R&D collaborations.

Another interesting finding was that firms involved in fundamental research tend to search each other out when not sharing a GAIO, but this effect is no longer present when two firms share a GAIO. This seems to implicate that GAIO enable the combination of fundamental and applied research firms. Again, the context of the policy matters: the GAIO policy is specifically aimed at involving industry more in public-private research. Hence the focus of the research agendas is not only on fundamental research, where governments often contribute most (Cohen, 2006), but also on taking fundamental research through to the applied stage. It thus makes sense that amongst GAIO, there is a tendency to couple firms involved in fundamental research to those involved in applied research. This would be in line with propositions by Winch and Courtney (2007) who indicate that intermediaries can provide a space for technology development that combines public long-term research needs and the near-to-market short-term research needs of firms (Winch and Courtney, 2007).

With regards to geographical proximity, we saw that this was not a significant predictor in the split models, or in the AME model with product term, but there was some negative significant effect in the AME model without product term (AME model 3). If we look again at the background models, or stylized facts, we see that geographical distance was a strong predictor of ties at t_0, but not so much at t_1. Taking these results together, it appears as if geographical distance was an important inhibitor of collaboration at t_0, but no more at t_1. This would imply that from t_0 to t_1 there has been a shift in that firms first collaborated with local partners in public-private partnerships, whereas now they literally look further. Given the notion that GAIO have been implemented at the national scale, this seems to be a sign that indeed implementation of intermediaries at the national level are able to override local clusters. Further research is warranted to substantiate this claim though.

One of the most clear results was that social proximity is an important influence within member groups of the GAIO, but not amongst firms who do not share a GAIO. This appears to contrast previous evidence from case studies, in which brokerage by SII is considered effective (Klerkx and Leeuwis, 2009/7; Bakici et al., 2013). However, it should be noted that these studies mainly consider ‘brokerage’ as connecting different types of actors, e.g. for-profit firms, government, academia and NGO. Brokerage amongst firms, or across structural holes, is not explicitly considered in these other studies.

As mentioned in theory, a possible explanation for this clustering is the notion that GAIO act as demand articulators and agenda setters (van Lente et al., 2003; Abbate et al., 2013; Kivimaa, 2014). In this specific case, a firm became affiliated with a GAIO if the firm was involved in a public-private research project at the time when the GAIO were started. This means that the initial members of the GAIO, were pre-emptively selected based on them being active already in public-private R&D collaborations at the time. With this (relatively narrow) group of actors, each GAIO established a research agenda, which they subsequently carried out by initiating and funding new R&D projects. The GAIO are then acting almost as implementers of bottom-up policy, steering research towards the avenues that their members support. Indeed, that intermediaries can act as implementers of bottom-up policy has been pointed out in literature before (Backhaus, 2010). It does not seem unreasonable to assume then, that those firms that were most influential or important in the knowledge network when the research agendas were set up, have had the most influence in determining what should be in these agendas. From the data, we see that firms find new partners relatively close in their embedded networks. This would indicate that a group of actors that is tight-knit, becomes even more compact. There thus seems to be a clear risk that a select group of actors are benefitting from the policy they partially shaped. This notion has been described in literature before (Fromhold-Eisebith and Eisebith, 2005). The selective nature of GAIO may thus put those that are not pre-emptively considered a part of the relevant research community, at a distinct disadvantage.

Another important result is that cognitive proximity is of no influence on new partnership formation if firms are members of the same GAIO. This indicates that GAIO facilitate collaboration over cognitive distance. This is in line with preliminary empirical evidence by Mahnke (2008), who perform a single case-study on a for-profit intermediary and show that intermediaries are able to bridge cognitive distance between agents (Mahnke et al., 2008). It should be noted however, that within GAIO, actors shared a slightly higher cognitive proximity on average than actors between GAIO[4]. This makes sense, as the twelve separate GAIO in this case study are industry-specific, which is often the case for GAIO or SII (van Lente et al., 2003; Winch and Courtney, 2007; Dalziel, 2010). This could contribute then, to the finding that member firms who are in those industry-specific GAIO clusters collaborate more amongst each other than in between clusters, as firms from unrelated knowledge bases are unlikely to collaborate (Nooteboom et al., 2007; Janssen, 2015). Taking this together, it appears that though firms may collaborate with the goal of developing new technologies or products, they still mostly form new partnerships with firms that are in their sector-specific subgroup.

What this may indicate, is that new partnership formation of firms under the GAIO policy occurs on the basis of a 'related R&D diversification strategy': firms are able to diversify by accessing knowledge that is new to them (not affected by cognitive distance), but still relatively related to their current knowledge base (amongst their sector-specific GAIO co-members) (Dosi, 1982; Frenken, 2017). Thus, their knowledge development trajectories are still constrained by path-dependency. Frenken (2017) poses that related diversification strategies are well-supported by policy for collaborations and R&D subsidies, which would be in line with what we see from this analysis. Another important point he makes is that large firms tend to benefit more from such policies and subsidized R&D projects than SME, which is again in line with what we see in this analysis: large firms are much more likely to form new partnerships than SME (Frenken, 2017).

Hence, it is plausible that the implementation of (industry-specific) GAIO in a national knowledge network facilitates a related diversification strategy amongst firms. Such a related diversification strategy will create 'related variety' within the knowledge available in the network or in the country. There are definitely positive outcomes of this process: On a regional level, related variety has been shown to enhance employment, economic growth and innovation (Frenken et al., 2007; Boschma and Iammarino, 2009; Castaldi et al., 2015).

At the same time however, Frenken (2017) poses: *"most likely these innovations will be closely related to existing strengths of incumbents [...] simply because innovation systems are built up by incumbent actors largely setting agendas on their own"* (Frenken, 2017). Indeed, it has been shown that related diversification leads to incremental innovation, whereas radical innovation is more likely to come from unrelated diversification (Castaldi et al., 2015). In the same line, it has been shown that policy demarcating a specific range of included sectors and research topics may hamper synergetic potential and create risks of

---

[4] Though as mentioned in section 4.3., there is still a good spread of the SR variable within GAIO. Also, the standard deviation of SR for the GAIO = 1 subgroup is higher than the standard deviation for the GAIO = 1 subgroup. So we deem the results of the analysis valid.

over-specialization (Fromhold-Eisebith and Eisebith, 2005). Indeed, a country that develops only via related diversification runs a *"serious risk of running out of [diversification] opportunities" (Frenken, 2017)*. GAIO themselves also run the risk of getting locked into the existing regime (Kivimaa, 2014).

Thus, ideally, governments attempt to steer at both a related and unrelated diversification strategy, which will provide the platform for both incremental and radical innovation, benefitting both the economy and society in the long run (Frenken, 2017). However, this is not an easy task. As a potential solution, Janssen (2015) has proposed the notion of 'cross-specialization', where governments specifically steer at knowledge recombination between specialized industries already present in the country (Janssen, 2015; Frenken, 2017). In this way, countries can build on their existing deep and specialized knowledge bases, which, when recombined, are also expected to create most radical innovation (Fleming, 2001; Janssen, 2015). Indeed, this is also what the GAIO studied in this paper, the TKI in The Netherlands, have attempted to do to implicitly. All TKI mention in their research agendas and through their communication channels that they aim to stimulate "cross-over collaborations" between the different TKI ("Topconsortia voor Kennis & Innovatie," 2016; Janssen et al., 2016). However, according to a recent interim evaluation of the TKI, it has proven difficult to get the TKI to collaborate to launch exactly those projects (Janssen et al., 2016). The results from this study are in line with those findings, indicating that each TKI may lead their members to focus more on each other than on those that remain outside their own group. Thus, within the TKI, new partnerships are forged that are not influenced by cognitive proximity, but between TKI, where cognitive distance between firms is higher on average, little new partnerships are formed.

Partly, this may be attributable to the way the TKI are set-up, with which goals in mind. In the case of the TKI, the main goal was to *"create excellent public-private partnerships in research and innovation"* that would have both *"economic and societal relevance"*. Involving industry to help shape the public research agenda together with academia and government would lead to more demand-driven research and more willingness of industry to contribute (financial) resources to public research ("Staatscourant nr. 18236," 2012). Given these goals, the use of GAIO seems an appropriate tool. Indeed, with regards to contribution of firms' financial resources to public research, the policy has proven more than successful in meeting its financial goals (Janssen et al., 2016). Also, from a macroeconomic view, in recent years The Netherlands has climbed the ranks of competitive economies worldwide, reaching the top 5 in several rankings ("Nederland is het beste land ter wereld, op Zwitserland na," 2016, "Nederlandse economie draait nog steeds mee in wereldtop," 2017). With regards to the aspired economic relevance then, The Netherlands is clearly on the right path and the TKI seem to provide the right support for that. With regards to the aspired societal relevance, the TKI currently do not seem to provide the support for those knowledge recombinations of which we expect the most radical innovation needed to solve complex societal challenges.

Thus, if one questions whether the implementation of industry-specific GAIO is 'good' for a country, that may depend on the policy goal in mind and subsequently how these GAIO are setup to operate. If the goal is to create economic growth and incremental innovation, spurred mostly by large firms, then the use of sectoral GAIO and bottom-up research agenda setting seems appropriate. On the other hand, when looking for technological breakthroughs and radical innovation to solve ‘grand societal challenges’, a policy that focuses more on unrelated diversification of firms is more likely to achieve those goals (Frenken, 2017). In the latter case, industry-specific GAIO may not be the most suitable policy tool to use.

6.2. Academic contribution

This study has made a few contributions to academic literature, the importance of which may be judged by the reader.

Firstly, an empirical contribution was made to the GAIO literature. As mentioned, still very little of SII is known, let alone GAIO (Dalziel, 2010; Abbate et al., 2013). This study has provided insights into one of the key functions of (systemic) intermediaries: network building. It was shown quantitatively that GAIO stimulate firms to start new partnerships, increasing the probability of a new tie forming between two firms in the same GAIO up to three times compared to two firms that do not share the same GAIO. This is in line with empirical findings by Kivimaa et al. (2014) who, in their case study on a GAIO, find that stakeholders assign a strong role to GAIO in terms of new network formation (Kivimaa, 2014).

It becomes more interestingly though, when the framework for GAIO, or SII, is coupled to that of the proximities theory by Boschma (2005). This gives more insight as to how specifically GAIO are able to increase collaboration amongst their member firms. As far as the author is aware, this is the first attempt to combine the literature streams on GAIO and the proximities theory by Boschma. This is relevant, as SII are often overlooked in research (Dalziel, 2010). One interesting finding is that the effect of cognitive proximity on collaboration, which is well-established in literature, is no longer a significant determinant when two firms are both members of the same GAIO. On these two counts, facilitating collaboration and bridging cognitive distance, we see an alignment of the quantitative data with qualitative data from previous studies, providing a stronger empirical foundation for the role of (government-affiliated) intermediaries.

An attempt was also made in this study to investigate the role of GAIO in bridging social distance, investigating their role as ‘brokers’ that can reach across structural holes in the network (Burt, 2004; Long et al., 2013). This study has provided no evidence to support this theory. Quite the contrary, we see an opposite effect: network effects of acting locally and triadic closure appear stronger when firms share a GAIO. Note that this could be due to the specific way that the GAIO operate or are set-up. Nevertheless, these findings that directly contradict previous research provide an interesting starting point for future research.

Lastly, methodologically, as far as the author is aware this is the first attempt at using Skill Relatedness as a measure for cognitive proximity at the dyad level or for interfirm collaboration. As the results regarding this proxy have been according to expectation, and its performance is stable in both the split models and models on the full dataset, it appears that using this proxy is viable for other research on R&D collaborations. Considering that it is available for all firms, and not dependent on a firm's ability to patent or publish, it may even provide a superior alternative to those most commonly used proxies for cognitive distance.

6.3. Shortcomings and future research

It is important to note the potential shortcomings of this research. As always, there are several, of which some weigh more heavily than others.

Looking at the research design, a notable problem arises that is difficult to work around: there is no 'control group'. As the policy is applied at the national level, not a similar collection of firms exists within the same contexts that the used firm set can be compared to. This is a common issue in researching and evaluating policy (Isserman and Merrifield, 1982). Hence, we can only compare amongst those within the intervention: those that share a GAIO and those that do not share a GAIO. There is no ground to make statements with regards to firms that are not involved in the knowledge network as a whole. To give some insight and provide as much context as possible, background analyses were performed comparing the 'before' and 'after' situation. However, this provides no evidence as to what the effects of the intervention as a whole are. Nevertheless, the cross sectional research design, as used here, provides a first hint at identifying associations that can be further studies with more rigorous data collection (Mann, 2003). Given the lack of quantitative empirical evidence on GAIO, or even SII, this study should be viewed as a first attempt to provide that, rather than as a conclusive study with regards to their effects or effectiveness. A starting point, rather than an end point.

Another limitation in the research design is that firms that entered at t_1 could not be included in the analysis. This was the case because the social network variables were derived from data on the firms at t_0. This is unfortunate as now it cannot be assessed whether the GAIO have any influence on attracting new firms to join the knowledge network. From the recent evaluation of this specific policy, we do know that new firms join the network at t_1 (Janssen et al., 2016). However, there is no research as to what specific determinants influenced those 'new' firms to join. For future research, it would be insightful to assess these newly joined firms as well. This would require an alternative source of data to model the underlying social networks. The latter would be a good idea anyway, as the used R&D network may not be completely representative of the underlying social networks that are built upon non-public collaborations and interpersonal networks. The use of qualitative research methods to complement the quantitative data may also provide relevant insights with regards to why (new) firms choose to join the network.

With regards to the statistical model used, it is clear that the nonlinear logistic regression model has its disadvantages. We have seen this from some discrepancies that the split models and models on the full dataset gave. Most likely these discrepancies have come about from the different mathematical methods used in the models. The full models use the AME. This means that GAIO is set to 0 or 1 for all observations in the set, independent of its original value, to calculate the AME of the variable of interest at different values of GAIO. In the split models, GAIO is not set to certain values, but instead its observed value is used to split the data, and then assess (interaction) effects of the variable of interest. As the discussion with regards to the interpretation of interaction effects in nonlinear models is a long-standing one and remains unsolved, it is perhaps worth exploring whether a different, linear, model can be used to confirm the findings of this study.

Also, there were some specific difficulties with regards to the data. One important problem was data quality. As the TKI each report their own data, there is little consistency over the twelve TKI in how they report data, specifically how they deliver the participant names. The latter caused duplicates in the firm entities, which, from a network perspective, can have far-stretching consequences. Using several fuzzy string matching methods, and quite some manual labour, these duplicates have been removed. The same was true for project names. Sometimes it was unclear whether a project that was registered in the next year was a new project, or a continuation of an old project but slightly re-named (or misspelled). A combination of fuzzy string matching on the project name, and matching on project participants was used, to remove as much as possible duplicate projects. Thus, to the best of my ability, the data has been cleaned. However, there is unfortunately no guarantee that no duplicate projects or participant names remain. The only thing that may alleviate these problems in the future is a better control of the input data.

A last opportunity for future research is to look at the societal impact of projects. As mentioned in this paper, the goal of combining different knowledge bases is to achieve radical innovation that has high economic and societal value. The latter is especially important considering the notion that policy should be direct towards the public good. The data used in this analysis contains some information with regards to the societal relevance of projects. This would provide an interesting opportunity to, for example, assess whether collaborations that occur over a larger cognitive distance, or social distance, truly generate projects with more relevance to society.

Clearly, there is a lot left to discover about GAIO and SII. More empirical research on GAIO is absolutely needed to get a complete picture of their potential role in facilitating innovation.

References

Abbate, T., Coppolino, R., Schiavone, F., 2013. Linking Entities in Knowledge Transfer: The Innovation Intermediaries. Journal of the Knowledge Economy 4, 233–243.

Ahuja, G., 2000. Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. Administrative science quarterly 45, 425–455.

Ahuja, G., Lampert, C.M., 2001. Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions. Strategic Management Journal 22, 521–543.

Ai, C., Norton, E.C., 2003. Interaction terms in logit and probit models. Economics letters 80, 123–129.

Asheim, B.T., Boschma, R., Cooke, P., 2011. Constructing Regional Advantage: Platform Policies Based on Related Variety and Differentiated Knowledge Bases. Regional studies 45, 893–904.

Audretsch, D.B., Feldman, M.P., 1996. R&D Spillovers and the Geography of Innovation and Production. The American economic review 86, 630–640.

Autant-Bernard, C., Billand, P., Frachisse, D., Massard, N., 2007. Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies*. Papers in regional science: the journal of the Regional Science Association International 86, 495–519.

Backhaus, J., 2010. Intermediaries as Innovating Actors in the Transition to a Sustainable Energy System. Central European Journal of Public Policy 86–109.

Bakici, T., Almirall, E., Wareham, J., 2013. The role of public open innovation intermediaries in local government and the public sector. Technology Analysis & Strategic Management 25, 311–327.

Bala, V., Goyal, S., 2000. A Noncooperative Model of Network Formation. Econometrica: journal of the Econometric Society 68, 1181–1229.

Balland, P.-A., 2012. Proximity and the Evolution of Collaboration Networks: Evidence from Research and Development Projects within the Global Navigation Satellite System (GNSS) Industry. Regional studies 46, 741–756.

Barabasi, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.

Batterink, M.H., Wubben, E.F.M., Klerkx, L., Omta, S.W.F. (onno), 2010. Orchestrating innovation networks: The case of innovation brokers in the agri-food sector. Entrepreneurship and Regional Development 22, 47–76.

Baum, J.A.C., Cowan, R., Jonard, N., 2010. Network-Independent Partner Selection and the Evolution of Innovation Networks. Management science 56, 2094–2110.

Berry, W.D., DeMeritt, J.H.R., Esarey, J., 2010. Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential? American journal of political science 54, 248–266.

Boschma, R., 2005. Proximity and Innovation: A Critical Assessment. Regional studies 39, 61–74.

Boschma, R., Iammarino, S., 2009. Related Variety, Trade Linkages, and Regional Growth in Italy. Economic geography 85, 289–311.

Brambor, T., Clark, W.R., Golder - Political analysis, M., 2005, 2005. Understanding interaction models: Improving empirical analyses. academic.oup.com.

Bryman, A., 2015. Social Research Methods. Oxford University Press.

Burt, R. s., 2004. Structural Holes and Good Ideas. The American journal of sociology 110, 349–399.

Cantner, U., Meder, A., 2007. Technological proximity and the choice of cooperation partner. Journal of Economic Interaction and Coordination 2, 45–65.

Cassiman, B., Veugelers, R., 2006. In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition. Management science 52, 68–82.

Castaldi, C., Frenken, K., Los, B., 2015. Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting. Regional studies 49, 767–781.

Choirat, C., Honaker, J., Imai, K., King, G., Lau, O., 2017. Zelig: Everyone's Statistical Software.

Cohen, E., 2006. Theoretical foundations of industrial policy. European Investment Bank Papers 11, 84–106.

Cohen, W.M., Levinthal, D.A., 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative science quarterly 35, 128–152.

Dalziel, M., 2010. Why do innovation intermediaries exist, in: Proceedings of DRUID Summer Conference, London. researchgate.net, pp. 16–18.

Dhanaraj, C., Parkhe, A., 2006. Orchestrating Innovation Networks. Academy of management review. Academy of Management 31, 659–669.

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

Doz, Y.L., Olk, P.M., Ring, P.S., 2000. Formation Processes of R&D Consortia: Which Path to Take? Where Does It Lead? Strategic Management Journal 21, 239–266.

Feldman, M.P., 1999. The New Economics Of Innovation, Spillovers And Agglomeration: Areview Of Empirical Studies. Economics of Innovation and New Technology 8, 5–25.

Feldman, M.P., Kelley, M.R., 2006. The ex ante assessment of knowledge spillovers: Government R&D policy, economic incentives and private firm behavior. Research policy 35, 1509–1521.

Firth, D., 1993. Bias reduction of maximum likelihood estimates. Biometrika 80, 27–38.

Fleming, L., 2001. Recombinant Uncertainty in Technological Search. Management science 47, 117–132.

Frenken, K., 2017. A Complexity-Theoretic Perspective on Innovation Policy. Complexity, Governance & Networks 0, 35–47.

Frenken, K., Van Oort, F., Verburg, T., 2007. Related variety, unrelated variety and regional economic growth. Regional studies.

Fromhold-Eisebith, M., Eisebith, G., 2005. How to institutionalize innovative clusters? Comparing explicit top-down and implicit bottom-up approaches. Research policy 34, 1250–1268.

Gassmann, O., Daiber, M., Enkel, E., 2011. The role of intermediaries in cross-industry innovation processes. R&D Management 41, 457–469.

GoogleMaps Geocoding API [WWW Document], 2017. . Developer's Guide | Google Maps Geocoding API | Google Developers. URL https://developers.google.com/maps/documentation/geocoding/intro

Graf, H., 2006. Networks in the Innovation Process. Books.

Greene, W., 2010. Testing hypotheses about interaction terms in nonlinear models. Economics letters 107, 291–296.

Gulati, R., 1995. Does Familiarity Breed Trust? The Implications of Repeated Ties for Contractual Choice in Alliances. Academy of Management journal. Academy of Management 38, 85–112.

Hanaki, N., Nakajima, R., Ogura, Y., 2010. The dynamics of R&D network in the IT industry. Research policy 39, 386–399.

Hannon, M., Skea, J., Rhodes, A., 2014. Facilitating and coordinating UK energy innovation through systemic innovation intermediaries, in: 5th International Conference on Sustainability Transitions. pp. 27–29.

Heinze, G., Ploner, M., 2016. logistf: Firth's Bias-Reduced Logistic Regression.

Hensher, D.A., Stopher, P.R., 1979. Behavioural Travel Modelling. Croom Helm.

Hoekman, J., Frenken, K., van Oort, F., 2009. The geography of collaborative knowledge production in Europe. The Annals of regional science 43, 721–738.

Hoetker, G., 2007. The use of logit and probit models in strategic management research: Critical issues. Strategic Management Journal 28, 331–343.

Howells, J., 2006. Intermediation and the role of intermediaries in innovation. Research policy 35, 715–728.

Huang, C., Shields, T.G., 2000. Interpretation of interaction effects in logit and probit analyses.

American politics quarterly 28, 80–95.

Hurmelinna-Laukkanen, P., Olander, H., Blomqvist, K., Panfilii, V., 2012. Orchestrating R&D networks: Absorptive capacity, network stability, and innovation appropriability. European Management Journal 30, 552–563.

Imai, K., King, G., Lau, O., 2008. Toward A Common Framework for Statistical Analysis and Development. Journal of Computational Graphics and Statistics.

Isserman, A.M., Merrifield, J., 1982. The use of control groups in evaluating regional economic policy. Regional science and urban economics 12, 43–58.

Jackman, S., 2015. pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University.

Janssen, M., 2015. Cross-specialization and structural holes: the case of the Dutch Topsectors. Papers in Evolutionary Economic Geography 15.

Janssen, M., den Hertog, P., Korlaar, L., Erven, B., Te Velde, R., Bogers, M., Vankan, A., 2016. Tussenevaluatie TKI-toeslagregeling. Dialogic Innovatie & Interactie, in opdracht van Ministerie van Economische Zaken Nederland.

Kanda, W., Gonzaléz, P. del R., Hjelm, O., Bienkowska, D., 2015. A function of innovation systems approach for analysing the roles of intermediaries in eco-innovation, in: Global Cleaner Production and Sustainable Consumption Conference,1-4 November,Sitges-Barcelona,Spain.

Karaca-Mandic, P., Norton, E.C., Dowd, B., 2012. Interaction terms in nonlinear models. Health services research 47, 255–274.

Kilelu, C.W., Klerkx, L., Leeuwis, C., Hall, A., 2011. Beyond knowledge brokering: an exploratory study on innovation intermediaries in an evolving smallholder agricultural system in Kenya. Knowledge Management for Development Journal 7, 84–108.

King, G., Zeng, L., 2001. Logistic Regression in Rare Events Data. Political analysis: an annual publication of the Methodology Section of the American Political Science Association 9, 137–163.

Kivimaa, P., 2014. Government-affiliated intermediary organisations as actors in system-level transitions. Research policy 43, 1370–1380.

Klein Woolthuis, R., Lankhuizen, M., Gilsing, V., 2005. A system failure framework for innovation policy design. Technovation 25, 609–619.

Klerkx, L., Leeuwis, C., 2009/7. Establishment and embedding of innovation brokers at different innovation system levels: Insights from the Dutch agricultural sector. Technological forecasting and social change 76, 849–860.

Klerkx, L., Leeuwis, C., 2008. Matching demand and supply in the agricultural knowledge infrastructure: Experiences with innovation intermediaries. Food Policy 33, 260–276.

Knoben, J., Oerlemans, L., 2006. Proximity and inter-organizational collaboration. International Journal of Management Reviews 8, 71–89.

Knockaert, M., Spithoven, A., Clarysse, B., 2014/1. The impact of technology intermediaries on firm cognitive capacity additionality. Technological forecasting and social change 81, 376–387.

Leeper, T.J., 2014. User manual R package “margins.”

Leeper, T.J., 2017. margins: Marginal Effects for Model Objects.

van Lente, H., Hekkert, M., Smits, R., van Waveren, B., 2003. Roles of Systemic Intermediaries in Transition Processes. International Journal of Innovation Management 07, 247–279.

van Lente, H., Hekkert, M., Smits, R., van Waveren, B., 2011. Systemic intermediaries and transition processes, in: Guy, S., Moss, T., Medd, W., Marvin, S. (Eds), Shaping Urban Infrastructures: Intermediaries and the Governance of SocioTechnical Networks. Earthscan, London, pp. 36–52.

Levén, P., Holmström, J., Mathiassen, L., 2014. Managing research and innovation networks: Evidence from a government sponsored cross-industry program. Research policy 43, 156–168.

Loecher, M., Ropkins, K., 2015. RgoogleMaps and loa: Unleashing R Graphics Power on Map Tiles. Journal of Statistical Software.

Long, J.C., Cunningham, F.C., Braithwaite, J., 2013. Bridges, brokers and boundary spanners in collaborative networks: a systematic review. BMC health services research 13, 158.

Mahnke, V., Wareham, J., Bjorn-Andersen, N., 2008. Offshore middlemen: transnational intermediation in technology sourcing. Journal of Information Technology Impact 23, 18–30.

Mann, C.J., 2003. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. Emergency medicine journal: EMJ 20, 54–60.

McAdam, R., Reid, R., 2001. SME and large organisation perceptions of knowledge management: comparisons and contrasts. Journal of Knowledge Management 5, 231–241.

Midi, H., Sarkar, S.K., Rana, S., 2010. Collinearity diagnostics of binary logistic regression model. Journal of Interdisciplinary Mathematics 13, 253–267.

Nagler, J., 1991. The effect of registration laws and education on US voter turnout. The American political science review 1393–1405.

Nederland is het beste land ter wereld, op Zwitserland na., 2016. Financieele Dagblad.

Nederlandse economie draait nog steeds mee in wereldtop., 2017. NOS.

Neffke, F., Otto, A., Weyh, A., 2016. Inter-industry labor flows. IAB-Discussion Paper.

Nicholas, J., Ledwith, A., Perks, H., 2011. New product development best practice in SME and large organisations: theory vs practice. European Journal of Innovation Management 14, 227–251.

Nooteboom, B., 2000. Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance. Journal of Management and Governance 4, 69–92.

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., van den Oord, A., 2007. Optimal cognitive distance and absorptive capacity. Research policy 36, 1016–1034.

Onukwugha, E., Bergtold, J., Jain, R., 2015. A primer on marginal effects--Part I: Theory and formulae. PharmacoEconomics 33, 25–30.

Paier, M.F., Scherngell, T., 2008. Determinants of Collaboration in European R&D Networks: Empirical Evidence from a Binary Choice Model Perspective.

Paier, M., Scherngell, T., 2011. Determinants of Collaboration in European R&D Networks: Empirical Evidence from a Discrete Choice Model. Industry and Innovation 18, 89–104.

Phelps, C., Heidl, R., Wadhwa, A., 2012. Knowledge, Networks, and Knowledge Networks. Journal of management 38, 1115–1166.

Ponds, R., Van Oort, F., Frenken, K., 2007. The geographical and institutional proximity of research collaboration*. Papers in regional science: the journal of the Regional Science Association International 86, 423–443.

Porter, M.E., 1986. Competition in Global Industries. Harvard Business Press.

Powell, W.W., Koput, K.W., Smith-Doerr, L., 1996. Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology. Administrative science quarterly 41, 116–145.

Powers, E.A., 2005. Interpreting logit regressions with interaction terms: an application to the management turnover literature. Journal of Corporate Finance 11, 504–522.

Rainey, C., 2016. Compression and Conditional Effects: A Product Term Is Essential When Using Logistic Regression to Test for Interaction*. Political Science Research and Methods 4, 621–639.

R Core Team, 2017. R: A Language and Environment for Statistical Computing.

Rigby, D., Zook, C., 2002. Open-market innovation. Harvard business review 80, 80–9, 129.

Rosenberg, N., 1990. Why do firms do basic research (with their own money)? Research policy 19, 165–174.

RStudio Team, 2015. RStudio: Integrated Development Environment for R.

Russo, M., Caloffi, A., Rossi, F., Righi, R., 2016. Designing performance-based incentives for innovation intermediaries:Evidence from regional innovation poles. CIMR Research Working Paper Series.

Rycroft, R.W., Kash, D.E., 2004. Self-organizing innovation networks: implications for globalization. Technovation 24, 187–197.

Schechter, C., 2017. Quote on Statalist forum.

Scherngell, T., Barber, M.J., 2009. Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. Papers in Regional Science 88, 531–546.

Schilling, M.A., Phelps, C.C., 2007. Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation. Management science 53, 1113–1126.

Schumpeter, J.A., 1934. The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle. Transaction publishers.

Staatscourant nr. 18236., 2012. .

StataCorp, L., 2013. STATA User manual margins command. StataCorp LLC.

Topconsortia voor Kennis & Innovatie [WWW Document], 2016. . URL https://www.topsectoren.nl/innovatie

Torre, A., Rallet, A., 2005. Proximity and Localization. Regional studies 39, 47–59.

Wallace, J.R., 2012. Imap: Interactive Mapping.

Warwick, K., Nolan, A., 2014. Evaluation of Industrial Policy: Methodological issues and policy lessons. OECD Publishing, OECD Science, Technology and Industry Policy.

Weitzman, M.L., 1998. Recombinant Growth. The quarterly journal of economics 113, 331–360.

Wernerfelt, B., 1984. A Resource-based View of the Firm. Strategic Management Journal, Vol. 5, 171–180.

Williamson, O.E., 1981. The Economics of Organization: The Transaction Cost Approach. The American journal of sociology 87, 548–577.

Williams, R., 2012. Using the margins command to estimate and interpret adjusted predictions and marginal effects. The Stata Journal 12, 308–331.

Williams, R., 2017. Using Stata's Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects.

Williams, R., 2017. Marginal Effects for Continuous Variables [WWW Document]. University of Notre Dame course material. URL https://www3.nd.edu/~rwilliam/xsoc73994/Margins02.pdf

Winch, G.M., Courtney, R., 2007. The Organization of Innovation Brokers: An International Review. Technology Analysis & Strategic Management 19, 747–763.

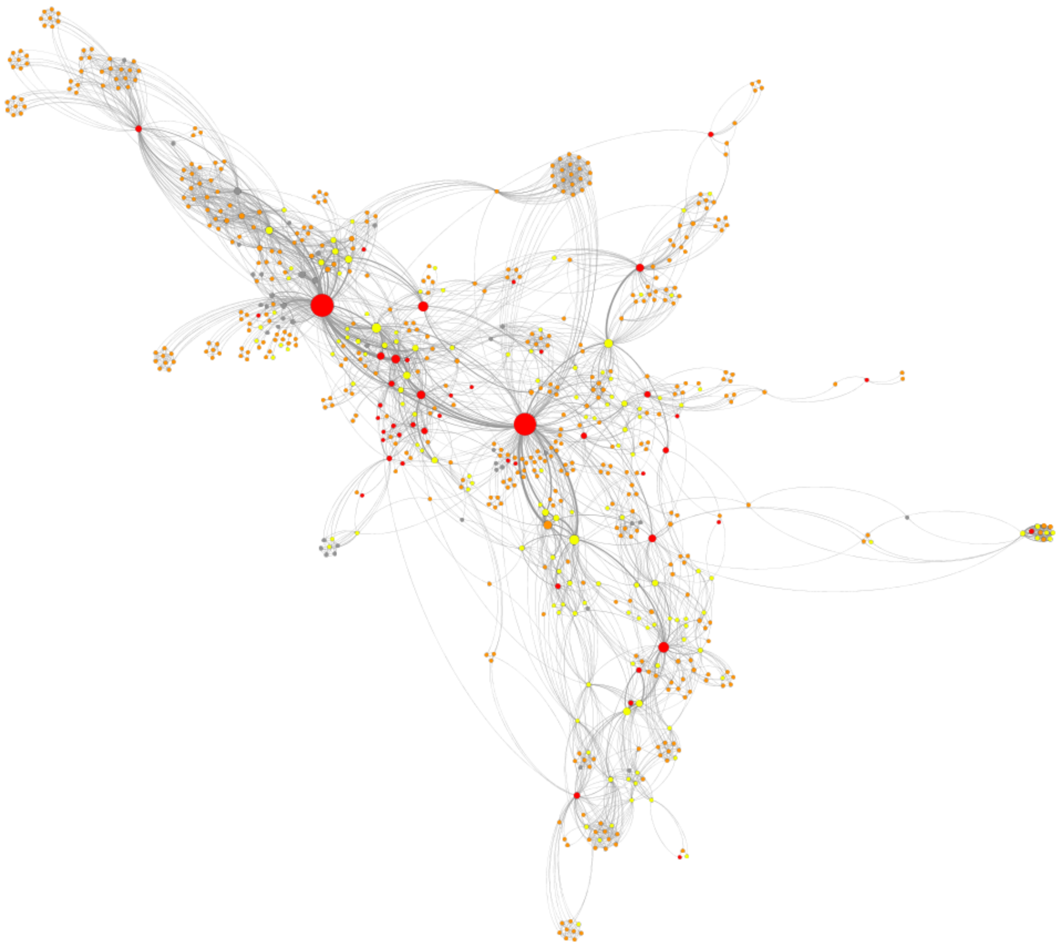
Wolfinger, R.E., Rosenstone, S.J., 1980. Who votes. New Haven. CT: Yale University From.

Zeileis, A., Hothorn, T., 2002. Diagnostic Checking in Regression Relationships. R News.

Appendices

Appendix A

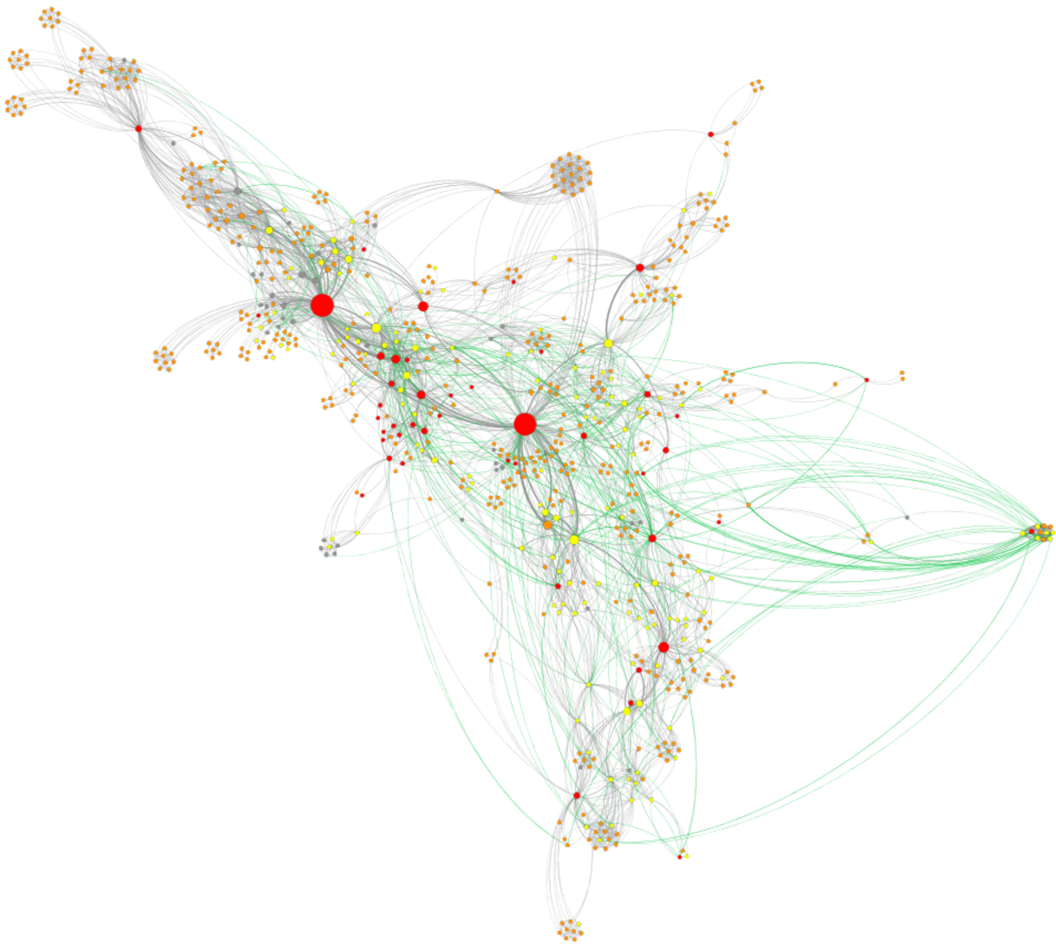
Visualizations of the Dutch knowledge network of public-private research partnerships per year. In 2013, all then current projects are taken into account and ties between organizations are indicated in grey. For each subsequent year, newly formed ties are indicated in green, whereas all ties from previous year(s) are indicated in grey. Public organizations are indicated in red, SMEs in orange, Large firms in yellow.



2013

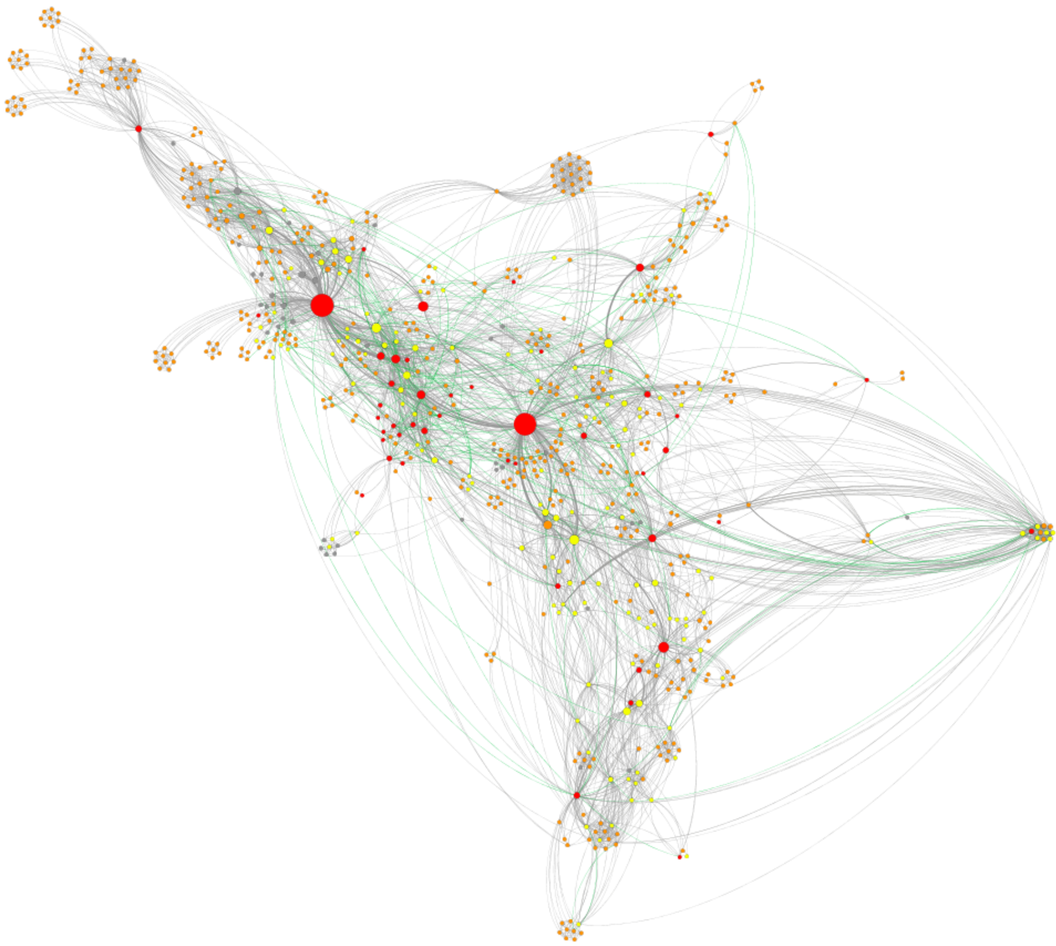
Appendix A continued

Visualizations of the Dutch knowledge network of public-private research partnerships per year. In 2013, all then current projects are taken into account and ties between organizations are indicated in grey. For each subsequent year, newly formed ties are indicated in green, whereas all ties from previous year(s) are indicated in grey. Public organizations are indicated in red, SMEs in orange, Large firms in yellow.



2014

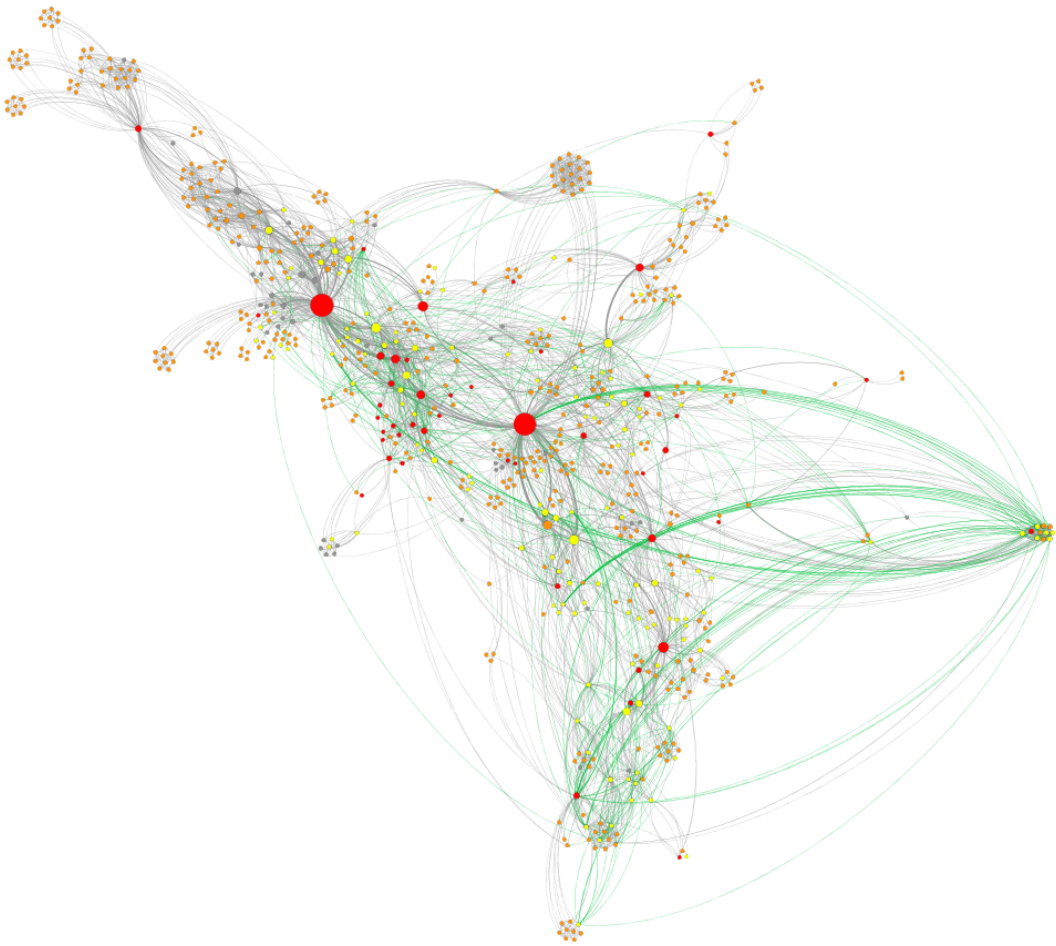
Visualizations of the Dutch knowledge network of public-private research partnerships per year. In 2013, all then current projects are taken into account and ties between organizations are indicated in grey. For each subsequent year, newly formed ties are indicated in green, whereas all ties from previous year(s) are indicated in grey. Public organizations are indicated in red, SMEs in orange, Large firms in yellow.



2015

Appendix A continued

Visualizations of the Dutch knowledge network of public-private research partnerships per year. In 2013, all then current projects are taken into account and ties between organizations are indicated in grey. For each subsequent year, newly formed ties are indicated in green, whereas all ties from previous year(s) are indicated in grey. Public organizations are indicated in red, SMEs in orange, Large firms in yellow.



2016

Appendix B

Comparison of results of the normal logit model with two logit models specifically for rare events data: King and Zeng's relogit model and a logit model with penalized likelihood estimation, Firth's method. AME model 3 was used for the comparison.

Table B.1: Results of standard binary logistic regression on AME Model 3.

Variable	Coefficient	SE	p-value	Sign.
GAIOyes	1,357	0,163	< 2e-16	***
SR	0,306	0,117	0,009	**
ND	-0,032	0,099	0,744	n.s.
Shared_partn	0,272	0,058	3,03e-06	***
Geog	-0,285	0,138	0,040	*
Firm_typesSME_SME	-1,293	0,203	1,78e-10	***
Firm_typesLarge_Large	1,681	0,147	< 2e-16	***
ResearchAppl_Appl	-0,245	0,159	0,124	n.s.
ResearchFund_Fund	0,496	0,185	0,007	**
Mean_proj	0,066	0,020	0,001	**
Constant	-6,692	0,369	< 2e-16	***

SE = Standard Error; Sign. = significance; Significance codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 'n.s.' 1

Table B.2: Results of King and Zeng's adapted logistic regression for AME model 3.

Variable	Coefficient	SE	p-value	Sign.
GAIOyes	1,358	0,163	< 2e-16	***
SR	0,304	0,117	0,009	**
ND	-0,030	0,099	0,764	n.s.
Shared_partn	0,273	0,058	2,68e-06	***
Geog	-0,281	0,138	0,043	*
Firm_typesSME_SME	-1,285	0,203	2,27e-10	***
Firm_typesLarge_Large	1,679	0,147	< 2e-16	***
ResearchAppl_Appl	-0,244	0,159	0,124	n.s.
ResearchFund_Fund	0,498	0,185	0,007	**
Mean_proj	0,066	0,020	9,62e-4	***
Constant	-6,684	0,369	< 2e-16	***

SE = Standard Error; Sign. = significance; Significance codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 'n.s.' 1

Appendix B continued

Comparison of results of the normal logit model with two logit models specifically for rare events data: King and Zeng's relogit model and a logit model with penalized likelihood estimation, Firth's method. AME model 3 was used for the comparison.

Table B.3: Results of Firth's penalized adapted logistic regression for AME model 3.

Variable	Coefficient	SE	p-value	Sign.
GAIOyes	1,358	0,161	2,22e-16	***
SR	0,304	0,115	8,18e-03	***
ND	-0,030	0,098	0,763	n.s.
Shared_partn	0,273	0,058	1,12e-05	***
Geog	-0,281	0,137	3,90e-02	*
Firm_typesSME_SME	-1,285	0,200	1,49e-11	***
Firm_typesLarge_Large	1,679	0,146	< 2e-16	***
ResearchAppl_Appl	-0,244	0,157	0,123	n.s.
ResearchFund_Fund	0,498	0,183	8,34e-03	**
Mean_proj	0,066	0,020	1,31e-03	***
Constant	-6,684	0,365	< 2e-16	***

SE = Standard Error; Sign. = significance; Significance codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 'n.s.' 1

Appendix C

Correlations amongst (numeric) independent variables, for the full dataset (Table C.1), and the two subsets: the GAIO-0 set (Table C.2) and the GAIO-1 set (Table C.3.).

Table C.1: Correlations for the full dataset (n = 171198).

Variable		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
[1]	GAIO	1							
[2]	SR	0,140	1						
[3]	ND	-0,334	-0,069	1					
[4]	Shared_partn	0,427	0,079	-0,594	1				
[5]	Geog	-0,007	-0,039	0,062	-0,042	1			
[6]	Mean_proj	0,073	0,001	-0,196	0,214	-0,003	1		
[7]	Mean_partn	0,125	0,003	-0,207	0,216	-0,049	0,662	1	
[8]	Dummy_same_industry	0,133	0,341	-0,065	0,061	-0,020	-0,018	0,010	1

Table C.2: Correlations for the GAIO-0 dataset (n = 150494).

Variable		[2]	[3]	[4]	[5]	[6]	[7]	[8]
[2]	SR	1						
[3]	ND	-0,025	1					
[4]	Shared_partn	0,015	-0,596	1				
[5]	Geog	-0,043	0,060	-0,031	1			
[6]	Mean_proj	-0,012	-0,182	0,190	0,004	1		
[7]	Mean_partn	-0,023	-0,181	0,161	-0,042	0,643	1	
[8]	Dummy_same_industry	0,311	-0,036	0,011	-0,034	-0,019	-0,013	1

Appendix C continued

Correlations amongst (numeric) independent variables, for the full dataset (Table C.1), and the two subsets: the GAIO-0 set (Table C.2) and the GAIO-1 set (Table C.3.).

Table C.3: Correlations for the GAIO-1 dataset (n = 20704).

Variable		[2]	[3]	[4]	[5]	[6]	[7]	[8]
[2]	SR	1						
[3]	ND	-0,013	1					
[4]	Shared_partn	0,043	-0,621	1				
[5]	Geog	-0,006	0,095	-0,089	1			
[6]	Mean_proj	0,004	-0,209	0,258	-0,033	1		
[7]	Mean_partn	0,026	-0,168	0,246	-0,081	0,722	1	
[8]	Dummy_same _industry	0,424	0,051	-0,005	0,044	-0,056	0,013	1

Appendix D

Results of two logit models as background analysis. The first model, the t_0-bg-model, is run on collaborations in the baseline network of 2013 (t_0). Only variables that were not derived from project data are used, e.g. Skill Relatedness, Geographical distance (in 100km), Firm_types and Dummy_same_industry. In Tables D.1 and D.2 descriptives and model results of the t_0 model are given. The second model, the t_1-bg-model, is run on new collaborations in the network at t_1 (GAIO-funded projects in 2014-2016). This is essentially the same dataset as the main analysis, only the independent and control variables in the model differ. Results of this model are in Table D.3. Descriptives for the t_1 model are not given, as these are the same as in the main analysis, and thus can be read from Table 3 on page 32.

Table D.1: Descriptives of data for t_0-bg-model (n = 173 166).

Numerical variables	Mean	Min	Max	SD
IV: SR	-0,033	-1	1	0,585
IV: Geog	0,894	0	3,19	0,509
Categorical variables	**Count of 0**	**Count of 1**	**Perc. 0 (%)**	**Perc. 1 (%)**
DV: Y	171198	1968	98,86	1,14
IV Firm_types: SME_Large	-	60970	-	35,21
IV: Firm_types: SME_SME	-	103285	-	59,65
IV: Firm_types: Large_Large	-	8911	-	5,15
CV: Dummy_same_industry	161375	11791	93,19	6,81

Min = minimum value; Max = maximum value; SD = Standard Deviation

Appendix D continued

Table D.2: Results of binary logit regression of t_0-bg-model (n = 173166)

Variable		Coefficient	SE	p-value	Sign.
IV	*SR*	0,813	0,050	<2e-16	***
	Geog	-0,663	0,051	<2e-16	***
	Firm_typesSME_SME	-0,023	0,051	0,653	n.s.
	Firm_typesLarge_Large	1,109	0,073	<2e-16	***
CV	*Dummy_same_industry*	0,846	0,065	<2e-16	***
Constant	*Constant*	-4,240	0,056	<2e-16	***
Model fit	*AIC*	20244			
	R2	0,061			

CV = control variables; IV = independent variables; SE = Standard Error; Sign. = significance; Significance codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 'n.s.' 1

Table D.3: Results of binary logit regression of t_1-bg-model (n = 171 198)

Variable		Coefficient	SE	p-value	Sign.
IV	*SR*	0,629	0,127	7,87e-07	***
	Geog	-0,308	0,141	0,0293	*
	Firm_typesSME_SME	-1,511	0,196	1,13e-14	***
	Firm_typesLarge_Large	2,074	0,140	< 2e-16	***
CV	*Dummy_same_industry*	-0,169	0,261	0,5171	n.s.
Constant	*Constant*	-6,198	0.1526	< 2e-16	***
Model fit	*AIC*	3261			
	R2	0,122			

CV = control variables; IV = independent variables; SE = Standard Error; Sign. = significance; Significance codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 'n.s.' 1

Appendix E

Model estimates for AME models 1-9 are given on the next three pages.

Table E.1. contains model estimates for AME models 1-3.

Table E.2. contains model estimates for AME models 4-6.

Table E.3. contains model estimates for AME models 7-9.

Table E.1: Results of binary logit regressions on AME Models 1 - 3

	AME Model 1				AME Model 2				AME Model 3			
	Coeff.	**SE**	**p-value**	**sign.**	**Coeff.**	**SE**	**p-value**	**sign.**	**Coeff.**	**SE**	**p-value**	**sign.**
GAIOyes	-	-	-	-	-	-	-	-	1,357	0,163	< 2e-16	***
SR	-	-	-	-	0,440	0,117	1,76e-4	***	0,306	0,117	0,009	**
ND	-	-	-	-	-0,258	0,095	0,006	**	-0,032	0,099	0,744	n.s.
Shared_partn	-	-	-	-	0,433	0,054	1,65e-15	***	0,272	0,058	3,03e-06	***
Geog	-	-	-	-	-0,298	0,140	0,033	*	-0,285	0,138	0,040	*
Firm_typesSME_SME	-	-	-	-	-1,255	0,202	4,87e-10	***	-1,293	0,203	1,78e-10	***
Firm_typesLarge_Large	-	-	-	-	1,660	0,147	< 2e-16	***	1,681	0,147	< 2e-16	***
ResearchAppl_Appl	-	-	-	-	-0,201	0,158	0,203	n.s.	-0,245	0,159	0,124	n.s.
ResearchFund_Fund	-	-	-	-	0,475	0,185	0,011	*	0,496	0,185	0,007	**
Mean_proj	0,185	0,029	2,09e-10	***	0,064	0,020	0,002	**	0,066	0,020	0,001	**
Mean_partn	0,048	0,014	6,28e-4	***	-	-	-	-	-	-	-	-
Dummy_same_industry	0,178	0,246	0,470	n.s.	-	-	-	-	-	-	-	-
Constant	-7,556	0,121	< 2e-16	***	-5,732	0,339	< 2e-16	***	-6,692	0,369	< 2e-16	***
AIC	3473,5				3082				3016,8			
McFadden's pseudo-R2	0,063				0,172				0,190			

*Coeff. = Coefficient; SE = Standard Error; sign. = Significance. Significance levels: 0,000 *** 0,001 ** 0,01 * 0,05 ‘n.s.’ 1*

Table E.2: Results of binary logit regressions on AME Models 4 - 6.

	AME Model 4				AME Model 5				AME Model 6			
	Coeff.	SE	p-value	sign.	Coeff.	SE	p-value	sign.	Coeff.	SE	p-value	sign.
GAIOyes	1,384	0,166	< 2e-16	***	3,847	0,743	2,27e-07	***	1,301	0,203	1,53e-10	***
SR	0,391	0,162	0,016	*	0,316	0,117	0,007	**	0,307	0,117	0,009	**
ND	-0,030	0,099	0,761	n.s.	0,109	0,105	0,301	n.s.	-0,059	0,115	0,609	n.s.
Shared_partn	0,272	0,058	2,86e-06	***	0,207	0,062	9,14e-4	***	0,185	0,199	0,354	n.s.
Geog	-0,284	0,138	0,040	*	-0,269	0,138	0,052	n.s.	-0,284	0,138	0,040	*
Firm_typesSME_SME	-1,291	0,203	1,86e-10	***	-1,260	0,202	4,78e-10	***	-1,290	0,203	1,94e-10	***
Firm_typesLarge_Large	1,679	0,147	< 2e-16	***	1,688	0,147	< 2e-16	***	1,682	0,147	< 2e-16	***
ResearchAppl_Appl	-0,243	0,159	0,126	n.s.	-0,277	0,159	0,082	n.s.	-0,242	0,159	0,129	n.s.
ResearchFund_Fund	0,498	0,184	0,007	**	0,524	0,185	0,005	**	0,499	0,185	0,007	**
Mean_proj	0,066	0,020	0,001	***	0,065	0,020	0,001	**	0,066	0,020	9,39e-4	***
*GAIO * SR*	-0,177	0,232	0,445	n.s.	-	-	-	-	-	-	-	-
*GAIO * ND*	-	-	-	-	-1,041	0,317	0,001	**	-	-	-	-
*GAIO * Shared_partn*	-	-	-	-	-	-	-	-	0,090	0,196	0,647	n.s.
Constant	-6,704	0,369	< 2e-16	***	-7,139	0,394	< 2e-16	***	-6,588	0,434	< 2e-16	***
AIC	3018,3				3004,4				3018,6			
McFadden's pseudo-R2	0,190				0,194				0,190			

*Coeff. = Coefficient; SE = Standard Error; sign. = Significance. Significance levels: 0,000 *** 0,001 ** 0,01 * 0,05 'n.s.' 1*

Table E.3: Results of binary logit regressions on AME Models 7 - 9.

	AME Model 7				AME Model 8				AME Model 9			
	Coeff.	**SE**	**p-value**	**sign.**	**Coeff.**	**SE**	**p-value**	**sign.**	**Coeff.**	**SE**	**p-value**	**sign.**
GAIOyes	1,275	0,266	1,58e-06	***	1,516	0,225	1,75e-11	***	1,585	0,219	4,35e-13	***
SR	0,305	0,117	0,009	**	0,299	0,116	0,010	*	0,309	0,117	0,008	**
ND	-0,032	0,099	0,743	n.s.	-0,006	0,099	0,955	n.s.	-0,003	0,101	0,979	n.s.
Shared_partn	0,273	0,059	2,80e-06	***	0,293	0,058	4,26e-07	***	0,293	0,058	4,60e-07	***
Geog	-0,339	0,197	0,085	n.s.	-0,291	0,138	0,035	*	-0,283	0,138	0,040	*
Firm_typesSME_SME	-1,293	0,203	1,78e-10	***	-1,474	0,290	3,65e-07	***	-1,290	0,204	2,30e-10	***
Firm_typesLarge_Large	1,682	0,147	< 2e-16	***	1,913	0,198	< 2e-16	***	1,68	0,147	< 2e-16	***
ResearchAppl_Appl	-0,245	0,159	0,124	n.s.	-0,255	0,160	0,112	n.s.	-0,157	0,208	0,450	n.s.
ResearchFund_Fund	0,495	0,184	0,007	**	0,503	0,183	0,006	**	0,912	0,252	3,03e-4	***
Mean_proj	0,066	0,020	0,001	**	0,067	0,020	8,68e-4	***	0,065	0,020	0,001	**
*GAIO * Geog*	0,107	0,274	0,695	n.s.	-	-	-	-	-	-	-	-
*GAIO * Firm_typesSME_SME*	-	-	-	-	0,369	0,395	0,351	n.s.	-	-	-	-
*GAIO * Firm_typesLarge_Large*	-	-	-	-	-0,497	0,286	0,082	n.s.	-	-	-	-
*GAIO * ResearchAppl_Appl*	-	-	-	-	-	-	-	-	-0,191	0,298	0,521	n.s.
*GAIO * ResearchFund_Fund*	-	-	-	-	-	-	-	-	-0,788	0,353	0,026	*
Constant	-6,649	0,384	< 2e-16	***	-6,850	0,383	< 2e-16	***	-6,897	0,382	< 2e-16	***
AIC	3018,7				3014,9				3015,9			
McFadden's pseudo-R2	0,190				0,192				0,192			

*Coeff. = Coefficient; SE = Standard Error; sign. = Significance. Significance levels: 0,000 *** 0,001 ** 0,01 * 0,05 'n.s.' 1*

Appendix F

Likelihood ratio test results for comparing the different models.

Models being compared	#df	LogLik	Delta df	Chisquare	Pr(>Chisq)	Sign.
Model 1 (CV)	4	-1732,7				
Model 2 (IV w/o GAIO)	10	-1531,0	6	403,45	<2,2e-16	***
Model 2 (IV w/o GAIO)	10	-1531,0				
Model 3 (IV)	11	-1497,4	1	67,169	2,49e-16	***
Model 3 (IV)	11	-1497,4				
Model 4 (SR*GAIO)	12	-1497,1	1	0,5817	0,446	n.s.
Model 2 (IV)	11	-1497,4				
Model 5 (ND*GAIO)	12	-1490,2	1	14,415	1,47e-4	***
Model 2 (IV)	11	-1497,4				
Model 6 (Shared_partn*GAIO)	12	-1497,3	1	0,2169	0,641	n.s.
Model 2 (IV)	11	-1497,4				
Model 7 (Geog*GAIO)	12	1497,3	1	0,1535	0,695	n.s.
Model 2 (IV)	11	-1497,4				
Model 8 (Firm_types*GAIO)	13	-1494,5	2	5,9121	0,052	n.s.
Model 2 (IV)	11	-1497,4				
Model 9 (Research*GAIO)	13	-1494,9	2	4,9644	0,084	n.s.

CV = control variables; IV = independent variables; df = degree of freedom; LogLik = Log likelihood; Sign = significance. Significance codes: 0 ‘***’ 0,001 ‘**’ 0,01 ‘*’ 0,05 ‘n.s.’ 1