



A concept for measuring network proximity of regions in R&D networks[☆]



Iris Wanzenböck^{a,b,*}

^a Copernicus Institute of Sustainable Development, Innovation Studies Group, Utrecht University, The Netherlands

^b AIT Austrian Institute of Technology, Center for Innovation Systems & Policy (CRISP), Donau-City-Straße 1, 1220 Vienna, Austria

ARTICLE INFO

Article history:

Available online 16 November 2017

Keywords:

Network proximity
Aggregated networks
First-order proximity
Higher-order proximity
R&D networks
Knowledge spillovers

ABSTRACT

This paper proposes a new measure for assessing the network proximity between aggregated units, based on disaggregated information on the network distance of actors. Specific focus is on R&D network structures between regions. We introduce a weighted version of the proximity measure, related to the idea that direct and indirect linkages carry different types of knowledge. First-order proximity arising from direct cross-regional linkages is distinguished from higher-order network proximity, resulting from indirect linkages in the R&D network. We use an macroeconomic application in which we analyse the productivity effects of R&D network spillovers across regions to illustrate the usefulness of a proximity measure for aggregated units.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Social Network Analysis (SNA) provides useful instruments for exploring the structure and dynamics of research and development (R&D) networks. A R&D network may be defined as a set of nodes representing knowledge producing actors and a set of linkages representing R&D collaborations between these actors. Many studies emphasise the importance of tie strength, local clustering or short path lengths for the transmission of knowledge in such networks (Newman, 2001; Fleming et al., 2007). A dense web of interaction in the core of the network together with selective relations to the network periphery is assumed to guarantee efficient knowledge and information diffusion throughout the network structure (see also Cowan and Jonard, 2003; Goyal, 2007).

This paper takes up the research interest on R&D networks proposing a new approach for assessing the network proximity at an aggregated level. The aim is to develop a new measure of network proximity for aggregated units that accounts for information on the network structure at the micro level of individual actors.

[☆] This work was supported by the Anniversary Fund of the Oesterreichische Nationalbank (OeNB), Grant No. 16301. I am grateful for the comments and valuable suggestions provided by Thomas Scherngell, Michael Barber, Manfred M. Fischer and Sebastian Woblistin.

* Correspondence to: Copernicus Institute of Sustainable Development, Innovation Studies Group, Utrecht University, Heidelberglaan 2, 3584CS Utrecht, The Netherlands.

E-mail address: i.wanzenbock@uu.nl

Despite the variety of SNA methods to analyse networks, most of the proposed concepts and interpretations are applicable only to individual actors. The peculiarities when observing network structures between aggregates such as regions or countries have been widely neglected in social network research for a long time. Only recently, the enhanced analytical opportunities in viewing networks in terms of a system of different levels of aggregation have gained recognition in the multilevel analysis of social networks (e.g. Lazega et al., 2008; Lomi et al., 2016). Major motivations of the multilevel approaches to networks are to disentangle the influences of structural factors working at different social levels, and also to reduce systematic errors in the inferences drawn from statistical data on social relations observed only at an aggregated level (Lazega and Snijders, 2015). This study follows the second motive by addressing the question of how to represent network proximity at some aggregate level when the network structure is determined by the R&D relations and knowledge flows at the micro level of actors, such as between individuals or organisations, engaging in R&D collaborations.

Moreover, the social network perspective has made substantive inroads in disciplines such as economics (e.g. Ahuja, 2000; Jackson, 2010; Hausmann and Hidalgo, 2011; Caldarelli et al., 2012) or economic geography (e.g. Ter Wal and Boschma, 2009; Balland, 2012; Scherngell, 2013; Huggins and Thompson, 2014; Bergé, 2016), that traditionally lay focus on the analysis of aggregated units such as regions or countries. With this paper we aim at further incorporating both the concepts of social network analysis and related interpretations into economical questioning and

reasoning at a macro or meso level. A typical question referring to the proximity in networks could be formulated in terms of how inter-dependent are regions or countries on each other due to, for example, policy networks, trade flows, global value chains or knowledge relations. Especially in the case of knowledge or R&D relations, quantitative measures for the strength or the reach of relations between geographical areas may serve as analytical vehicle to indicate not only the amount of knowledge flowing between these areas but also the spatial patterns of knowledge diffusion through networks (e.g. Owen-Smith and Powell, 2004; Scherngell and Barber, 2009). Hence, measuring network proximity of aggregates may be a valuable extension of the common relationship, distance or closeness measures in SNA (e.g. Wasserman and Faust, 1994; Opsahl et al., 2010).

In doing so, we devote particular attention to R&D network linkages running across regions. So-called R&D networks of regions are an interesting case for the application of the proposed network proximity measure, especially because of the considerable number of studies drawing on SNA concepts and tools in this research field (see among others the studies of Autant-Bernard et al., 2007; Ponds et al., 2010; Barber et al., 2011; Hazir, 2013; Scherngell, 2013; Sebestyen and Varga, 2013). These studies often regard regions as a single network node, while disregarding information on 'node-internal' linkages or on the indirect linkages running through the network. Such aggregate-level approaches become increasingly criticised for their simplified, often unrealistic, assumptions on the flow of knowledge through network linkages within and across regions (see e.g. Breschi and Lissoni, 2001, 2009; Bergé et al., 2015; Bergé, 2016). Hence, there is an increasing need to reflect more explicitly on the representation of R&D networks in form of aggregated networks and the associated drawbacks from a social network analytical perspective.

This study aims at addressing the shortcomings of aggregate-level network approaches, by *first*, setting out with a measure for network proximity between regions constructed at the actor level. We tackle the problem of aggregation by accounting for the structure of the underlying network. All direct and indirect network linkages running within and across regions are considered before transforming the information to the 'higher' level of aggregation. A region is viewed as an aggregated unit consisting of the actors that are, according to their spatial attributes, located in this region. However, the measure may be equally applicable to other (spatial and non-spatial) observed aggregates of social systems such as large organisations, economic sectors or countries.

Second, we propose a weighted version of the proximity measure. Related to arguments that direct and indirect linkages fulfil different functions in R&D networks (e.g. Granovetter, 1973; Uzzi, 1997), we distinguish between the notions of first-order network proximity (i.e. given the strength of direct linkages) and higher-order network proximity (i.e. given the reach of indirect linkages) between regions. The differentiation between first-order and higher-order network proximity is based on the idea that particular kinds of knowledge are relevant for research and innovation (Lundvall et al., 2016). It is not only the know-how (i.e. the capability and skills of knowing how to do different kind of things), but also relevant knowledge of the 'know-who' type (i.e. knowing who can do peculiar things, and how to access this knowledge) determines R&D and innovation success. We argue that R&D networks provide the structure for both types of knowledge flows.

Third, we demonstrate in an illustrative example a way to employ the network proximity measure in economic analysis. We construct a region-by-region proximity matrix to reflect the arrangement of a set of regions in the R&D network space, and to assess network spillovers between those regions. By using spatial econometric modelling techniques, we analyse the relationship between regional economic productivity and the effects of cross-

regional knowledge spillovers arising from direct and indirect linkages in R&D networks. Different specifications of the network-based weight matrix are discussed.

The structure of the paper is as follows: After defining the network under consideration in Section 2, Section 3 sets forth the conceptualisation of the network proximity measure. We present our general approach to measure network proximity between aggregated units in Section 4. In Section 5 we introduce the weighted version of the network proximity measure, enabling us to distinguish between first-order and higher-order network proximity of regions. Section 6 provides the empirical application context of the network proximity measure. The final section closes with a discussion, some ideas for applications as well as opportunities for further development of the proposed measure.

2. Network definition

A R&D network of actors in its most basic form may be viewed as an undirected graph of the form $G(V, L)$, in which the set of nodes $V = \{v_1, \dots, v_M\}$ stands for the M actors participating in R&D collaborations. The set of edges $L = \{l_1, \dots, l_k\}$ corresponds to the set of R&D collaborations between these actors. An R&D collaboration between two actors v_u and v_q ($u, q = 1, \dots, M$) is represented by an edge $(v_u, v_q) = l_k \in L$. We do not consider the direction of knowledge flows between actors, (v_u, v_q) denotes an unordered pair, and since no actor can collaborate with itself, $(v_u, v_u) \notin L$. The network we are considering can be represented by a matrix

$$A = (a_{uq})_{1 \leq u, q \leq M} \quad (1)$$

which is a symmetric adjacency matrix of dimension $M \times M$, in which $a_{uq} = 1$ if $(v_u, v_q) \in L$, and zero otherwise.

Network nodes not adjacent in the network may be reachable via a path in the network. A network path between a pair of actors (v_u, v_q) is defined as an alternating sequence of nodes and edges in which each edge is traversed only once and each of the nodes is visited only once. The number of edges of a path denotes its length. Then, the length of the shortest path between two nodes, v_u or v_q , also referred to as the network distance, is denoted by $d(u, q)$. The network distance indicates the minimal number of edges to be traversed in order to reach node v_q starting at node v_u . We set $d(u, q) = \infty$ if two nodes are not connected with each other, that is, if there is no path along edges connecting them. The network distances $d(u, q)$ for any dyad (v_u, v_q) can be displayed in form of a distance matrix

$$D = (d_{uq})_{1 \leq u, q \leq M} \quad (2)$$

with $d_{uu} = 0$. Obviously, if the actors share a direct link, i.e. a R&D collaboration, they have a network distance of $d_{uq} = 1$ in the distance matrix. If $1 < d_{uq} < \infty$, the corresponding actor pair is reachable via a network path but is only indirectly connected.¹

In this article we are particularly interested in the network structure of R&D linkages across regions. Given N regions, the M actors are partitioned so that each actor is located in exactly one region (i.e. multiple regional assignments or regional attributes are not possible). By V_i we denote the set of actors that belong to region i ($i = 1, \dots, N$), and accordingly, by R_i the index set of the actors that belong to V_i , i.e. $R_i = \{u \in \{1, \dots, M\} : v_u \text{ belongs to region } i\}$. Further, let M_i denote the number of actors located in region

¹ An alternative approach would be to indicate the number of walks from node v_u and node v_q by using the K th-power of the adjacency matrix A , so that A^K give the number of walks of length K for all (v_u, v_q) . By adding, for example, the matrices A and A^2 one would observe the number of walks of length $K \leq 2$. However, conceptual problems might arise as a walk from v_u to v_q is not necessarily equal to the (shortest) paths from v_u to v_q (Wasserman and Faust, 1994).

i. For convenience, let us assume that the actors are ordered ascendingly according to which region i they belong. That is, $V_1 = \{v_1, \dots, v_{s_1}\}$, $V_2 = \{v_{s_1+1}, \dots, v_{s_2}\}$, and so on.

The adjacency matrix can then be partitioned with respect to the regional assignment into $N \times N$ a-priori defined blocks.² Each block on the off-diagonal represents a non-aggregated cross-regional adjacency matrix A_{ij} between the regions i and j , for $i \neq j$ of the size $M_i \times M_j$ defined as

$$A_{ij} = (a_{uq})_{u \in R_i, q \in R_j} \quad (3)$$

with elements a_{uq} indicating whether any two actors located in region i and region j are connected by a joint R&D collaborations (i.e. an inter-regional linkage). Eq. (3) is a non-aggregated adjacency matrix of the unweighted network for the region pair (i, j) . The sub-matrices A_{ii} contain the linkages between the actors located within a distinct region (i.e. intra-regional linkages).

In analogy to Eq. (3), we define a non-aggregated distance matrix D_{ij} between the regions i and j , for $i \neq j$, in terms of

$$D_{ij} = (d_{uq})_{u \in R_i, q \in R_j} \quad (4)$$

being a partition of the distance matrix D in Eq. (2). Since we are only interested in the cross-regional network distances of actors, we do not consider the sub-matrix D_{ii} , containing the network distances of actors within region i .

Based on the network representation in Eq. (3) and the cross-regional distance matrix in Eq. (4) for all regions, we can speak of an R&D network of regions, where the regions are defined as aggregates of the actors that are according to their spatial attributes located in this region. In the following, we will use this actor-level information to define the network proximity between aggregated units. Before doing so, we discuss the main building blocks of our proximity measure and show how the proposed concept differs compared to previous approaches.

3. Conceptualising network proximity at an aggregated level

The notion of social networks has been increasingly exploited to analyse cross-regional or cross-country networks and the spatial structure of the linkages (e.g. Scherngell and Barber, 2009; Hausmann and Hidalgo, 2011; Caldarelli et al., 2012). However, most large-scale empirical investigations, including a large set of regions or countries, are based on network representations in which the nodes of the network represent the geographical units. Accordingly, the R&D network of regions is characterised in terms of a region-by-region adjacency matrix. Such an aggregate-level approach might be appropriate when focus is only on the direct linkages crossing regions. However, as soon as we focus on the real network properties of R&D linkages – involving both direct and indirect linkages between the regions – analytical restrictions emerge when each region is regarded only as one single network node. Their major drawbacks are as follows:

First, we systematically cut off selected node–linkage sequences of the underlying network structure, namely those linkages established by actors within regions. We do not know how R&D linkages are distributed across the actors, or whether there are either only a few key players holding all the external links, while others are isolated. It is implicitly assumed that linkages are equally distributed across all actors within a region, and that regions show high internal

network density. However, this assumption stands in stark contrast to empirical observations recently made in innovation research, emphasising the heterogeneity of regions or clusters resulting in very different manifestations of region-internal network structures (e.g. Giuliani, 2007; Wanzenböck et al., 2015).

Second, if regions are observed as single nodes, we cannot evaluate whether there is a region-internal ‘network channel’ between those actors holding the cross-regional linkages, i.e. we do not know if actors located in one regions are connected, or reachable, in the network. The actors could, for instance, operate in different components of the network. This is problematic as a number of studies are built on the assumptions that knowledge can flow more easily across regions because of the R&D linkages attached to these regions (e.g. Autant-Bernard et al., 2007; Huggins and Thompson, 2014). However, region-internal disruptions of such network channels, for instance because actors located in the same region do not collaborate with each other, clearly do not support this view.³ To guarantee efficient knowledge diffusion, the network has to be not only connected within and between regions, but also provide short path length and dense network structures between regions (e.g. Newman, 2001; Fleming et al., 2007). Only actor-level linkages provide us the information necessary to evaluate such structural properties of the network in a suitable way.

Third, and related to the former, is the fact that we have no information on the length of the network paths within the regions. Network distance is a useful indicator to quantify the distance knowledge needs to travel to reach other regions. If the region-internal ‘network channels’ are not observed, the length of the shortest path between two regions might differ considerably compared to a network constructed at the actor-level. For example, in a network of three regions i, j and k , a link between region pair (i, j) and region pair (j, k) would bring also regions (i, k) in closer proximity in the network, separated only by the two cross-regional linkages attached to the intermediary region j . Hence, it becomes apparent that the proximity between regions observed for such aggregate-level network topographies will always be less or equal the network proximity when region-internal linkages are considered.

Our measure of regional network proximity is designed to overcome these drawbacks for aggregated networks. We argue that a real network perspective on cross-regional R&D collaborations calls for a more fine-grained view on the structure of the underlying network of actors. Two levels are of importance and need to be distinguished: the one where the network is constructed in terms of relations between actors, and the aggregate one where the concept of network proximity is applied and further analysis of the connectivity structure of aggregated units can take place. Regional network proximity is regarded as a link-based concept derived from the underlying network linkages at the actor level as given by Eq. (3) and the distances as given by Eq. (4). By using information on actor linkages between regions (inter-regional linkages) as well as within the region (intra-regional linkages), we aim at measuring the network proximity between regional units in a more appropriate way than region-level approaches.

Fig. 1 illustrates the basic idea of a concept of network proximity between regions. It is based on the network distance of actors located in different regions, and reveals the insufficiencies when network proximity is defined at the regional level. Two different

² Note that we use exogenously defined characteristics of nodes to partition the network. This is in contrast to other methods such as structural equivalence (Sailer, 1979), modularity (Newman, 2006) or block modelling (Ziberna, 2014) relying on structural network properties as criterion to detect subgroups, clusters or communities in networks.

³ The aggregate-level approach to regional network proximity seems to be in line with one of the basic assumptions often taken in regional studies, in which geographical proximity is regarded as sufficient condition for knowledge spillovers to occur between actors. However, this argument is increasingly called into question by theoretical consideration and supported by empirical evidence (see e.g. Breschi and Lissoni, 2001, 2009; Boschma, 2005; Fitjar and Huber, 2015). Both emphasize that knowledge spillovers more likely follow the channels of social networks, eliciting a reconsideration of how network proximity between regions is measured.

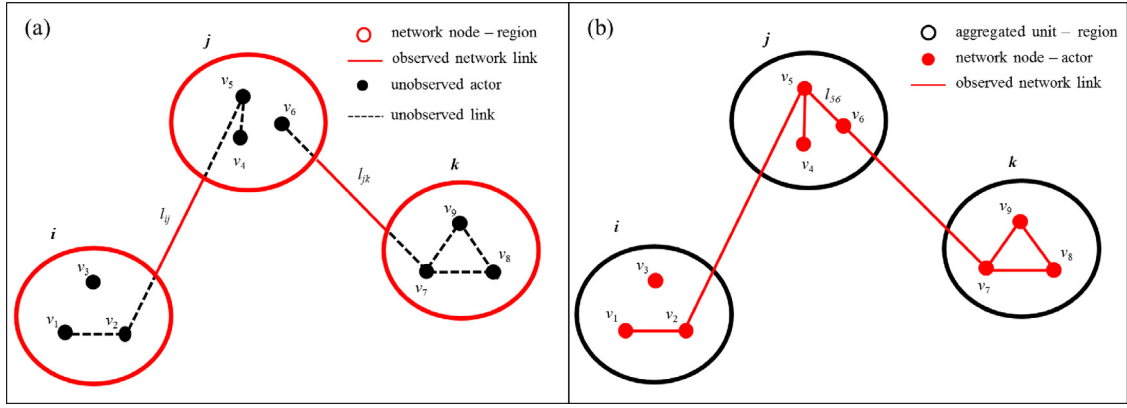


Fig. 1. Network proximity between regions in R&D networks. Notes: Panel (a) is based on a network representation at the regional level, where region-internal linkages are disregarded. Network proximity between region pair (i, k) is observed with path length two. In contrast, in panel (b) the network is represented at the level of actors. Network proximity between region pair (i, k) is obtained only due to the region-internal link (l_{56}). The shortest path between region i and k is $\{v_2, v_5, v_6, v_7\}$ with length $d_{27} = 3$.

situations for the case of three regions (i, j and k) are displayed: First, panel (a) shows a situation where the regions are the units of observation. The individual regions (here indicated as i, j and k) are captured as single nodes in the network, as indicated by the red circle. The actor-level structure of the network (the nodes and lines marked in black) remains unobserved in this representation. The existence of at least one network linkage between the regional dyads (i, j) and (j, k) , here indicated by l_{ij} and l_{jk} , suffices to connect the non-adjacent region pair (i, k) . According to this network representation, the non-adjacent region pair (i, k) would be indirectly connected in the network, and counted to be in proximity of path length two, though nodes v_2 in region i and v_7 in region j are not reachable.

In panel (b), we show links that are observed at the actor level (the red nodes and linkages), leading to a regional network proximity that accounts for the link structure at the micro level. Here, the regions are considered as aggregates of the actors represented by the set of nodes $V_i = \{v_1, v_2, v_3\}$, $V_j = \{v_4, v_5, v_6\}$ and $V_k = \{v_7, v_8, v_9\}$. Opposed to panel (a), a link between actors v_5 and v_6 has been added in this example, i.e. region pair (i, k) is therefore indeed connected in the network. Owing to the network linkage (l_{56}) in region j we observe a cross-regional network path of $\{v_2, v_5, v_6, v_7\}$ between the dyad (v_2, v_7) of length $d_{27} = 3$. For a situation where regions are indirectly linked at the actor level, we can speak of higher-order proximity (see Section 5 for details). However, without the region-internal linkage of actors in region j , region pair (i, k) would remain disconnected according to the actor-level representation of the network. Hence, regions with a fragmented internal structure will obtain also lower network proximity to other regions, as they can take less advantage of the internal network density. To resume, situations indicated by panel (a) may lead to inadequate measures of network proximity, while observing links at the actor level as in panel (b) allows for a more realistic approximation of network proximity between regions.

4. Measuring network proximity between aggregated units

In the previous section, we have shown that the existence of a network path between two aggregated units can be considered as a precondition to count these two units in network proximity. For our case, this implies that there must be, by definition, at least one network linkage or a sequence of linkages created at the actor level that connects these two regions. This basic information on whether actor pairs located in two different regions are either directly connected or indirectly linked in the network can be read off from the non-aggregated distance matrix D_{ij} as defined in Eq. (4). Conse-

quently, sub-matrix D_{ij} already gives a first indication of whether the regions i and j are reachable in the network.

On this basis, a *measure of network proximity* between aggregated units, in our case any two regions, may be understood as a function which assigns to a pair of regions i and j the value

$$p_{ij} = \mathcal{L}(D_{ij}^*) \quad (5)$$

where D_{ij}^* denotes a transformed variant of the non-aggregated distance matrix in Eq. (4) which is required to indicate proximity between the regions i and j , for $i \neq j$, and \mathcal{L} is any matrix norm $\mathcal{L} : \mathbb{R}^{M_i \times M_j} \rightarrow \mathbb{R}_{\geq 0}$ specifying the approach of how the information on actor-level network distances contained in D_{ij}^* is aggregated. Note that the measure of network proximity is defined for two distinct regions i and j , so that $p_{ii} = 0$ as we are not interested in the intra-regional network distances between actors, that is the network proximity of a region with itself. From Eq. (5), it is easy to see that network proximity between aggregated units depends on the one hand on how the network distances between actor pairs are evaluated and on the other hand on the method of aggregating this information. We will elaborate on this further.

4.1. Distance-decay effect between aggregated units

Connectedness or reachability is not sufficient that knowledge can effectively diffuse across regions. Actors that are closer to each other have a higher, or a different kind of influence than nodes that are more distant (Borgatti, 2005). The length of network channels within regions might differ considerably, so that the question of how network distances are valued might be even more relevant for analyses at the aggregated level. For the case of R&D networks, for instance, we can assume that knowledge with a high degree of tacitness (e.g. previous experiences, specific technological opportunities or new combinations) is more difficult to transmit or only spills over after repeated interaction. It may lose its specificity already with a small number of intermediary actors lying in between (Sorenson et al., 2006; Fafchamps et al., 2010). For two regions indirectly connected via the actors of a third region, this implies that a long network path through the third region would reduce network proximity and knowledge flows between the two regions.

To consider such distance-decay effects, we introduce a transformed, so-called decay-weighted, network distance matrix D_{ij}^* , taking the general form of

$$D_{ij}^* = (d_{uq}^{-\alpha})_{u \in R_i, q \in R_j} \quad (6)$$

where d_{uq} represents the shortest path length d_{uq} , for all $u \in R_i$, $q \in R_j$, with the parameter $\alpha \geq 0$ imposing decreasing weights on

higher network distances between the actor pairs.⁴ Accordingly, the elements of the transformed matrix D_{ij}^* take values ranging from one for directly linked actor pairs (i.e. $d_{uq} = 1$), to zero for actor pairs that are non-reachable in the network. The magnitude of α reflects the extent to which the flow of knowledge decreases with growing distance. If $\alpha = 0$ there is no distance decay effect, based on the assumption that knowledge can flow unrestrictedly through the network. Often a value of $\alpha = 1$ is applied to measure network spillovers (e.g. Leenders, 2002), so that the distance decay is inversely related to the path length d_{uq} . Generally stated, the higher the magnitude of α , the steeper is the distance decay function, and the more significant the decay effect on longer network paths.⁵ The value of α is to be set by the researcher depending on the empirical context, the observed properties of the network (e.g. small work properties) as well as theoretical considerations regarding the nature and flow of knowledge (e.g. degree of tacitness) through the network.

4.2. Aggregating actor-level information

In Eq. (5) we introduced \mathcal{L} as a specific linking function applied to the decay-weighted distances between the actors located in two regions. \mathcal{L} could basically be any matrix norm $\|\cdot\|_p$. For measuring the proximity between regions in R&D networks we consider a scaled Manhattan norm with $\mathcal{L} = \|\cdot\|_1$, involving the linkages and network distances of all regional actors to be the most appropriate. The advantage of this aggregation measure is that it accounts for and gives equal weight to all linkages established between two regions.⁶ Hence, each further R&D link established at the actor level contributes to the network proximity between regions. The smaller the network distance between the actors, the higher is the regional network proximity between these regions, with increasing probability that knowledge can flow through network channels.

Bearing these considerations in mind, the general version in Eq. (5) can be reformulated so that we obtain the measure of network proximity in terms of

$$p_{ij} = \frac{1}{\lambda} \sum_{\substack{u \in R_i \\ q \in R_j \\ u \neq q}} |d_{uq}^{-\alpha}| \quad (7)$$

⁴ The function $f(d) = d^{-\alpha}$ is a distance decay function, by which we understand a monotonously decreasing function $f: \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ with $f(1) = 1$ and $\lim_{d \rightarrow \infty} f(d) = 0$. The matrix D_{ij}^* could be reasonably defined by $D_{ij}^* = (f(d_{uq}))_{u \in R_i, q \in R_j}$ for any distance decay function f . For the sake of simplicity, we restrict the discussion to the prototype $f(d) = d_{uq}^{-\alpha}$.

⁵ Moreover, a distance threshold, i.e. an upper barrier for the maximum path length to be considered between two actors, could be imposed. Accordingly, all non-adjacent actors exceeding this threshold can be set to $d_{uq} = \infty$ before distance weighting. For example, following Leenders (2002), a path length of three is most often a reasonable choice for defining the scope of influence arising from indirect relations in social networks.

⁶ Following Everett and Borgatti (1999), the problem of aggregation can be related to methods proposed in the hierarchical clustering literature (e.g. Rokach, 2010 for an overview). In contrast to the Manhattan norm, the minimum as well as the maximum method would calculate the proximity between two regions by focusing only on one specific linkage (i.e. the shortest or the longest) connecting the actors in two different regions. By using the minimum, also referred to as single linkage criterion, network proximity between two regions is calculated based on the distance of two distinct actors, one in each region, that are nearest to each other, i.e. $\mathcal{L} = \min(\cdot)$. Hence, it would ignore the region-internal structure of actors and their linkages in the network. Moreover, regional network proximity based on the network distance of regional actors that are farthest apart from each other, indicated by the supremum norm with $\mathcal{L} = \|\cdot\|_{\infty}$, may be reminiscent of the complete linkage criterion in hierarchical clustering. Such a specification, however, seems not appropriate for applications on R&D networks, as it might not hold that all actor pairs between two regions are connected via a network path.

where $d_{uq}^{-\alpha}$ denotes the decay-weighted elements of the network distance between the actor pair u in region i and q in region j . λ is a scaling factor with $\lambda \leq (M_i \cdot M_j)$.

Based on Eq. (7), we see that not only those actors holding the direct cross-regional linkages contribute to regional proximity, also actors inter-linked via indirect linkages can increase the network proximity between regions. Depending on the empirical case, λ may be defined as $\lambda = 1$, as $\lambda = M_i \cdot M_j$, or as a proportion of regional actors holding network linkages on the total number of actors in the two regions. The latter two cases correspond to an average or weighted linkage concept. It can be rather used for comparisons, for instance, when it is reasonable to have a relative measure of proximity that is independent of the regional network size. The former, in contrast, is more useful when interest is on the absolute amount of proximity, that is the absolute magnitude of 'network channels' between regions. By setting $\lambda = 1$ we assume that external knowledge can potentially flow between two regions with each further link at the actor level, so that each (direct and indirect) linkage at the actor level contributes to the network proximity of regions.

Our measure of network proximity shows some similarities with standard closeness matrices typically calculated on the inverse sum of the network distances at the micro level (e.g. Freeman, 1977; Wasserman and Faust, 1994). However, network proximity between aggregated units differs from closeness centrality on conceptual grounds. Network proximity considers only those dyads in the network that are attributable to two specific aggregated units, i.e. a pair of regions. Basically, only one network linkage between an actor pair would suffice to consider two regions in proximity, despite the fact that all other actors might operate in different components of the networks. Closeness centrality, in contrast, assesses the positioning of a specific node by taking into account all dyadic distances in the network, and usually can be calculated only for connected networks (Opsahl et al., 2010).⁷

5. Weighted R&D networks: the notions of first-order and higher-order network proximity

In R&D networks, as in many other network types, the strength of linkages may play a crucial role for the type of knowledge and information running through the network structure (Opsahl et al., 2010; Newman, 2004). This is why R&D networks are increasingly seen as weighted networks, with link weights being interpreted as indicator for the amount of knowledge potentially flowing between the actors. In this section we extend the general, unweighted, version of the proximity measure to the case of weighted R&D networks, in which the links carry the information on the R&D collaboration intensity between the regions.

At the micro (actor) level, efforts have been recently made to combine tie weights with the distance between nodes when assessing closeness. For instance, Opsahl et al. (2010) extend common approaches (e.g. Dijkstra, 1959; Newman, 2001) for identifying the shortest path in weighted networks. They suggest a node centrality measure which allows weighting the relative importance of tie strength compared to path length. In contrast to these previous measures for weighted networks, our approach builds more on

⁷ Another relevant network measure for the analysis of R&D networks of regions would be the clustering coefficient, which can provide specific insights in existing clique structures within and between regions. Particularly useful for R&D network research are the recently developed measures for weighted networks (e.g. Opsahl and Panzarasa, 2009) and those for two-mode network data (e.g. Opsahl, 2013). Local clustering, for instance, will give further insights in whether there are dense cliques of actors within a region (i.e. whether the regional actors are also partners with each other) or between a specific set of regions (i.e. whether there are several triangles formed by actors across regions). In this study, however, we focus on regional network proximity in more general terms.

the idea that direct and indirect linkages fulfil different functions in networks, and therefore involve different types of knowledge flows (e.g. Granovetter, 1973; Uzzi, 1997; Ahuja, 2000; Fafchamps et al., 2010). It is argued that strong direct ties enable the flow of more tacit and purposeful knowledge which is trust-based and related to concrete skills. While such direct R&D connections enable the transmission of specific 'know-how' between partners, considered as knowledge about or the capability of doing specific things (Lundvall et al., 2016), indirect R&D network ties to distant actors may constitute more an information repository for actors which can be highly useful when it comes to the early recognition of potential new knowledge sources. Following innovation research, the latter can be related to the 'know-who' or 'who knows what' type of knowledge (Lundvall et al., 2016), which characterises very specific information about the availability and location of external knowledge sources crucial for successful innovation.

When we talk about network proximity at an aggregated level, we are particularly interested in the strength of direct ties that run between the aggregated units. We therefore differentiate between region pairs that are directly linked in the R&D network, referred to as *first-order proximity*, and region pairs indirectly connected given the network paths at the actor level, referred to as *higher-order proximity*. The aim is to pin down the network proximity between aggregated units so that the orders of proximity allow for interpretations similar to those usually associated with the different functions of direct and indirect linkages in social networks.

To introduce the weighted R&D network, consider for any directly connected actor pair (v_u, v_q) the non-negative real number $w_{uq}^{(\phi)}$ which denotes the weight attached to the link $l_k = (v_u, v_q)$ according to the number of R&D collaborations. Superscript ϕ , for $\phi \geq 1$, indicates the minimum link weight assuring that an actor pair shows a certain frequency or strength of interaction in the R&D network (measurable e.g. in terms of the number of common projects, personal meetings during a collaboration, or some effective collaboration output such as co-patents or joint research papers). Moreover, $w_{uq}^{(\phi)} = w_{qu}^{(\phi)}$, as we do not consider any possible direction of knowledge flows between the partners in the R&D collaboration. A weighted R&D network is accordingly specified in terms of the weighted adjacency matrix $W = (w_{uq}^{(\phi)})_{1 \leq u, q \leq M}$, with $w_{uu}^{(\phi)} = 0$ for all u , and $w_{uq}^{(\phi)} = 0$ if v_u and v_q are not connected. In analogy with the case of the unweighted network (Eq. (3)), we define the corresponding cross-regional sub-matrix, for $i \neq j$, by

$$W_{ij} = (w_{uq}^{(\phi)})_{u \in R_i, q \in R_j}. \quad (8)$$

Then, the *weighted measure of network proximity* between regions is a function of the strength of relations between adjacent actors in the network and the network distance between non-adjacent actors in the form of

$$p_{ij}^{(w)} = \mathcal{L}_w(W_{ij}) + \mathcal{L}(D_{ij}^*) \quad (9)$$

where both \mathcal{L}_w and \mathcal{L} are matrix norms on $\mathbb{R}_{\geq 0}^{M_i \times M_j}$. \mathcal{L}_w can be regarded as the regional linking function applied to the weighted adjacency matrix W_{ij} , and $\mathcal{L}(D_{ij}^*)$ corresponds to the non-weighted representation of regional network proximity⁸ as in Eq. (5). The value of $p_{ij}^{(w)}$ depends on the choice of these matrix norms, the cross-regional R&D intensities between adjacent actors, and the cross-regional – decay-weighted – network distances between non-adjacent actors. For the weighted measure, too, we restrict ourselves to the case of $i \neq j$ so that $p_{ii}^{(w)} = 0$.

In analogy to the unweighted version of network proximity in Eq. (7), we exemplify the weighted measure by using a scaled Manhattan norm for $\mathcal{L}_w = \mathcal{L} = \|\cdot\|_1$, equally weighting the network data matrices W_{ij} and D_{ij}^* , respectively, to account for all the established linkages between two regions when aggregating the actor-level information. This leads us to a weighted measure of network proximity between regions in terms of

$$p_{ij}^{(w)} = \frac{1}{\lambda} \left(\sum_{\substack{u \in R_i \\ q \in R_j \\ u \neq q}} |w_{uq}^{(\phi)}| + \sum_{\substack{u \in R_i \\ q \in R_j \\ u \neq q}} |d_{uq}^{-\alpha}| \right) \quad (10)$$

where ϕ indicates the minimum link weight as in Eq. (8), and α denotes the distance decay effect placed on higher network distances as in Eq. (6), and the scaling factor λ controls for the size of the network, or the number of connected actors in the two regions as in Eq. (7). The network proximity in weighted form is accordingly defined as the sum of the R&D collaboration intensities between the pair of actors u located in region i and q located in region j , and the sum of the network distances between these actors.

From the expressions in Eqs. (9) and (10) we see that the weighted network proximity measure consists of two building blocks that can be interpreted in terms of different orders of proximity: If the first term $\mathcal{L}_w(W_{ij}) > 0$, there is at least one pair of actors that is connected via a linkage (with weights greater or equal ϕ). Consequently, two regions are said to be in *first-order proximity*, if we can observe at least one direct network linkage (meeting the minimum weight criterion) between the actors located in two different regions. Moreover, if the second term $\mathcal{L}(D_{ij}^*) > 0$, there is at least one indirect network linkage between the regions i and j , i.e. a connection via common research partners in a third region, and the regions are considered to be in *higher-order proximity*. It depends on the network structure at the actor level, whether we can observe first-order proximity, only higher-order proximity, or both, between a regional pair.

Note further that first-order proximity between regions is a special case of our weighted network proximity measure, namely when indirect cross-regional network linkages are not analysed. Such an 'aggregated network proximity measure' has been usually applied to indicate the network interdependencies and empirically study spillover effects arising across regions (see e.g. the studies of Ponds et al., 2010; Basile et al., 2012; Hazir, 2013; Paci et al., 2014). Interdependencies emanating from indirect network channels (i.e. higher-order proximity) are neglected in this case, i.e. the second term in Eq. (10) would equal zero.

The weighted version of the network proximity measure offers a more realistic and flexible approach to measure the network proximity between regions. Tailored to the empirical network under investigation, variations of the measure can be employed with respect to (i) the regional linking function (i.e. the way of aggregating the actor-level information), (ii) the order of proximity (first- to higher-order proximity) and the minimum link weight to be considered, and (iii) the decay effect imposed on higher network distances.

The next section demonstrates in which way the proposed network proximity measure between regions can be used in empirical applications. In our illustrative example, we are interested in the question of whether the R&D network proximity between regions influences the economic productivity of these regions. Particular focus is laid on the comparison of different specifications of the regional proximity measure.

⁸ To avoid double counts in Eq. (9), all linkages with a network distance of one are excluded in D_{ij}^* by construction.

6. Empirical application to cross-regional knowledge spillovers in R&D networks

Knowledge spillovers between regions or countries are an actively debated research issue in economics and an important question for applied macroeconomic research (e.g. [Coe and Helpman, 1995](#); [Aghion and Jaravel, 2015](#)). Following this literature stream, network linkages are one of the mechanisms to transfer knowledge across geographical space ([Owen-Smith and Powell, 2004](#)), and are assumed to induce knowledge spillovers across regions (e.g. [Breschi and Lissoni, 2001](#); [Huggins and Thompson, 2014](#)). We are interested in the question of which regions are actually able to benefit from network spillovers in terms of increasing regional economic productivity. R&D network proximity of all regional pairs (i.e. the aggregated units) is calculated in order to estimate the effects of network-based knowledge spillovers on regional productivity. A network-based weight matrix as typically employed in spatial econometric modelling approaches (see e.g. [LeSage and Pace, 2009](#)) is used to quantify cross-regional spillovers arising from network proximity. In social network research, such a weight matrix captures potential interaction effects between the units of observation, also referring to so-called peer effects (e.g. [Leenders, 2002](#); [Bramoullé et al., 2009](#)). Before we provide details on how we consider the region-by-region network proximities, we discuss our modelling approach to relate knowledge spillovers to regional productivity.

6.1. Empirical model and variables

The empirical model to be estimated is a cross-section variant of the spatial knowledge capital model as used in [Fischer et al. \(2009\)](#). In matrix notation, stacked over the N regions, the model takes the form of

$$y = \beta x + \theta P x + \gamma Z + \varepsilon \quad (11)$$

with y being a $N \times 1$ vector denoting observations on regional total factor productivity (TFP) and x representing regional knowledge stocks ($N \times 1$). P is the key element of the model, denoting the network-based weight matrix ($N \times N$) with information on the proximity of all regional units in the R&D network. The matrix elements on the off-diagonal, defined either by p_{ij} in an unweighted representation as in Eq. (5) or by $p_{ij}^{(w)}$ in weighted form as in Eq. (10), reflect the extent to which knowledge located in region j influences the productivity of region i due to cross-regional network linkages (all $p_{ii}, p_{ii}^{(w)} = 0$). Accordingly, the spillover term Px can be understood as a measure for the accumulated – ‘network proximity weighted’ – knowledge that can potentially spill over from other regions. Expressed as scalar, the term $\sum_j^N p_{ij} x_j$ represents a linear combination of the knowledge stocks observed in all other regions (except the own region). β and θ are the regression coefficients for the region-internal knowledge stocks, and for the region-external knowledge mediated through network linkages. The latter reflects the marginal effects on productivity of an increase in the region-external knowledge stocks. We further include the matrix Z ($N \times 2$) containing additional factors such as the regional size or the degree of urbanisation (i.e. agglomeration effects) that potentially influence the TFP relationship. γ is the corresponding column vector (2×1) containing the coefficients for regional population (γ_1) and regional population density (γ_2). ε is an ($N \times 1$) error term with zero mean and variance $\sigma_\varepsilon^2 I_N$.

Our empirical setting consists of $N=254$ European NUTS-2 regions in the EU-27 member states. The dependent variable is regional TFP in manufacturing, defined as usual in terms of the fraction of regional economic output in the manufacturing sector not explained by the conventional production inputs, regional

labour (measured in terms of hours worked) and physical capital employed in the manufacturing sector. Due to the definition of TFP as ‘residual’ of economic production (see e.g. [Hall and Jones, 1996](#)), the values are not directly observable but must be calculated. For this empirical application, we calculate relative TFP values for all regions in the sample in the year 2007 (a formal definition of the measure and the description of the data used is given in [Appendix A](#)). In this way, we are able to observe the cross-sectional variations in TFP and measure whether a region’s proximity to other regions in the R&D network is – due to knowledge spillovers – positively related to higher productivity. The independent variables referring to the region-specific stock of knowledge are proxied in terms of a region’s patent stock. We calculate the stock on the basis of patenting activities during 2004 and the discounted stock of patents of previous periods (see also [Appendix A](#) for further details on the calculation and data sources). The control variables (regional population and population density) are observed in 2004, with regional data provided by Eurostat.

6.2. Specifying different network-based weight matrices

In our empirical application we compare different specifications of the network-based weight matrices. The aim of this comparison is twofold: *First*, we want to compare weight matrix specifications based on the network proximity measure as proposed in Eqs. (7) and (10) with the often used aggregate-level or region-level approaches where the regions are regarded as single nodes in the network (see Section 3 for more details). *Second*, we aim at comparing the weighted network proximity with the unweighted version and evaluate whether the strength of network linkages between regions (first-order proximity) delivers relevant information for measuring knowledge spillovers as opposed to a situation in which only the actors with their network distances are considered.

For this purpose, we use five different specifications of the region-by-region network weight matrix P , which corresponds to five different specifications of the empirical model in Eq. (11): In our first specification (*Model 1*), the region-by-region weight matrix is constructed based on the aggregate-level approach. Only the existence of direct linkages between regions is considered, leading to a non-weighted, dichotomous specification of the network matrix P , in which $p_{ij} = 1$ if regions are directly linked in the network, and zero otherwise. This binary definition of P does not correspond with the perception of network proximity as proposed in this article. We consider it for comparative reasons as it has been one of the most commonly used representations of the network arrangement of regions (e.g. [Maggioni et al., 2007](#); [Broekel and Hartog, 2013](#); [Varga et al., 2014](#)). We additionally consider decay-weighted network distances in our second specification (*Model 2*), which is still calculated at the aggregate level. In a similar manner as for the network distances at the actor-level, we set $p_{ij} = d_{ij}^{-\alpha}$ with $\alpha = 1$ so that $p_{ij} = 0.5$ if two regions are indirectly connected by network linkages that run via a third region, $p_{ij} = 0.333$ if two other regions lie in between them, and so on. Both model specifications follow a regional study perspective by implicitly assuming that all the regional actors can equally participate in external knowledge flows as knowledge will be fully disseminated within a region given the geographical proximity.

In contrast, the following specifications of the weight matrix are all based on a non-aggregated representation of regional network proximity, which means that the region-internal structure and distribution of links is taken into account. More specifically, the third specification (*Model 3*) refers to the unweighted measure of network proximity as in Eq. (7), while the fourth and fifth specification (*Model 4* and *Model 5*) show the results for the weighted R&D network proximity as in Eq. (10). In *Model 4* we consider

the special case of first-order proximity (i.e. only the number of direct linkages), where higher-order proximity between regions is excluded by definition. This model is particularly interesting since many empirical studies focusing on network-based spillover effects implement a weighted regional adjacency matrix equivalent to what we label first-order proximity here (e.g. Ponds et al., 2010; Basile et al., 2012; Paci et al., 2014). In our last specification (Model 5), we account for higher-order proximity between the regions, including not only the number of direct R&D collaborations but also the indirect linkages established between the actors in the network.

For the non-aggregated specifications, we use the Manhattan norm approach (i.e. $\| \cdot \|_1$) to aggregate the actor-level linkages and network distances. Moreover, we set the distance decay parameter $\alpha = 1$ to allow for smaller weights on indirect linkages; also the minimum link weight for first-order proximity (ϕ) and the scaling factor (λ) for the number of regional actors are set to one for this comparison.⁹

As common Table 1 in spatial or network econometric modelling, we standardise the resulting network matrices P for the empirical estimation. In contrast to standard row- or column normalisation approaches (see e.g. Leenders, 2002), we use a single rescaling factor based on the maximum eigenvector denoted by δ_{max} , which allows keeping the proportionality in the proximity structure between the regional units (Kelejian and Prucha, 2010). Accordingly, the entries of all network matrix specifications are defined as $P = p_{ij}/\delta_{max}$.

The R&D network under consideration is a network based on collaborative R&D projects funded by EU Framework Programme (FP) running in the years 2002 to 2004. A link is counted between two organisations for each year they are listed as partners in a R&D project, so that we obtain a valued adjacency matrix with entries representing the number of R&D projects between the organisations.¹⁰ Given our empirical context, we are able to provide evidence on the productivity effects of EU funded R&D collaborations due to the network proximity they create across European regions. Table A.1 provides relevant statistics of the R&D network at the organisational level, Fig. A.1 compares the distribution of the different proximity measures, and Fig. A.2 illustrates the spatial pattern of cross-regional proximity in the pan-European network. Fig. A.2 illustrates that the European R&D network creates the highest network proximity mostly between capital regions in Western Europe. However, as observable in both figures, the distribution of network proximity gets less hierarchical when indirect linkages are taken into account (as for the unweighted as well as the weighted higher-order proximity measure), compared to a situation where only the direct linkages (i.e. first-order proximity) are considered.

6.3. Estimation results

In discussing the empirical results we focus only on the comparison of the network proximity measure. The coefficient of the

spillover parameter θ is of main interest. We see that the coefficient is insignificant in Model 1 and Model 2 based on a reduced specification of regional network proximity. The aggregate level approaches seem to deliver insufficient information for measuring network proximity and thus for the empirical investigation of cross-regional knowledge spillovers. Model 1 and 2 take account only of whether a single network link exists between regional units but ignore the structure of linkages between actors.

Interestingly, once we use a non-aggregated measure of network proximity based on the network linkages established between actors across the regions (Models 3 to 5), we obtain positive and significant results for the network-based spillover parameter θ . The observed knowledge spillover effects are rather similar for all non-aggregated specifications of regional network proximity, irrespective of whether a representation considering first-order proximity (Model 4), first-order and higher-order proximity (Model 5) or an unweighted network representation is chosen. Also the model fit improves when the proposed measures are included. We can therefore conclude that an actor-level perspective on network proximity adds crucial information on regional interdependencies due to R&D network linkages. The number or strength of ties between regional actors seem to be fundamental in understanding the inter-regional transmission of knowledge through R&D networks. In our example, however, it appears that the direct linkages as measured by first-order proximity deliver more essential information to the model than the reach of indirect linkages as for higher-order proximity. Hence, the direct linkages across regions seem to moderate also the spillover impact of indirect linkages.

Based on the results of our empirical example, we may conclude that first-order proximity of regions, which is observable easily and less cumbersome to calculate from the information contained in the non-aggregated regional adjacency matrix, is a suitable proxy for measuring the network proximity of regions. The matrices resulting from the weighted proximity measure are highly correlated with those of the non-weighted representation (see Table A.1b). It is noteworthy at this point that EU funded projects typically involve a high number of participants, leading to a high mean degree of the network nodes and a high number of directly connected actors across regions. The picture regarding the significance of indirect network channels might therefore change when exploring other types of network structures.

7. Discussion and concluding remarks

The aim of this paper was to develop a new measure of network proximity for aggregated units by integrating the information as given by the network structure at the micro (non-aggregated) level. The proposed concept is tied particularly to R&D networks, in which linkages established on the basis of joint R&D projects are viewed in terms of channels for knowledge flows across actors and regions. Geographical regions, regarded as aggregated spatial units composed of individual actors, are used as instantiation of the general principles of the measure.

Our measure attempts to address the problem of aggregation in social networks. It differs from the previous approaches to construct R&D networks of regions in several aspects: Most importantly, *first*, a region is not captured as single node in the network but thought to represent an aggregate of actors located in this region and participating in R&D collaborations. We define the proximity measure on the basis of network paths between actor pairs located in different regions. The existence of at least one network linkage or a sequence of linkages, created at the actor level and that connects two aggregates, is a necessary condition to measure the network proximity between two aggregated units.

⁹ To check the robustness of the regression results, we tested for higher distance decay parameters in the construction of the measure. Moreover, we calculated proximity values based on 'relative proximity'. Hereby, we scale the network proximity between regions by the ratio between those firms linked via R&D collaborations and the total number of firms in the two regions. In both cases the substantive results of the regression in terms of effect size and significance do change only marginally. Concerning the latter, we additionally checked the correlations between a network weight matrices based on the 'absolute' proximity ($\lambda = 1$) as reported in Table 1 and a 'relative' proximity. A Mantel test based on Spearman rank correlations confirms the significantly positive relation between the matrices.

¹⁰ Data on EU funded R&D networks are drawn from the EUPRO database, which comprises information on research projects funded by the EU FPs (complete for FP1–FP7) and all participating organisations. See the studies of e.g. Scherngell and Barber (2009) and Wanzenböck et al. (2015) for further details on the database.

Table 1
Estimation results for different specifications of network proximity between regions.

| | (1) aggregate level, only direct links | (2) aggregate level, indirect links | (3) non-aggregated, unweighted | (4) Non-aggregated, weighted, only first-order | (5) Non-aggregated, weighted, first- and higher-order |
|-----------------------------------|---|--|-----------------------------------|--|---|
| Knowledge stock (β) | 0.088*** (0.027) | 0.091*** (0.028) | 0.065*** (0.011) | 0.068*** (0.011) | 0.065*** (0.011) |
| Knowledge spillover (θ) | −0.027 (0.032) | −0.030 (0.033) | 0.059*** (0.016) | 0.066*** (0.015) | 0.059*** (0.016) |
| Population (γ_1) | −0.001*** (0.000) | −0.001*** (0.000) | −0.001*** (0.000) | −0.001*** (0.000) | −0.001*** (0.000) |
| Population density (γ_2) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| R squared | 0.142 | 0.144 | 0.170 | 0.183 | 0.171 |

Notes: OLS estimates with White robust standard errors in parentheses. The dependent variable is regional TFP as defined in Eq. (12) in the Appendix; knowledge stock refers to a region's patent stock; knowledge spillover reflect the network-based lags of the regional patent stocks. We checked variance inflation factors (VIFs) for the variables.

*** Significant at the 0.01 significance level.

Second, the measure considers network distances across the aggregates (i.e. inter-regional linkages of the actors) but also within them (i.e. intra-regional linkages). We show that this constitutes a unique feature of our measure, as only the latter allows assessing whether regions are truly indirectly connected, and whether knowledge can flow through 'network channels' also over longer network distances. We propose a scaled Manhattan norm as the most suitable way to aggregate the actor-level information.

Third, a distance decay function is incorporated to place higher weights on shorter network distances. By this, depending on how this decay effect is specified, the network proximity measure is able to account for the importance of short network paths between the aggregated units. We argue that especially for the case of R&D networks, the decreasing probability and the changing nature of knowledge flows coming with greater network distance should be considered.

On this basis, we further propose to differentiate between first-order proximity (i.e. directly linked region pairs), and higher order proximity (i.e. indirectly connected regions according to the underlying network structure). A weighted network proximity measure consisting of two building blocks is introduced: first-order proximity captures the strength of ties between adjacent actors, while higher-order proximity captures the network distance between non-adjacent actors in two regions. By this, our approach for regional network representations corresponds well with the common characterisations of embeddedness in social networks of individuals (Granovetter, 1992). It builds on the idea that direct and indirect linkages fulfil different functions in networks. First-order proximity is associated with tacit knowledge flows induced within R&D collaborations. Higher-order proximity facilitates more the identification of alternative knowledge sources. In particular in the absence of geographical proximity, the structure of indirect linkages can be a crucial source of additional information for regional actors (e.g. Singh, 2005; Ter Wal, 2014).

In an illustrative example we demonstrated how the proximity measure can be applied to quantify network spillovers between aggregated units. By transforming the network proximity into an region-by-region weight matrix, as typically applied in spatial econometric models, we are able to identify productivity effects of cross-regional knowledge spillovers induced by R&D network proximity. The application shows that the measure clearly outperforms approaches where networks are constructed at the aggregated level; significant estimates for cross-regional knowledge spillovers are obtained for the proposed proximity measures.

The identification and consideration of different levels of influence is one of the cornerstones of multilevel network approaches. The proposed network proximity measure clearly relates to this relatively young but growing literature stream (see Lazega and Snijders, 2015; Lomi et al., 2016; Sun and Liu, 2016) by introducing a more realistic approximation of the micro-level network when macro-level interdependencies and their economic performance are modelled. We demonstrated that the analysis of multilevel

structures need not necessarily refer to the study of individual behaviour; modelling interrelations and the performance of macro units may corresponds in a similar way with the idea of multilevel thinking. However, investigating the geographical context of network nodes and considering them as additional structuring principle for network ties has been widely under-represented in multilevel network studies. Hence, applying the recently developed tools and models in the multilevel network field (Wang et al., 2013, 2016), together with a more nuanced contextualisation of network nodes, i.e. considering not only the organisational but also the geographical or the institutional level, would bring new insights into how linkages are formed and networks structured.

An additional issue to be further investigated concerns the evolution of network proximity, and with that, the application of the network proximity measure for dynamic analysis. Due to different network mechanisms, induced by, for example, information spillovers about valuable partners and new knowledge sources, higher-order proximity between regions may also trigger future first-order network proximity. Asking the question of how higher-order proximity contributes to first-order proximity between regional pairs, or other types of aggregates, would provide interesting insights in how R&D networks and network proximity evolve over time (Balland et al., 2015).

Finally, important to note is that the measure will be further developed in different aspects: We refrained from considering weights for the network distances in the general version of the network proximity measure. Incorporating information on link strength, also for indirect linkages in the sense of Opsahl et al. (2010), for instance, seems to be a natural extension of a proximity measure for aggregated units. In this way, the measure could be related more closely to concepts of network paths as used for the topology of multilayer networks (e.g. Boccaletti et al., 2014; Kivela et al., 2014). This would enable placing different weights on region-internal (i.e. intra-layer) and region-external (inter-layer) linkages. As the costs of region-internal collaborations may be considerable smaller than the costs associated with region-external collaboration, such a differentiation would provide new opportunities to analyse the influence of social network structures spanning geographical units such as regions.

Appendix A.

Details on the definition of the variables and data

Regional total factor productivity (TFP) levels are based on the index number approach as introduced by Caves et al. (1982), for time t defined by

$$y_{it} = \ln \left(\frac{Q_{it}}{\bar{Q}_t} \right) - (1 - \bar{s}_{it}^L) \ln \left(\frac{C_{it}}{\bar{C}_t} \right) - \bar{s}_{it}^L \ln \left(\frac{L_{it}}{\bar{L}_t} \right), \quad (12)$$

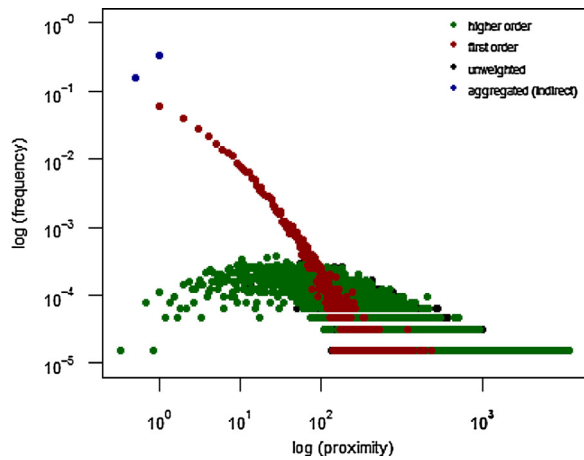


Fig. A.1. Proximity distribution of the different network proximity measures, 2004.

where y_{it} denotes the total factor productivity in manufacturing of region i at time t . Q_{it} , C_{it} , and L_{it} , refer to the output, physical capital stock and labour in the manufacturing sector, respectively. An upper bar denotes the geometric mean of the values of other regions at time t . \bar{s}_{it}^L is the share of labour income on gross value-added (GVA) in the manufacturing sector in region i at time t .

Regional output is measured by data on GVA in 2007, i.e. regional gross output less intermediate consumption (in euro, constant prices of 2000, deflated) in the manufacturing sector, regional labour inputs are measured in terms of the total number of annual hours worked in the manufacturing sector in 2007. As common practice, we use the perpetual inventory method to generate the physical capital stocks (see Hall et al., 2010). For the full sample investment data go back to 1990. We construct the initial level of the physical capital stock in 1999 based on the average geometric growth rate covering the period 1990 to 1998 (see e.g. Hall and Jones, 1996). The stock is weighted by a constant depreciation rate of ten per cent, and deflated data on gross fixed capital forma-

Table A.1

Summary statistics of the network structure and the network proximity weight matrices.

(a) The structure of the R&D network at the organisational level

| | 2002 | 2003 | 2004 |
|--------------------------------------|---------|---------|---------|
| # of organisations (= network nodes) | 17,625 | 16,961 | 17,768 |
| # of projects | 11,599 | 10,157 | 9,422 |
| # of edges | 329,955 | 343,689 | 468,245 |
| Mean degree | 37.44 | 40.53 | 52.71 |
| Max. degree | 3,421 | 3,995 | 7,045 |
| SD degree | 123.06 | 130.09 | 188.51 |
| Network density | 0.002 | 0.002 | 0.003 |
| Average distance | 3.03 | 3.01 | 2.90 |
| Share of main component (in %) | 93 | 97 | 97 |

Correlations of the network proximity matrices (av. 2002–2004)

| | (1) | (2) | (3) | (4) | (5) |
|---|-----|-------|-------|-------|-------|
| (1) Aggregate, direct | – | 0.989 | 0.252 | 0.214 | 0.251 |
| (2) Aggregate, indirect | | – | 0.250 | 0.211 | 0.249 |
| (3) Non-aggregated, unweighted | | | – | 0.923 | 0.997 |
| (4) Non-aggregated, unweighted, first-order | | | | – | 0.929 |
| (5) Non-aggregated, unweighted | | | | | – |

Notes: Correlation tests of the $N \times N$ weight matrices based on the Mantel test statistic (using the Pearson correlation coefficient).

tion are used to indicate the regional flows of investment in 2007. The share of regional labour income in the manufacturing sector is measured by the annual compensation of employees (i.e. wages and salaries and employers' social contributions) in 2007. Data are drawn from the Cambridge Economics database.

The regional knowledge stock is proxied in terms of patent applications in 2004 and the discounted stock of patents of previous periods. As common practice in the literature, we consider a time-lag to transform the already developed technological knowledge into regional TFP. The following expression $x_{i,2005} = (1 - r)x_{i,2004} + h_{i,2004}$ is used to derive the (logged) regional knowledge stocks, where $x_{i,2004}$ is the patent stock observed at the

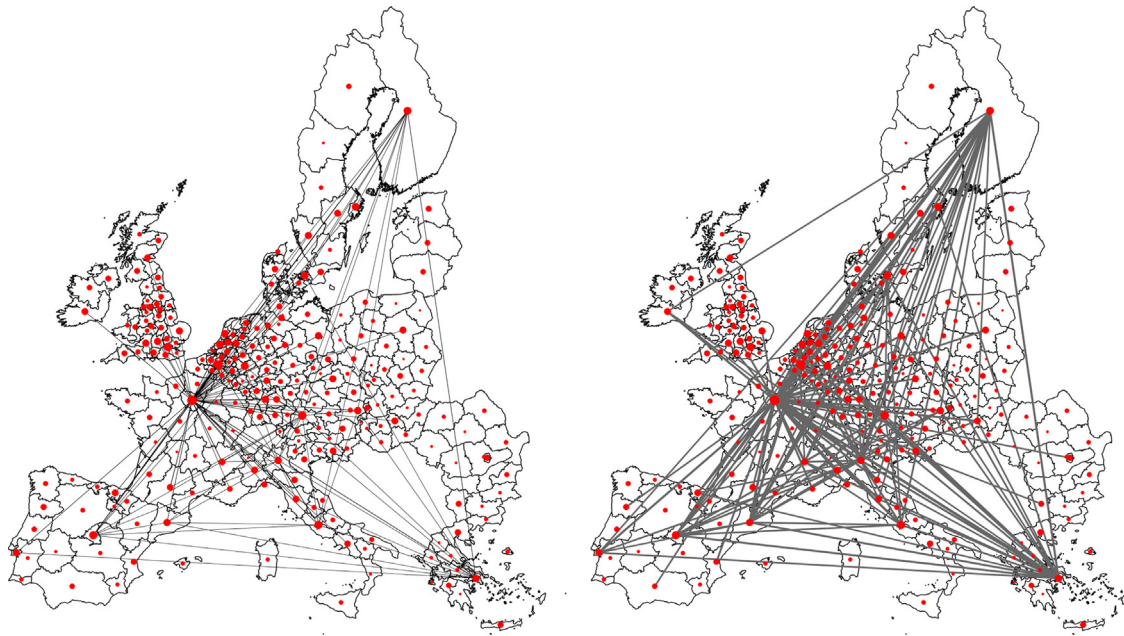


Fig. A.2. Regional proximity in the European R&D network, 2004. Notes: Node size corresponds to the number of direct linkages in the R&D network of a region, line width corresponds to network proximity of two region. Left plot displays first-order proximity, right plot displays higher-order proximity. For illustrative purposes, only the strongest 0.5% of all edges in terms of proximity are depicted.

beginning of the year 2004, and $h_{i,2004}$ denotes the patenting activities during 2004. We assume a constant rate of depreciation of $r = 12$ per cent (see Fischer et al., 2009), and consider the depreciated number of patent applications over the period from 1995 to 2003 to construct the initial stock in 2004. Regional patent applications refer to the number of patents assigned at the European Patent Office (EPO) or the World Intellectual Property Organisation (WIPO). To trace the geographical location of where the patentable knowledge has been created, we consider the address of the inventor, and in case of inventor teams, we follow the fractional counting approach.

References

- Aghion, P., Jaravel, X., 2015. Knowledge spillovers, innovation and growth. *Econ. J.* 125 (583), 533–573.
- Ahuja, G., 2000. Collaboration networks, structural holes and innovation: a longitudinal study. *Admin. Sci. Q.* 45 (3), 425–455.
- Autant-Bernard, C., Mairesse, J., Massard, N., 2007. Spatial knowledge diffusion through collaborative networks. *Pap. Reg. Sci.* 86 (3), 341–350.
- 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. *Reg. Stud.* 46 (6), 741–756.
- Balland, P.-A., Boschma, R., Frenken, K., 2015. Proximity and innovation: from statics to dynamics. *Reg. Stud.* 49 (6), 907–920.
- Barber, M.J., Fischer, M.M., Scherngell, T., 2011. The community structure of research and development cooperation in Europe: evidence from a social network perspective. *Geogr. Anal.* 43 (4), 415–432.
- Basile, R., Capello, R., Caragliu, A., 2012. Technological interdependence and regional growth in Europe: proximity and synergy in knowledge spillovers. *Pap. Reg. Sci.* 91 (4), 697–722.
- Bergé, L.R., 2016. Network Proximity in the Geography of Research Collaboration. *Papers in Regional Science*.
- Bergé, L., Wanzenböck, I., Scherngell, T., 2015. Centrality of regions in R&D networks: conceptual clarifications and a new measure. *Cahiers du GREThA 2015-31*. Groupe de Recherche en Economie Théorique et Appliquée.
- Boccaletti, S., Bianconi, G., Criado, R., Genio, C.I.d., Gómez-Gardenes, J., Romance, M., Sendina-Nadal, I., Wang, Z., Zanin, M., 2014. The structure and dynamics of multilayer networks. *Phys. Rep.* 544 (1), 1–122.
- Borgatti, S.P., 2005. Centrality and network flow. *Soc. Netw.* 27 (1), 55–71.
- Boschma, R., 2005. Proximity and innovation: a critical assessment. *Reg. Stud.* 39 (1), 61–74.
- Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. *J. Econom.* 150 (1), 41–55.
- Breschi, S., Lissoni, F., 2001. Knowledge spillovers and local innovation systems: a critical survey. *Ind. Corp. Change* 10 (4), 975–1005.
- Breschi, S., Lissoni, F., 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *J. Econ. Geogr.* 9 (4), 439–468.
- Broekel, T., Hartog, M., 2013. Determinants of cross-regional R&D collaboration networks: an application of exponential random graph models. *Ind. Innov.* 20 (3), 277–295.
- Caldarelli, G., Cristelli, M., Gabrielli, A., Pietronero, L., Scala, A., Tacchella, A., 2012. A network analysis of countries? Export flows: firm grounds for the building blocks of the economy. *PLoS ONE* 7 (10), 1–11.
- Caves, D.W., Christensen, L.R., Diewert, W.E., 1982. Multilateral comparisons of output, input, and productivity using superlative index numbers. *Econ. J.* 92 (365), 73–86.
- Coe, D.T., Helpman, E., 1995. International R&D spillovers. *Eur. Econ. Rev.* 39 (5), 859–887.
- Cowan, R., Jonard, N., 2003. The dynamics of collective invention. *J. Econ. Behav. Organ.* 52 (4), 513–532.
- Dijkstra, E., 1959. A note on two problems in connexion with graphs. *Numer. Math.* 1, 269–327.
- Everett, M.G., Borgatti, S.P., 1999. The centrality of groups and classes. *J. Math. Sociol.* 23 (3), 181–201.
- Fafchamps, M., van der Leij, M.J., Goyal, S., 2010. Matching and network effects. *J. Eur. Econ. Assoc.* 8 (1), 203–231.
- Fischer, M.M., Scherngell, T., Reismann, M., 2009. Knowledge spillovers and total factor productivity: evidence using a spatial panel data model. *Geogr. Anal.* 41 (2), 204–220.
- Fitjar, R.D., Huber, F., 2015. Global pipelines for innovation: insights from the case of Norway. *J. Econ. Geogr.* 15 (3), 561–583.
- Fleming, L., King, C., Juda, A.L., 2007. Small worlds and regional innovation. *Organ. Sci.* 18 (6), 938–954.
- Freeman, L.C., 1977. A set of measures of centrality based on betweenness. *Sociometry* 40 (1), 35–41.
- Giuliani, E., 2007. The selective nature of knowledge networks in clusters: evidence from the wine industry. *J. Econ. Geogr.* 7 (2), 139–168.
- Goyal, S., 2007. *Connections: An Introduction to the Economics of Networks*. Princeton University Press, Princeton.
- Granovetter, M., 1992. Problems of explanation in economic sociology. *Networks and Organizations: Structure, Form, and Action*, vol. 25, pp. 56.
- Granovetter, M.S., 1973. The strength of weak ties. *Am. J. Sociol.*, 1360–1380.
- Hall, B.H., Mairesse, J., Mohnen, P., 2010. Measuring the returns to R&D. *Handbook of the Economics of Innovation*, vol. 2. North-Holland, pp. 1033–1082.
- Hall, R.E., Jones, C.I., 1996. *The Productivity of Nations*, No. 5812. National Bureau of Economic Research.
- Hausmann, R., Hidalgo, C.A., 2011. The network structure of economic output. *J. Econ. Growth* 16 (4), 309–342.
- Hazir, C.S., 2013. Multilateral R&D collaboration: an ERGM application on biotechnology. In: Scherngell, T. (Ed.), *The Geography of Networks and R&D Collaborations*. Springer-Physica Verlag, Berlin-Heidelberg/New York, pp. 221–237.
- Huggins, R., Thompson, P., 2014. A network-based view of regional growth. *J. Econ. Geogr.* 14 (3), 511–545.
- Jackson, M.O., 2010. *Social and Economic Networks*. Princeton University Press, Princeton.
- Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *J. Econom.* 157 (1), 53–67.
- Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A., 2014. Multilayer networks. *J. Complex Netw.* 2 (3), 203–271.
- Lazega, E., Jourda, M.-T., Mounier, L., Stofer, R., 2008. Catching up with big fish in the big pond? Multi-level network analysis through linked design. *Soc. Netw.* 30 (2), 159–176.
- Lazega, E., Snijders, T., 2015. *Multilevel Network Analysis for the Social Sciences: Theory, Methods and Applications*. Methodos Series. Springer International Publishing.
- Leenders, R.T.A.J., 2002. Modeling social influence through network autocorrelation: constructing the weight matrix. *Soc. Netw.* 24 (1), 21–47.
- LeSage, J., Pace, R., 2009. *Introduction to Spatial Econometrics*. Statistics: A Series of Textbooks and Monographs. CRC Press, Boca Raton, New York.
- Lomi, A., Robins, G., Tranmer, M., 2016. Introduction to multilevel social networks. *Soc. Netw.* 44, 266–268.
- Lundvall, B.-Å., Johnson, B., 2016. The learning economy. In: Lundvall, B.-Å. (Ed.), *The Learning Economy and the Economics of Hope*. Anthem Press, New York, London, pp. 107–129 (chapter 5).
- Maggioni, M.A., Nosvelli, M., Uberti, T.E., 2007. Space versus networks in the geography of innovation: a European analysis. *Pap. Reg. Sci.* 86 (3), 471–493.
- Newman, M.E., 2001. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci.* 98 (2), 404–409.
- Newman, M.E.J., 2004. Analysis of weighted networks. *Phys. Rev. E* 70, 056131.
- Newman, M.E., 2006. Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* 103 (23), 8577–8582.
- Opsahl, T., 2013. Triadic closure in two-mode networks: redefining the global and local clustering coefficients. *Soc. Netw.* 35 (2), 159–167.
- Opsahl, T., Agneessens, F., Skvoretz, J., 2010. Node centrality in weighted networks: generalizing degree and shortest paths. *Soc. Netw.* 32 (3), 245–251.
- Opsahl, T., Panzarasa, P., 2009. Clustering in weighted networks. *Soc. Netw.* 31 (2), 155–163.
- Owen-Smith, J., Powell, W.W., 2004. Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community. *Organ. Sci.* 15 (1), 5–21.
- Paci, R., Marrocu, E., Usai, S., 2014. The complementary effects of proximity dimensions on knowledge spillovers. *Spat. Econ. Anal.* 9 (1), 9–30.
- Ponds, R., Oort, F.v., Frenken, K., 2010. Innovation, spillovers and university-industry collaboration: an extended knowledge production function approach. *J. Econ. Geogr.* 10 (2), 231–255.
- Rokach, L., 2010. A survey of clustering algorithms. In: Maimon, O., Rokach, L. (Eds.), *Data Mining and Knowledge Discovery Handbook*, 2nd ed. Springer Science+Business Media, New York, Dordrecht, Heidelberg, London, pp. 269–298.
- Sailer, L.D., 1979. Structural equivalence: meaning and definition, computation and application. *Soc. Netw.* 1 (1), 73–90.
- Scherngell, T. (Ed.), 2013. *The Geography of Networks and R&D Collaborations*. Springer-Physica Verlag, Berlin-Heidelberg/New York.
- Scherngell, T., Barber, M.J., 2009. Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. *Pap. Reg. Sci.* 88 (3), 531–546.
- Sebestyen, T., Varga, A., 2013. Research productivity and the quality of interregional knowledge networks. *Ann. Reg. Sci.* 51 (1), 155–189.
- Singh, J., 2005. Collaborative networks as determinants of knowledge diffusion patterns. *Manag. Sci.* 51 (5), 756–770.
- Sorenson, O., Rivkin, J.W., Fleming, L., 2006. Complexity, networks and knowledge flow. *Res. Policy* 35 (7), 994–1017.
- Sun, Y., Liu, K., 2016. Proximity effect, preferential attachment and path dependence in inter-regional network: a case of China's technology transaction. *Scientometrics* 108 (1), 201–220.
- Ter Wal, A.L.J., 2014. The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure. *J. Econ. Geogr.* 14 (3), 589–620.
- Ter Wal, A., Boschma, R., 2009. Applying social network analysis in economic geography: framing some key analytic issues. *Ann. Reg. Sci.* 43 (3), 739–756.
- Uzzi, B., 1997. Social structure and competition in inter-firm networks: the paradox of embeddedness. *Admin. Sci. Q.* 42 (1), 35–67.

- Varga, A., Pontikakis, D., Chorafakis, G., 2014. Metropolitan Edison and cosmopolitan Pasteur? Agglomeration and interregional research network effects on European R&D productivity. *J. Econ. Geogr.* 14 (2), 229–263.
- Wang, P., Robins, G., Pattison, P., Lazega, E., 2013. Exponential random graph models for multilevel networks. *Soc. Netw.* 35 (1), 96–115.
- Wang, P., Robins, G., Pattison, P., Lazega, E., 2016. Social selection models for multilevel networks. *Soc. Netw.* 44, 346–362.
- Wanzenböck, I., Scherngell, T., Lata, R., 2015. Embeddedness of European regions in European Union-funded research and development (R&D) networks: a spatial econometric perspective. *Reg. Stud.* 49 (10), 1685–1705.
- Wasserman, S., Faust, K., 1994. *Social Network Analysis: Methods and Applications. Structural Analysis in the Social Sciences*. Cambridge University Press.
- Ziberna, A., 2014. Blockmodeling of multilevel networks. *Soc. Netw.* 39, 46–61.