G Model
SON-1054; No. of Pages 12

ARTICLE IN PRESS

Social Networks xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

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 xxx

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.

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. 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

☆ 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

https://doi.org/10.1016/j.socnet.2017.10.003
0378-8733/

Please cite this article in press as: Wanzenböck, I., A concept for measuring network proximity of regions in R&D networks. Soc. Netw. (2017), https://doi.org/10.1016/j.socnet.2017.10.003

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 $V = \{v_1 \ldots, v_M\}$ stands for the M actors participating in R&D collaborations. The set of edges $L = \{l_1, \ldots, l_k\}$ corresponds to the set of R&D collaborations between these actors. An R&D collaboration between two actors v_u and v_q $(u, q = 1, \ldots, M)$ is represented by an edge $\left(v_u, v_q\right) = l_k \in$ 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

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

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

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

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 $(i = 1, \ldots, N)$, and accordingly, by R_i the index set of the actors that belong to V_i, i.e. $R_i = \{u \in \{1, \ldots, M\} : v_u \text{ belongs to region } i\}$. 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 $\mathcal{K}$th-power of the adjacency matrix A, so that $A^{\mathcal{K}}$ give the number of walks of length $\mathcal{K}$ 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 $\mathcal{K} \leq 2$. 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).