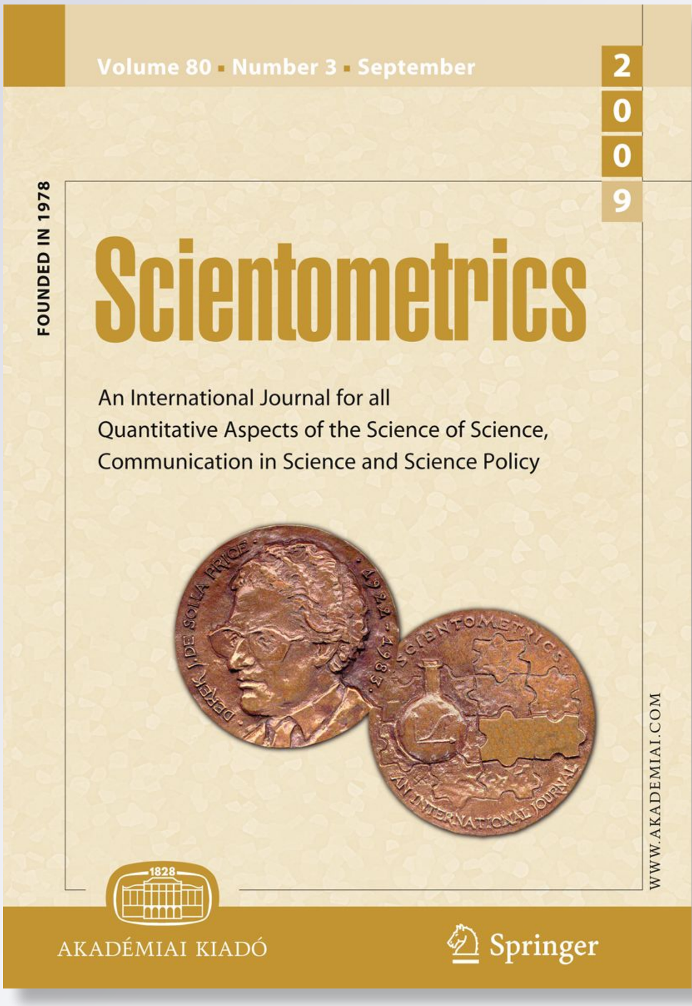
Beyond the basemap of science: mapping multiple structures in research portfolios: evidence from Hungary

Sándor Soós & George Kampis

Scientometrics
An International Journal for all Quantitative Aspects of the Science of Science, Communication in Science and Science Policy

ISSN 0138-9130

Scientometrics
DOI 10.1007/s11192-012-0713-x



Springer

Scientometrics
DOI 10.1007/s11192-012-0713-x

Beyond the basemap of science: mapping multiple structures in research portfolios: evidence from Hungary

Sándor Soós · George Kampis

Received: 1 November 2011

Abstract As a novel tool for evaluating research competences of R&D actors, science overlay maps have recently been introduced in the scientometric literature, with associated measures for assessing the degree of diversification in research profiles. In this study, we continue the elaboration of this approach: based on science overlay maps (called here *m-maps*), a new type of map is introduced to reveal the competence structure of R&D institutions (*i-maps*). It is argued, that while *m-maps* represent the multidisciplinarity of research profiles, *i-maps* convey the extent of interdisciplinarity realized in them. Upon *i-maps*, a set of new measures are also proposed to quantify this feature. With these measures in hand, and also as a follow-up to our previous work, we apply these measures to a sample of Hungarian Research Institutions (HROs). Based on the obtained rankings, a principal component analysis is conducted to reveal main structural dimensions of researh portfolios (of HROs) covered by these measures. The position of HROs along these dimensions then allows us to draw a typology of organizations, according to various combinations of inter- and multidisciplinarity characteristic of their performance.

Keywords Science overlay maps · Science mapping · Interdisciplinarity · Multi disciplinarity · Network analysis · PCA · Diversity index · Integration index · Polarization index · Hungary

Introduction

In recent years, evaluating research performance of the actors in S&T markets based on global maps of science has become a rapidly evolving trend (Boyack 2009; Rafols et al. 2010). A well-known approach, referred to as "science overlay maps" has been introduced by

S. Soós (✉)
Department of Science Policy and Scientometrics, Library of the Hungarian Academy of Sciences, Budapest, Hungary
e-mail: soossand@caesar.elte.hu

G. Kampis
History and Philosophy of Science, Lorand Eötvös University, Budapest, Hungary

Published online: 08 April 2012

Rafols et al. (2010). At the heart of this method is the utilization of a so-called basemap of science, a reference system for comparing the publication profile of actors. The basemap, with the aim of modelling the current organization of the science system, is made up of ISI Subject Categories (SCs) or "research fields", and conveys their relations, that is, their proximity network constructed from bibliographic couplings (in terms of cited SCs). Given a particular actor *A* (a research body or a country), this map enables the analyst to project the publication portfolio of *A* on the network, by identifying nodes, that is, Subject Categories, where *A* bears research output or impact (retrieved from the ISI databases), and weighting them according to their role in the profile of *A*. Thereby a so-called overlay map, or a customized version of the global map is created, as a kind of "fingerprint" characterizing the position, or (from a strategic perspective) stregths and weaknesses of *A* in scientific markets.

The overlay map is a rich source of structural information for the assessment and comparisons of research performance. Most importantly, the network representation makes it feasible to formulate novel measures of how diversified a given portfolio is. As the inventors of the method have pointed out (cf. Leydesdorff 2010), since the underlying reference system contains the relative position of research fields, it supports diversity metrics that take into account the disparity in the profile, that is, the relatedness and distance of active fields beyond variation and evenness (the number and the relative weight of fields in the portfolio). To capture the notion of "the more [in number], the more balanced, and more distant areas are active, the greater the diversity is", the cited authors have utilized the Stirling index (Stirling 2007) and its variants (Soós and Kampis 2011). Hereafter we refer to these family of indices as "diversity measures".

In a previous study, elaborating on the results briefly exposed above, we developed a new, modified version of the Stirling index, along with its application to the set of top S&T performers in Hungary (Soós and Kampis 2011). We found that, as compared to the diversity indices, the new measure revealed an additional dimension of research "heterogeneity", namely, the extent of polarization in the composition of research output (polarization index). This study was intended to make an initial step in exploring the structural aspects of performance that can be uncovered by exploiting the information content of such overlay maps.

Such a new aspect, or novel structural dimension, relative to the original use of overlay maps so far, is addressed in the research presented below. This new approach is best described in relation to the concept and the principal features of overlay maps and diversity measures. Overlay science maps are designed to quantify and visualize the relationship between an organization's research activity/success, and the current reference system of science, whereby the connections between fields (Subject Categories) are given by global citation patterns, independently of the organization's profile in question. More specifically, as an overlay map is generated by highlihgting the Subject Categories from an actor's profile, the map and the relations are invariant for all these maps (that explains the term "basemap"), derived from the ISI databases.

However, in characterizing the competences of an actor in S&T, equally relevant might be the issue of relationships between research fields *within* its own portfolio. Arguably, these types of relationships are being suppressed when using the overlay technique. In particular, if an institution *A* has a set of publications simultaneously assigned to the SCs *molecular biology* and *food science*, this output will be present in the overlay map as two distinct activities, but the exposed link, or the joint competence/topic will be lost for the end result. In conventional terms, we might formulate this observation that the map will reflect the *multidisciplinarity* of the research profile, at the cost of hiding the *interdisciplinarity* of the output. This behavior is due to the fact that original overlay maps reflect the distribution of the publication record of *A* (papers) over Subject Categories, while interdisciplinary relations could, in this sense, be revealed via the opposite perspective: the distribution of Subject Categories over the publication record.

The bibliometric characterization of multi- and interdisciplinarity has recently been developed into a stikingly active research topic, with rapid developments. The overlay technique utilized in this study is part of a comprehensive conceptual and methodological framework proposed in a series of papers by Leydesdorff, Meyer, Porter and Rafols (Porter et al. 2007; Leydesdorff 2007; Rafols and Meyer 2010; Leydesdorff and Rafols 2011; Rafols et al. 2011). Not only does this framework reinforce the distinction highlighted above, i.e. the orthogonal relationship between the (1) diversity and (2) interrelatedness of fields underlying a research profile, but also provides a theoretically sound taxonomy of the various conceptualizations for inter- and multidisciplinarity, along with operationalizing them by fitting a distinct set of novel bibliometric measures to each. Within this taxonomy (an in-depth articulation of which can be found in Rafols et al. 2011), that focuses on the patterns exhibited in publication records (at varying levels of aggregation), diversity measures are clearly associated with the degree multidisciplinarity. On the other hand, the notion of interdisciplinary research (IDR) is decomposed into two differing perspectives:

(1) On one account, IDR is conceived as *knowledge integration*, an indicator of which is the degree of overall interrelatedness of the units of analysis constituting a portfolio. Integration is here the very aspect contrasted with diversity the previous discussion. Since the constitution of profiles is measured in terms of ISI Subject Categories in our case, interconnections are to be assessed between SCs as the units of analysis.
(2) On the other account, however, interdisciplinarity is viewed as *intermediation* (Leydesdorff 2007) between knowledge domains, embodied in publication sets positioned between more established clusters of journals or fields. In this case, instead of overall relatedness, the interdisciplinarity is recognized as a capacity to relate otherwise non-interacting fields (or units of competence) via a research profile. It should be noted that, based on the usually non-paradigmatic character of intermediary research, this aspect of IDR has been argued to be recognizable at a lower level of aggregation, that is, below the established fields or even subfields. The proposed units of analysis has been, therefore, journals or individual papers far below the aggregation level of Subject Categories (ibid).

Based upon the arguments above, our aim in this paper is to experiment with a method that (1) utilizes the overlay mapping technique and (2) is also able to express the degree of interdisciplinarity for any depicted S&T player, while taking into account the multidimensional nature of this concept. To this end, drawing on the framework proposed by the authors referred to above, we introduce an additional model of research portfolios, that captures the interplay between their constituent fields, leading also to a network of Subject Categories as well. To distinguish this type of network, we call it the *i-network* (as a shorthand for “interdisciplinarity networks”) of the actor, as contrasted to the overlay map of the same portfolio, referred to as the *m-network* (for “multidisciplinarity network”, the names being drawn from the argument above). We define *i-network*s so that they also inherit the relational information from the basemap of science (the basis for *m-networks*), that is, the global relation of fields. Upon this novel types of structural models, we build an enriched set of measures in order to assess the interdisciplinarity of publication profiles, just as diversity measures have been used, in relation to overlay maps, to capture the degree of multidisciplinarity. The indices of interest will be discussed in the context of the taxonomy imposed on IDR research, and also contrasted with the indicators proposed to serve that taxonomy.

Given the above variety of structural measures, we are primarily interested in their comparative analysis. The issue addressed here is multifold: First, we would be interested in (1) how a certain group of S&T actors performs assessed by diversity versus integration (or intermediation) measures, or, more generally, in the relationship of inter- and multidisciplinarity. A further, more ambitious (but interrelated) set of questions is (2) what are the major dimensions of the structure and diversification of research that can be identified upon this pool of measures, and (3) what typology of S&T institutions can be drawn along these dimensions. In this study, we focus on the same sample of organizations that we started to explore in a previous study (Soós and Kampis 2011): the top ranking Hungarian R&D organizations (in terms of their output), in order to contribute to the issues (1)–(3). Accordingly, the paper is organized as follows.

In the next section, we formally introduce the concept of an i-network. Based upon this concept, we define two sets of interdisciplinarity measures with respect to research portfolios to exploit the information content of i-networks.

In the subsequent section, the results of applying the new measures to the Hungarian sample are discussed. The analysis is an operationalization of issues (1)–(3) in the following sense:

(1) First, we compare the rankings imposed on the set of Hungarian organizations by the newly introduced metrics, against the rankings obtained by diversity measures in our previous study. The goal, beyond characterizing the sample organizations, is to detect the relation of inter- and multidisciplinarity within this sample.
(2) Based on the rankings, a principal component analysis (PCA) is performed of the whole series of measures, in order to extract the major dimensions covered by these tools.
(3) The resulted factors from the PCA are used to set up a typology of sample institutions as well, by ordering and clustering them via these factors expressing the main dimensions of research diversity and integration.

Methods and materials

The model: i-networks

To reveal the organization of research fields within a given publication record, we employed the concept of *interdisciplinarity networks* or *i-networks* characteristic of the profile of an S&T actor. The model relies upon the Subject Category (SC) assignments in the ISI databases, whereby journals (and, therefore, journal publications), considered to be interdisciplinary, are represented under each related category. The i-network for organization A, extracted from its publication record in the ISI, is then a weighted proximity network of the fields constituting the profile of A. The nodes of this network is the set of Subject Categories drawn from the publication record, and the ties represent the association of SCs, i.e., their co-occurence in the profile. Edge weights convey the strength of association in terms of (normalized) co-occurence frequencies.

That is, the i-network of the research profile for organization A is a graph

$$\mathrm{i-network}\ (A) =_{df} \langle SC_A, E_A, W_A \rangle \ , \quad \text{where}$$

$$SC_A =_{df} \{ \text{ SC is an ISI Subject Category} \mid \text{some publications in the record of } A \text{ are assigned to SC } \}$$

$$E_A =_{df} \{ \langle SC_1, SC_2 \rangle \mid \text{some publications in the record of } A \text{ are assigned to both } SC_1, SC_2 \in SC_A. \}$$

Edge weights, the elments of W_A for any two related Subject Categories SC_i and SC_j in SC_A are calculated as their co-occurence frequency normalized to the sum of co-occurence frequencies within the profile of A:

$$w_{ij} =_{df} p_{ij} = \frac{\#\mathrm{co-occurenceof}SC_i\mathrm{and}SC_j}{\#\mathrm{all\ co-occurences\ of\ SCs\ in}\ SC_A}.$$

The relative frequency of co-occurences was chosen for weighting ties in order to indicate the share of a particular interdisciplinary connection from to the total amount of interplay within the portfolio.

The concept of i-network, as defined above, is closely related to the method proposed by Rafols et al. (2011) (a work in parallel) for the measurement of IDR dimensions via science overlay maps. By the same scheme, the corresponding procedure underlying IDR measurements is to first construct a network of SCs reflecting their interrelatedness within a profile, and then to superimpose this new map on the basemap reflecting the positions of subfields on the global science map. A main conceptual difference lies in the particular relation when setting up such networks. In our case (i-networks), the indicator of connectedness is the intersection of SCs within the profile determined by the assignments of source publications (predefined in WoS databases). The alternative is, roughly speaking, to use the co-occurence of SCs at the level of references within the same profile (determined by reference patterns of source publications, instead of WoS co-classifications). Although relying on references, that is, the knowledge base of the publication profile, is undoubtedly more accurate in revealing knowledge integration than using WoS co-classification (which has recognized shortcomings for analytic purposes[1]), we nevertheless decided to the use WoS co-classification in this study for two, interrelated reasons. Among our original goals was to complement the data structure originally proposed in (Rafols–Porter–Leydesdorff 2010) for assessing multidisciplinarity at the institutional level (i.e. the distribution of Subject Categories over source publications), with a data structure of the same level of availability and simplicity, that serves the parallel assessments of interdisciplinarity. Especially when large-scale samples, such as a decade-long dataset of national output used in our study, are investigated, there is a huge gap, both in terms of computational demand and data availability, between evaluating multidisciplinarity at the source level, and interdisciplinarity at the references level. This trade-off between accuracy and general feasibility also led us to experiment with i-networks based on co-assignments. (Even so, comparative studies of interdisciplinarity measurements based on the two respective indicators, to address issues of sensitivity and robustness, are in order.)

In the parallel publication Rafols et al. (2011) alternative i-maps are superimposed on the basemap of science, generating a further overlay. In principle, this could be done with our i-maps as well; yet the reason why we keep maps separate from the (multidisciplinarity-conveying) overlays is to support the analytic purposes outlined below. By

[1] We thank one of the referees for pointing out this issue.

overlaying i-maps on the backbone, the relative positions of the fields on the visualization are determined by the basemap, that is, via the global similarity of fields. However, since we are particularly interested in the topology of i-maps (the topological features induced by interdisciplinary relations and not others), we force our visualizations to convey this topology or configuration specific to the individual i-maps (to support the interpretation of the topological measures introduced below).

The measurement of profile interdisciplinarity

Given the subject networks defined above, our primary goal was to construct a measure (or a set of measures) that is capable of grasping the degree of interdisciplinarity encoded in those graphs. In doing so, we consider diversity measures introduced in relation to m-networks, that is, overlay maps as a starting point. These measures of multidisciplinarity are designed to sum up the connections in the m-network within active fields, weighted by their distance and the relative size of the SCs involved. In a parallel manner, the amount of interdisciplinarity exhibited by a portfolio can be conceived as being proportional to

(1) the number of connections within the i-network (that is, the number two different fields interacting with each other),
(2) the weight of ties (the strength of the interactions) and
(3) the proximity/distance of interacting fields according to the basemap of science, i.e. the m-network.

In the framework decomposing interdisciplinarity, this conceptualization responds to the *knowledge integration* perspective. According to this conceptualization, we defined the degree of interdisciplinarity for a profile as follows:

Field coherence

$$\sum_{i=1,j=1,i\neq j}^{n} p_{ij}d_{ij},$$

where p_{ij} is the relative frequency of SCs i and j co-occuring in the profile, or the weight of the tie connecting i and j in the i-network, and d_{ij} is the distance of i and j according to the m-network. This definition implies that the measure above is dependent upon both models of a research profile: it conveys the pattern of the interrelation of fields (p_{ij}), but also draws on the relative position of them in the global map of science (d_{ij}). As an informal definition, the formula says that „the more strongly the distant areas are related in the portfilio, the greater its "interdisciplinarity", in the sense of strongly connecting disparate items of knowledge.
It is worth noting that in the work of Rafols et al. (2011), knowledge integration is operationalized in a rather similar manner, resulting in a measure called *Coherence*—the name of the above index, in part, derives from its direct relation to this alternative. As explained with respect to mapping differences concerning i-maps versus their alternatives, the main contrast between the two different integration measures stems from the interpretation of p_{ij}. The weight of a field–field connections is, in the other version, based on co-reference frequencies of SCs within the publication profile instead of co-assignments of source publications. As a further difference, the measure applied in (Rafols et al. 2011) is normalized with the diversity of the profile, which, basically, aims to correct for size differences of portfolios. As to our version, we omit this step for a reason discussed in a subsequent section (see "Results and Discussion").

The indicator of field coherence quantifies the degree of knowledge integration at the level of fields (or rather subfields), that is, Subject Categories. However, one might wonder whether, and to what extent linking knowledge items is being realized at higher levels of aggregation: to what degree the different *disciplines*, comprised of Subject Categories, are related by research portfolios. Arguably, linkages within disciplines are less "interdisciplinary" than those between disciplines, or larger areas of science.

The basemap used in this study (Leydesdorff and Rafols 2009), provides us with further information as to the position of Subject Categories occupied in the overall system of science. Based on the citation matrix underlying its topology, the cited authors performed a classification of SCs (or, technically speaking, a PCA), aggregating them into large disciplines jointly costituting the top-level taxonomy of science. Since, via this grouping, each subject category is assigned to a top-level subject class or *discipline*, it is possible to sharpen our definition of interdisciplinarity with a further formula that takes into account the disciplinary structure as well. The idea behind is that a link between two SCs in the i-network is more "valuable" in terms of bringing together diverse subjects, if those SCs belong to different disciplines. Formalizing this, we add.

Disciplinary coherence

$$\sum_{i=1,j=1}^{n} p_{ij} d_{ij} c_{ij}$$

to our toolkit of measurements, where

$$c_{ij} = \begin{cases} 0, & \text{iff Discipline}(i) = \text{Discipline}(j), \\ 1, & \text{otherwise.} \end{cases}$$

The formula is a modification of Field coherence introduced above: the technical difference is that we further weighted actual distances (d_{ij}) with a "filter" variable (c_{ij}) accounting for the discipline (that is, the aggregate category) of the fields engaged in the relation. Conceptually, Disciplinary coherence captures those links and distances in the i-network, that involve a switch in discipline, and ignores other links connecting SCs being uniform at a higher level, thus belonging to the same general category.
The two measures presented above both focus on the quantity of interdisciplinary connections (taking into account the disparity exemplified by each). An equally relevant aspect, however, is the topological information encoded in i-networks, and neglected so far. Not only do these graphs provide the basis for the "enumeration" of such connections, but, through their topological features, convey information on the particular pattern (or "configuration") of how the ingredients (fields) are connected by the profile. In (Rafols et al. 2011), pattern-related properties of interdisciplinarity graphs were utilized to characterize a further aspect of the notion, namely, that of *intermediation* (as contrasted to *integration*). Leydesdorff (2007) provided an analysis of betweenness centrality as an indicator of the role of units (in his case, journals) in connecting originally disconnected areas of research. In (Rafols et al. 2011), the application was further elaborated, and complemented with a further measure, clustering coefficient.

In our study, to pave the path to the experiment central to our aims (the exploration of the structural dimensions of research profiles that can be identified upon multi- and interdisciplinarity measurements), it is also of fundamental interest to find indicators for the configuration of knowledge domains. Therefore, in part drawing on previous indicators of intermediation, we adopted a set of measures from social network analysis, assumed to be capable of conveying a relevant aspect of interdisciplinary patterns. As can be seen from the discussions below, our topological selection is expected to add both to the integration and the intermediation view of IDR.

Range

The range of interdisciplinary connections is proposed here to investigate how far the relations within an i-network can take us in terms of pairwise integrated research fields. The most natural candidate to grasp this property formally is by the length of the maximal (shortest) path connecting subfields in the portfolio, i.e. the diameter of the network. The diameter, at first glance, might serve as a proxy for the (maximal) range of knowledge integration achieved in the portfolio. However, on closer inspection, this measure seems to contribute to both the (1) integration and the (2) intermediation view of interdisciplinarity. (1) A low diameter indicates that the constituent subfields are "closely linked" (quickly reachable from each other), hence a higher degree of coherence. On the other hand, (2) not only does a high diameter value suggest that (at least) some constituent fields are less readily connected, but also that these SCs are linked through several other subfields, where the latter by necessity mediate between them. In this sense, high diameter values also express a higher scope or range of intermediation exhibited in the portfolio.

- Diameter. The maximum number of steps between any two Subject Categories SC_1 and SC_2 minimally required to reach SC_1 from SC_2 (and vice versa) via the i-network.

Multimodality

A further set of topological measures has been selected to explore the macro-level pattern of knowledge integration within the research profile. An indicative feature of the extent to which the various fields, underlying a body of publications, are interlinked, is the fragmentation of the i-network. If, as can be expected, SCs in the network tend to cluster into distinct aggregates of varying sizes, then knowledge integration can be said "multimodal". That is, integration occurs in certain subspaces of the disciplinary backgorund, that are otherwise still isolated. Multimodality shall inform us, therefore, about how many (coherent) subgroups the constituent fields are being organized into, and in what distribution. A natural way to operationalize this aspect is to focus on the connected components of the i-network, using the following indicators.

- Number of connected components (#components). The number of subnetworks, or independent subgroups of related fields.
- Share of maximal component (Share): The relative size of the largest subgroup of related fields.
- Size distribution of components (Shannon): Quantified by the Shannon-Wiener entropy measure: $-\sum_{i=1}^{n} q_i \log q_i$. The latter shows the balance in the distribution of network elements among network components (q_i stands for the relative size of component i among n components). Small values are the sign of an imbalanced group structure, whereby the vast majority of fields are related (directly or indirectly) forming a single giant component, while higher values result from a more even distribution, with more similar sized, independent aggregates. Both Share and Shannon can be used as proxies of how balanced the knowledge integration is (into subdomains that are interdisciplinary by themselves) within the given space of Subject Categories.

Intermediation

Finally, the two topological measures entertained in the above referred, related studies are also incorporated in a specific way, in order to explore the extent of intermediation exhibited by portfolios. As noted above, intermediation has been proposed as best detected at levels of aggregation lower than that of fields or subfields (SCs). However, results of this study point to the direction that its derived measures can be valuable indicators for SC-profiles as well, as they do contribute to the structural dimensions gained in our experiment (the proper interpretation of these indices with respect to IRD dimensions, therefore, found its place in the discussion of the results):

- Normalized maximum betweenness centrality (Max betweenness): The maximum of betweenness centralities (*bw*) in the i-network was used as a sign of the tendency for profiles to possess central fields, that connect many others. Since bw is size-dependent, we applied size-normalized centrality values calculated as $\frac{bw}{(n-1)(n-2)}$, where n is the number of nodes in the network.
- Clustering coefficient (cc): The average local clustering coefficient for the i-networks was calculated as a quantitative measure of the degree of integration between SCs connected at least indirectly. Since this network indicator quantifies the extent to which neighbours of a node are themselves directly connected, cc indicates the tendency of transitivity in linking fields, or, to what extent intermediation "results" in integration.

To provide a snapshot of our system of indicators, relative to the reference system set up by the conceptual dimensions of IDR, we first contrasted the topology-related measures above with IDR perspectives (Table 1). A more comprehensive taxonomy, including Field coherence and Disciplinary coherence as well, is provided in Table 2, along with a comparison between i-networks and overlay maps (m-networks). The diversity and polarization measures listed for the m-networks are defined and analized in detail in our previous work (Soós and Kampis 2011).

Materials: the Hungarian sample

For the purposes of comparison, and also to elaborate on the structural description, we used the very same body of bibliographic data, as analyzed in our previous study of research performance diversity (Soós and Kampis 2011). We compiled the publication record of Hungarian research organizations (HROs) for the recent decade, that is, covering the period 2000–2009. Data were retreived from the TRTM databases through the ISI WoS portal. The resulting dataset was subjected to a thorough cleaning procedure, which consisted of the normalization of institutional names. Since institutional affiliations are represented in publications at various organizational levels (such as the university level or the faculty level),

Table 1 Topological measures and their relation to conceptual dimensions of IDR

Topological measures	Default interpretation	Coherence	Intermediation
Diameter	Range of integration	x	x
Connected components	Multimodality of integration	x	
Max. betweenness	Extent of intermediation		x
Average (local) clustering coefficient	Effects of intermediation		x

Table 2 Comparison of the i-network and the science overlay map modelling a given research profile

	i-network of profile A	m-network (overlay map) of profile A
Nodes of the network	Subject Categories used in ISI databases and being present in profile A	Subject Categories used in ISI databases
Content of ties in the network	Association of SCs in profile A	Global poximity of SCs based on their citation pattern
Specificity to profiles	i-network is specific to profile A	m-network is only specific to profile A w.r.t the node weights
Conceptual dimensions	Integration and intermediation	Diversity and polarization
Measures	*Field/Disciplinary coherence* $\sum_{i=1,j=1}^{n} p_{ij}d_{ij}$, $\sum_{i=1,j=1}^{n} p_{ij}d_{ij}c_{ij}$, *Topology measures*diameter, component statistics, betweenness, clustering coefficient	*Diversity measures*Stirling index (Rao–Stirling diversity) $\sum_{ij(i \neq j)} d_{ij}p_ip_j$, Modified Rao–Stirling diversities $\sum_{ij(i \neq j)} g_{ij}p_ip_j$, $\sum_{ij(i \neq j)} g_{ij}^{W} p_ip_j$
Interpretation of measures	Interdisciplinarity in profile A	Multidisciplinarity in profile A

making organizations comparable required to aggregate publication entries at a selected, more-or-less uniform level. For a definite part of our data, organizations were referred to at the topmost level (e.g. MTA, Hungarian Academy of Sciences, an umbrella term for many research institutions) that could not be disaggregated, while others were cited at some lower levels (e.g. as some institution belonging to the MTA). Because of this feature of the dataset, and also to avoid imposing ad hoc hypotheses on the equivalence of organizational units, we used the top level for each organization. This resulted in a set of altogether 6,154 research units, including Hungarian universities, governmental institutions concerned with research and development, and various companies exhibiting R&D activity.

In a subsequent step, this maximal list was reduced to a sample containing the "biggest" actors in Hungary, based on a ranking of the listed organizations according to the size of their publication record. In particular, actors were included that posessed a minimum of 100 publications per organization per year within the 10-year window of analysis. In the final set, 27 HROs were subjected to analysis. Organizations included in this sample are listed in Appendix 1.

Results and discussion

One-dimensional comparisons of inter- and multidisciplinarity measures

As the first step of observing our measures in action, for each HRO profile we obtained the corresponding i-network based on their aggregated publication record over 2000–2009. Upon calculating the values of the two main interdisciplinarity measures introduced above, we first set up a ranking of HROs according to their Field coherence, and Disciplinary coherence, respectively. Since our primary interest lies in the comparison of inter- and multidisciplinarity for the selected set HROs, beyond contrasting the two rankings, we also examine them against a diversity measure (that is, in this context, multidisciplinarity measure) applied in our previous study, the *Rao-Stirling diversity*. This particular diversity index was chosen for comparison because it incorporates the same indicator for the proximity of SCs that was built into both versions of coherence measures (as provided by the basemap of science, namely, d_{ij}).

The pairwise rank-order comparisons are shown in Fig. 1a, d. Striking from the plots is that the two interdisciplinarity measures, Field coherence and Disciplinary Coherence are in a relatively high agreement (Fig. 1a), while a poor relation between either of those and the multidisciplinarity indicator can be observed (Fig. 1b, c). Indeed, the sample-based rank correlation between the two coherence indicators is $\rho = 0.84$, while these exhibit a weak negative pairwise correlation with Rao–Stirling diversity, $\rho \approx -0.1$ in both cases.

This first result immediately suggests that, for the HROs under study, being active in a wide range of diverse (distant) fields is not strongly related to being highly interdisciplinary in the sense of combining those fields in the portfolio.

As to the actual rankings, most telling is the spread of HROs in the scatterplot of Field coherence against *Rao–Stirling diversity*, a measure of inter- and multidisciplinarity, respectively, both relying on the same calculation of SC distances (Fig. 1b). Ranks are assigned to institutes in a reverse order: the HRO with the maximal value has the maximal rank number (27). The dotted line represents the points where the two rankings would be

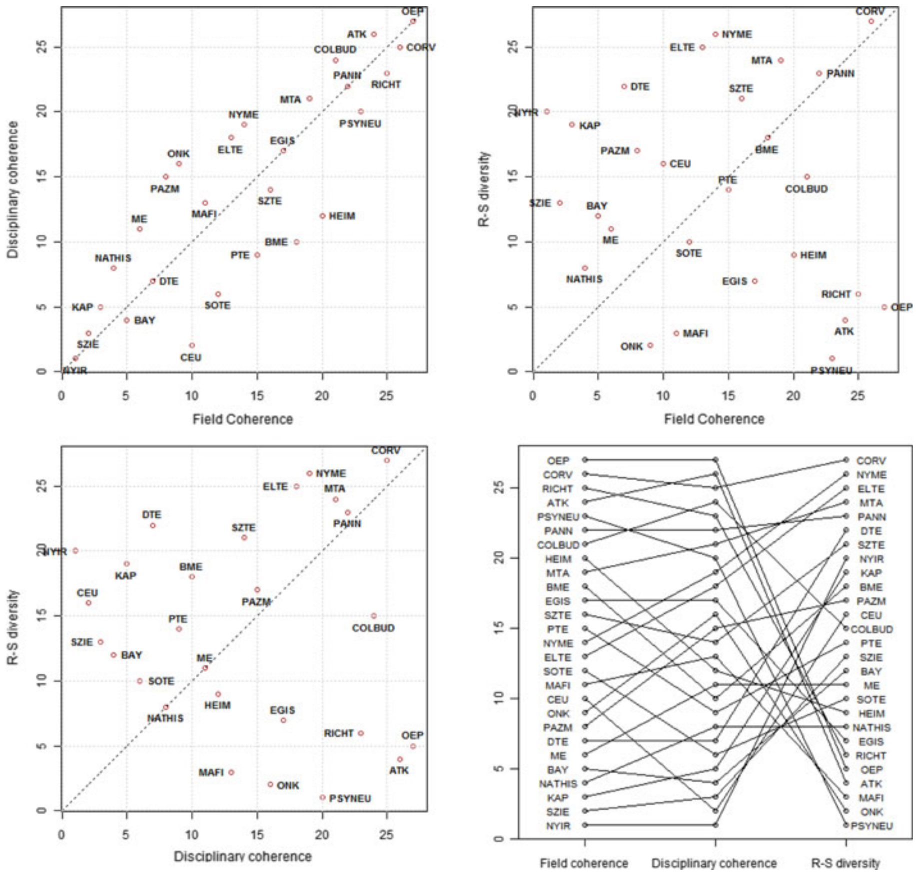


Fig. 1 **a**, **d** Pairwise comparisons of rankings yielded by Disciplinary coherence, Field coherence and Rao–Stirling diversity (**a**–**c**), and changes in rank of HROs along the tree measures (**d**)

identical. As can be seen, there are only a few organizations that occupy the same position when it comes to inter- and multidisciplinarity: two universities, CORV owns the topmost ranking on both scales, followed by PANN. The very same, and relatively high position is assigned to BME with an educational profile historically dedicated to engineering sciences, while in the middle of both rankings resides the University of Pécs (PTE). In the upper triangle, notably, mostly HROs performing in higher education (universities etc.) are present, while the lower triangle is ruled by research institutes, firms, hospitals, such that the reference line seem to discriminate between, roughly speaking, universities and non-universities (except for the special cases residing on the line, highlighted above). More precisely, universities with a generalist (educational) profile turn out to be rather multi-disciplinary and less interdisciplinary, while specialized universities (such as SOTE) and non-universities appear as more interdisciplinary and less multidisciplinary. MTA as a special aggregated unit clustered together with generalist universities along this mapping.

Multidimensional approach to inter- and multidisciplinarity

In order to reveal the relationship between the broader set of measures applied to our sample, including also topological measures applied to i-networks (and thereby to gain a deeper insight into the emerging typology of HROs), we extended our analysis beyond one-dimensional rankings. As the first step, for each HRO, the values of topological measures applied to their *i-network*s are calculated. This enriched set of indices, that is, the two measures of interdisciplinarity (Field coherence, Disciplinary Coherence) plus the topological measures was complemented by the values of diversity measures or *m-network* indicators applied to Hungarian HROs in our previous study (Soós and Kampis 2011). As a result, the multivariate description of our organizations can also encompass the three diversity indices *Rao–Stirling diversity* (*R–S diversity*), together with its path-based versions *R–S diversity: path* and *R–S diversity: weighted path* as profile diversities using different interpretations of "distance" between Subject Categories on the basemap.

Given this set of portfolio characteristics, we conducted a principal component analysis (PCA) on the 11 inter- and multidisciplinarity proxies to uncover their correlational structure. Though this PCA was based on the rankings obtained by the individual measures, instead of using the rank correlation matrix as the direct input, we used a modified set of original, rank-transformed variables. In this step, rankings were transformed to normally distributed values, to meet the requirement of PCA with respect to normality. The primary reason for this method relying on the original matrix of cases (HROs) versus indices was to obtain scores for HROs along the resulting variables (components) as well, something that can be utilized for constructing a typology of institutions with respect to the dimensions of inter- and multidisciplinarity.

With this setting, the PCA resulted in four principal components (with an associated eigenvalue above 1), jointly explaining a considerable amount, about 87 % of the total variance in the original dataset. Based on the factor structure, a measure was assigned to a component with a loading on it higher than 0.5 (it should be noted, that much higher values were gained for those variables trespassing this threshold—in the range from 0.7 to 0.9—indicating a rather clear component structure). The variables pertaining to each component is summarized in Table 3 (factor loadings are presented in Appendix 3). In this table, beyond constituent measures, components are also characterized with their "definitive non-constituents", that is, with variables showing a strong negative loading on them (indicated with a "−" sign).

Table 3 Main components resulted from the PCA and the associated measures

Dimensions (rotated components)	Measures
RC1	Diameter, diversity (R–S), diversity (R–S, path), diversity (R–S, weighted path)
RC2	# comp, shannon, –share,
RC3	Field coherence, disciplinary coherence
RC4	cc, max betweenness, –share

Based on this arrangement of measures, the major structural dimensions of research profiles can, in terms of multi- and interdisciplinarity, be interpreted as follows.

Amount of multidisciplinarity (RC1, enumerative)

The first component (with the greatest explanatory power) contains the three diversity measures we hereby refer to as multidisciplinarity indicators (relying on the m-network), that, most interestingly, turn out to be fairly (positively) correlated with the diameter of the i-network, as a topological feature of the co-occurence of fields. Scoring high along this dimension therefore implies for a HRO to have a diverse research profile, ranging over numerous distant areas, but also that these areas are less directly related within (and via) this profile. Conversely, the smaller the diversity exhibited on the global map of science, the more fields tend to form a "small world" network in the profile, being more closely connected to each other. In other words, this component reflects a tendency whereby less multidisciplinary profiles contain fields in closer combinatons. A (partial) explanation for the emergence of this dimension can be that distant fields are less likely to be combined in the practice of S&T, and vice versa. According to our previous discussion on diameter as an indicator of the range of knowledge integration, however, this coupling of variables suggests that diversity also bound to the type of intermediation where fields are interlinked through many others. This topological aspect of RC1 makes it a really complex dimension worth studying over further samples.

Multimodality (RC2, topological)

The second component combines two variables that convey topological features of the i-network: the number and the size distribution of its connected components (the latter expressed by the Shannon-Wiener entropy measure). Part of its description is that the above combination is negatively correlated with the share of the maximal component. These joint topological properties can be interpreted as the extent of fragmentation within an interdisciplinarity network, that is, this variable summarizes the multimodality of knowledge integration. Scoring high along this dimension indicates that fields in the portfolio tend to cluster into more, similar-sized groups, whereby no single cluster outsizes the rest. On the contrary, low scores convey a topology where a major or "giant" cluster is being formed, connecting the majority of underlying fields (though, of course, not directly, in most cases). An example to both cases from our sample, a (relatively) fragmented i-network as contrasted to an i-network with high overall integration, is shown in Fig. 2a, b.

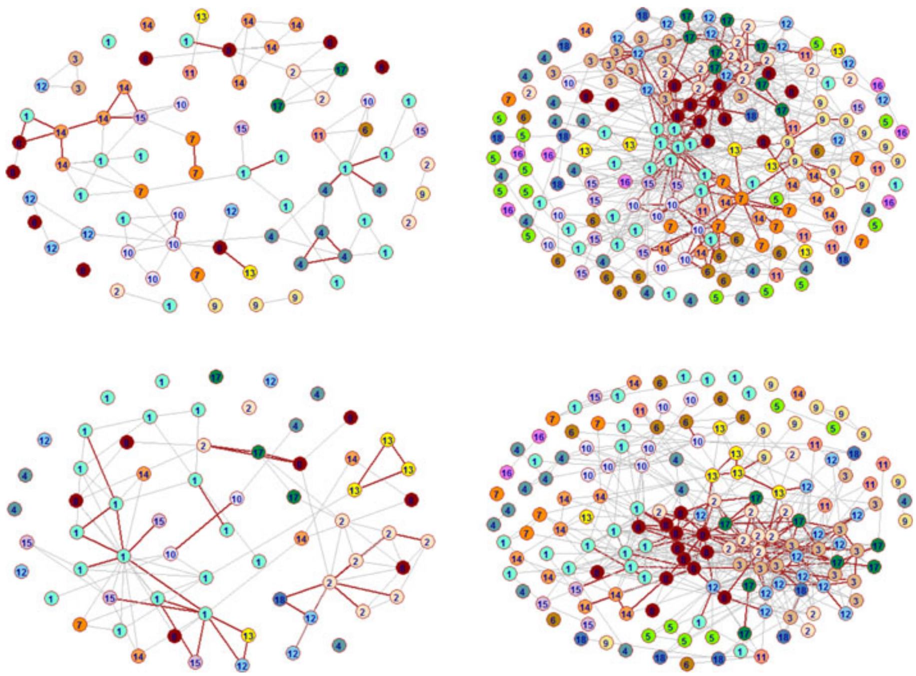


Fig. 2 **a**, **d** The i-network of **a** KAP, **b** ELTE (*top left*, *top right*), **c** BAY and **d** BME (*bottom left*, *bottom right*), respectively. *Red* (*dark*) ties represent weights above the average. Nodes, that is, Subject Categories with the same color code/number belong to the same discipline. Numbers refer to the following disciplines: **1.** Biomedical Sci **2.** Materials Sci **3.** Computer Sci. **4.** Clinical Med **5.** Econ. Polit. and Geography **6.** Psychology **7.** Ecological Sci **8.** Chemistry **9.** Geosciences **10.** Cognitive Sci. **11.** Health and Social Issues **12.** Engineering **13.** Env Sci and Tech **14.** Agri Sci **15.** Infectious Diseases **16.** Social Studies **17.** Physics **18.** Business and Management

Strength of coherence (RC3, enumerative)

The third component correlates with the two interdisciplinarity measures, Field coherence and Disciplinary coherence introduced in this study. This finding suggest that these quantities form a distinct dimension of research profiles: since, in principal, both measures operate by quantifying (and weighting) the amount of knowledge integration in a portfolio, we might call this component the amount or stregth of coherence.

Intermediation (RC4, topological)

The fourth component, like the second one (RC2), collects variables that describe two further topological features of an i-network, namely, the clustering coefficient and the normalized maximum betweenness centrality of the graph. This component also shows a negative correlation with the size (share) of the maximal component of the network. Configured as it is, the RC4 component can be recognized as conveying the degree to which an i-network fits into a specific, but well-known network topology. This topology can be described as containing dense areas (resulting in high clustering coefficients) linked through a small number of subjects (with, thereby, high centrality values). In contrast to RC2, where the overall connectedness of constituent fields are being grasped, RC4 rather captures the pattern of connections: scoring high on this component, that is, having a greater fit to the topology described for a profile, indicates the presence of some intermediary fields, that hold together otherwise distinct subject clusters. On the contrary, lower scores report differing organization of fields in the profile, typical in the cases when a giant cluster unifies the Subject Categories: this effect is mirrored in the finding that the size of the maximal component is conversely related to this component. Given its ability to convey information on the pattern of how subjects are connected, this dimension seems to convey the degree of *intermediation* exemplified in research profiles. Two extremes from our sample are shown in Fig. 2c represents the very topology described above, while the profile in Fig. 2d entirely lacks these structural properties.

It should be noted that the above variables emerging from the PCA can be divided into two categories with respect to the type of information conveyed. RC1 and RC3 reflects the degree of multi- and interdisciplinarity (respectively) in an *enumerative* manner: roughly speaking, both variables calculate the quantity of diversity (RC1) or that of interrelatedness (RC2, with a topological aspect as well). By contrast, RC2 and RC4 provides structural or *topological* information on how the interdisciplinary network is being organized. As we shall see in the next section, this dichotomy will be quite useful for the interpretation of the taxonomy of Hungarian Research Organizations, based on the resulted dimensions for profile description. For the sake of brevity, we will refer to the first category, including RC1 and RC3, as *enumerative dimensions*, while RC2 and RC4 will be termed as *topological dimensions*.

The typology of HROs

With the aid of the multidimensional description of the structure of publication portfolios, our final (but also primary) goal was to set up an empirical typology of Hungarian Research Institutions constituting our sample. Such a typology would reflect the main groups of HROs, characterized by more or less distinctive research patterns, or "strategies".

To this end, for each HRO, their scores on the four combined variables, RC1–RC4 were obtained from the PCA results. Based on their new coordinates in the component space (that is, the score matrix), a (hierachical) clustering of sample institutions were conducted, mirroring their relative position along the dimensions RC1–RC4. This exercise allowed us to (1) describe each portfolio in the quantitative and topological dimensions, and (2) to group them according to these features. Moreover, as the main rationale for this method, resultant groups could also be characterized via the scores of their members (the region occupied in the component space), so that clusters are interpreted ("explained") in terms of the respective contributions of quantitative and topological dimensions.

The result of these joint exercises (i.e. arranging HROs in the component space and their corresponding clustering) are shown in Fig. 3a–b. On the plot, each institution is represented via its name and the id-number of the cluster it belongs to. To make use of the dichotomy of dimensions described in the previous section, the relevant component space (that is, RC1-RC4) is depicted via two scatterplots: Fig. 3a represents the *enumerative dimensions*, plotting RC1 against RC3, while Fig. 3b is for the *topological dimensions* (RC2 against RC4). For the most informative cluster structure, we cut the cluster tree (drawn by group average agglomerative clustering) at the level where the most inlcusive, but outlier-free (or still cohesive) groups were yielded.

For improved readability, result are also reported in a tabulated form (Table 4). In the table, each cluster is listed along with its members. More importantly, clusters are further explained by the role of the underlying dimensions in their formation, using the following simplifications:

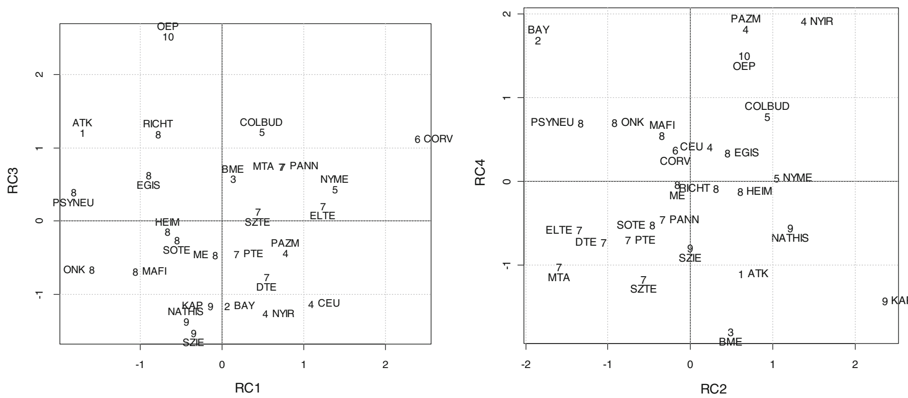


Fig. 3 **a**, **b** Position and clustering of HROs along the four dimensions extracted from the set of inter- and multidisciplinarity measures (RC1 vs. RC3, and RC2 vs. RC4). Numbers indicate cluster membership

Table 4 Composition of HRO clusters and their characterization in terms of the four dimensions of inter- and multidisciplinarity

Cluster	HROs	RC1	RC3	RC2	RC4
1	ATK	−	++	+	−
2	BAY	0	−	− −	++
3	BME	0	+	+	− −
4	CEU; NYÍR; PAZM	+	−/− −	0/+	+/++
5	COLBUD; NYME	+/++	+	+	0/+
6	CORV	++	+	−	+
7	DTE; ELTE; MTA; PANN; PTE; SZTE	+	∓	−/− −	−
8	EGIS; HEIM; MAFI; ME; ONK; PSYNEU; RICHT; SOTE	−/− −	∓	∓	∓
9	KAP; NATHIS; SZIE	0/−	−/− −	0/++	−/− −
10	OEP	−	++	+	++

- "++" for dimension *D* indicates high positive scores for dimension *D* (exceeding the center of its positive range); "−" indicates strong negative scores (exceeding the center of its negative range);
- "+" for dimension *D* indicates moderate positive scores for dimension *D* (around the center of its positive range); "−" indicates moderate negative scores (around the center of its negative range);
- "0" for dimension *D* indicates that all cluster members score near 0 for dimension *D*.

Given this notation, the claim that a cluster exhibits "++" on RC1 implies that all its members score high with respect to the "amount of multidisciplinarity", so that member institutions can be characterized by highly multidisciplinary profiles. By contrast, having "±" as the corresponding evaluation says that cluster members are distributed in a broader range of negative to positive scores, therefore institutions in the group exhibit varying degrees of multidisciplinarity, in which case the latter doesn't qualify as a distinctive feature for the cluster (though still count as a useful description of it).

As the table shows, by the discussed method ten clusters have been formed out of 27 research organizations. Of the ten clusters, the largest one encompasses eight organizations (cluster 8). Most strikingly, all HROs are being collected by this category that share a profile with the biomedical sciences, irrespective of their organizational status, the size or the sector they belong to: universities (SOTE), large firms (RICHT), smaller governmental institutions (PSYNEU) and hospitals (HEIM) are equally included. Their distribution along the main dimensions shows a low amount of multidisciplinarity with varying amounts of interplay between fields (RC1, RC3). There is some heterogeneity in the category as to the structure of this interplay as well, but the integration is relatively high: SOTE is one extreme (highest integration), EGIS is the other (lowest).

The next collection of HROs, following the biomedical cluster in size, can also be characterized by a single unifying feature: it contains (large) governmental institutions from the academics, that maintain a generalist profile: mostly non-specialized universities (covering various disciplines in their educational portfolio) and MTA belongs here. As the component maps show, it is a rather coherent group along two dimensions, one enumerative (RC1) and one topological (RC4). As can be expected, all members are multidisciplinary. They all exhibit an overall integration within the subjects covered, with RC4 indicating a dense i-network, as the other dimension of close resemblance of these HROs. This latter observation is reinforced by the RC2 scores, indicating a giant component formed by the interrelated SCs. On the other hand, the amount of interdisciplinarity, shown via RC3, is non-equally distributed. MTA and, interestingly, PANN, being in a very similar position, are positively scored as to connecting distant areas of research; most members, however, came out with negative values. It follows that the extensive number of diverse subject areas, densely connected in university profiles, still form subgroups made of less distant, or occasionally co-occuring subjects.

In the mid league of clusters we can find three groups, each with three or two members. Cluster 9 is, again, reflects a thematic unity: HROs working upon the closely related fields of biosystematics, ecological and agricultural sciences are being represented here. As coherent as it is in terms of profiles, the institutional types included make this set a rather diverse collection with a college (KAP), a museum (NATHIS) and a university (SZIE). Characteristic of these HROs is a high similarity with respect to dimensions RC1 and RC2: They share a low amount of multidisciplinarity, and a very low amount of interdisciplinarity. On the topological side, the overall integration is also quite moderate, with more distinct subnets of SCs, and low clustering tendencies (KAP is the less integrated, along RC2 and RC4).

The next cluster of the same size (Cluster 4) is less heterogenous by intstitutional type, but less homogenous in profile as the previous one: two universities, one international and one catholic (CEU and PAZM, respectively), and a college (NYÍR) belongs to this one. They all reside in the the same quarter in both diagrams, but otherwise are relatively scattered. Their profile is well-distributed over the global map of science, so that multidisciplinarity is a clear feature of it; however, the amount of interdisciplinarity is in the negative: for example, CEU's is the most multidisciplinary, but among the less interdisciplinary portfolios. Their i-networks show a moderate to very little integration, with several isolated subgroups of interrelated SCs, and many central topics (high betweenness centrality) to wich otherwise non-related fields are connected (star-like topology). This pattern is especially salient in the case of NYÍR and PAZM. Cluster 5, on the contrary—as comprised of COLBUD and NYME, an academic research institution and a relatively specialized university, respectively—while described by a similar topology, also conveys a high amount of interdisciplinarity as well, especially for COLBUD.

The rest of the HRO types are each exemplified by just one organization. Two universities form their own cluster: BME, bearing primarily (and historically) an educational profile in engineering, is characterized by a low level of multidisciplinarity, accompanied by a relatively high level on interdisciplinarity (RC3) with the interacting subjects being densely connected (RC4). CORV, originally operating in economics and social science (and later incocorporating an agricultural segment), is both highly multi- and interdisciplinary on RC1 and RC3; however, the extent of integration is much lower, with a giant component in the i-network that is held together by some central subjects to which the others are being connected. Two further R&D bodies, a governmental one (ATK) and a foundation-based network (BAY) each occupies a special niche in the component space: BAY is heavily scored on RC4 excusively with negative or small values by the other factors, that implies weakly connected sets of interrelated subject categories, any of which represents a relatively small range of fields (moderate multi- and low interdisciplinarity). ATK, inversely, is highly interdisciplinary, in spite of being active only on a few and quite close fields, which capacity is furhter enriched by those fields being well integrated in dense (small) network components. Almost the same can be observed for OEP, the last one-membered category (and also an outlier).

We note that dimensions of multi- and interdisciplinarity are obviously affected by institution size (reflected in the sample size of the bibliometric dataset)—as each discipline is likely to have a critical size. How many of them can be supported by an institution is thus an extensionalist measure (still, when fileds are integrated, raw size can be said characteristic of the amount of interdisciplinarity as well). The same "size effect" can be noticed when comparing generalist universities with specialist colleges or research institutes. In end effect, we can expect large universities to be generalist and hence (multiple) discipline-driven but less interdicplinary as reflected in department structure; smaller colleges to be specialist but more easily interdisciplinary as dictated by the needs of the given specialization. Our results, to a large extent, corresponds to these intuitions, which we consider an asset—without a priori information, the presented methods can "find" the departments of the large universities and identify the "required heterogeneity" of the seemingly narrow-profile shools, such as medical colleges. (Hence portfolio analysis can be conceived as a size predictor in "blind" applications). On the other hand, we note that for our sample, the rank correlation between Field coherence and its size-normalized version (the raw indicator divided with Rao–Stirling diversity) amounted to 0.8, indicating the robustness of the original measure in this particular case.

Conclusion

In the study presented above we addressed the trend in science mapping that utilizes ISI Subject Category maps (SC maps or "basemaps") for an in-depth, structural characterization of research profiles. To elaborate on measures previously introduced to grasp the degree of diversification exhibited by publication portfolios, we introduced a further type of SC map and two corresponding measures we call "interdisciplinarity map" or i-map, partially derived from the basemap, and "interdisciplinarity indicators", respectively. The goal of this extension was to quantify, beyond diversification, the degree of interplay, or integration of fields in institutional profiles. On our account, the former can be viewed as the degree of multidisciplinarity, while the latter as that of interdisciplinarity of research organizations.

It should be noted that the choice of the terms "interdisciplinarity" and "multidisciplinarity" was not intended to convey deep conceptual commitments here. In recent decades, an extensive body of literature has emerged focusing on these concepts, attempting to clarify their meaning, including their very scientometric operationalizations and applications (cf. Porter et. Al 2007, Rafols and Meyer 2010, just to mention the approaches most closely related to the presented methodology, or Wagner et al. 2011, for the most recent comprehensive review). Our work, however, is not supposed to add to the general interpretation of these constructs, but, conversely, to use the related interpretations to characterize important features of research profiles.

Given these models, our primary goal was to investigate the relationship between the inter- and multidisciplinarity (or diversification and integration) of institutional research profiles. As our further goal was to enrich the practice of research assessment with tools more sensitive to structural aspects of publication output, we tested our measures in a regional sample of interest, covering Hungarian Research Institutions (HROs) with their ten-year publication profile each. The analysis involved the following steps: (1) the pairwise comparisons of HRO rankings yielded by multidisciplinarity and interdisciplinarity measures, (2) the principal component analysis of measures, complemented with topological indicators of the i-networks in order to derive main dimensions of the organization of fields in profiles, and (3) setting up a typology of HROs with respect to the obtained dimensions.

Results confirmed, in the first place, that distinguishing between different types of patterns underlying research profiles has several pay-offs. Even simple, rank-order comparisons showed that multi- and interdisciplinarity of profiles are disentangled features, with weak negative correaltion. Simple comparisons discriminated between universities with a generalist educational profile (as rather multidisciplinary) and non-universities plus specialist universities (as rather interdisciplinary). The latter combination, i.e. being both specialist and interdisciplinary at the same time, is explained by being active on a smaller number and/or less distant research areas, but in such a way that these are nevertheless well combined in rich ways within the portfolio, resulting in stratified institutional competencies.

The PCA revealed four major dimensions jointly described by our measures. These included the two quantities introduced in this paper for interdisciplinarity, three quantities of diversification or multidisciplinarity, complemented with a handful of further measures each describing a topological aspect of i-networks (a total of 11 variables). Two of them are interpreted as the "amount of multi- and interdisciplinarity", respectively, while the remaining two integrated topological information on interdisciplinarity networks (called "multimodality" and "intermediation"). Arranging HROs along these dimensions yielded a typology with ten groups with several, though informative, singletons: BME, an institution historically dedicated purely to engineering science but later integrating social science education, and CORV, for which, roughly, the converse applies, form their own group. Most other clusters also showed characteristic features, independent of the variables upon which they were formed: an extensive set is that of generalist universities again (along with MTA, the "superorganization"), while the other major cluster gathers actors in biomedical sciences (irrespective of their organizational status). Also distinctive are the patterns of HROs concerned with the related fields of biosystematics, ecology, and agricultural science. Rare combinations of high inter- and multidisciplinarity scores also form separate clusters (such as COLBUD and NYME), while topological features of the combination of fields also differentiates a cluster (CEU, NYÍR, PAZM).

Since our experiments with the proposed measures constitute a first case study, future work would naturally extend to the inclusion of larger samples into our analysis to examine the general trends in the relationship of inter- and multidisciplinarity. A further dimension of research we intend to explore is the elaboration on measures in order to improve both their accuracy and expressive power.

Acknowledgments We acknowledge the financial support from the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under the FET Open grant number 233847 (Dynanets project, www.dynanets.org), as well as the support of the Hungarian Government (Anyos Jedlik programme) through the TexTrend project (www.textrend.org), contract no. NKFP_07_A2 (2007)/TEXTREND. The supports are gratefully acknowledged.

Appendix 1

See Table 5.

Table 5 Abbreviations for Hungarian Research Institutions (HROs)

Abbrev.	Institution
ATK	Research Institute for Animal Breeding and Nutrition
BAY	Bay Zoltan Foundation for Applied Research
BME	Budapest University of Technology & Economics
CEU	Central European University
COLBUD	Collegium Budapest Institutie for Advanced Study
CORV	Corvinus University Budapest
DTE	University of Debrecen
EGIS	EGIS Pharmaceutical Ltd
ELTE	Eötvös Loránd University
HEIM	Heim Pál Children's Hospital
KAP	University of Kaposvár
MAFI	Geological Institute of Hungary
ME	University of Miskolc
MTA	Hungarian Academy of Sciences
NATHIS	Hungarian Natural History Museum
NYIR	College of Nyiregyháza
NYME	University of West Hungary
OEP	National Health Insurance Fund Administration of Hungary
ONK	National Institute of Oncology
PANN	Pannon University
PAZM	Péter Pázmány Catholic University
PSYNEU	National Institute of Psychiatry & Neurology
PTE	University of Pécs
RICHT	Gedeon Richter Chemical Works Ltd
SOTE	Semmelweis University (of Medicine)
SZIE	Szent István University
SZTE	University of Szeged

Appendix 2

See Table 6.

Table 6 Values of measures obtained for the HROs included (the last three variables have been adopted from Soós–Kampis 2011)

HRO	Field coherence	Disciplinary coherence	#comp	Shannon	Share	cc	Diameter	Max betweenness	Diversity (R-S)	Diversity (R-S, path)	Diversity (R-S, weighted path)
ATK	0.345	0.297	3	0.709	0.591	0.429	5	0.184	0.292	0.449	0.299
BAY	0.230	0.154	1	0.000	0.828	0.651	10	0.338	0.376	0.743	0.415
BME	0.302	0.177	7	0.536	0.848	0.486	7	0.063	0.404	0.768	0.458
CEU	0.271	0.139	6	0.905	0.795	0.565	8	0.256	0.396	0.932	0.598
COLBUD	0.329	0.263	7	1.080	0.667	0.710	7	0.255	0.395	0.839	0.490
CORV	0.389	0.277	4	0.337	0.771	0.495	13	0.263	0.446	1.061	0.653
DTE	0.255	0.162	3	0.130	0.916	0.495	9	0.085	0.415	0.812	0.502
EGIS	0.300	0.194	5	0.792	0.721	0.633	5	0.248	0.340	0.620	0.355
ELTE	0.276	0.196	2	0.061	0.969	0.497	9	0.075	0.431	0.904	0.553
HEIM	0.315	0.180	6	0.917	0.673	0.523	8	0.259	0.358	0.598	0.372
KAP	0.226	0.159	9	1.765	0.286	0.407	7	0.257	0.406	0.749	0.474
MAFI	0.272	0.182	3	0.589	0.750	0.622	6	0.245	0.289	0.640	0.347
ME	0.251	0.179	4	0.419	0.722	0.532	8	0.182	0.367	0.750	0.445
MTA	0.302	0.210	1	0.000	1.010	0.471	7	0.073	0.428	0.869	0.529
NATHIS	0.226	0.166	8	1.227	0.610	0.447	9	0.314	0.343	0.706	0.413
NYIR	0.195	0.129	6	1.530	0.268	0.763	4	0.322	0.413	0.887	0.561
NYME	0.291	0.204	8	0.962	0.660	0.617	8	0.144	0.433	0.902	0.563
OEP	0.745	0.571	3	0.995	0.294	0.767	2	0.213	0.311	0.635	0.414
ONK	0.269	0.188	2	0.230	0.836	0.582	6	0.265	0.253	0.454	0.256
PANN	0.331	0.227	4	0.260	0.855	0.520	8	0.109	0.422	0.827	0.514

Table 6 continued

HRO	Field coherence	Disciplinary coherence	#comp	Shannon	Share	cc	Diameter	Max betweenness	Diversity (R-S)	Diversity (R-S, path)	Diversity (R-S, weighted path)
PAZM	0.259	0.188	5	1.140	0.449	0.696	8	0.395	0.403	0.884	0.554
PSYNEU	0.345	0.207	1	0.000	0.826	0.581	5	0.258	0.242	0.471	0.243
PTE	0.295	0.175	4	0.228	0.896	0.502	11	0.092	0.389	0.743	0.445
RICHT	0.361	0.240	4	0.647	0.706	0.542	6	0.188	0.339	0.623	0.356
SOTE	0.276	0.160	4	0.295	0.908	0.568	6	0.076	0.361	0.642	0.383
SZIE	0.205	0.149	5	0.455	0.791	0.493	8	0.135	0.380	0.697	0.444
SZTE	0.296	0.187	4	0.186	0.931	0.432	8	0.111	0.415	0.817	0.500

Appendix 3

See Table 7.

Table 7 Detailed results of the PCA on selected measures

	RC1	RC2	RC3	RC4
Field coherence	−0.071	−0.142	0.943	−0.041
Disciplinary coherence	−0.099	−0.066	0.943	0.012
# comp	0.233	0.877	−0.201	−0.194
Shannon	0.004	0.972	−0.064	0.161
Share	0.186	−0.832	0.014	−0.443
cc	−0.040	−0.029	0.152	0.885
Diameter	0.626	−0.275	−0.280	−0.231
Max betweenness	−0.102	0.298	−0.255	0.732
Diversity (R–S)	0.939	0.038	0.015	−0.224
Diversity (R–S, path)	0.965	0.028	−0.071	0.082
Diversity (R–S, weighted path)	0.968	0.152	−0.034	0.030
Prop variance explained	0.296	0.238	0.181	0.154
Cumulative variance explained	0.296	0.535	0.716	0.870

References

Boyack, K. W. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics, 79*(1), 27–44.

Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Soceity for Information Science and Technology, 58*(9), 1303–1319.

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology, 60*(2), 348–362.

Leydesdorff, L., & Rafols, I. (2011). Indicators of interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics, 5*(1), 87–100.

Porter, A., Cohen, A., Roessner, D., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics, 72*(1), 117–147.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2011). How journal rankings can suppress interdisciplinarity. The case of innovation studies in business and management. *Arxiv preprint*. arXiv:1105.1227.

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics, 82*(2), 263–287.

Rafols, I., Porter, A., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology, 61*(9), 1871–1887.

Soós, S., & Kampis, G. (2011). Towards a typology of research performance diversity: The case of top Hungarian players. *Scientometrics, 87*(2), 357–371.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society, Interface, 40*(15), 707.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., et al. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A Review of the Literature. *Journal of Informetrics, 5*(1), 14–26.