Check for updates

RESEARCH ARTICLE

REVISED Mapping leadership and communities in EU-funded research through network analysis

[version 2; peer review: 1 approved, 2 approved with reservations]

Fabio Morea[1], Alberto Soraci[1,2], Domenico De Stefano[3]

[1]Area Science Park, Trieste, Friuli-Venezia Giulia, Italy
[2]Department of Economics, University of Messina, Messina, Italy
[3]Universita degli Studi di Trieste, Trieste, Friuli-Venezia Giulia, Italy

v2 **First published:** 18 Dec 2024, **4**:268
https://doi.org/10.12688/openreseurope.18544.1
Latest published: 07 Aug 2025, **4**:268
https://doi.org/10.12688/openreseurope.18544.2

Abstract

Background

Horizon 2020 and Horizon Europe are flagship programs of the European Union aimed at supporting research and innovation, fostering collaboration among companies, academic institutions, and research organizations. Comprehensive data on projects, objectives, participants, funding details, and results of Horizon projects is available through the open access portal CORDIS (Community Research and Development Information Service). This paper introduces a novel methodology for utilizing CORDIS data to reveal collaborations, leadership roles, and their evolution over time.

Methods

The methodology is based on network analysis. Data is downloaded from the CORDIS portal, enriched, segmented by year and transformed into weighted networks representing collaborations between organizations. Centrality measures are used to assess the influence of individual organizations, while community detection algorithms are used to identify stable collaborations. Temporal analysis tracks the evolution of these roles and communities over time. To ensure robust and reliable results, the methodology addresses challenges such as input-ordering bias and result variability, while the exploration of the solution space enhances the accuracy of identified collaboration patterns.

Results

Open Peer Review

Approval Status ✓ ? ?

	1	2	3
version 2 (revision) 07 Aug 2025			? view
version 1 18 Dec 2024	✓ view	? view	

1. **Laila Zemīte**, Riga Technical University, Riga, Latvia
2. **Andreas V Olympios**, University of Cyprus, Nicosia, Cyprus
3. **Chanel Tri Handoko**, Sebelas Maret University, Surakarta, Indonesia

Any reports and responses or comments on the article can be found at the end of the article.

To illustrate the approach, the methodology is applied to a specific case: analyse the evolution of collaborations in hydrogen valleys, in the broader frame of “hydrogen energy” research and innovation projects funded by Horizon programmes.

Conclusions

The proposed methodology effectively identifies influential organizations and tracks the stability of research collaborations. The insights gained are valuable for policy-makers and organizations seeking to foster innovation through sustained partnerships. This approach can be extended to other sectors, offering a framework for understanding the impact of EU research funding on collaboration and leadership dynamics.

Keywords
Innovation, Horizon 2020, Horizon Europe, Network Analysis, Community Detection, Hydrogen, Hydrogen Valley

This article is included in the Horizon Europe gateway.

This article is included in the Artificial Intelligence and the Social Sciences and Humanities collection.

Corresponding author: Fabio Morea (fabio.morea@areasciencepark.it)

Author roles: Morea F: Conceptualization, Data Curation, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Soraci A**: Writing – Original Draft Preparation, Writing – Review & Editing; **De Stefano D**: Methodology, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No [101111927](North Adriatic Hydrogen Valley [NAHV]). Authors F.M. and A.S. acknowledge the financial support of the North Adriatic Hydrogen Valley (NAHV) project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views and opinions expressed in this paper are solely those of the author(s) and do not necessarily reflect the official positions of the European Union or the Clean Hydrogen Joint Undertaking. Neither the European Union nor the granting authority assumes responsibility for the content presented.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Morea F *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Morea F, Soraci A and De Stefano D. **Mapping leadership and communities in EU-funded research through network analysis [version 2; peer review: 1 approved, 2 approved with reservations]** Open Research Europe 2025, **4**:268 https://doi.org/10.12688/openreseurope.18544.2

First published: 18 Dec 2024, **4**:268 https://doi.org/10.12688/openreseurope.18544.1

REVISED **Amendments from Version 1**

The updated version (V2) incorporates substantial revisions informed by the reviewers' comments and our own efforts to enhance the quality of the manuscript.

The introduction has been expanded to include additional references to prior studies that have utilised CORDIS data, highlighting the limited application of network analysis in this context. Moreover, the section addressing the policy framework on hydrogen valleys has been extended to incorporate recent developments supported by the Clean Hydrogen Partnership. This update has also led to the inclusion of additional bibliographical references.

Although the methodology remains unchanged from the previous version, the corresponding section has been reformulated to improve readability. A high-level Algorithm 1, and it is accompanied by detailed pseudocode boxes (Algorithm 2 to Algorithm 5) describing in detail the most relevant parts.

The captions of several figures have been revised to offer more precise explanations of the visual elements, to highlight key findings, and to assist the reader in interpreting the results more effectively. This applies in particular to Figure 4, Figure 5, and Figure 7 through Figure 12. Additionally, a new figure has been included in the introduction, resulting in changes to figure numbering in V2 compared to V1.

Any further responses from the reviewers can be found at the end of the article

1 Introduction

Understanding the long-term dynamics of innovation is relevant for effective policy-making and strategic research planning. While innovation is fundamentally driven by collaboration between companies, academic institutions, and research organizations, its long-term dynamics remain elusive due to lack of structured data and analytical challenges.

In this paper, we present a novel methodology that employs network analysis techniques to assess the evolution of collaborations driven by Horizon 2020 and Horizon Europe programmes. To illustrate the proposed approach, we examine a case study in the hydrogen energy sector—specifically focusing on hydrogen valleys.

Section 1 introduces the context of the methodology, the case study (hydrogen valleys), and it outlines the key research questions. Addressing these questions requires a complex approach, employing network analysis and tackling challenges such as input-ordering bias and result variability, as detailed in Section 2 (Methodology). Section 3 presents a case study focused on the development of hydrogen-related research and the current support to hydrogen valleys, using data from Horizon 2020 and Horizon Europe projects from 2015 to 2029. The results allow to identify organisation with leading roles, a community structure, and its evolution over time.

1.1 Analysis of CORDIS data with network analysis techniques

Researchers have applied diverse data science techniques to CORDIS data since 2010 to gain insights into collaboration and innovation patterns. While tabular analysis remains a common method, more advanced approaches, such as text mining, have increasingly been adopted: for example, text-mining approaches have been used to analyze project descriptions from CORDIS data in 1 and 2.

The application of network analysis tools to the innovation networks derived from collaboration to research project is not new and it has been used to study: topological properties of collaboration networks[3]; the relation between network structure and the EU broad policy objectives[4]; innovative dynamics at country-level[5,6]; technological diversity[7,8] proposed the use of a rank-size approach for clustering complex networks derived from the collaboration in European project at organisational level. In the work by 9, CORDIS data were used to construct a network and identify evidence of "closed clubs" within Horizon 2020 projects—specifically, a correlation between network centrality and an organisation's success in securing funding[10,11] used network visulisation techniques to highlight the research linkages among participants to EU-funded projects.

A limited number of recent studies have adopted community detection techniques on networks derived from CORDIS data. For example in 12 researchers built a multipartite network of projects and, using a community detection algorithm, identified field-specific collaboration clusters that varied in size and composition across scientific domains. Another study[13] introduces a novel link prediction model for analyzing and forecasting H2020 collaboration networks, while also identifying communities and assessing their regional distribution.

In this stream of research, contributions often focus on specific network analysis tasks; however, none of them combine the various social network analysis tools to fully exploit the potential information embedded in such rich network data.

In the present contribution, we instead provide a comprehensive and integrated framework for the use of social network analysis tools to thoroughly explore EU project-derived networks—such as centrality and coreness analysis of partners, and the detection of "robust" clusters of firms and research organisations. Furthermore, we propose an innovative approach to overcome the current limitations of the most widely adopted community detection algorithms.

The methodological advancements proposed in our study are threefold. First, the study applies community detection algorithms on weighted network, capturing both the existence and the intensity of collaborative ties among participants. Second, it adopts the European Science Vocabulary (EuroSciVoc)[14], a taxonomy that provides a structured and semantically precise classification of research projects. Third, it incorporates a temporal dimension that enables the tracking of evolving collaboration patterns over time.

1.2 Hydrogen valleys

Hydrogen Valleys, as a concrete example of how collaborative efforts fuel innovation, illustrate the broader principle that understanding the long-term dynamics of innovation is

essential for informed policy-making and strategic research planning.

A "Hydrogen Valley" is a geographical area where several hydrogen applications are combined into an integrated ecosystem that produces, exchanges, and consumes a significant amount of hydrogen, covering the entire hydrogen value chain: production, storage, distribution, and final use[15].

Across Europe, at least 90 Hydrogen Valleys are operational, planned, or proposed[16]. In Europe this growth is largely driven by initiatives like the Clean Hydrogen Partnership (or Clean Hydrogen Joint Undertaking - JU), an Institutionalised Partnership settled under the art 187 of the TFUE[17] which sees the joint commitment of EU Commission and other public and private organizations to support research and innovation (R&I) activities in hydrogen technologies in Europe.

The Clean Hydrogen Joint Undertaking leverages Horizon Europe funding to support research, innovation, and scale-up projects aligned with the EU's climate and energy objectives. Moreover, various national and regional funding programs further support the development of HVs. Furthermore, the Clean Hydrogen Joint Undertaking has recognized the relevance of the HV has a tool to filling the gaps in the wider hydrogen sector at EU level. This effort is illustrated in Figure 1.

Altogether, the Clean Hydrogen JU has supported to date 20 hydrogen valleys projects across 18 European countries, out of which 19 projects are still ongoing. Together they represent investment for more than 1.2 bill EUR with a JU total funding of 250 MEUR. Furthermore, this number will increase in the coming months due to the new call for proposal issued by the Clean Hydrogen JU in 2025[18]. The relevance and the characteristics of the Hydrogen Valleys has been also pointed out by the Commission Staff Working Document "Towards a roadmap for accelerating the deployment of Hydrogen Valleys across Europe"[19] by recognized their nature of "hydrogen ecosystems" with a strong potential to speed up the development of a renewable hydrogen sector in the EU"

More Hydrogen Valleys are under development also beyond the scope of Clean Hydrogen Joint Undertaking, in Europe and abroad[20]. However, due to their relatively recent constitution[21] few studied have been developed to map the

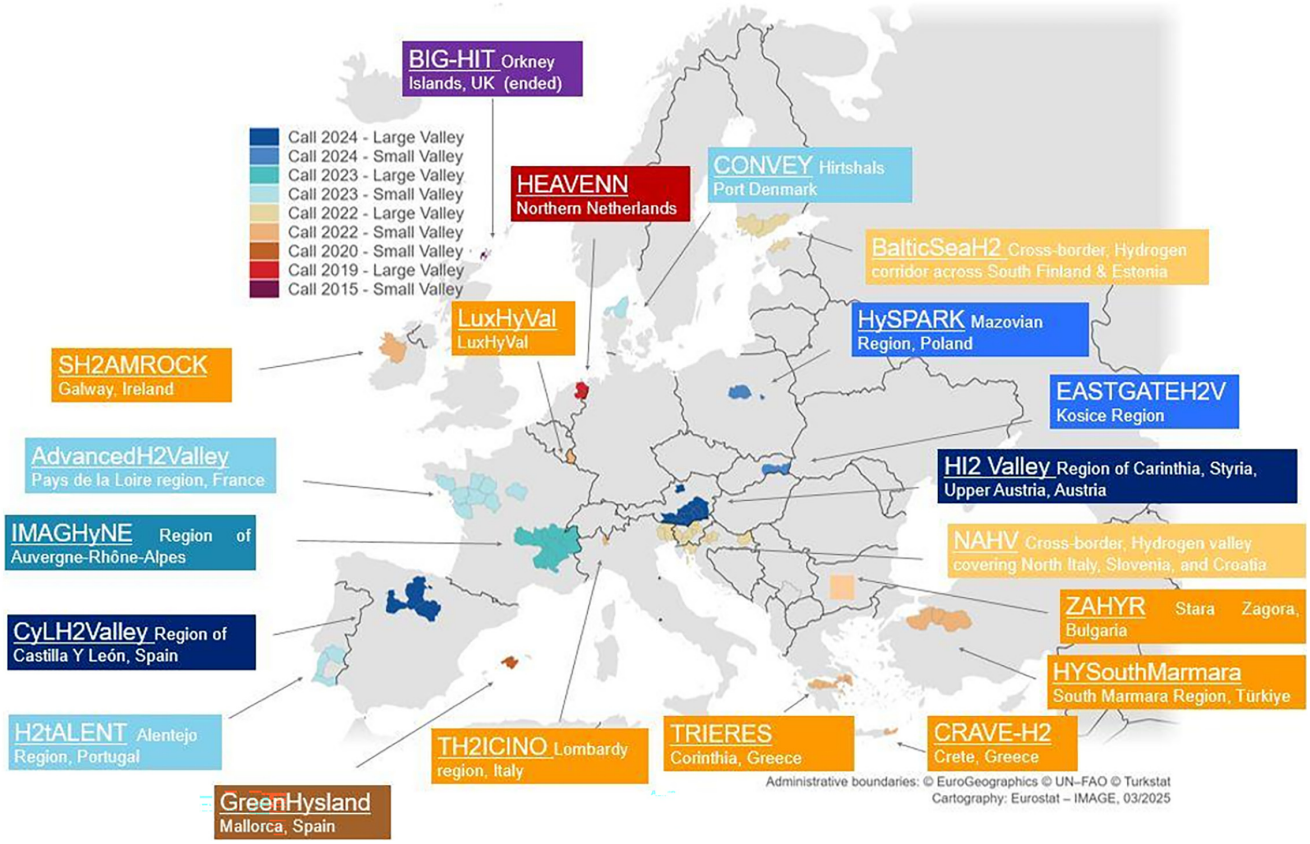

Figure 1. Hydrogen Valleys supported by the Clean Hydrogen Partnership (source: Clean Hydrogen Partnership).

hydrogen valley impact on a territory, and its development over the years. This small number of studies has been also assessed in 2023 by a Polish team[22] by applying the bibliographic method and a quantitative analysis of the collected publications. The researchers identified 284 publications dealing with hydrogen valleys, with a systematic increase in the number of cited papers. Furthermore, from a deeper analysis, they found that the most frequently cited publications are studies presenting different types of technological solutions and concepts. This study indicated the existence of a significant research gap in the field of research on the creation and development of Hydrogen Valleys. Among the several, some other studies have been focused on the techno-economic assessment of green hydrogen valley[23] providing multiple end-users and their LCA impact[24]. For example, the South Africa Hydrogen Valley report[25] has examined the socioeconomic impact from the Hydrogen Valley project across multiple dimensions in terms of GDP, job market and tax revenue.

The **North Adriatic Hydrogen Valley** or NAHV project[26] is one of the first transnational hydrogen valleys developed in Europe. It embraces the EU territories of Croatia, Slovenia and the Italian region Friuli Venezia Giulia and involves 37 partners based mainly on those countries. The project is co-financed by Horizon Europe programme and supported by the Clean Hydrogen Partnership[27].

NAHV has been planned under the willingness to activate a long-lasting hydrogen economy. The heuristic structure of the NAHV is based on a systemic vision and the development of a holistic model of the valley. As showed in Figure 2, this heuristic approach considers not only innovation related to vertical "hard pillars" (which are typical of a hydrogen valley, such as production, transport, storage and distribution of hydrogen) but also to the "soft pillars", which are regulatory framework, training and capacity building and involvement of civil society. Equally important is the level of cooperation among the various actors within the valley. NAHV aims also to develop and test a pathway of industrial transformation

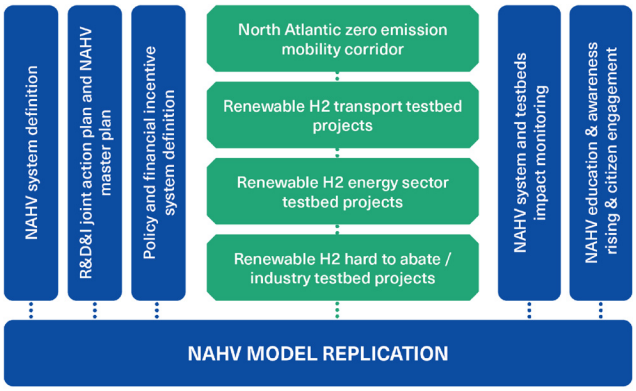

Figure 2. A schematic representation of the North Adriatic Hydrogen Valley project. Testbeds are primarily driven by industrial partners, while collaboration with policymakers, universities, and research organisations is crucial for enabling innovation.

linked to green transitions[28] that will produce system innovation leading to a radical change of routines, and transform economic activities and industry ecosystems.

1.3 Collaboration-driven innovation in Hydrogen Valleys

Hydrogen valleys are innovative ecosystems focused on hydrogen technology, leveraging regional and national initiatives to develop sustainable energy solutions. These ecosystems align well with the Quintuple Helix innovation model, which involves collaboration between five key actors: academia, industry, government, civil society, and the environment. By looking closer the hydrogen valley modus operandi, it is possible to observe that it is based on two main paradigms: the open innovation model[29], where the exchange of know-how among the several actors of the valley is stimulated and (b) the quintuple helix[30] model that is a model in which innovation is pushed by the integration in the classical triple helix model of the needs expressed by the civil society and media, along with a targeted effort to solve a specific challenges[31]. In the context of Hydrogen valleys, the primary challenge is typically addressing climate change by driving the decarbonization of the economy. According to this vision a trans-regional innovation ecosystems[32,33] is needed to stimulate the interaction, creativity and innovation and to develop a strong and inclusive macro-regional ecosystems that support business growth, as well as the entrepreneurial environment necessary to foster more innovative regional economies[34] and to increase competitiveness of small- and medium-sized enterprises (SMEs). To reach this wide and ambitious scope, it is necessary to put in action a combined strategy[35] based on fostering collaborations within the Hydrogen Valleys, creating new businesses and attracting investment to address gaps in the hydrogen value chain[36].

The Staff working document of the EU Commission cited above has recognized the need to move forward for the development of HV by defining 5 building block: 1) Research and innovation; 2) Regulatory framework; 3) Crowding-in investments and funding; 4) Education, training and skills; 5) Developing Hydrogen Valleys globally, with the scope of increasing the number of HV in Europe and globally. To reach this ambitious goal a wide number of initiatives, funds, projects and actors in Europe and beyond have been mobilized. The EU Commission with its several directorates has established several platforms such as CORDIS and Kohesio[37] to collect and systematize the results of the several funded projects. Furthermore, other platforms have been established by the managing authorities of other EU funded initiatives. In this wide panorama, the CORDIS platform stands out as one of the most organized sources, offering regularly updated data.

1.4 Research questions

Understanding the long-term dynamics of innovation is crucial for effective policy-making and strategic research planning. From the perspective of a research organization, understanding the dynamics of innovation at the regional or international level can be a good opportunity. Consider, for example, an organization involved in a new hydrogen valley project: it must not only

focus on the development of its own project activities, but it may benefit from identifying partners and competitors that are engaged in other hydrogen valleys. This broader awareness may help identifying synergies, risks, and market opportunities.

Policymakers adopt yet another perspective. They have a strategic vision to align local, national, and European resources, aiming for medium- and long-term impact of the innovation ecosystems. Achieving this requires a clear understanding of how collaborations emerge between industry and research, who the key stakeholders are, and how these collaborations evolve over time.

This paper addresses the elusiveness of innovation proposing a methodology that is suitable for two approaches: project-specific (e.g., examining the role of project partners to gain deeper insights into their role within the broader landscape of EU-funded research in the same sector) and a policy-oriented (e.g., analysing how Horizon grants are shaping competitive ecosystems in a given sector). In both cases, two research questions are essential:

1. Which organisations are the most influential in driving research and innovation? Identifying these key players — whether companies, academic institutions, or research organisations — is important for other organisation that want to be in contact with them in future projects. Moreover, policy makers may be interested in assessing whether the policy put in place in their territory has produced an improvement of leadership roles over the years.

2. Are EU-funded projects fostering partnerships that extend beyond the duration and scope of individual projects? If such long-term communities exist, they serve as vital indicators of the open innovation and quintuple helix models effectiveness. Stable communities suggest robust exchanges of ideas and collaboration, which are essential for sustaining innovation and achieving the long-term innovation goals.

The methodology presented in Section 2 aims to provide quantitative answers to the research questions above. Key requirements of this approach are that results must be independent of contingent factors (such as software implementation or ordering of the input data) and tested for validity. Additionally, when multiple algorithmic options are available, the selection must be data-driven and performance-oriented, ensuring that the chosen algorithm yields the most interpretable and reliable outcomes.

2 Methods

The proposed methodology builds on well-established network analysis techniques, such as centrality measures and community detection, while incorporating novel elements to enhance the stability and quality of results, and to produce a temporal analysis. Notably, a new method for exploring the *solution space* is introduced, along with a comprehensive quality check process to select the most appropriate algorithm, validate the results, and manage input-ordering bias. When multiple solutions emerge, a *consensus community detection* approach is employed to consolidate them into a single, coherent outcome.

The methodology foresees 6 main steps: data acquisition further discussed in Section 2.1, data preparation and enrichment (Section 2.2), network analysis (Section 2.3), centrality measures (Section 2.4 community detection (Section 2.5), and temporal analysis of communities (Section 2.6). An overview of the process is provided in Algorithm 1

Algorithm 1. Methodology for analysing research collaborations using CORDIS data

1: **Data acquisition:**
  1.1 Access CORDIS and download Horizon 2020 dataset
  1.2 Access CORDIS and download Horizon Europe dataset
  1.3 Merge corresponding tables for Organisations $\mathbf{O}$*, Projects $\mathbf{P}$* and Topics $\mathbf{T}$*
2: Enrich data with AI generated keywords and categories
3: Filter and subset data
  3.1 filter data by relevant topics $\mathbf{T} \subset \mathbf{T}$*
  3.2 subset $\mathbf{P} \subset \mathbf{P}$*($\mathbf{T}$) *projects related to selected topics*
  3.3 subset $\mathbf{O} \subset \mathbf{O}$*($\mathbf{P}$) *organisations involved in selected projects*
4: **Calculate weights matrix** $\mathbf{W}$ // refer to Algorithm 2
5: for each year do // Network analysis
  4.1: Create Network G_y
  4.2: Calculate Centrality measures
  4.3: Identify stable partition of G_y // refer to Algorithm 3
5: **Temporal Analysis** // refer to Algoritthm 5
6: **Output**: community labels and centrality measures associated to each organisation

2.1 Data acquisition

The source of data for this study is CORDIS (Community Research and Development Information Service), the European Commission's primary public repository and portal to disseminate information on all EU-funded research projects and their outcomes[38]. It provides open access to comprehensive data on projects, including objectives, participants, funding details, and results. For the purpose of this study, two datasets have been extracted: Horizon 2020 (projects starting from 2014 to 2020)[39] and Horizon Europe (projects starting from 2021 to 2027)[40]. The two datasets have the same structure, composed of several tables, the most relevant being the organisations table denoted as $\mathbf{O}$* and the projects table denoted as $\mathbf{P}$*.

Each project is characterized by a unique identifier (`projID`), its acronym, title, start date, end date. Moreover, each project is associated with a long textual field, describing its objectives, and with structured categorical information on the call and funding scheme. A separated table associates each

project with one or more topics, encoded according to the European Science Vocabulary (*EuroSciVoc*), the multilingual taxonomy representing scientific fields developed by the EU's Publications Office.

There is a many-to-many relationship between organisations and projects (i.e. a project has several organisations, and an organisation can be part of several projects). In table $\mathbf{O}^*$ each organisation is identified by a unique identifier (`orgID`), a name, detailed geolocation information. In addition, the table records the role of each organisation in each project (coordinator, participant, or associated partner), as well as the total costs incurred by the organisation within each project (`totalCost`) and the corresponding eligible contribution from the Horizon programme (`netEcContribution`).

Data representation can be optimized by merging the corresponding tables for Horizon 2020 and Horizon Europe, to obtain the following core data structures:

- table $\mathbf{P}^*$ listing all projects, their `projID`, start and end dates, reference to the Horizon calls for proposal, keywords, and a long textual descriptions of project objectives;
- table $\mathbf{O}^*$ listing all organisations, their `orgID` location, their role in the project, and monetary values `totalCost` and `netEcContribution`;
- table $\mathbf{T}^*$ associating each project in P to one or more keywords of the *EuroSciVoc* taxonomy.

2.2 Data preparation

The initial step in data preparation involves identifying the subset of projects $\mathbf{P} \subset \mathbf{P}^*$ on which we aim to focus, based on the topics from that are relevant for the objectives of the analysis. Consequently, the other tables can be filtered to ensure that only pertinent data is retained in the form of $\mathbf{O} \in \mathbf{O}^*$ (organisation name and location). The core information for our analysis the weights table denoted as $\mathbf{W}$, which encodes the annual effort each organisation $o \in \mathbf{O}$ puts into each project $p \in \mathbf{P}$ in a given year. A proxy for the values in W can be derived from either `netEcContribution` or `totalCost`. The former indicates the total amount of public funding received by an organisation upon project completion and serves as a useful proxy for how much the Horizon grant promotes collaboration. The latter reflects the total cost incurred by the organisation to complete the project. For non-profit organisations involved in research projects, these values often coincide, as the Horizon grant may cover 100% of the costs. However, for private companies investing in pilot projects, such as in the NAHV project, the two values can differ significantly, sometimes by an order of magnitude. In the case study presented in Section 3, the weight is based on `netEcContribution`, expressed in thousands of euros.

The data is then segmented on a yearly basis. Denoting by $\mathbf{W}_y$ the matrix for a given year y, it is a rectangular matrix of size $|\texttt{projID}| \times |\texttt{orgID}|$. The value of each project is assumed to be equally distributed over its duration, from its start date to its end date. Consequently, the contribution of a project to the matrix $\mathbf{W}_y$ is divided proportionally across the years during which the project is active.

Textual fields in table $\mathbf{P}$ contain valuable information, but in a format unsuitable for network analysis. The project objectives are typically described in hundreds of words, and user-defined keywords lack a standardized vocabulary. To make this information usable, the last step in data preparation involves creating Boolean attributes for each project. These attributes encode relevant information in a simplified format, such as whether the project is focused on technology or market uptake, or whether hydrogen is a primary focus or one of several applications. This was accomplished using a script that interacts with the API of a Large Language Model, which processes the lengthy textual fields and generates a dataframe, indexed with `projID` and one or more Boolean variables, which can be merged with $\mathbf{P}$.

Algorithm 2. Calculate weights matrix W

Input:
- P = set of projects, each with start_date, end_date, total_cost, participants

Output:
- W[o, p, y] = weighted involvement of org o in project p during year y

$y_{min} \leftarrow \min(\textit{project start year } \forall\ p \in P)$

$y_{max} \leftarrow \max(\textit{project end year } \forall\ p \in P)$

for $y = y_{min}$ to y_{max} **do**
  for *each project* $p \in P$ **do**
    $f_{p,y} \leftarrow (\textit{duration of } p \textit{ within year } y)/365$
    for *each organisation o involved in project p* **do**
      $\mathbf{W}(o, p, y) \leftarrow f_{p,y} *$ *total cost of p*
    end for
  end for
end for

2.3 Network analysis

The first step in the analysis is to build a network, by identifying its nodes, edges, and weights in the form of an *adjacency matrix*.

A network (or graph) is a set $G = \{V, E\}$ composed of $n_v = |V|$ vertices and $n_e = |E|$ edges. For the purpose of our analysis, a *one-mode network* would be a suitable model, in which nodes represent organisations and edges encode the information about projects. However, in our dataset $\mathbf{W}$ is a rectangular matrix of size $|\texttt{projID}| \times |\texttt{orgID}|$, which generates a *two-mode network*, i.e., a network with two distinct sets of nodes where connections are established only between nodes from different sets (e.g., organisations connected through joint participation in projects). A one-mode network G can

be obtained from a two-mode network and described by a square matrix, denoted as $\mathbf{A} = \mathbf{W}\ ^{T}\mathbf{W}$, where $\mathbf{W}^{T}$ is the transposed of $\mathbf{W}$. In our case, $\mathbf{A}$ is symmetric, and G is consequently an undirected graph. Figure 3 illustrates the transformation process from the two-mode network of organisations by projects to the one-mode network consisting only of organisation-by-organisation ties.

Since the data is calculated on a yearly basis from 2015 to 2029, the complete model consists of set of networks $\mathcal{G} = \{G_{2015} \ldots G_{2029}\}$ that represents collaborations between organisations within Horizon research project.This spans from the launch of the first projects in 2015 to the anticipated completion of the longest ongoing Horizon Europe project in 2029. The generic network is denoted as G_y where the subscript y refers to a specific year.

Algorithm 3. Network Analysis

Input:
- W weight matrix for year y

Output:
- C_y = community partition of the network in year y
- centrality for each organisation o

for $y = y_{min}$ to y_{max} **do**

$\mathbf{A}_y \leftarrow$ WyT $\mathbf{W}_y$ *Calculate adjacency matrix for one-mode network*

$G_y \leftarrow$ network nodes and edges from $\mathbf{A}_y$

Calculate centrality measures of G_y and associate values to each organisation

Community detection $C_y \leftarrow$ *a valid and stable partition of* G_y *// refer to Algorithm 4*

end for

2.4 Centrality measures analysis

The network structure can be evaluated using centrality measures—degree, strength, and coreness—to determine the importance and influence of nodes, each offering distinct insights.

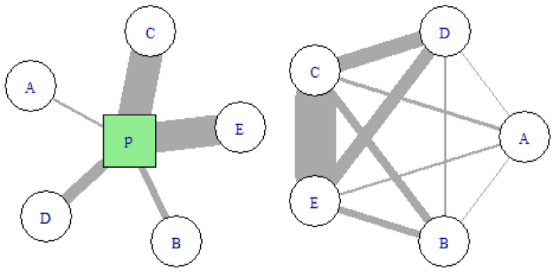

Figure 3. Schematic view of a project as two-mode network (left) and as one-mode network (right). Uppercase letters A, B, C, D, E represent organisations. The green square denoted with P represents a project. Edge width is proportional to the weight, i.e. the value each organisation brings to the project. The sum of weights in the two-mode network is equal to the sum of weights in the one-mode network.

The **degree** of a node v, denoted as $deg(v)$ is the simplest centrality measure, calculated as the number of edges incident to that node. In this context, the degree reflects the number of projects in which an organization is involved during a given year. A higher degree indicates that the node plays a more locally central role within the network, as it has a greater number of direct interactions with other nodes; in the context of our research, a node with a high degree represents an organisation that is, or has been, a partner in many large projects, reflecting its extensive collaborative involvement.

Strength of a node is the sum of the weights of the edges incident to that node: $s(v) = \sum_{u \in N(v)} w_{vu}$ where $N(v)$ is the set of neighbours of v, and w_{vu} represents the weight of the edge between nodes v and u. In weighted networks, strength provides a more nuanced measure of a node's connectivity: with reference to Figure 3, all nodes have the same degree, but for example $s(C) \gg s(A)$. In this context, the strength reflects the monetary value of the projects in which an organization is involved during a given year.

The k-coreness (or **coreness**)[41] of a node is a measure of the node's position within the network's hierarchical structure, based on its connectivity. Specifically, a node has a k-coreness of k if it belongs to the k-core of the network. The k-core is a maximal subgraph in which every vertex has at least degree k, i.e. within this subgraph, each node is connected to at least k other nodes. In this context, coreness can be interpreted as the capacity of an organisation to partner with other organisations that, in turn, possess the same level of collaborative capacity.

Centrality measures are computed individually for each $G_y \in \mathcal{G}$ and saved as attributes of the nodes in G_y. This allows to track changes and compare the network structure across different years, providing insights into the dynamics of the collaborations and the shifting roles of organisations over time.

2.5 Community detection

Community detection is a crucial step in the analysis, as it helps to understand the role of an organisation within the network. A **community** C is defined as a subnetwork of G that satisfies a condition: nodes that belong to C are more densely connected within each other than with the rest of G. A **partition** $\mathcal{P}$ is a set of k disjoint subnetworks $C_1, \ldots, C_k$ whose union is equal to G.

A community detection algorithm $\mathcal{A}(G, \rho) \rightarrow \mathcal{P}$ is a function that takes as input a graph G and one or more parameters ρ, and returns a partition $\mathcal{P}$. Several community detection algorithms are discussed in literature[42,43]. Ideally, any algorithm should produce a single, valid partition each time it is applied with the same parameters. In practice, however, for large, dense networks, several issues may arise. Some community detection algorithms may produce **invalid results** such as communities that are internally disconnected or fail to meet the basic criterion of being more densely connected within than with the rest of the network. This issue should be

addressed with a thorough quality check on $\mathcal{P}$. A less known issue is the **Input-Ordering Bias**: Although networks are mathematically non-ordered entities, their implementation in any software algorithm is inevitably ordered (in the form of a list of edges, or a matrix as $\mathcal{A}$ mentioned above). The order in which nodes and edges are stored in the practical implementation of the network can affect the results, as illustrated in 44.

A more elusive issue is the *variability* of results. When $\mathcal{A}$ is based on a heuristic or randomized approach to identify the optimal partition, it may yield different results each time they are executed: $\mathcal{P}_i \neq \mathcal{P}_j$. All these issues can be addressed through Algorithm 4, based on the repeated execution of $\mathcal{A}(G^*, \rho)$, where G^* is shuffled version of G_y to reduce the effect of input ordering bias) and to explore the *solution space*. Specifically, the **solution space** $\mathbb{S} = \{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_{ns}\}$ is the set of all unique partitions that $\mathcal{A}$ produces across t trials. Our research aims to analyse $\mathbb{S}$, and to determine the minimum number of trials t_c required to confidently assert the completeness of $\mathbb{S}$.

After selecting the most suitable algorithm, the analysis may yield either a single, ideal solution or a dominant solution—one that is more frequently observed than others. In these cases, this solution can be identified and analysed. However, if no single solution emerges as dominant, a "consensus community detection approach" can be applied to consolidate the variable results into a single, reliable solution. This process is illustrated in 4 and discussed thoroughly in 44.

Algorithm 4. Consensus Community Detection

```
// explore solution space
for t = 1 to t_max do
    G* ← random permutation of G_y
    P_t ← A(G*, ρ)
    if P_t is identical to another partition in 𝕊 then
        Update frequency count in 𝕊
    else
        Append P_t to 𝕊
    end if
end for

// Identify a single solution according to solution space type
if solution space type is single then
    return the unique partition
else if solution space type is dominant or multiple then
    the most frequent partition
else
    handle sparse or lack of valid community detection
```

Knowledge about the solution space, along with the awareness that a single or dominant solution exists, emerges gradually as a result of an experiment of repeated trials $\mathcal{A}(G, \rho) \rightarrow \mathcal{P}_i$, as illustrated in Figure 4. In the example, the probability of solutions in $\mathbb{S}$ are represented as point estimates (solid line for $\overline{p}$) and intervals (ribbon between (p_{lower} and p_{upper}). A prevalent solution (shown in red) emerges soon; as t increases additional solutions are discovered and beyond $t = 50$, the probability distributions are virtually unaffected by any new solutions. When the experiment is repeated, a slightly different situation may be observed, but convergence towards a dominant solution consistently occurs.

Probabilities associated with each solution is $\mathbb{S}$ calculated under a Bayesian framework, using a Beta Binomial model: this ensures a thorough yet efficient analysis, and optimizes computational resources.

2.6 Temporal evolution of communities

As introduced in Section 1, it is important to provide insights into the underlying dynamics of collaboration over time, which can be achieved by tracking and categorize the evolution of communities over time. The information is represented as a family of networks, denoted as G_y where each network corresponds to a specific year. Within each yearly network, communities are identified, and our primary objective is to understand how these communities evolve over time. The method involves comparing each community $C_i \in G_y$ with each $C_j \in G_{y+1}$. Through this comparison, we determine whether C_i and C_j are disjoint or have a non-null intersection. If they intersect, the relationship between the two communities is further classified as either continuing (i.e. C_i shares most of its members with C_j), or as part of a merge or split. To ensure accurate tracking across different years, we assign global community labels that remain consistent for identical or continuing communities.

Algorithm 5. temporal evolution of communities

```
for y = y_1 to (y_2 − 1) do
    for c_i ∈ C_y1 do
        for c_j ∈ C_y1+1 do        assess correlation between c_i and c_j as continue, split, merge, mix, disjoint
        end for
    end for
end for
Generate global community labels
```

3 Results

This section applies the methodology outlined in Section 2 to a subset of hydrogen-related projects, including the NAHV project, which is the focus of this study; the data encompasses projects from Horizon 2020 and Horizon Europe, covering the period from 2015 to 2029. The analysis aims to investigate leadership roles, community structures, and their evolution over time. Furthermore, the analysis demonstrates how AI-generated keywords can be integrated to provide additional insights into the distinction between market-oriented and technology-oriented projects.

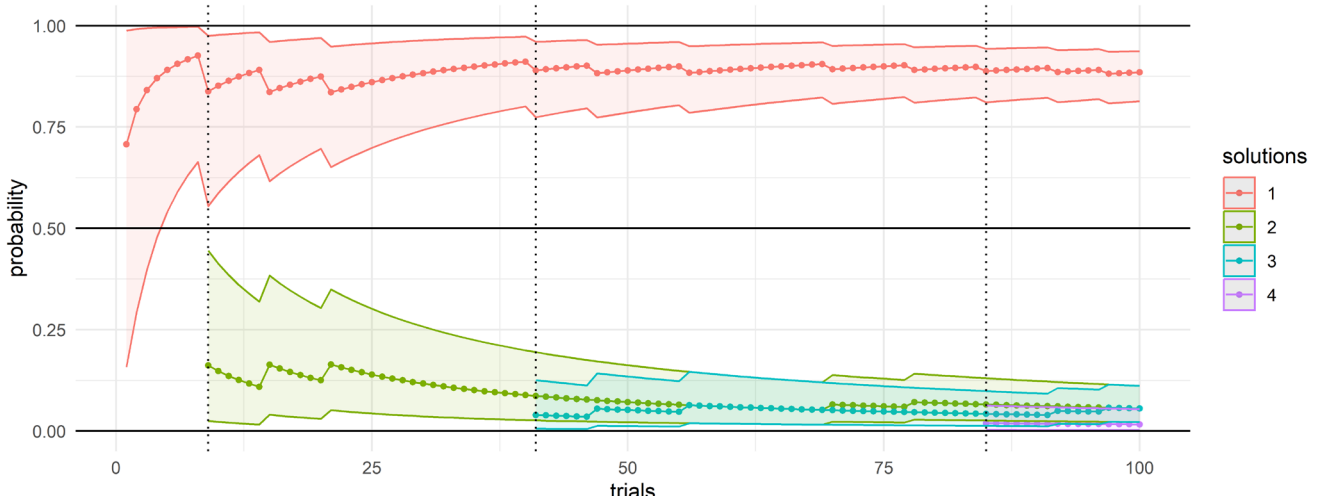

Figure 4. Confidence intervals associated with different solutions in the solution space. The diagram illustrates how repeated runs of the community detection algorithm progressively reveal the a range of solutions. A dominant solution (red) emerges early and remains stable, while new solutions become rare beyond 50 trials—indicating convergence.

The data has been prepared using the *EuroSciVoc* topic *hydrogen energy*, and the AI-generated categories are market vs technology, as explained in Section 2.2. The family of networks $\{G_{2015} \ldots G_{2024}\}$ has a remarkable evolution over time. G_{2015} is composed of 83 organisations, and the number steadily increases to 648 by 2024. The number of collaborations, represented by the network's edges, grew proportionally during this period, and the total value of projects, measured by the sum of weights in the network, saw an even more significant increase, rising from 547 to 6628. Looking ahead to the period from 2025 to 2029, the future of collaborations will be shaped by contracts currently in operation and influenced by upcoming calls and new projects. As shown in Figure 5, after 2024 the number of partners and collaborations decreases, while the total investment remains stable for the next two years.

Between 2016 and 2024 there is an increase in the number of organisations from 83 to 648 and a proportional rise in collaborations as represented by weights, rising from 547 to 6628. Current data for projects continuing in years 2025 to 2029 suggest further growth in the total value of projects.

3.1 Leadership roles

The methodology for AI-generated categories, discussed in 2.2, is applied to the hydrogen projects dataset. This classification approach distinguishes between projects focused on market development — encompassing policy, market uptake, business models, and hydrogen valleys — and those centred on technological development. The NAHV project clearly falls into the former category. This classification enables the creation of two distinct families of networks, denoted as G_y^M and G_y^T. The total value of the projects in each group is shown in Figure 6, revealing an interesting trend: while investment in market-oriented projects has been consistently present since 2016, there is a notable increase in such investments starting in 2023. This increase reflects the launch of Horizon Europe initiatives such as the Clean Hydrogen Partnership, which prioritise large-scale deployment and market integration, notably through Hydrogen Valley projects.

Figure 7 illustrates how centrality measures and AI-generated categories provide insights into the roles of organisations in hydrogen-related innovations, highlighting differences between those engaged in market-oriented and technology-oriented projects.

The figure shows in grey the range of values found each year, and highlights organisations participating in the NAHV project. The diagrams provide an overview of the evolution of coreness in G_y^M (left) and G_y^T (right). The grey shaded area represents the range of coreness values observed each year. The organisations involved in the NAHV project are highlighted with coloured dots, connected by lines that illustrate their progression over time. A notable observation is that, while all NAHV partners are part of G_y^M (which is expected, given NAHV's classification as a market project), only a few are involved in G_y^T.

Coreness values can fluctuate over time: an increase suggests that an organisation is gaining influence through participation in relevant projects, while a decrease may indicate that, following the completion of a project, the organisation has not yet initiated a new one within the Horizon framework. The analysis also includes the years 2024 to 2029, reflecting ongoing projects that are scheduled to conclude during this period. It is important to note that this is not a forecast, but rather a record of existing projects with future completion dates.

Moreover, centrality of all organisations can be visualized in a degree-coreness plane, where each organisation is represented

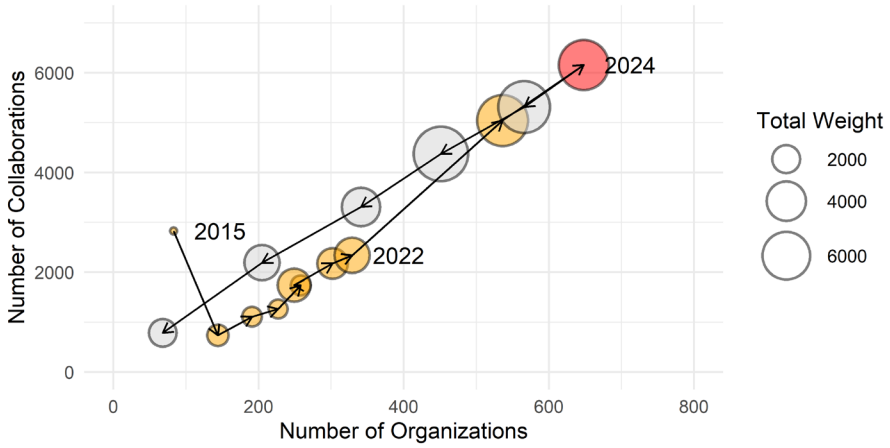

Figure 5. Evolution of the networks G_y *from 2015 to 2029,* represented by the number of organisations (horizontal axis), the number of collaborations (vertical axis) and total investment per year (bubble size). Between 2015 and 2024 the network steadily expanded. The apparent decline after 2025 reflects the fact that only projects already funded and ongoing are included; future Horizon Europe calls and new projects are not yet represented in the data, which explains the decrease in visible partners and collaborations.

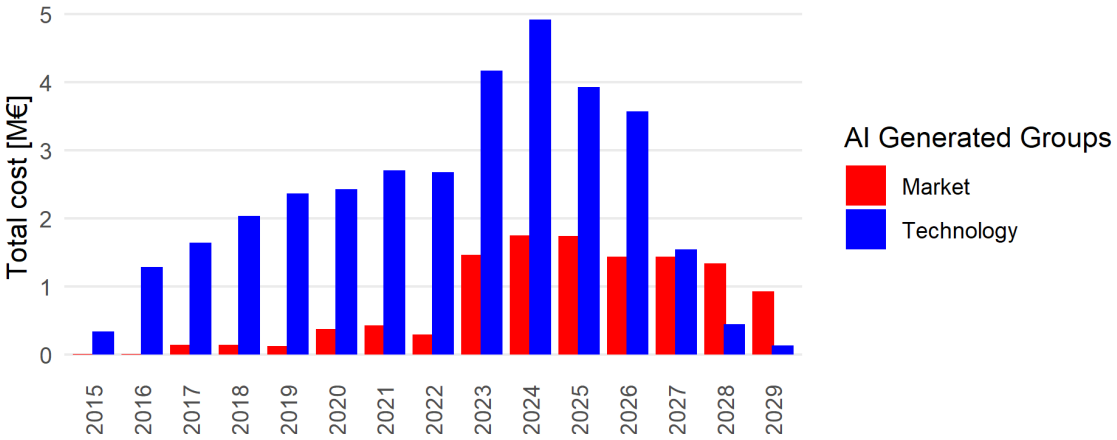

Figure 6. Comparison between technology-oriented and market-oriented project. The value is measured by `netEcContribution`; groups are identified by the AI generated labels: G_y^M and G_y^T.

by a bubble. Degree serves as a proxy for an organisation's capacity to form partnerships with many other organisations, while coreness reflects its ability to partner with influential organisations. Although coreness is limited by degree, the size of the bubble represents the organisation's strength, which serves as a proxy for its capacity to attract substantial funding and invest in large projects. An example is shown in Figure 11, which also includes an information about communities.

3.2 Communities

In the context of this paper, communities refer to relevant subgroups of organisations that exhibit strong collaborative ties. The focus is on analysing a single partition and its evolution over time. The analysis is conducted on the family of complete networks, G_y.

As explained in Section 2.5, the first step involves exploring the solution space $\mathbb{S}_y^A$ in terms of completeness, the number of solutions, and their validity across various algorithms. The tested algorithms include Louvain (LV)[45], Leiden (LD)[46], Label Propagation (LP)[47], Edge Betweenness (EB)[48], Leading Eigenvector (EV)[49], Walktrap (WT)[50], and Infomap (IM)[51]. The tests were conducted in R[52], using the iGraph[53] and communities[54] libraries.

The number of solutions $|\mathbb{S}_y^A|$ are illustrated in Figure 8: most algorithms produced multiple solutions across most years, with the exceptions of 2015 and 2029, where the networks were simpler. Interestingly, WT consistently provided a unique solution. However, EB had a different issue, generating partitions that were not similar to each other, with similarity coefficients

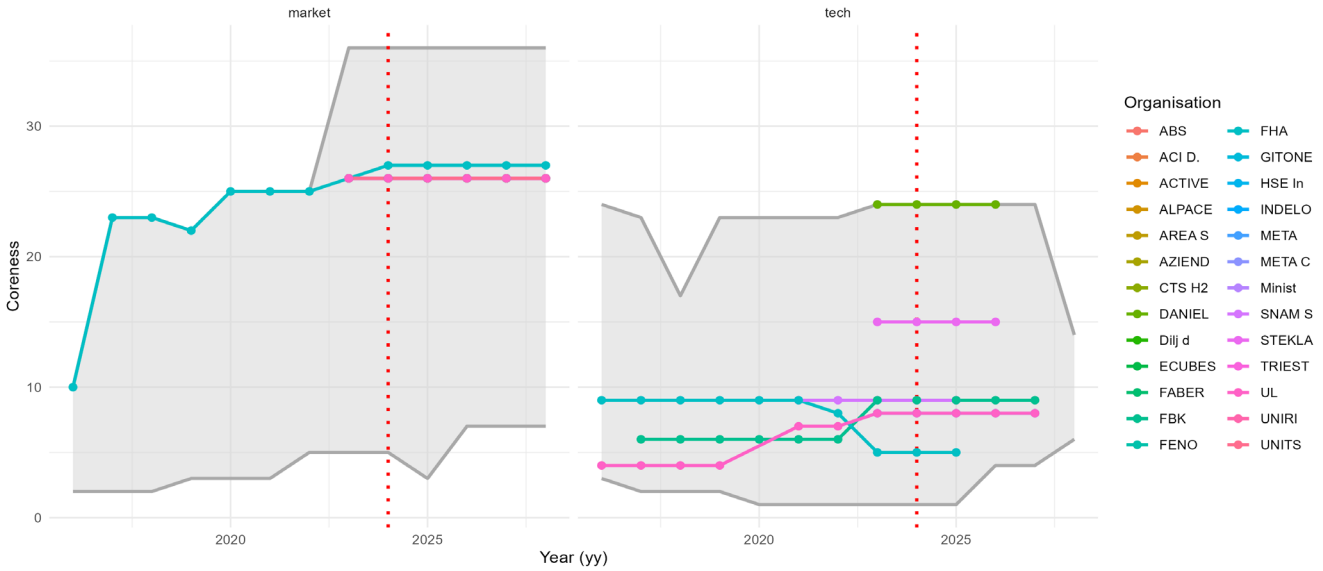

Figure 7. Longitudinal coreness of NAHV partner organisations (2015–2029). Coloured lines track the coreness of each NAHV partner over time, compared to the full yearly range (grey background). Trends reflect each organisation's evolving centrality within the network. Left panel: market-oriented projects. Right panel: technology-oriented projects.

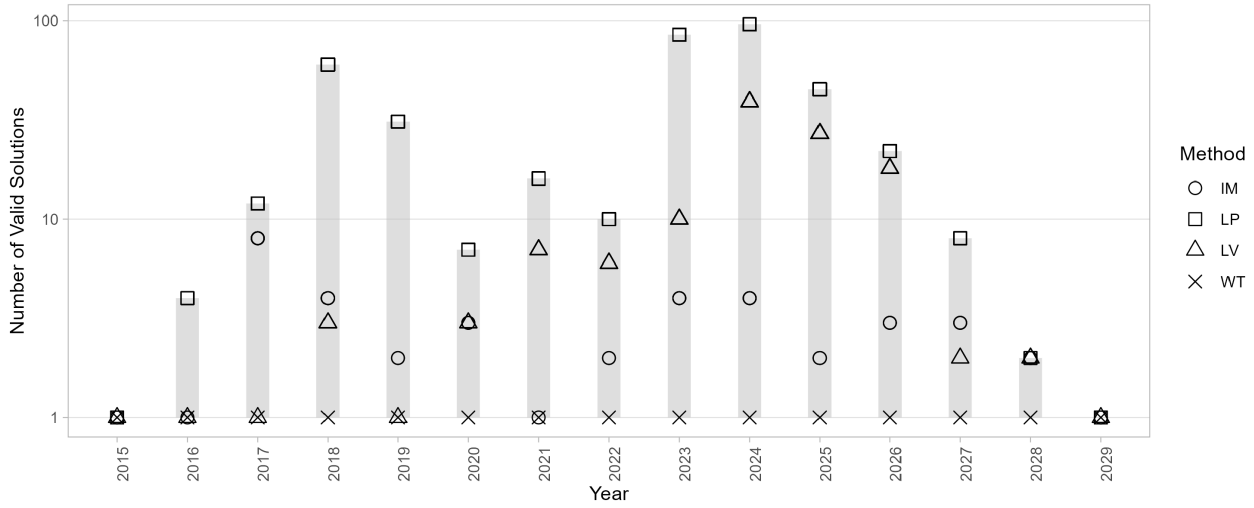

Figure 8. Number of solutions $|\mathbb{S}_y|$ generated by different community detection algorithms. Each marker represents the number of distinct partitions identified in a given year by one algorithm. Ranges and peaks reflect variations in network complexity and algorithm behaviour across the study period.

often below 0.5. EV produced multiple valid solutions in most years but failed to find a partition in 2024.

A more in-depth observation of the solution space for 2024 (the most critical year) reveals additional details, as shown in Figure 9. LV generates a large set of solutions with $|\mathbb{S}_{2024}^{LV}| = 69$, which are also highly diverse, as evidenced by the similarity plot showing a bimodal distribution with peaks around 0.6 and 0.9. Moreover, approximately one-third of the solutions are invalid due to containing one or more internally disconnected communities. This issue is well documented for the LV algorithm[46], where disconnected or fragmented communities can frequently occur and compromise the validity of the partition. LD, specifically designed to address the problem of disconnected communities, also produced invalid solutions for this network: its partition consists of a few large communities and a high proportion of singletons, with a mixing parameter exceeding 0.5, indicating that nodes within a community were more connected to nodes outside their community than to those within it.

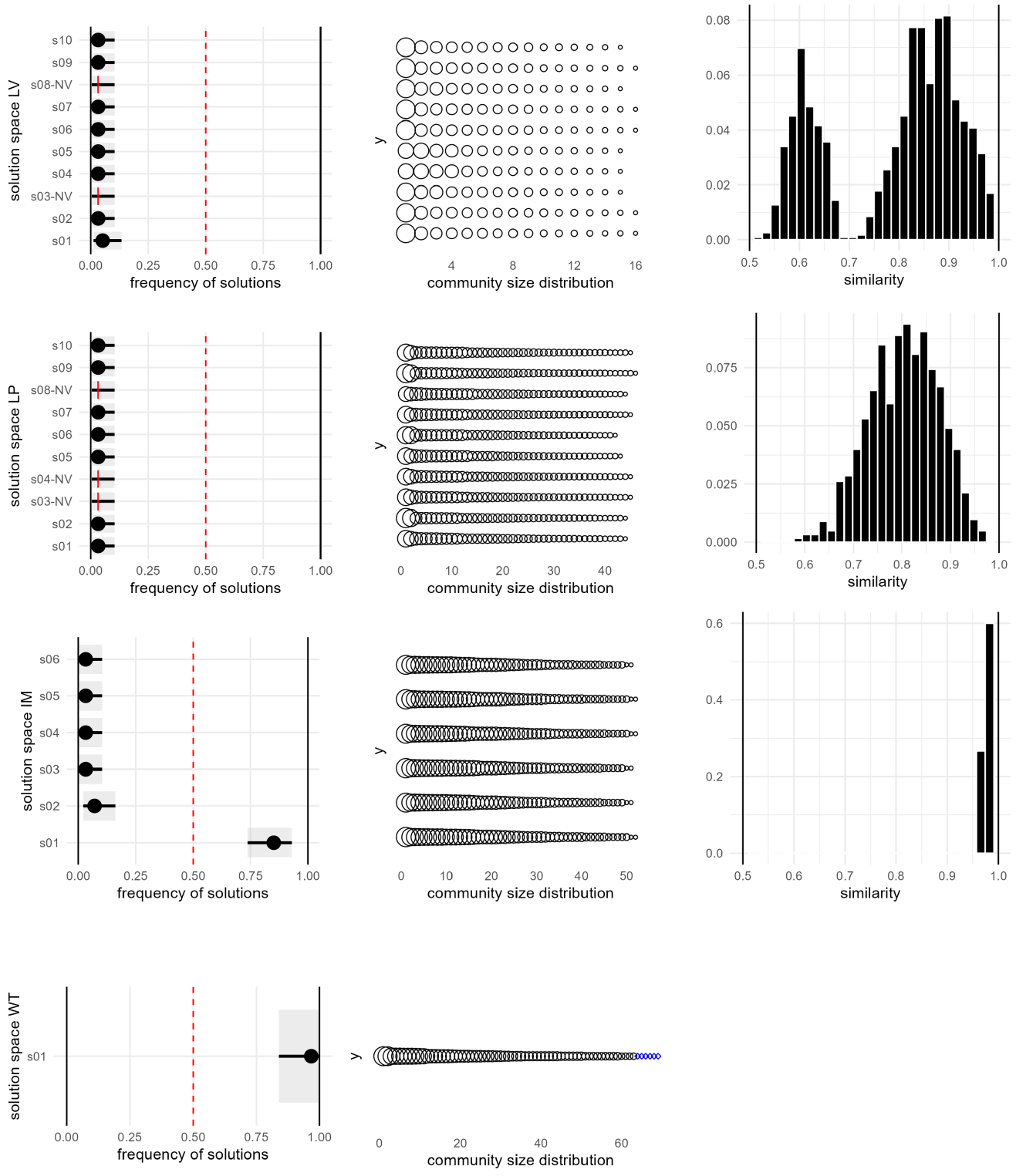

Figure 9. Solution space diagrams for selected community detection algorithms (2024 network). Left: frequency and confidence intervals for the solutions identified by each algorithm. In the case of LV and LP only the 10 most frequent solutions are shown. Middle: distribution of community size for each solution. Proper communities are represented as black circles; single-node communities are represented as blue diamonds. Right: pairwise similarity between solutions.

LP generates a different solution with each run; in this case, with t = 100, there are 96 valid and distinct solutions, while 4 are invalid due to internally disconnected communities. Better results can be achieved with IM that produces $|\mathbb{S}^{IM}| = 1$ in most years, and a dominant one solution in the most complex cases such as 2024. Overall, WT is the only algorithm that consistently provided valid results across all years, with valid partitions and $|\mathbb{S}^{WT}| = 1$, consequently, it will be used for community detection and temporal evolution analysis.

The selected partition $\mathcal{P}_y$ can be represented as a network in which nodes belonging to the same community are grouped and coloured to distinguish between communities. For example, the result for P_{2018} is shown in Figure 10. Some communities appear disconnected from the others; these consist of organisations that, in that year, participated in only one project and thus form a separate component.

Each bubble represents an organisation in the degree-coreness plane. Degree is a proxy for the organisation's ability to win projects and attract funding, while coreness reflects its capacity to partner with other influential organisations. The largest bubbles in the top-right of the diagram highlight organisations that excel in both areas.

Community detection can be integrated with centrality measures to provide a more comprehensive understanding of network structure.

Figure 11 illustrates the results for $\mathcal{P}_{2018}$. In this visualisation, each bubble represents an organisation, with its position determined by coreness and degree centrality. Bubble colours correspond to the community, and bubble size is proportional to the strength (the total weight of its connections in the given year). The diagram clearly demonstrates that degree

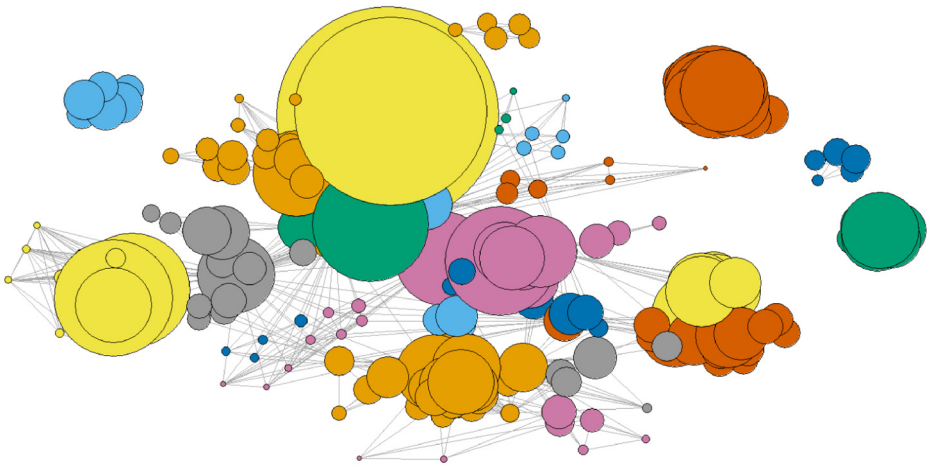

Figure 10. Network G_{2018}, with communities detected using the WT algorithm. Nodes represent organisations, and edges represent project-based collaborations. Communities are grouped and colour-coded. The figure illustrates both the internal structure of major communities and the fragmentation of less-connected organisations.

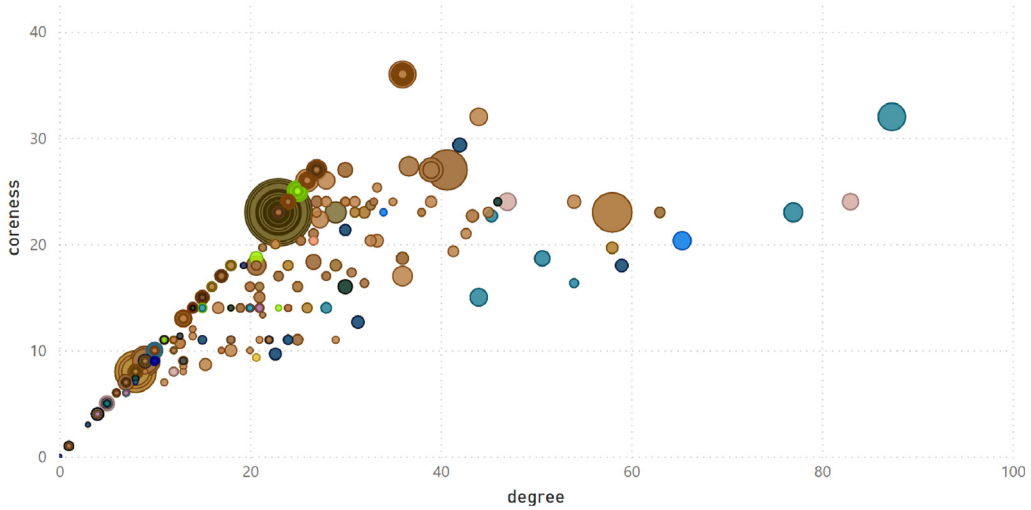

Figure 11. Centrality measures of all organisations in G_{2024}. Each bubble represents an organisation, with position defined by degree (horizontal axis, number of collaborations) and coreness (vertical axis, influence within the network). Bubble size reflects strength (total funding), and colour indicates community membership. The plot highlights how organisations differ in their ability to be part of large projects, work with influential partners, and secure large funding.

is the upper limit for coreness, as no points appear above the line with a slope of 1. In practical terms, this indicates that an organisation that participates in many projects (high degree) is not necessarily the most influential in the network. Strength, which reflects the total weight of an organisation's connections, remains independent of both degree and coreness, meaning that large bubbles can appear anywhere on the graph. This means that some organisations, despite having relatively few connections (low degree), are linked to relevant partners (high coreness) and stand out for their ability to secure significant EU research funding (high strength).

3.3 Evolution of communities

Temporal evolution was calculated as described in Section 2.6, using the WT algorithm to ensure the robustness of the results. In the diagram shown in Figure 12, the horizontal axis represents time, with each community depicted as a coloured rectangle. The height of each rectangle is proportional to the sum of the strengths of its members. For the purpose of clarity, only the six largest communities are represented. The analysis reveals a significant pattern: initially a large number of small communities is formed, and their configuration changes rapidly. This reflects the fact that in Horizon 2020 public funding was primarily allocated to technology-oriented research projects, and partnerships changed rapidly. However, with the onset of Horizon Europe, and specifically the Clean Hydrogen partnership, three large communities emerge, and are set to continue collaboration through 2029.

The significant size of the largest communities reflects the increasing allocation of resources to Horizon Europe projects in the hydrogen sector. This growth is driven by the development of hydrogen valleys, which have fostered long-term, stable partnerships, contributing to both the stability of collaborations and the continued expansion of these communities.

4 Discussion

This paper presented a comprehensive methodology for analysing the long-term impact of EU-funded research projects, with a case study focused on the hydrogen sector. By employing network analysis, centrality measures, and community detection, we demonstrated how to identify pivotal organizations, monitor their influence over time, and delineate the evolution of collaborations fostered by Horizon 2020 and Horizon Europe projects.

Specifically, the analysis shows that Horizon grants in hydrogen sector projects have prompted the emergence of communities that have significantly evolved over the years. Initially characterized by small, rapidly changing groups, these groups gradually shifted towards the formation of larger, more stable partnerships, particularly in the domain of hydrogen valleys.

The methodology presented in this paper, particularly the use of centrality measures, community detection algorithms and temporal network analysis, offers opportunities to improve the visualizations in the Horizon Dashboard[55]. For instance, the "Organization Profile" visualization, which allows users to explore an organization's performance in EU research and innovation programs, could benefit from incorporating centrality measures and tracking community memberships of each organization over time. This could be done by analyzing the entire Horizon network or by focusing on specific subsets categorized according to the top-level EuroSciVoc fields: *Natural Sciences*, *Engineering and Technology*, *Medical and Health Sciences*, *Agricultural Sciences*, *Social Sciences*, and *Humanities*.

One limitation of this study is that it is focused exclusively on the CORDIS dataset, which restricts its scope to EU-funded initiatives, potentially neglecting initiatives funded by national or local authorities, or private funds. Incorporating data from

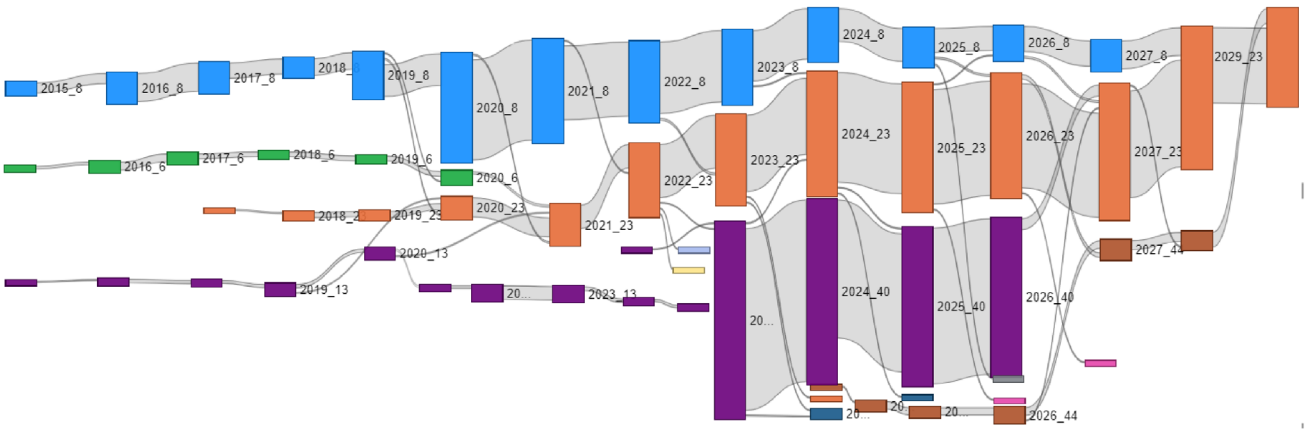

Figure 12. Evolution of the largest communities from 2015 to 2029. For clarity, only the six largest communities are displayed. Bar height reflects total strength of each community. The figure highlights an early phase of fragmentation followed by consolidation into larger, stable partnerships during Horizon Europe. The 2025–2029 data reflects funded projects to date and excludes future calls.

other sources could provide a more comprehensive perspective, but such datasets are currently unavailable.

5 Conclusions and future work

The methodology presented in this paper can provide valuable insights for both policymakers and project coordinators. Policy makers, such as those supporting Hydrogen Valley projects across Friuli Venezia Giulia, Croatia, and Slovenia, can leverage this methodology to monitor the engagement of organizations in their respective territories. As the Hydrogen Valley projects launched in 2022–2023 come to an end and new Horizon calls for proposals become available, policy makers will be able to assess whether today's leading organisations can form new partnerships and maintain their leadership positions. In addition, policy makers at European level can use the methodology to gain insights into the evolution of communities, to assess their openness to newcomers and the stability of collaboration between industrial and academic partners. From an alternative perspective, project coordinators within industrial or research organisations may use the methodology presented in this paper to identify potential partners who may be influential in facilitating future collaborations.

Future work will focus on several key areas. We will continue to monitor the centrality measures of the NAHV partners over the course of the project and beyond. In addition, we will apply the methodology to further case studies on different topics, in particular pandemic preparedness and electron microscopy, both of which are topics of interest for Area Science Park and have been published in 56. Finally, future research will explore how the rich textual information from project deliverables available in CORDIS can be effectively leveraged to enhance network models.

Ethics and consent

Ethical approval and consent were not required.

Data availability

The underlying data for this research is available through data.europa.eu, the official portal for European data, under a CC0 license:

Repository name: CORDIS - EU research projects under Horizon 2020 (2014–2020). DOI: https://doi.org/10.2906/11211709810 8/12[39]. The repository contains the following underlying data: `H2020 projects` file, which includes participating organisations, legal basis information, topic information, project URLs and classification with the European Science Vocabulary (EuroSciVoc).

Repository name: CORDIS - EU research projects under Horizon Europe (2021–2027). DOI: https://doi.org/10.2906/112117098108/20[40] The repository contains the following underlying data: HORIZON projects file, which includes participating organisations, legal basis information, topic information, project URLs and classification with the European Science Vocabulary (EuroSciVoc).

The data created by this research is available in Zenodo under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0). Repository name: Horizon Projects Network DOI: https://doi.org/10.5281/zenodo.13765372 The repository contains the following data:

- **O.csv**: Organisations' unique identifiers and attributes.
- **P.csv**: Projects' unique identifiers and attributes.
- **W.csv**: Describes the participation of each organization in a project in each specific year. The participation is weighted according to the "total cost" of the project, shared proportionally based on the project's duration within the year.
- **activity type.csv** provides a description of the codes used to define the activity types associated with each organization.
- **2015.graphml**, **2016.graphml**, ..., **2029.graphml**: a set of files representing network model for collaboration between organizations in any given year from 2015 to 2029. Each graph encodes the centrality measures (degree, strength, coreness) and community labels.
- **format-info.txt** provides information on the format of **.csv** and **.graphml** files used in this dataset.
- **sample-networks-hydrogen-energy.pdf** showcases a visual representation of the networks from 2015 to 2029.

Software availability

The software developed for this research is composed of two parts:

`communities` **package**[54]

Source code available at: https://github.com/fabio-morea/communities

Archived software available at: https://doi.org/10.5281/zenodo.13825471

License: Creative Commons Attribution 4.0 International (CC-BY 4.0)

`horizon intelligence` **code**[57]

Source code available at: https://github.com/fabio-morea/communities

Archived software available at: https://doi.org/10.5281/zenodo.11276687

License: Creative Commons Attribution 4.0 International (CC-BY 4.0)

References

1. Malloci FM, Penadés LP, Boratto L, *et al.*: **A text mining approach to extract and rank innovation insights from research projects.** In: *Web Information Systems Engineering -- WISE 2020*. Springer International Publishing, Cham, 2020; 143–154.
Publisher Full Text
2. Esztergár-Kiss D: **Horizon 2020 project analysis by using topic modelling techniques in the field of transport.** *Trans Telecommun J.* 2024; **25**(3): 266–277.
Publisher Full Text
3. Barber MJ, Krueger A, Krueger T, *et al.*: **Network of European Union-funded collaborative research and development projects.** *Phys Rev E Stat Nonlin Soft Matter Phys.* 2006; **73**(3 Pt 2): 036132.
PubMed Abstract | **Publisher Full Text**
4. Breschi S, Cusmano L: **Unveiling the texture of a European Research Area: emergence of oligarchic networks under EU framework programmes**. In: *Knowledge flows in European industry*, Routledge. 2006; 294–324.
Reference Source
5. Balland PA, Boschma R, Ravet J: **Network dynamics in collaborative research in the EU, 2003–2017.** *Eur Plan Stud.* 2019; **27**(9): 1811–1837.
Publisher Full Text
6. Maruccia Y, Solazzo G, Del Vecchio P, *et al.*: **Evidence from network analysis application to innovation systems and quintuple helix.** *Technol Forecast Soc Change.* 2020; **161**: 120306.
Publisher Full Text
7. Muscio A, Ciffolilli A, Lopolito A: **Technological diversity in collaborative projects: insights into European research policy.** *J Econ Policy Reform.* 2022; **25**(3): 322–343.
Publisher Full Text
8. Cerqueti R, Iovanella A, Mattera R: **Clustering networked funded European research activities through rank-size laws.** *Ann Oper Res.* 2023; **342**: 1–29.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
9. Enger SG: **Closed clubs: network centrality and participation in Horizon 2020.** *Sci Public Policy.* 2018; **45**(6): 884–896.
Publisher Full Text
10. Popescu-Apreutesei LE, Iosupescu MS, Fotache D, *et al.*: **European Union machine learning research: a network analysis of collaboration in higher education (2020–2024).** *Electronics.* 2025; **14**(7): 1248.
Publisher Full Text
11. Protogerou A, Caloghirou Y, Siokas E: **Research networking and technology fusion through EU-funded collaborative projects.** *Sci Public Policy.* 2013; **40**(5): 576–590.
Publisher Full Text
12. Genova VG, Giordano G, Ragozini G, *et al.*: **An analytic strategy for data processing of multimode networks.** *Adv Data Anal Classif.* 2024; **18**: 745–767.
Publisher Full Text
13. Kosztyán ZT, Király F, Katona AI: **Analysis and prediction of the Horizon 2020 R&D&I collaboration network.** *Expert Syst Appl.* 2024; **255**(Part B): 124417.
Publisher Full Text
14. Publications Office of the European Union: **Euroscivoc - European science vocabulary**. Accessed: 2024-09-24.
Reference Source
15. Clean Hydrogen Partnership: **Hydrogen valleys: insights into the emerging hydrogen economies around the world**. 2021.
Reference Source
16. Bampaou M, Panopoulos KD: **An overview of hydrogen valleys: current status, challenges and their role in increased renewable energy penetration.** *Renew Sustain Energy Rev.* 2025; **207**: 114923.
Publisher Full Text
17. **Treaty on the functioning of the European Union.** *Official Journal of the European Union.* 2012; **C326**. Consolidated version.
Reference Source
18. Clean Hydrogen Partnership: **HORIZON-JU-CLEANH2-2025 - Call for Proposals 2025**. April 2024; 2025-05-03.
Reference Source
19. European Commission: **Towards a roadmap for accelerating the deployment of Hydrogen Valleys across Europe: challenges and opportunities**. Commission Staff Working Document SWD(2024) 159 final, June 24, 2024; Accessed: 2025-05-03.
Reference Source
20. Weichenhain U, Kaufmann M, Hölscher M, *et al.*: **Going global: an update on hydrogen valleys and their role in the new hydrogen economy**. Clean Hydrogen Partnership, 2021.
Reference Source
21. Majka A, Klimczyk M, Kucharski K, *et al.*: **Hydrogen valley as a hub for technological cooperation between science, business, local government and NGOS: an overview of approaches in Europe.** *Torun Int Stud.* 2023; **1**(17): 5–15.
Publisher Full Text
22. Frankowska M, Cheba K: **Exploring the research landscape of hydrogen valleys: a bibliometric analysis.** *J Sustain Dev Transport Logist.* 2023; **8**(2): 348–359.
Publisher Full Text
23. Petrollese M, Concas G, Lonis F, *et al.*: **Techno-economic assessment of green hydrogen valley providing multiple end-users.** *Int J Hydrogen Energy.* 2022; **47**(57): 24121–24135.
Publisher Full Text
24. Concas G, Cocco D, Lecis L, *et al.*: **Life cycle analysis of a hydrogen valley with multiple end-users.** *J Phys Conf Ser.* 2022; **2385**: 012035.
Publisher Full Text
25. Department of Science, Bambili Energy Innovation, Anglo American Platinum, and ENGIE: **South Africa hydrogen valley final report**. 2021; Accessed: September 12, 2024.
Reference Source
26. European Commission: **North Adriatic hydrogen valley - project website**.
Publisher Full Text
27. European Commission: **Call HORIZON-JTI-CLEANH2-2022-2 - Horizon Europe joint technology initiative on clean Hydrogen**. Accessed: 2024-08-26.
Reference Source
28. Komninos N: **Transformation of industry ecosystems in cities and regions: a generic pathway for smart and green transition.** *Sustainability.* 2022; **14**(15): 9694.
Publisher Full Text
29. Chesbrough HW: **Open innovation: the new imperative for creating and profiting from technology**. Harvard Business School Publishing Company, 2003.
Reference Source
30. Carayannis EG, Campbell DF: **Triple helix, quadruple helix and quintuple helix and how do knowledge, innovation and the environment relate to each other?: A proposed framework for a trans-disciplinary analysis of sustainable development and social ecology.** *Int J Soc Ecol Sustain Dev.* 2010; **1**(1): 41–69.
Publisher Full Text
31. Casaramona A, Sapia A, Soraci A: **How TOI and the quadruple and quintuple helix innovation system can support the development of a new model of international cooperation.** *J Knowl Econ.* 2015; **6**(3): 505–521.
Publisher Full Text
32. Hegyi FB, Mariussen A: **Connecting urban and regional Innovation Ecosystems to enhance competitiveness.** *Osuwa.* 2021.
Publisher Full Text
33. Pires SM, Polido A, Teles F, *et al.*: **Territorial innovation models in less developed regions in Europe: the quest for a new research agenda?** *Eur Plan Stud.* 2020; **28**(8): 1639–1666.
Publisher Full Text
34. Agrawal A, Cockburn I, Galasso A, *et al.*: **Why are some regions more innovative than others? The role of small firms in the presence of large labs.** *J Urban Econ.* 2014; **81**: 149–165.
Publisher Full Text
35. Soraci A: **Innovation as a strategy of reaction to delocalization: shipbuilding sector, a study case.** Promoting Entrepreneurship by Universities, Lahti University of Applied Sciences, 2009.
36. Federal Economic Development Agency for Northern Ontario: **Regional innovation ecosystems.** 2024.
Reference Source
37. European Commission: **Kohesio: discover EU projects in your region**. 2025; Accessed: 2025-05-03.
Reference Source
38. European Commission: **CORDIS: Eu research results.** Accessed: Sept. 2024.
Reference Source
39. EU Publications Office: **CORDIS - EU research projects under Horizon 2020 (2014–2020).** 2015.
http://www.doi.org/10.2906/112117098108/12
40. EU Publications Office: **CORDIS - EU research projects under HORIZON EUROPE (2021–2027).** 2022.
http://www.doi.org/10.2906/112117098108/20
41. Batagelj V, Zaversnik M: **An o(m) algorithm for cores decomposition of networks.** *ArXiv*. cs.DS/0310049, 2003.
Publisher Full Text
42. Diboune A, Slimani H, Nacer H, *et al.*: **A comprehensive survey on community detection methods and applications in complex information networks.** *Soc Netw Anal Min.* 2024; **14**(1): 93.
Publisher Full Text
43. Khawaja FR, Zhang Z, Memon Y, *et al.*: **Exploring community detection methods and their diverse applications in complex networks: a comprehensive review.** *Soc Netw Anal Min.* 2024; **14**: 115.
Publisher Full Text
44. Morea F, De Stefano D: **Enhancing stability and assessing uncertainty**

in community detection through a consensus-based approach. 2024.
Publisher Full Text

45. Blondel VD, Guillaume JL, Lambiotte R, *et al.*: **Fast unfolding of communities in large networks.** *J Stat Mech.* 2008; **2008**(10): P10008.
Publisher Full Text

46. Traag VA, Waltman L, Van Eck NJ: **From Louvain to Leiden: guaranteeing well-connected communities.** *Sci Rep.* 2019; **9**(1): 5233.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

47. Raghavan UN, Albert R, Kumara S: **Near linear time algorithm to detect community structures in large-scale networks.** *Phys Rev E Stat Nonlin Soft Matter Phys.* 2007; **76**(3 Pt 2): 036106.
PubMed Abstract | **Publisher Full Text**

48. Newman MEJ, Girvan M: **Finding and evaluating community structure in networks.** *Phys Rev E Stat Nonlin Soft Matter Phys.* 2004; **69**(2 Pt 2): 026113.
PubMed Abstract | **Publisher Full Text**

49. Newman MEJ: **Finding community structure in networks using the eigenvectors of matrices.** *Phys Rev E Stat Nonlin Soft Matter Phys.* 2006; **74**(3 Pt 2): 036104.
PubMed Abstract | **Publisher Full Text**

50. Pons P, Latapy M: **Computing communities in large networks using random walks.** In: *Computer and Information Sciences-ISCIS 2005: 20th International Symposium, Istanbul, Turkey, October 26–28, 2005. Proceedings 20*. Springer, 2005; 284–293.
Publisher Full Text

51. Rosvall M, Bergstrom CT: **An information-theoretic framework for resolving community structure in complex networks.** *Proc Natl Acad Sci U S A.* 2007; **104**(18): 7327–7331.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

52. R Core Team: **R: a language and environment for statistical computing.** 2022.
Reference Source

53. Csardi G, Nepusz T, Traag V, *et al.*: **igraph: network analysis and visualization in R.** 2023.
Reference Source

54. Morea F: **https://github.com/fabio-morea/communities**. **August, 2024.**

55. European Commission: **Horizon Dashboard.** Accessed: 2024-10-16.
Reference Source

56. Morea F: **Horizon projects network dataset.** Dataset, 2024.
Publisher Full Text

57. Morea F: **https://github.com/fabio-morea/horizon-intelligence**. August. 2024.

Open Peer Review

Current Peer Review Status: ✔ ? ?

Version 2

Reviewer Report 09 September 2025

https://doi.org/10.21956/openreseurope.22476.r58795

© **2025 Handoko C.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Chanel Tri Handoko
Sebelas Maret University, Surakarta, Indonesia

The authors introduce a robust approach to network analysis using CORDIS data, applying AI-driven classification and community detection algorithms to reveal the evolution of project collaborations from 2015 to 2029. The North Adriatic Hydrogen Valley (NAHV) is used as a case study to highlight methodological advancements and practical insights. The work is contextualized against contemporary literature, emphasizing gaps and innovations in mapping EU research collaborations. Methodological steps, including data acquisition, network construction, centrality and coreness analysis, and consensus approaches for community detection, are clearly outlined. Results underscore the increasing scale and stability of research communities driven by EU hydrogen initiatives, with implications for policy and future research.
1. The introduction concisely positions the study within the broader context of EU-funded innovation research, drawing from text mining and network analysis literature. The manuscript highlights prior limitations of network methods applied to CORDIS data and references several recent surveys and application papers. To improve engagement, the manuscript could discuss specific recent findings from the referenced text-mining and community detection studies in more detail, especially regarding methodological choices (e.g., how consensus community detection advances the field relative to alternative methods)
2. The stepwise methodology is well-structured, including pseudocode and detailed algorithmic descriptions for each stage: data preparation, network construction, centrality and coreness computation, community detection, and longitudinal tracking. The choice of specific algorithms (Louvain, Leiden, Label Propagation, Edge Betweenness, Leading Eigenvector, Walktrap, Infomap) is justified, and the strengths/limitations of each are empirically explored in the results section. The technical discussion of how consensus and quality checks are implemented could be expanded, especially on algorithm selection and validation, perhaps with more concrete implementation notes or reproducibility factors.
3. The manuscript might briefly summarize the data cleaning and validation steps performed before analysis, including potential issues with missing or inconsistent records from CORDIS.
4. The manuscript may benefit from including summary statistics (mean, variance, etc.) for centrality measures, and quantitative measures of algorithmic performance or computational cost comparison across algorithms

Is the work clearly and accurately presented and does it engage with the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

Are all the source data and materials underlying the results available?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: renewable energy, public partner partnership

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Version 1

Reviewer Report 07 January 2025

https://doi.org/10.21956/openreseurope.20048.r48915

© 2025 Olympios A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Andreas V Olympios
University of Cyprus, Nicosia, Cyprus

This paper develops a methodology to analyse collaborations over time among participants in Horizon 2020 and Horizon Europe, using the hydrogen energy sector as a case study. I would like to thank the authors for their interesting work, and provide my comments regarding the strengths and areas for improvement:

- The introduction begins clearly but a more comprehensive review of the literature is

needed. It would be useful to include an in-depth analysis of previously used methods and their limitations before discussing the novel contributions of this specific paper.
- While hydrogen is identified as a case study, the introduction focuses almost entirely on this specific topic, which creates a disconnect. If hydrogen is just the case study, the discussion in the introduction should be broader, emphasising the general methodology rather than the specific application.
- Algorithm 1 is presented in a mixed format, combining textual descriptions and pseudocode. This could be improved by separating the process description into a table and presenting the pseudocode independently for clarity.
- The data acquisition strategy appears logical and well-thought-out.
- Captions for all figures should be detailed enough to make the figures self-explanatory. Some captions, such as for Figure 3, lack sufficient detail.
- Figure 3 requires further explanation, and the image quality appears compressed, making the text difficult to read.
- The font sizes used in figures are inconsistent. For example, compare the y-axis labels in Figures 4 and 5.
- Figure 4 timeline begins in 2015, but the caption states 2016.
- The downward trend from 2025–2029 in Figure 4 is unclear. The paper mentions a decrease in partners and collaborations but does not explain why. If the decline is due to incomplete data (e.g., new projects not yet appearing), the figure may be misleading and should be clarified to avoid confusion.
- The classification of projects into market-oriented and technology-oriented categories reveals interesting trends. However, the notable increase in market-oriented projects in 2023 is not discussed in detail, leaving a gap in interpretation.
- Figure 6 is interesting, but the small font size makes the organization names difficult to read. It would be good to increase the size or restructure the figure for clarity.
- For Figure 8, the paper mentions that one-third of the solutions are invalid according to the literature, but the supporting references or discussion are missing at this point.
- Figures 9 and 10 lack legends and are challenging to interpret. Adding legends or annotations that correspond to the organizations or data points would make them informative.
- The professionalism of the paper could be improved by ensuring consistency in figure design, font sizes, and formatting.

Overall, the paper introduces a methodology that can provide insights into collaborations in the hydrogen energy sector, but there are some areas for improvement, particularly in terms of literature review, presentation of figures, and depth of analysis.

Is the work clearly and accurately presented and does it engage with the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

Are all the source data and materials underlying the results available?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Energy systems, heat pumps, buildings, green hydrogen, combined heat and power, optimisation.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 09 Jun 2025

Fabio Morea

Thank you for your detailed review. Your feedback has been very helpful in identifying specific areas for improvement, and we have revised the manuscript accordingly. The updated version will be submitted shortly. Specifically, we expanded the introduction to provide a broader framing of the methodology and added a more comprehensive literature review on network analysis. Algorithm 1 has been reformatted for clarity, and several improvements were made to the figures and captions.

Best regards,
the Authors

Competing Interests: No competing interests were disclosed.

Reviewer Report 26 December 2024

https://doi.org/10.21956/openreseurope.20048.r48910

© 2024 Zemīte L. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Laila Zemīte
Riga Technical University, Riga, Latvia

The research article "Mapping leadership and communities in EU-funded research through network analysis" by Fabio Morea, Alberto Soraci, and Domenico De Stefano introduces a novel

methodology for analysing collaborations and leadership roles in EU-funded research projects using network analysis. The study focuses on the hydrogen energy sector, particularly the North Adriatic Hydrogen Valley (NAHV) project, and utilizes data from the CORDIS portal, which provides comprehensive information on Horizon 2020 and Horizon Europe projects. The methodology involves downloading and enriching data from CORDIS, segmenting it by year, and transforming it into weighted networks representing collaborations between organizations. Centrality measures are used to assess the influence of individual organizations, while community detection algorithms identify stable collaborations. Temporal analysis tracks the evolution of these roles and communities over time. The results demonstrate the emergence of key leaders and stable communities within the hydrogen energy sector, highlighting significant collaboration. The insights gained are valuable for policymakers and organizations seeking to foster innovation through sustained partnerships. The approach can be extended to other sectors, offering a framework for understanding the impact of EU research funding on collaboration and leadership dynamics. The article also discusses the concept of hydrogen valleys, which are integrated ecosystems that produce, exchange, and consume hydrogen, covering the entire hydrogen value chain. The NAHV project, involving 37 partners from Croatia, Slovenia, and the Italian region of Friuli Venezia Giulia, aims to develop a long-lasting hydrogen economy through a systemic and holistic approach. The study's findings reveal that Horizon grants in the hydrogen sector have led to the formation of stable, long-term partnerships, particularly in the domain of hydrogen valleys. The methodology presented can improve visualizations in the Horizon Dashboard and provide valuable insights for both policymakers and project coordinators. Future work will focus on monitoring the centrality measures of NAHV partners and applying the methodology to other case studies.

The paper presents an excellent framework for analysing collaborative networks within the hydrogen sector. The paper employs advanced centrality measures and community detection algorithms to identify influential organizations and analyse the stability of collaborations over time. By focusing on the North Adriatic Hydrogen Valley (NAHV) project, the study demonstrates how EU-funded initiatives effectively foster collaboration among industry, academia, and government, while addressing critical sustainability challenges. The methodology for analysing collaborative dynamics and the entire hydrogen value chain is both innovative and meticulously detailed. It skillfully integrates technical approaches, such as network analysis, with practical applications, including policy impact evaluation. The robust data acquisition process and thorough analysis of collaboration patterns further strengthen the paper's contribution to the field.

The article's current title does not fully reflect its focus on hydrogen projects, and its abstract lacks emphasis on the hydrogen sector. A revised title, such as **“Mapping Leadership and Communities in EU-funded Hydrogen Research Through Network Analysis”**, will better align with the paper's subject matter and objectives. Additionally, the abstract should be refined to highlight the centrality of hydrogen research in the analysis.

Section 1.1 limits its scope by primarily discussing the NAHV project. Expanding this section to include a broader examination of existing, planned, and proposed hydrogen valleys across the EU would significantly enhance the article's value. Incorporating insights from sources like the Clean Hydrogen Partnership’s database (e.g., https://h2v.eu/hydrogen-valleys) will provide a more comprehensive overview of the hydrogen valleys landscape. This expanded approach will highlight regional variations, common challenges, and synergies, contributing to the research objectives and aligning with the EU's broader hydrogen strategy.

The study's reliance on the CORDIS dataset, while robust, restricts its scope by excluding data from other funding sources such as local Member State calls, the Connecting Europe Facility (CEF), or the Interregional Innovation Investment (I3) instrument. To address this limitation the methodology could be enhanced by integrating data from these additional funding sources. Including insights from diverse datasets would expand the applicability of the findings, offering more actionable recommendations for policymakers and stakeholders, while capturing the full spectrum of hydrogen-related projects within and beyond EU programs. Please include details on the methodologies and techniques for integrating different datasets to encompass additional funding sources.

Is the work clearly and accurately presented and does it engage with the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

Are all the source data and materials underlying the results available?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: My areas of research focus on renewable energy, hydrogen technologies, energy storage, smart energy systems, electricity grid systems, risk analysis in energy, and the regulation of electricity and gas transmission. My work also includes expertise in gas infrastructure, energy security, and sustainable energy development. Additionally, I contribute to innovation in smart grids and energy market regulations.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 09 Jun 2025

Fabio Morea

Thank you for your constructive review.

Your feedback helped us improve the structure and clarity of the article. The revised

manuscript (to be submitted shortly) reflects your suggestions and incorporates recent literature. Key changes include a revised title and abstract (updated to better reflect the contents of the paper) and an expanded introduction chapter to provide a broader overview of hydrogen valleys. This has made the introduction significantly longer—hopefully not excessively so.
Regarding additional funding sources (e.g. CEF, I3, national/regional grants), we fully agree that this is an important direction. However, due to the differing structure, scope, and accessibility of these datasets, we see this as a separate line of work, which we plan to explore in a future study.

Thank you again for your valuable contribution

Best regards,
the Authors

Competing Interests: No competing interests were disclosed.