Advances in Data Analysis and Classification
https://doi.org/10.1007/s11634-023-00556-4

REGULAR ARTICLE

An analytic strategy for data processing of multimode networks

Vincenzo Giuseppe Genova[1] · Giuseppe Giordano[2] · Giancarlo Ragozini[3] · Maria Prosperina Vitale[2]

Received: 29 November 2022 / Accepted: 20 July 2023

Abstract

Complex network data structures are considered to capture the richness of social phenomena and real-life data settings. Multipartite networks are an example in which various scenarios are represented by different types of relations, actors, or modes. Within this context, the present contribution aims at discussing an analytic strategy for simplifying multipartite networks in which different sets of nodes are linked. By considering the connection of multimode networks and hypergraphs as theoretical concepts, a three-step procedure is introduced to simplify, normalize, and filter network data structures. Thus, a model-based approach is introduced for derived bipartite weighted networks in order to extract statistically significant links. The usefulness of the strategy is demonstrated in handling two application fields, that is, intranational student mobility in higher education and research collaboration in European framework programs. Finally, both examples are explored using community detection algorithms to determine the presence of groups by mixing up different modes.

✉ Giuseppe Giordano
ggiordano@unisa.it

Vincenzo Giuseppe Genova
vincenzogiuseppe.genova@unipa.it

Giancarlo Ragozini
giragoz@unina.it

Maria Prosperina Vitale
mvitale@unisa.it

[1] Department of Economics, Business, and Statistics, University of Palermo, Viale delle Scienze Ed. 13, 90128 Palermo, PA, Italy

[2] Department of Political and Social Studies, University of Salerno, Via Giovanni Paolo II n. 132, 84084 Fisciano, SA, Italy

[3] Department of Political Science, University of Naples Federico II, Via Leopoldo Rodinò n. 22/a, 80138 Naples, NA, Italy

Published online: 29 August 2023

Keywords Multipartite networks · Data simplification · Data filtering · Normalization · Real-data examples

1 Introduction

In recent years, the use of complex networks to capture the richness of social phenomena has become more frequent, going beyond traditional statistical techniques focused on attribute variables. Multipartite networks are an example of such a data structure in which a multitude of scenarios are represented with different modes or nodes and types of relations, such as multilayer (Interdonato et al. 2020; Magnani and Wasserman 2017; Dickison et al. 2016; Kivelä et al. 2014; Batagelj et al. 2007), multiplex (Genova et al. 2022; Giordano et al. 2019; Bródka et al. 2018), multilevel (Zhu et al. 2016), and multimode networks (Everett and Borgatti 2019; Borgatti and Everett 1992), and various temporal snapshots in a longitudinal perspective (Boccaletti et al. 2014). Network data can be enriched with additional features, such as nodes' attributes (Giordano and Vitale 2011) and links' weights (Menichetti et al. 2014), with the aim of describing real-life phenomena in more detail.

To the best of our knowledge, few papers have been devoted to the analysis of multimode networks representing a generalization of the conceptual basis and the matrix formalism of two-mode networks and bipartite graphs. In Fararo and Doreian (1984) seminal paper, a tripartite network (a special case of a multimode network) is defined as consisting of three types of nodes, and ties are present only between nodes of distinct types. This structure can be extended to any number of modes, which gives rise to multimode networks (Everett and Borgatti 2019).

In this scenario, the present contribution aims at introducing an analytic strategy for handling complex networks with a procedure that relies on network simplification, normalization, and filtering. Specifically, we propose a three-step procedure for the data processing phase: 1) *data simplification*, which transforms a multipartite network into bipartite weighted networks, passing through hypergraphs, without losing any relevant information; 2) *data normalization*, which exploits log-linear models (Agresti 2007) to highlight interesting association patterns; and 3) *data filtering*, which retains the most significant links through studentized residuals from the log-linear model. The resulting networks can be examined with network analysis.

Such complex data structures may arise for a variety of systems in many application fields, such as in folksonomy (users, texts, tags, and topics; Giordano et al. (2021); Saoud and Platoš (2018)), bibliographic data (papers, journals, keywords, references) Batagelj and Cerinšek (2013), and genomic networks (genes, diseases, and patients). Moving from these scenarios, we exploit the usefulness of the proposed procedure in two real-life data settings involving higher education institutions in which weighted directed and undirected graphs are defined for intranational student mobility flows (Dotti et al. 2014) and scientific collaboration in projects (Garas and Argyrakis 2009), respectively. The simplified networks for both examples are then analyzed with a flow-based community detection algorithm (Blöcker and Rosvall 2020; Edler et al. 2017) that partitions filtered bipartite, directed and undirected, weighted networks.

This paper is organized as follows. Section 2 briefly describes real-data examples that demonstrate the usefulness and practical implications of the proposed approach. Section 3 reports the definition and the notation of multimode networks and hypergraphs. Section 4 presents the technical details of the proposed analytic strategy with the three-step procedure for complex network data processing. Section 5 discusses the results of the proposed strategy applied to the two real-data settings. Section 6 reports the main findings of clustering network algorithms performed on the simplified, normalized, and filtered bipartite weighted networks derived for student mobility and collaboration data. The last section provides concluding remarks and suggestions for future research.

2 Description of real-data examples

Intranational mobility networks. Several researchers have reported the factors that influence students' decisions to stay in or leave their home province to undertake university studies (Prazeres 2013) and consider whether drivers of international student migration also apply to student mobility within a nation (Findlay et al. 2018). In addition to the persistence of interregional economic disparities between the South and the Centre-North in Italy, the relevance of intranational university students' mobility is emphasized (Dotti et al. 2014). Student mobility in higher education is worth investigating due to the peculiar characteristics of the Italian university system (Columbu et al. 2022; Santelli et al. 2022; Columbu et al. 2021; Genova et al. 2021; Santelli et al. 2019). This phenomenon follows the traditional South-to-North migration chain (Genova et al. 2019), as well as the recent North-to-North mobility (Rizzi et al. 2021). Because of the availability of data at the individual level in the MOBYSU.IT (2016) database,[1] a very wide spectrum of data structures can be derived and further investigated with network analysis tools: i) *directed unipartite weighted networks*, with provinces or universities as nodes and students' flows as the links' weights (Columbu et al. 2022, 2021); ii) *bipartite weighted networks*, with student flows from the provinces of origin and the universities of enrollment or the educational programs (Santelli et al. 2022; Genova et al. 2019); and iii) *multiplex unipartite weighted networks*, with provinces or universities as nodes and the educational programs as layers (Primerano et al. 2021).
The analysis we report reconstructs *multimode networks* in which *students*, *regions*, *provinces*, *universities*, and *educational programs* are linked through a set of affiliation networks generated by different kinds of relations.

Scientific collaboration networks. Scientific collaboration between institutions (universities, research centers, and private companies) in projects is another area of interest widely investigated with network science tools. The main focus is related to the impact of interregional collaboration flows on knowledge diffusion and innovation captured

[1] Data drawn from the "Anagrafe nazionale degli studenti e dei laureati (ANS)" of the Italian Ministry of University and Research were processed according to the national research project "From high school to the job market: analysis of the university careers and the university North–South mobility." The MOBYSU.IT database includes information on students' demographic characteristics, their high school backgrounds, and their bachelor and master program choices at university since the 2008–2009 academic year.

by weighted matrices that show the intensity of exchanges among partners in terms of multiplex collaborative ties, that is, joint projects, patents, and publications (Meliciani et al. 2022; Maggioni et al. 2013).
Open data derived from funding schemes provided by the European Framework Programs (EU-FP7) are often considered to explore the international cooperation networks of partners that collaborate to address scientific problems. Specifically, researchers have investigated topological and structural collaboration network patterns, discovering that institutions or countries acting as central hubs in specific thematic areas (Garas and Argyrakis 2009) and research collaboration networks enter into multilayer data structures to measure multiple connections between academic and non-academic actors in higher education institutions (Kosztyán et al. 2021).
In this scenario, the CORDIS[2] database (European Commission 2022; Amoroso et al. 2018) consisting of institutions from European Union (EU) and non-EU countries and several projects funded under the framework programs (FPs) is used to reconstruct multimode networks in which *countries*, *institutions*, *framework programs*, *projects*, and *fields* represent different modes.

3 Multimode networks and hypergraphs

Formally, a multimode network $\mathcal{M}$ can be conceived of consisting of the pair $(\mathcal{V}, \mathcal{L})$. $\mathcal{V}$ is the collection of M set of nodes $\{\mathcal{V}^m\}_{m=1,\dots,M}$, with $\mathcal{V}^m \equiv \{v_1^m, \dots, v_i^m, \dots, v_I^m\}$ the set of I nodes of the $m-th$ mode. The sets of nodes V^m are mutually disjointed; that is, $m \neq m' \Rightarrow V^m \cap V^{m'} = \emptyset$.

In the multimode network approach reported by Everett and Borgatti (Everett and Borgatti 2019), links are defined only between each pair of nodes of different types. In general, we have $\frac{M(M-1)}{2}$ affiliation networks describing the binary relationship between a pair of sets of nodes. Then, the set of links $\mathcal{L}$ can be given by the possible combinations $\mathcal{L}^{m,m'} \subseteq \mathcal{V}^m \times \mathcal{V}^{m'}, \forall m \neq m'; m, m' = 1, \dots M$. A link $(v_i^m, v_j^{m'}) \in \mathcal{L}^{m,m'}$, with $m \neq m'$, is an ordered pair that indicates whether v_i^m, the i-th node of the m-th mode, is linked to $v_j^{m'}$, the j-th node of a different mode m'.

Generalizing the original idea of tripartite graphs (Fararo and Doreian 1984), it is possible to define a unique adjacency matrix $\mathbf{A} = (a_{ij}^{mm'})$ given by the combination in a block matrix of sociomatrix $\mathbf{A}^{mm'}$ corresponding to the two-mode networks $\mathcal{B}^{mm'} = (\mathcal{V}^m, \mathcal{V}^{m'}, \mathcal{L}^{m,m'})$, with $a_{ij}^{mm'} = 1$ if $(v_i^m, v_j^{m'}) \in \mathcal{L}^{m,m'}$, and $a_{ij}^{mm'} = 0$ if $(v_i^m, v_j^{m'}) \notin \mathcal{L}^{m,m'}$.

For the sake of simplicity, using the *intranational mobility network* example described in Sect. 1, we exploit the methodological formalization of the related multimode networks as follows: $\mathcal{V}^1 \equiv \mathcal{S} \equiv \{s_1, \dots, s_i, \dots, s_I\}$ the set of students; $\mathcal{V}^2 \equiv \mathcal{U} \equiv \{u_1, \dots, u_j, \dots, u_J\}$ the set of universities; $\mathcal{V}^3 \equiv \mathcal{E} \equiv \{e_1, \dots, e_k, \dots, e_K\}$ the

[2] The Community Research and Development Information Service–CORDIS– is the European Commission's primary source of results from the R&D projects funded by the EU's framework programmes since 1990. It includes details about various research initiatives, grants, and collaborative projects across a wide range of disciplines –science, technology, and innovation– as well as projects' information – title, objectives, organizations, funding details, and outcomes.

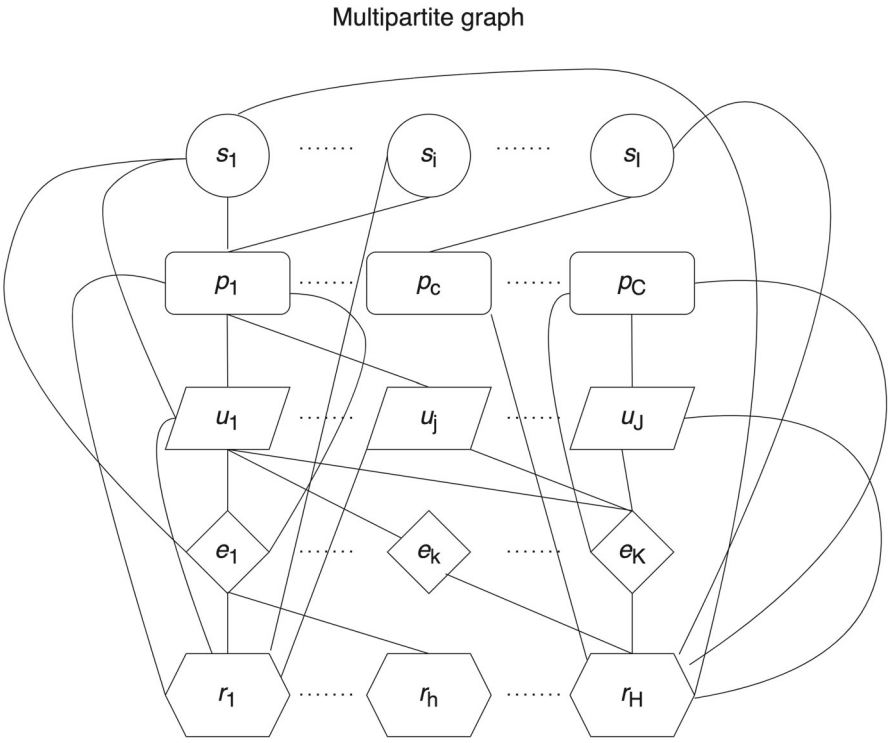

Fig. 1 Multipartite graph visualization for multimode networks with five modes (A toy example)

set of educational programs; $\mathcal{V}^4 \equiv \mathcal{R} \equiv \{r_1, \ldots, r_h, \ldots, r_H\}$ the set of regions; and $\mathcal{V}^5 \equiv \mathcal{P} \equiv \{p_1, \ldots, p_c, \ldots, p_C\}$ the set of provinces. Figure 1 shows the multipartite graph described above.

Although all the bipartite networks express affiliation links, the meaning of the links differs among the networks. For example, $\mathcal{B}^{\mathcal{SU}}$ contains the information about the enrollement of students in universities, while $\mathcal{B}^{\mathcal{SR}}$ the residence of students in regions; $\mathcal{B}^{\mathcal{RE}}$ considers if educational programs are present in regions. The network $\mathcal{B}^{\mathcal{RP}}$ reports the information of the list of provinces in each region. Then, some links are related to the choices of students and to their characteristics, while others are related to the structure of the Italian university system or geographic information. The corresponding adjacency matrix $\mathbf{A}$ is described as follows:

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{A}^{\mathcal{SU}} & \mathbf{A}^{\mathcal{SE}} & \mathbf{A}^{\mathcal{SR}} & \mathbf{A}^{\mathcal{SP}} \\ \mathbf{A}^{\mathcal{US}} & \mathbf{0} & \mathbf{A}^{\mathcal{UE}} & \mathbf{A}^{\mathcal{UR}} & \mathbf{A}^{\mathcal{UP}} \\ \mathbf{A}^{\mathcal{ES}} & \mathbf{A}^{\mathcal{EU}} & \mathbf{0} & \mathbf{A}^{\mathcal{ER}} & \mathbf{A}^{\mathcal{EP}} \\ \mathbf{A}^{\mathcal{RS}} & \mathbf{A}^{\mathcal{RU}} & \mathbf{A}^{\mathcal{RE}} & \mathbf{0} & \mathbf{A}^{\mathcal{RP}} \\ \mathbf{A}^{\mathcal{PS}} & \mathbf{A}^{\mathcal{PU}} & \mathbf{A}^{\mathcal{PE}} & \mathbf{A}^{\mathcal{PR}} & \mathbf{0} \end{bmatrix}.$$

3.1 Projecting multimode networks

As in the case of bipartite and tripartite networks, it is possible to define the projection of multimode network on one (or more) mode(s).

Let us consider a tripartite network $\mathcal{T} = (\mathcal{V}^1, \mathcal{V}^2, \mathcal{V}^3, \mathcal{L}^{1,2,3})$ and project it on one of the three modes, say, $\mathcal{V}^3$. The result of this projection is a weighted bipartite network, ${}_3\mathcal{B}^{12} = (\mathcal{V}^1, \mathcal{V}^2, {}_3\mathcal{L}^{1,2}, \mathcal{W})$, with ${}_3\mathcal{L}^{1,2} \subseteq \mathcal{V}^1 \times \mathcal{V}^2$ and the function $w : {}_3\mathcal{L}^{1,2} \to \mathbb{N}$, in which the weights of the links $w(v_i^1, v_j^2) = w_{ij}$ are the numbers of common links in $\mathcal{V}^3$ between v_i^1 and v_j^2. This can be obtained also by considering the network multiplication procedure defined in (Batagelj and Cerinšek 2013), that is, ${}_3\mathcal{B}^{12} = \mathcal{B}^{13} \times \mathcal{B}^{32}$, which also corresponds to the affiliation matrices multiplication, ${}_3\mathbf{A}^{12} = \mathbf{A}^{13}\mathbf{A}^{32}$.

Given the tripartite graph $\mathcal{T}$, it is possible to define a double projection, say, on $\mathcal{V}^3$ and $\mathcal{V}^2$, ending up in one mode ${}_{(2,3)}\mathcal{N}^1 = (\mathcal{V}^1, {}_{(2,3)}\mathcal{L}^1, \mathcal{W})$, with ${}_{(2,3)}\mathcal{L}^1 \subseteq \mathcal{V}^1 \times \mathcal{V}^1$ and the function $w : {}_{(2,3)}\mathcal{L}^1 \to \mathbb{N}$, in which the weights of the links $w(v_i^1, v_{i'}^1) = w_{ii'}$ are the sums of the weights for the common links between v_i^1 and $v_{i'}^1$ in bipartite network ${}_3\mathcal{B}^{12}$. In such a case, the graph is also directed and has self-loops. In terms of the network multiplication, we have that ${}_{(2,3)}\mathcal{N}^1 = {}_3\mathcal{B}^{12} \times \mathcal{B}^{21} = \mathcal{B}^{13} \times \mathcal{B}^{32} \times \mathcal{B}^{21}$, and in matrix form ${}_{(2,3)}\mathbf{A}^1 = \mathbf{A}^{13}\mathbf{A}^{32}\mathbf{A}^{21}$.

Analogously, we can consider other types of projections in more than two dimensions. If we consider the multipartite network with four modes $\mathcal{Q} = (\mathcal{V}^1, \mathcal{V}^2, \mathcal{V}^3, \mathcal{V}^4, \mathcal{L}^{1,2,3,4})$, and its projection on one of the four modes (say, $\mathcal{V}^4$), we get a weighted tripartite 3-uniform hypergraph ${}_4\mathcal{HT}^{123} = (\mathcal{V}^1, \mathcal{V}^2, \mathcal{V}^3, {}_4\mathcal{L}^{123}, \mathcal{W})$, with ${}_4\mathcal{L}^{123} \subseteq \mathcal{V}^1 \times \mathcal{V}^2 \times \mathcal{V}^3$, the collection of hyper-edges, with the generic term (v_i^1, v_j^2, v_k^3) defined as

$$(v_i^1, v_j^2, v_k^3) \in {}_4\mathcal{L}^{123} \iff (v_i^1, v_h^4) \in \mathcal{L}^{1,4} \wedge (v_j^2, v_h^4) \in \mathcal{L}^{2,4} \wedge (v_k^3, v_h^4) \in \mathcal{L}^{3,4}.$$

$\mathcal{W}$ is the set of weights obtained with the function $w : {}_4\mathcal{L}^{123} \to \mathbb{N}$, and $w(v_i^1, v_j^2, v_k^3) = w_{ijk}$ is the number of common links in $\mathcal{V}^4$ among v_i^1, v_j^2, and v_k^3. This network data structure can be described as a three-way array $\mathbb{A} = (a_{ijk})$, with $a_{ijk} \equiv w_{ijk}$, called three-way networks (Batagelj et al. 2007).

In addition, we can define a new kind of projection, namely, the *conditional projection*, in which the link in the projection exists if some condition on other links of other modes is satisfied. Consider a pentapartite network $\mathcal{P}(\mathcal{V}^1, \mathcal{V}^2, \mathcal{V}^3, \mathcal{V}^4, \mathcal{V}^5, \mathcal{L}^{1,2,3,4,5})$, and call ${}_4\mathcal{HT}^{123}_{|\mathcal{B}^{5m}}$ the projection of $\mathcal{P}$ into a tripartite hypergraph on $\mathcal{V}^4$ by conditioning on the presence of the links in the one of the bipartite $\mathcal{B}^{5\,m}$, $m \neq 5$. The links of this hypergraph ${}_4\mathcal{L}^{123}_{|\mathcal{B}^{5m}}$ can be defined as follows:

$$\begin{aligned}(v_i^1, v_j^2, v_k^3)_{|\mathcal{B}^{5m}} \in {}_4\mathcal{L}^{123}_{|\mathcal{B}^{5m}} \iff (v_i^1, v_h^4) \in \mathcal{L}^{1,4} \wedge (v_j^2, v_h^4) \in \mathcal{L}^{2,4} \wedge \\ \wedge (v_k^3, v_h^4) \in \mathcal{L}^{3,4} \wedge (v_c^5, v_j^m) \in \mathcal{L}^{5,m}.\end{aligned}$$

Returning to our example related to student mobility, we consider only three of the five modes (students, universities, and regions), and students as the mode for the

projection, that is, the tripartite network $\mathcal{T} = (\mathcal{S}, \mathcal{U}, \mathcal{R}, \mathcal{L}^{\mathcal{S},\mathcal{U},\mathcal{R}})$. A toy example of $\mathcal{T}$ with six students, three universities and two regions is depicted in Fig. 2 (upper panel). Projecting the tripartite network, we obtain a weighted bipartite network ${}_{\mathcal{S}}\mathcal{B}^{\mathcal{R}\mathcal{U}} = (\mathcal{R}, \mathcal{U}, {}_{\mathcal{S}}\mathcal{L}^{\mathcal{R},\mathcal{U}}, \mathcal{W})$, with ${}_{\mathcal{S}}\mathcal{L}^{\mathcal{R},\mathcal{U}} \subseteq \mathcal{R} \times \mathcal{U}$ and $w : {}_{\mathcal{S}}\mathcal{L}^{\mathcal{R},\mathcal{U}} \to \mathbb{N}$, in which the weights of the links $w(r_h, u_j) = w_{hj}$ are the numbers of students from region r_h who enroll in the university u_j. Note that the projection gives rise to new networks in which the links could have a different meaning with respect to the original ones. In the toy example, the three universities are connected to the two regions with weights equal to 3, 2, and 1, respectively (Fig. 2, middle panel). Note that the sum of the weights is equal to the total number of students. If we further project on the universities, we obtain a one-mode network of regions ${}_{(\mathcal{S},\mathcal{U})}\mathcal{N}^{\mathcal{R}} = (\mathcal{R}, {}_{(\mathcal{S},\mathcal{U})}\mathcal{L}^{\mathcal{R}}, \mathcal{W})$, with ${}_{\mathcal{S},\mathcal{U}}\mathcal{L}^{\mathcal{R}} \subseteq \mathcal{R} \times \mathcal{R}$ and the function $w : {}_{\mathcal{S},\mathcal{U}}\mathcal{L}^{\mathcal{R}} \to \mathbb{N}$, in which the weights of the links $w(r_h, r_{h'}) = w_{hh'}$ are the numbers of students residing region r_h enrolled in a university located in region $r_{h'}$. In the toy example, there are two students residing in region r_2 who enroll in a university located in region r_1 (namely, s_5 and s_6), while the other students are enrolled in a university located in the same region in which they reside (Fig. 2, bottom panel).

If we consider the multipartite network with four modes, $\mathcal{Q} = (\mathcal{S}, \mathcal{P}, \mathcal{U}, \mathcal{E}, \mathcal{L}^{\mathcal{S},\mathcal{P},\mathcal{U},\mathcal{E}})$, given by students, provinces, universities, and educational programs and its projection on students, we get a weighted tripartite 3-uniform hypergraph ${}_{\mathcal{S}}\mathcal{H}\mathcal{T}^{\mathcal{P}\mathcal{U}\mathcal{E}} = (\mathcal{P}, \mathcal{U}, \mathcal{E}, {}_{\mathcal{S}}\mathcal{L}^{\mathcal{P}\mathcal{U}\mathcal{E}}, \mathcal{W})$ with ${}_{\mathcal{S}}\mathcal{L}^{\mathcal{P}\mathcal{U}\mathcal{E}} \subseteq \mathcal{P} \times \mathcal{U} \times \mathcal{E}$, the collection of hyperedges, with generic term (p_c, u_j, e_k), which is the link connecting the c-th province, the j-th university, and the k-th educational program. $\mathcal{W}$ is the set of weights obtained with the function $w : {}_{\mathcal{S}}\mathcal{L}^{\mathcal{P}\mathcal{U}\mathcal{E}} \to \mathbb{N}$, and $w(p_c, u_j, e_k) = w_{cjk}$ is the number of students moving from province p_c to university u_j to attend a specific educational program e_k.

Starting from the toy example in Fig. 2, three educational programs are added to obtain the quadripartite network $\mathcal{Q}$ portrayed in Fig. 3 (upper panel). Projecting the 3-uniform hypergraph represented on the students Fig. 3 (bottom panel) is obtained. The hypergraph has four hyperedges: (u_1, e_1, p_1) with $w_{111} = 2$, (u_1, e_2, p_1) with $w_{121} = 1$, (u_2, e_2, p_2) with $w_{222} = 2$, and (u_3, e_3, p_2) with $w_{332} = 1$. None of the other possible links are present.

Finally, considering the conditional projection for the student mobility data, we are interested in describing the mobility trajectories of movers, that is, students who enroll in universities outside their region of residence, with respect to stayers, that is, students who enroll in universities within their region of residence. We start from the quadripartite $\mathcal{Q}$ and add information on the regions; that is, we add two nodes, r_1 and r_2, and the relative links (Fig. 4, top panel). Projecting on the students and conditioning on the regions, we can define two hypergraphs, one for the movers, ${}_{\mathcal{S}}\mathcal{H}\mathcal{T}^{\mathcal{P}\mathcal{U}\mathcal{E}}_{|\mathcal{B}^{\mathcal{U}\mathcal{R}}}$, and one for the stayers, ${}_{\mathcal{S}}\mathcal{H}\mathcal{T}^{\mathcal{P}\mathcal{U}\mathcal{E}}_{|\mathcal{B}^{\mathcal{U}\mathcal{R}}}$. The hypergraph of movers, ${}_{\mathcal{S}}\mathcal{H}\mathcal{T}^{\mathcal{P}\mathcal{U}\mathcal{E}}_{|\overline{\mathcal{B}^{\mathcal{U}\mathcal{R}}}}$, is defined as follows:

$$(p_c, u_j, e_k)_{|\overline{\mathcal{B}^{\mathcal{U}\mathcal{R}}}} \in_{\mathcal{S}} \mathcal{L}^{\mathcal{P}\mathcal{U}\mathcal{E}}_{|\overline{\mathcal{B}^{\mathcal{U}\mathcal{R}}}} \iff (s_i, p_c) \in \mathcal{L}^{\mathcal{S},\mathcal{P}} \wedge (s_i, u_j) \in \mathcal{L}^{\mathcal{S},\mathcal{U}} \wedge (s_i, e_k) \in \mathcal{L}^{\mathcal{S},\mathcal{E}}$$

$$\wedge (s_i, r_h) \in \mathcal{L}^{\mathcal{S},\mathcal{R}} \wedge (r_h, u_j) \notin \mathcal{L}^{\mathcal{R},\mathcal{U}}$$

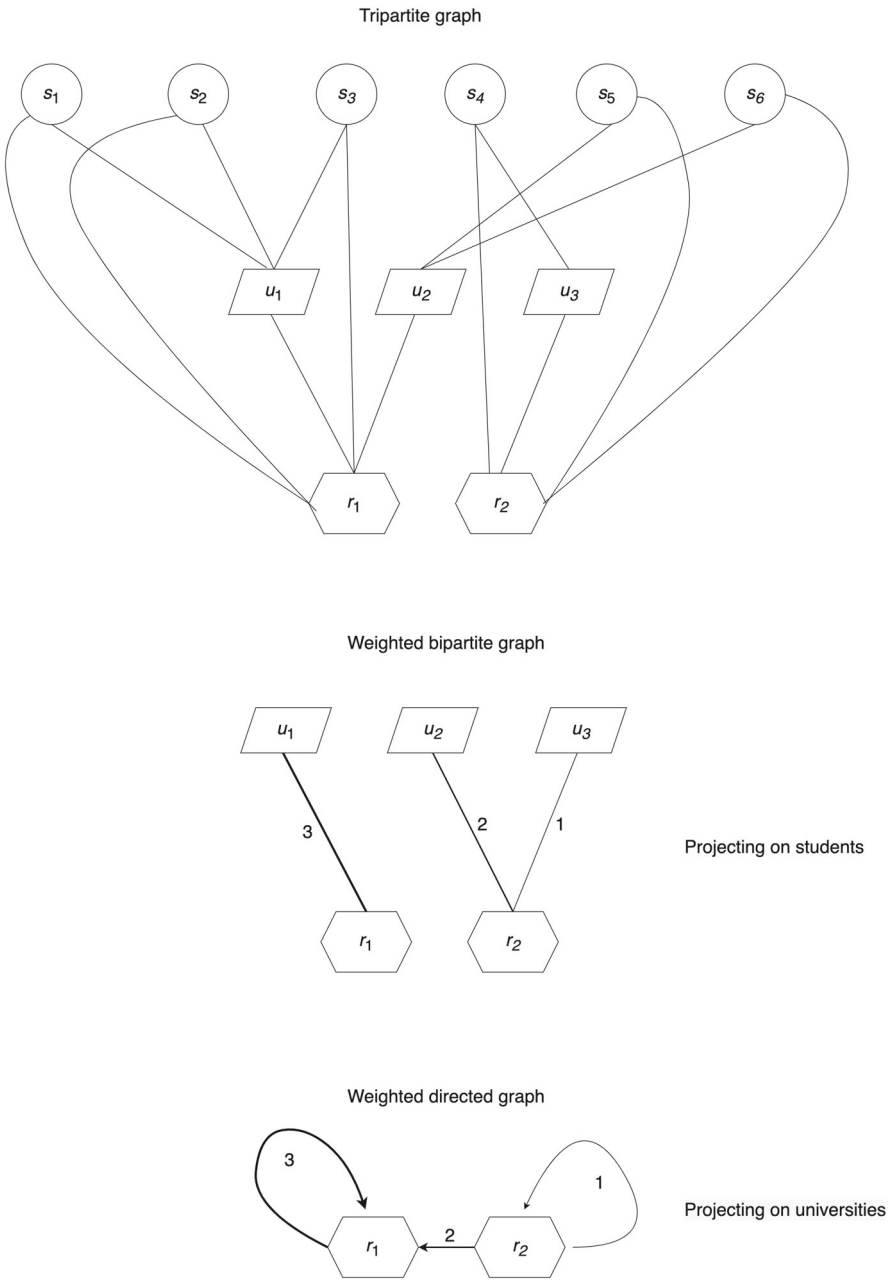

Fig. 2 Toy example: Tripartite graph visualization and its bipartite weighted projections on students and universities

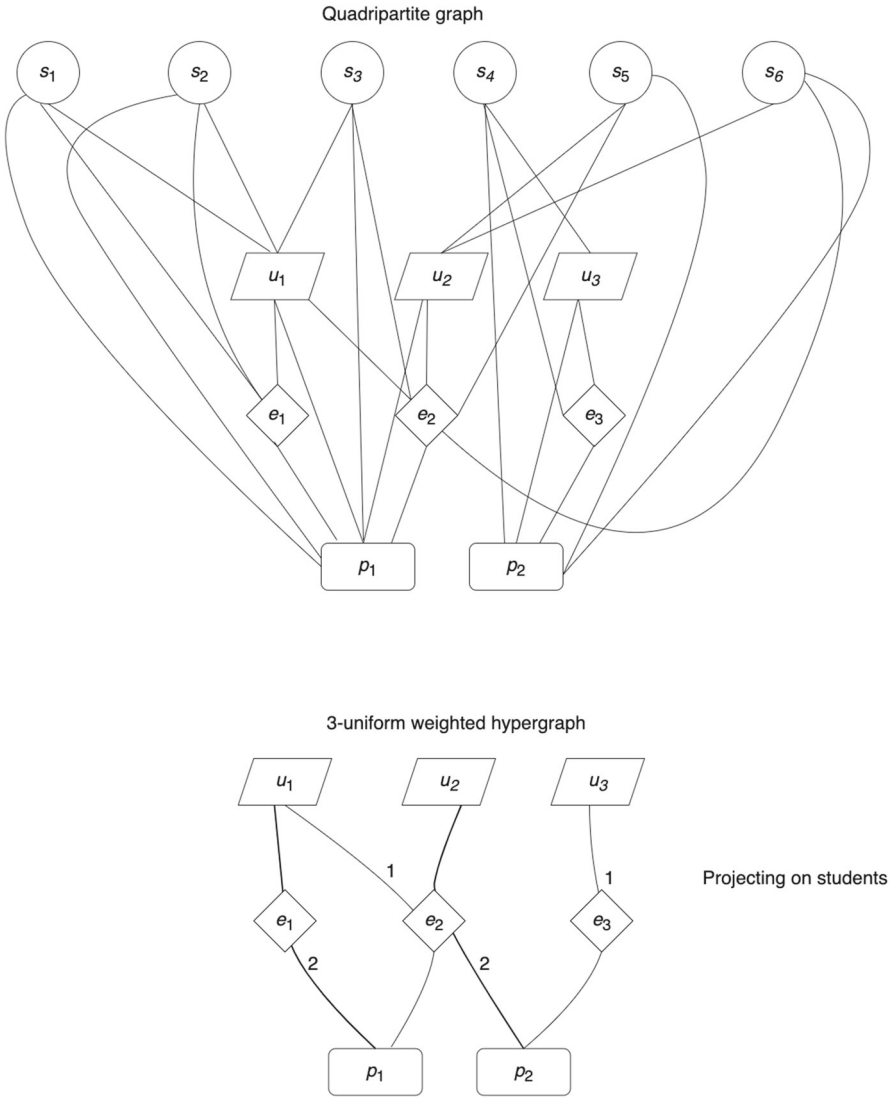

Fig. 3 Toy example: Multipartite graph visualization with four modes and the corresponding projection in a hypergraph

The hypergraph of stayers, ${}_{\mathcal{S}}\mathcal{HT}^{\mathcal{PUE}}_{|\mathcal{B}^{\mathcal{UR}}}$, is defined as follows:

$$(p_c, u_j, e_k)_{|\mathcal{B}^{\mathcal{UR}}} \in_{\mathcal{S}} \mathcal{L}^{\mathcal{PUE}}_{|\mathcal{B}^{\mathcal{UR}}} \iff \wedge(s_i, p_c) \in \mathcal{L}^{\mathcal{S},\mathcal{P}} \wedge (s_i, u_j) \in \mathcal{L}^{\mathcal{S},\mathcal{U}}$$

$$\wedge(s_i, e_k) \in \mathcal{L}^{\mathcal{S},\mathcal{E}} \wedge (s_i, r_h) \in \mathcal{L}^{\mathcal{S},\mathcal{R}} \wedge (r_h, u_j) \in \mathcal{L}^{\mathcal{R},\mathcal{U}}.$$

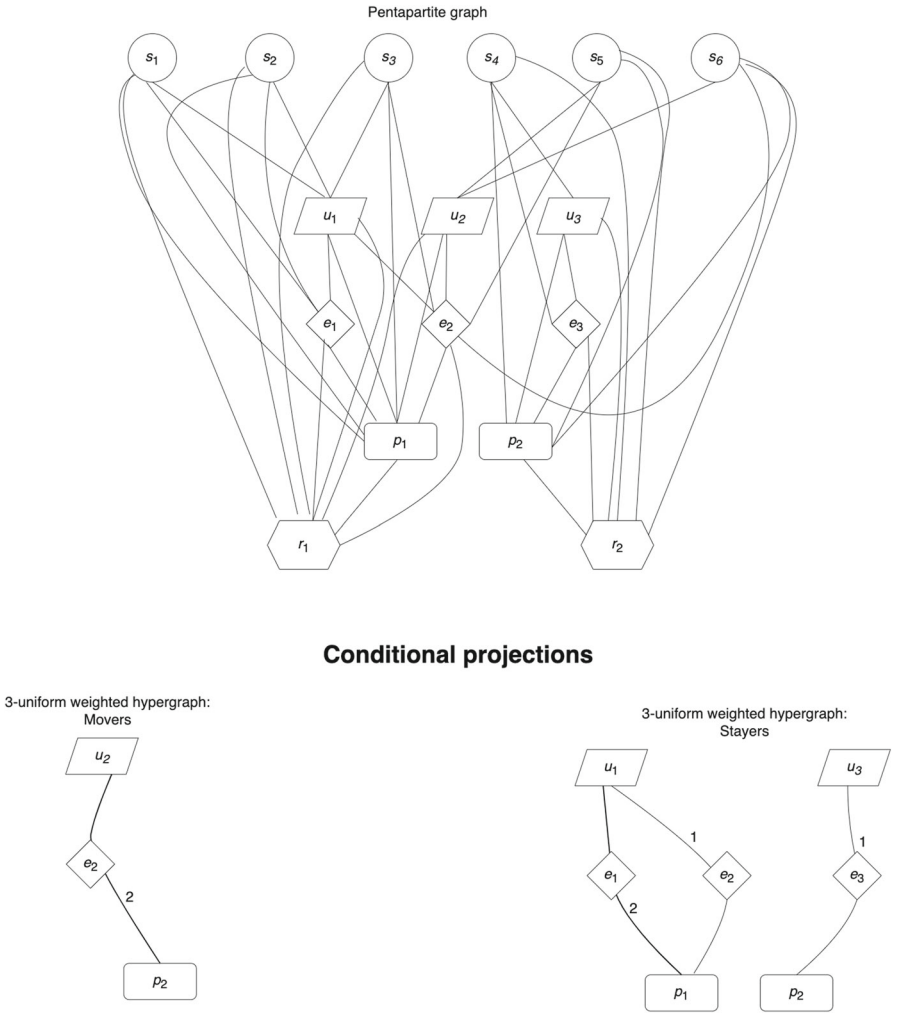

Fig. 4 Toy example: Multipartite graph visualization with four modes and its tripartite weighted projections on students and universities

In the toy example (Fig. 4, upper panel), the two conditional projections give rise to two hypergraphs. On the left side (Fig. 4, bottom panel), the hypergraph represents the flow of two students (s_5 and s_6) who reside in region r_2 who enroll in university u_2 located in province p_1 in region r_1. On the right side (Fig. 4, bottom panel) is the hypergraph of the stayers.

4 Data processing

To handle the complex network structures described above, we proposed a three-step procedure for the data processing phase: 1) *data simplification*, 2) *data normalization*, and 3) *data filtering*.

4.1 Data simplification

In statistical terms, array $\mathbb{A}$ can be interpreted as a three-way contingency table, and then the statistical techniques for evaluating the association among variables, that is, the modes, can be exploited (Agresti 2007). Because a three-way contingency table is a cross-classification of observations by the levels of three categorical variables, we define a network structure where the sets of nodes are the levels of the categorical variables. Specifically, if two modes are jointly associated (or independent) on the third mode, we assume that the tripartite hypergraph can be logically simplified into a bipartite graph.

Given the weighted tripartite 3-uniform hypergraph ${}_4\mathcal{HT}^{123} = (\mathcal{V}^1, \mathcal{V}^2, \mathcal{V}^3, {}_4\mathcal{L}^{123}, \mathcal{W})$ defined in the previous section, we can join a pair of sets of nodes, $\mathcal{V}^2$ and $\mathcal{V}^3$, in a new set $\widehat{\mathcal{V}^2\mathcal{V}^3} = \mathcal{V}^2 \times \mathcal{V}^3$. The hypergraph can then be simplified in a bipartite network ${}_4\mathcal{B}^{1\widehat{23}} = (\mathcal{V}^1, \widehat{\mathcal{V}^2\mathcal{V}^3}, {}_4\mathcal{L}^{1,\widehat{23}}, \mathcal{W}^*)$, with ${}_4\mathcal{L}^{1,\widehat{23}} \subseteq \mathcal{V}^1 \times \widehat{\mathcal{V}^2\mathcal{V}^3}$. The new edges $(v_i^1, (v_j^2, v_k^3))$ connect node v_i^1 with the pair (v_j^2, v_k^3), and the weights $\mathcal{W}^*$ are the same as in the hypergraph; that is, $w^*_{i(j,k)} = w_{ijk}$. Considering the matrix formulation, note that the elements contained in the three-way array $\mathbb{A}$ are preserved, but are reorganized in rectangular matrix $\mathbf{A}$ of I rows and $(J \times K)$ columns, in the so-called flag matrix.

Considering the student mobility case study, for instance, we could join the pair of nodes in $\mathcal{U}$ and in $\mathcal{E}$, and then we handle the relationships between these *dyads* and the nodes in $\mathcal{P}$. Following this assumption, the sets of nodes in $\mathcal{U}$ and $\mathcal{E}$ are put together in a set of joint nodes, namely, $\widehat{\mathcal{UE}}$. The hypergraph ${}_{\mathcal{S}}\mathcal{HT}^{\mathcal{PUE}}_{|\mathcal{B}^{\mathcal{UR}}}$ can be represented as the bipartite network ${}_{\mathcal{S}}\mathcal{B}^{\mathcal{P}\widehat{\mathcal{UE}}}_{|\mathcal{B}^{\mathcal{UR}}}$. The set of hyperedges ${}_{\mathcal{S}}\mathcal{L}^{\mathcal{PUE}}_{|\mathcal{B}^{\mathcal{UR}}}$ is thus simplified into a set of edges ${}_{\mathcal{S}}\mathcal{L}^{\mathcal{P},\widehat{\mathcal{UE}}}_{|\mathcal{B}^{\mathcal{UR}}} \subseteq \mathcal{P} \times \widehat{\mathcal{UE}}$. The new edges $(p_c, (u_j; e_k))$ connect province p_i with educational program e_k running in given university u_j. The weights are the same as in the hypergraph, that is, $w^*_{c(j,k)} = w_{cjk}$.

Starting from the hypergraph visualized in Fig. 5 (upper panel), the four hyperedges are simplified in four edges. For example, the hyperedges (u_1, e_1, p_1) with weight equal to 2 end up in the edge $((u_1, e_1), p_1)$ with the same weight. Conditional projections could be simplified in the same way (Fig. 5, bottom panel).

4.2 Data normalization and filtering

Several procedures have been introduced to normalize weighted network data (Primerano et al. 2021; Giordano et al. 2019; Giordano and Primerano 2018; Batagelj and Cerinšek 2013; Slater 2009). Here we propose a normalization procedure adapted for handling bipartite weighted networks. The chosen method should be derived from some aspects of the phenomenon under investigation. For example, as in the classical

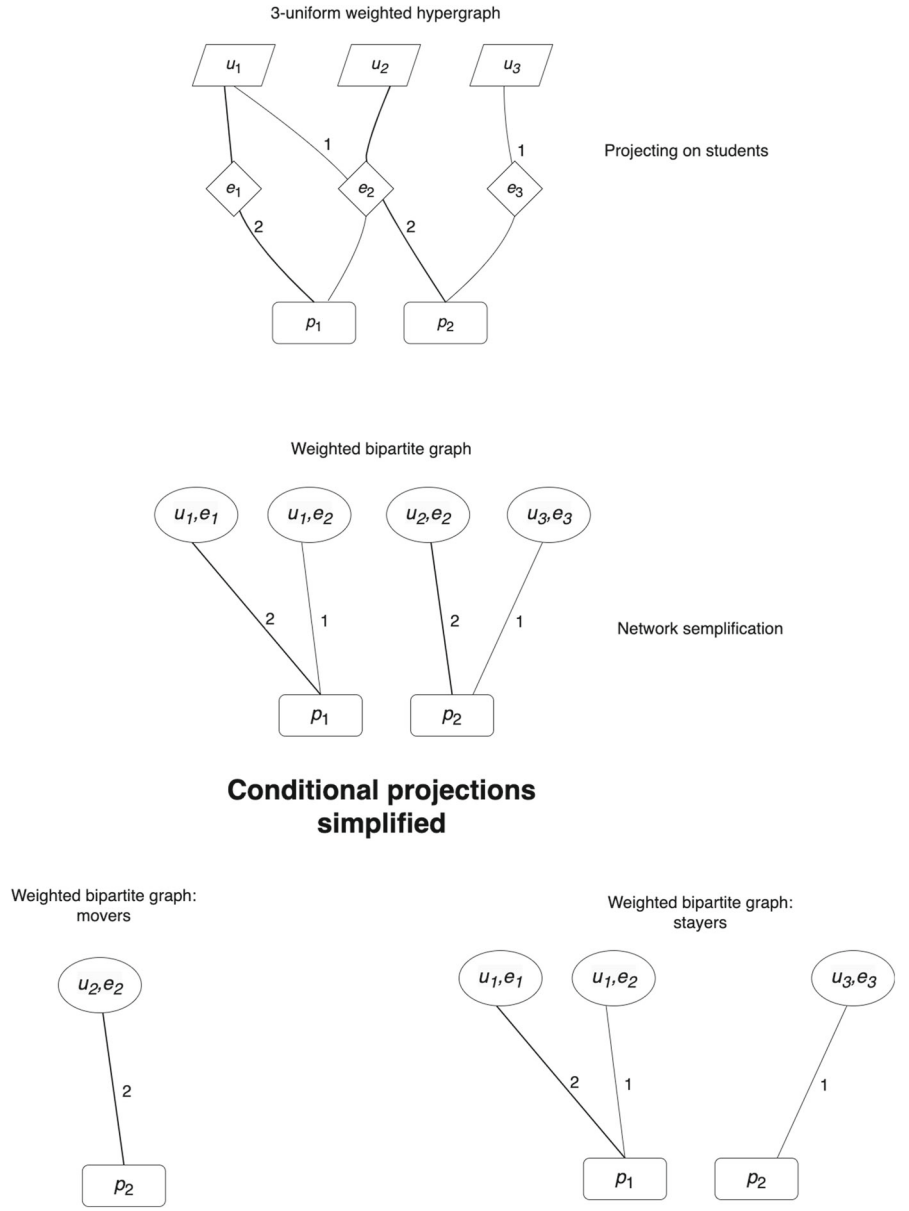

Fig. 5 Toy example: Conditional projection of the resulting weighted bipartite graph in Fig. 3

contingency table, rows (or columns) conditional distributions can be derived because the interest is in making comparisons among the row or column categories.

Another possibility is to derive normalization weights from secondary data available on the categories of interest, for instance, the gross number of inhabitants of each province, the number of young people aged 19–25, the age group in which people usually enroll in university, or the number of collaborations in a project. An important concept to be considered is that a weighting scheme always has implications for the profile similarity between pairs of nodes. Choosing different weighting schemes affects the way relationships are perceived and similarity/dissimilarity measures are computed. This, in turn, has direct effects on community detection purposes.

Recently, researchers have discussed two possible approaches. The first approach is based on the opportunity to detect nodes' similarity at the micro-level, that is, when we are interested in a weighting procedure that preserves local (micro) network properties (such as the neighboring proximities). The second approach relies on macroscopic properties of the network, preserving topological characteristics, such as the scale-free properties (Dey et al. 2020).

More complex schemes can be derived from the necessity to obtain complex normalization criteria. For instance, in the case of adjacency matrices, a dual (by rows and columns) normalization could be envisaged. Following Slater (2009) and Barthélemy and Suesse (2018), Primerano et al. (2021) apply a multidimensional iterative proportional fitting procedure (MIPFP) to student mobility data. This procedure starts from the original adjacency matrix (e.g., *universities*-by-*universities*), where the row and column totals represent the overall outgoing and incoming students for a given university, respectively, and then performs a reshaping algorithm. This procedure defines a value for each edge, ranging from 0 to 1, accounting for nodes' attractiveness (columns marginal) and nodes' repulsion (rows marginal). Thus, in the end, each weight is a value that takes into account the overall number of incoming and outgoing edges. Namely, the edges' weights inversely depend on the number of students. Higher weights are associated with universities having a small number of flows; lower weights are attached to universities characterized by a relatively large number of outgoing and incoming students. In a different context (travelers' activity on social media moving from one tourist place to another), Giordano et al. (2021) propose a normalization scheme that can define an asymmetric "performativity indicator" related to the number of actions performed by individuals in a destination, given the number of actions performed in the place of origin.

The need for the normalization process is mainly due to high unbalancing among the different categories present in each network mode. For example, in the intranational mobility data, the flows of students vary considerably, depending on provinces' size or contiguous territories gravitating around universities, universities' size, and bureaucratic constraints. In the case of scientific collaboration projects, such an unbalancing can be due to the different numbers of joint projects in which the institutions are involved.

Beyond the data normalization process, network filtering methods are generally based on the application of a threshold to the link's weight. Several methods have been proposed to filter the network's significant mobility patterns within the backbone extraction framework. A *bistochastic filter* approach (Foti et al. 2011) was proposed

to examine migration flows even if it seems that this method can alter the network structure due to the edge weight matrix manipulation. Similarly, Yongwan and Griffith (2011) reported the application of an eigenvector spatial filtering procedure that requires non-trivial data manipulation.

4.3 A proposal for network data normalization and filtering

In addition to the several approaches proposed in the literature, we propose a model-based approach through the standardized residuals of a log-linear model for contingency tables (Agresti 2007). This normalization procedure is applied to the **A** matrix mentioned in Sect. 4.1.[3] Recalling that an adjacency matrix can be seen as a contingency table, a log-linear model of independence for a contingency table can be estimated as follows:

$$\log(\mu_{c(j,k)}) = \lambda + \lambda_c^{\mathcal{P}} + \lambda_{(j,k)}^{\mathcal{UE}}, \tag{1}$$

where $\mu_{c(j,k)}$ is the expected link weight for the pair $c(j,k)$; and $\lambda_c^{\mathcal{P}}$ and $\lambda_{(j,k)}^{\mathcal{UE}}$ are the row and column effects for the **A** matrix, respectively. Following this model-based approach, the expected link weight $\mu_{c(j,k)}$ represents the co-occurrences for the pair $c(j,k)$ if the number of co-occurrences is proportional to the row and column effects of **A**. Under this assumption, $\lambda_c^{\mathcal{P}}$ and $\lambda_{(j,k)}^{\mathcal{UE}}$ can be interpreted as the repulsiveness effect of c and the attractiveness effect of (j,k), respectively. Significant deviations from $\mu_{c(j,k)}$ can suggest an extra-flow *w.r.t.* a null hypothesis of random co-occurrences. To evaluate these deviations, standardized residuals for contingency tables are used:

$$res_{c(j,k)} = \frac{w^*_{c(j,k)} - \mu_{c(j,k)}}{\sqrt{\mu_{c(j,k)}(1 - w_{c+})(1 - w_{+(j,k)})}} \tag{2}$$

where $w^*_{c(j,k)}$ are the observed weights of the **A** matrix for the pair $c(j,k)$; $\mu_{c(j,k)}$ are the estimated weights for the pair $c(j,k)$ under the model reported in Eq. (1); and w_{c+} and $w_{+(j,k)}$ are the marginal fraction of the $c-th$ row and the $(j,k)-th$ column of the **A** matrix, respectively.

The usefulness of the residuals in Eq. (2) is twofold. On the one hand, they allow us to normalize the weights $w^*_{c(j,k)}$ of the **A** matrix, taking into account the size of the source node, the size of the target node, and the whole size of the system. On the other hand, by construction, $res_{c(j,k)} \sim N(0,1)$, and this allows us to filter the network using a statistical threshold value. Significant deviations from the expected value $\mu_{c(j,k)}$ estimated in Eq. (1) can be interpreted as an over-expression of the co-occurrences $w^*_{c(j,k)}$ with respect to a system in which the link weights are proportional to the marginals of the **A** matrix. In statistical terms, this approach is able to discover over-expression in the network with respect to a null hypothesis of random co-occurrences. Residuals with values exceeding 2 (when the **A** matrix is small) or 3 (when the **A** matrix is large) indicates a violation of the null hypothesis. For example, in the case

[3] The normalization and filtering procedure notation outlined here specifically applies to the students' mobility data example; however, its applicability extends to different contexts and applications.

Table 1 Data processing of tripartite and bipartite network structures per student cohort

Year	Tripartite		Bipartite		Filtered bipartite	
	N. nodes	N. links	N. nodes	N. links	N. nodes	N. links
2008–2019	195	17,734	567	8, 867	122	116
2011–2012	195	17,204	551	8, 602	140	135
2014–2015	196	18,586	554	9, 293	140	127
2017–2018	196	20,886	562	10, 443	154	172

of a large $\mathbf{A}$, setting a threshold of 3 is roughly equivalent to a right-tail test with $\alpha = 0.01$, and this enables us to select links with weights exceeding those estimated under the null hypothesis.

Unlike a backbone extraction with an unconditional threshold in which the network structure strongly depends on the chosen threshold (Zachary 2014; Coronnello et al. 2009), the proposed approach preserves structural and multi-scale features of the network, controlling for the size of the source node ($n_{c.}$), the size of the target node ($n_{.j}$), and the whole size of the system (N). Our proposal is an unsupervised and data-driven method for evaluating statistically significant links between nodes. It aims to reveal preferential patterns, that is, patterns that show a significant deviation in the observed co-occurrences from the null hypothesis of random co-occurrences.

5 Real-data processing phase

Returning to the real-life examples, we exploit the usefulness of the proposed three-step procedure in the two datasets described in Sect. 2.

Intranational student mobility. Considering the MOBYSU.IT database, data on students enrolled in Italian universities in the academic years 2008-2009, 2011-2012, 2014-2015, and 2017-2018 are extracted. The procedure described in Sects. 3 and 4 is applied to the five modes of network data structures in which *students*, *regions*, *provinces*, *universities*, and *educational programs* are linked through a set of affiliation networks generated by different kinds of relations. This complex network data structure is analyzed with the final aim of revealing the presence of communities, mixing up provinces of origin, universities of destination, and specific degree programs that explain mobility choices.
Table 1 shows the changes in the student mobility network data structures before and after the analytic strategy adopted to simplify, normalize, and filter relevant information.

Scientific collaboration networks. By considering the CORDIS data, including details of R&D projects funded by EU-FP7, a multimode network is derived in which *countries*, *institutions*, *projects*, *framework programs*, and *fields* define the different modes. Specifically, we gathered information on collaboration among European institutions from 2007 to 2018. From these data, we select information about the countries of the institutions, the institutions' ID, the project ID, the framework programs, and the

Table 2 Data processing of tripartite and bipartite network structures per R&D collaboration in scientific fields

Thematic areas	Tripartite		Bipartite		Filtered bipartite	
	N. nodes	N. links	N. nodes	N. links	N. nodes	N. links
Engineering and Natural Sciences	4,897	20,854	4,875	10,427	322	333
Social Sciences and Humanities	1,128	3742	1,107	1,871	36	30
Medical Health and Agricultural	1,453	6,076	1,433	3,038	123	121

scientific field. Each feature is considered a mode of the network structure, where the edges' weights are the number of earned projects between pairs of institutions. The institution's ID acts as the primary key projected to obtain different networks. For instance, we consider the countries that are involved in different projects under several grant programs; in doing so, we could condition the networks based on the thematic field. This data structure is compatible with the one described in the simplification procedure. The two modes of the tripartite network, $Project$ and $Framework$, are logically nested and yield a new conditional bipartite network. It allows us to analyze the countries' and the institutions' scientific collaboration in specific framework programs and research projects with respect to a specific thematic field. The normalization and filtering procedures obtain the links characterized by the over-expression of collaboration between countries and $(Project\ ID \times Framework\ Programs)$. Table 2 shows the changes in scientific collaboration data structures before and after the analytic strategy adopted to simplify, normalize, and filter relevant information for the three scientific fields.

6 Network results

The derived data structures are explored with network analysis tools. Specifically, to reveal the presence of communities in the filtered networks, we use the Infomap community detection algorithm (Blöcker and Rosvall 2020; Edler et al. 2017). The algorithm is more suitable for examining the flows' patterns in network structures instead of modularity optimization that looks only at topological aspects of the network (Blondel et al. 2008; Newman and Girvan 2004). To analyze mobility and project data, flow-based approaches are likely to identify the most important features, revealing communities characterized by similar patterns.

The rationale of Infomap (*map equation*) takes advantage of the duality between finding communities and minimizing the length (*codelength*) of a random walker's movement on a network. The partition with the shortest path length best captures the community structure in the bipartite data. Formally, the algorithm defines a module partition $\mathbf{M}$ of n nodes into m modules such that each node is assigned to one and only one module. The algorithm looks for the best $\mathbf{M}$ partition that minimizes the expected *codelength*, $L(M)$, of a random walker, given by the following map equation (Edler et al. 2017):

$$L(M) = q_{\curvearrowright} H(\mathscr{Q}) + \sum_{i=1}^{m} p^i_{\circlearrowright} H(\mathscr{P}^i). \tag{3}$$

In Equation (3), $q_{\curvearrowright} H(\mathscr{Q})$ is the entropy of the movement between modules weighed for the probability that the random walker switches modules on any given step ($q_{\curvearrowright}$), and $\sum_{i=1}^{m} p^i_{\circlearrowright} H(\mathscr{P}^i)$ represents the entropy of movements within modules weighed for the fraction of within-module movements that occur in module i, plus the probability of exiting module i ($p^i_{\circlearrowright}$), such that $\sum_{i=1}^{m} p^i_{\circlearrowright} = 1 + q_{\curvearrowright}$.

6.1 Student mobility communities

By adopting the Infomap community detection algorithm on filtered bipartite weighted networks, for intranational mobility data, we obtained good relative codelength savings for student cohorts and a sort of stabilization phenomenon over time in terms of a reduction in the number of clusters in the mobility trajectories showing the central position of the most prestigious universities located in the Center-North (Table 3). For the sake of simplicity, the cluster solutions of the first and the last student cohorts are examined in detail.

- The **22 clusters** for the 2008-2009 academic year show two main groups of attractive universities in North Italy, highlighting the main trajectories in the Italian student mobility flows (South-to-North and South-to-Center), and several small groups (Fig. 6, upper panel). The biggest group (23 units) contains the universities of Bologna and Padua attracting students for *Social Sciences*, Pisa for *Engineering*, and Ca' Foscari Venice for *Arts and Humanities*. The second group (21 units) highlights the attractiveness of La Sapienza Rome university, mainly for *Arts and Humanities*, Polytechnic of Turin for *Engineering*, and Cattolica Rome and Milan for *Business, Administration* and *Law*. Several small clusters show, on the one hand, the authority role played by public and private Northern universities and on the other hand, the internal and external student flows among provinces and universities due to the geographical proximity of some Italian regions.
- The **15 clusters** for the 2017-2018 academic year is characterized by four main clusters of provinces and universities-educational fields highlighting the dichotomy between scientific and humanistic fields attracting South-to-North and North-to-North student flows (Fig. 6, bottom panel). The biggest group (52 units) contains, on the one hand, the universities of Ferrara, Marche, and Modena-Reggio attracting students for *Engineering* and *Natural Sciences* degree programs; on the other hand, Chieti-Pescara and Trento [the Italian Adriatic coast route] for *Social Sciences*. The second largest group (28 units) includes Bocconi, Cattolica and Florence for *Business, Administration* and *Law*; Polytechnic of Milan and Pisa for *Engineering*.

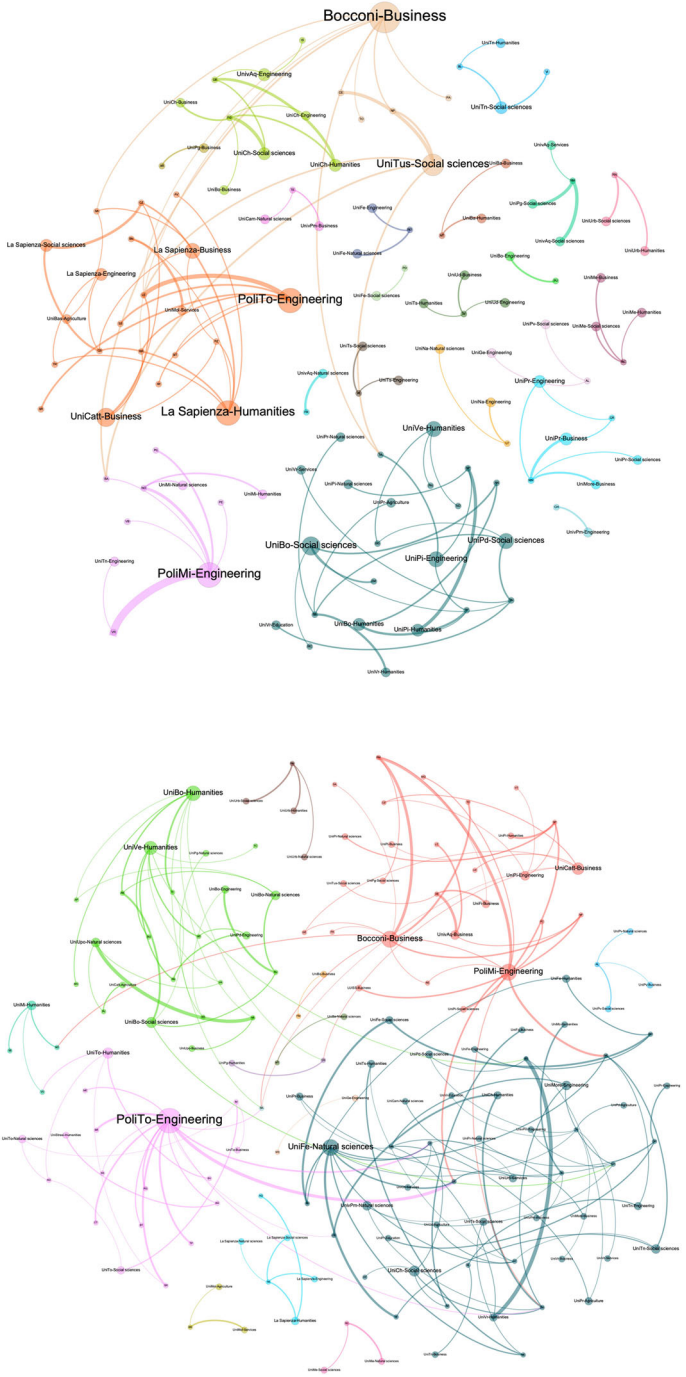

Fig. 6 Communities for student cohorts for the 2008-2009 academic year (upper panel) and the 2017-2018 academic year (bottom panel)

Table 3 Infomap community detection results per student mobility data cohort

Cohort	Clusters	Codelength	Relative codelength savings
2008–2009	22	0.81	85.41%
2011–2012	18	0.75	86.78%
2014–2015	21	0.76	86.19%
2017–2018	15	1.19	79.15%

Table 4 Infomap community detection results per field of the EU-FP7 data

Field	Clusters	Codelength	Relative codelength savings
Engineering and Natural Sciences	24	1.29	62.45%
Medical Health and Agricultural	15	4.55	30.35%
Social Sciences and Humanities	8	2.51	45.00%

6.2 Collaboration network communities

The results of the Infomap community detection algorithm on CORDIS data applied to the three scientific fields show good relative codelength savings for Engineering and Natural Sciences, whereas the values are moderate for the other two fields. The number of communities is higher for the Natural Sciences and Medical Health and Agricultural fields than for Social Sciences and Humanities field, in line with the different amounts of funds devoted to R&D projects for each field. Looking at the communities reported for the three fields, we notice the following.

- The **24 clusters** for countries and framework programs, conditioned to the Engineering and Natural Science field (Fig. 7, upper panel on the left) are described by six big groups around the main active countries, Germany (DE), the UK, France (FR), Spain (ES), Italy (IT), and the Netherlands (NL). The biggest group (72 units), including Germany and Norway (NO), is described by participation in projects related mainly to ICT and *Nanosciences, Nanotechnologies, Materials and New Production Technologies* (NMP).
- The **15 clusters** for countries and framework programs, conditioned to the Medical, Health, and Agricultural field (Fig. 7, upper panel on the right), are characterized by two main groups with around 20 units. The biggest group incorporates the UK and Sweden (SE) and projects related to Joint Technology Initiatives (JTI) and Health.
- The **8 clusters** for countries and framework programs, conditioned on the Social sciences and Humanities field (Fig. 7, bottom panel), present a low number of units including only one country per community and participation in multifaceted projects.

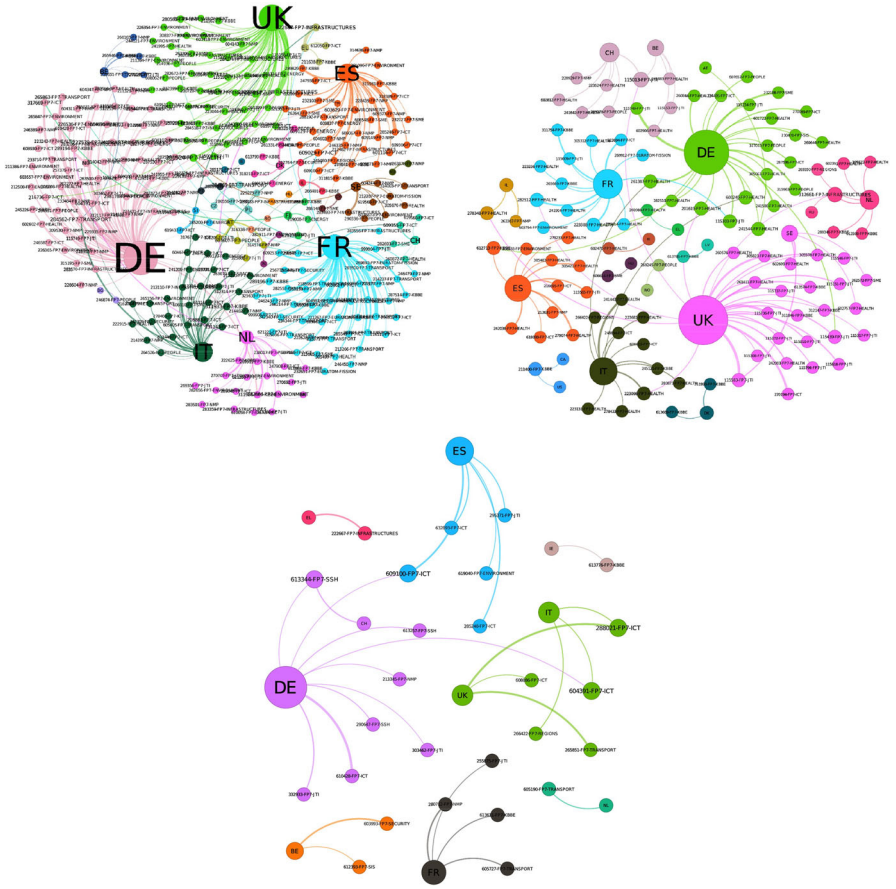

Fig. 7 Communities for countries and framework programs, conditioned to the Engineering and Natural Science field (upper panel, left), the Medical, Health, and Agricultural field (upper panel, right), and the Social Sciences and Humanities field (bottom panel)

7 Discussion and conclusions

Starting from complex network data structures, the present contribution discusses an original analytic strategy for handling multipartite networks. Simplification, normalization, and filtering procedures are adopted to analyze the complexity of these networks with a model-based approach to extract statistically significant links in the derived bipartite networks. Formal definition as well as the usefulness of the three-step procedure for data processing are illustrated by considering two real-life data examples regarding intranational student mobility flows and collaboration in scientific projects.

A flow-based community detection algorithm is adopted to partition the simplified network data structures. The partitioning solutions mixed up the different types of modes, enriching the interpretation of the examples under analysis. First, the procedure brings to light the main characteristics of Italian student mobility flows by

confirming the preferential attractiveness routes of Northern universities but also the dichotomy between scientific and humanist fields.[4] Second, the clustering solution for collaboration in EU-FP7 projects shows that some European countries play a central role in the participation in projects in specific thematic fields.

The three-step procedure is able to retain relevant information of complex data structures, highlighting the main network features. In this way, the procedure also provides feedback for policy makers to manage phenomena spanning a very wide spectrum of fields. Specifically, regarding the findings of the two real datasets, it offers suggestions for a more comprehensive assessment of universities' performance in achieving their core missions to attract students and in activating collaboration with other institutions within funded EU-FP7 projects. Finally, the proposed approach could be generalized to freely available and transparent data, such as CORDIS.

As future lines of research, a comparison with alternative approaches can be considered to deal with multiway weighted network data (Batagelj and Cerinšek 2013; Batagelj et al. 2007) in order to identify important and meaningful groups in these large and complex networks by discussing also other community detection algorithms proposed for bipartite weighted networks (Beckett 2016).

Acknowledgements The paper was supported by the Italian Ministerial grant PRIN 2017 "From high school to job placement: micro-data life course analysis of university student mobility and its impact on the Italian North–South divide," n. 2017HBTK5P-CUPB78D19000180001.

References

Agresti A (2007) An introduction to categorical data analysis. John Wiley & Sons, Hoboken, New Jersey

Amoroso S, Coad A, Grassano N (2018) European r&d networks: a snapshot from the 7th eu framework programme. Econ Innov New Technol 27(5–6):404–419

Barthélemy J, Suesse T (2018) mipfp: an r package for multidimensional array fitting and simulating multivariate Bernoulli distributions. J Stat Softw 86:1–20

Batagelj V, Cerinšek M (2013) On bibliographic networks. Scientometrics 96(3):845–864

Batagelj V, Ferligoj A, Doreian P (2007) Indirect Blockmodeling of 3-Way Networks. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 151–159

Beckett SJ (2016) Improved community detection in weighted bipartite networks. Royal Soc Open Sci 3(1):140–536

Blöcker C, Rosvall M (2020) Mapping flows on bipartite networks. Phys Rev E 102(5):052–305

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

[4] The findings for student mobility are in line with the main results of the national research project grant PRIN 2017, "From high school to job placement: micro-data life course analysis of university student mobility and its impact on the Italian North–South divide." For details: https://www.unipa.it/persone/docenti/a/massimo.attanasio/prin-2017/.

Boccaletti S, Bianconi G, Criado R et al (2014) The structure and dynamics of multilayer networks. Phys Rep 544(1):1–122
Borgatti SP, Everett MG (1992) Regular blockmodels of multiway, multimode matrices. Social networks 14(1–2):91–120
Bródka P, Chmiel A, Magnani M et al (2018) Quantifying layer similarity in multiplex networks: a systematic study. Royal Soc Open Sci 5(8):171747
Columbu S, Porcu M, Primerano I et al (2021) Geography of Italian student mobility: a network analysis approach. Socioecon Plann Sci 73(100):918
Columbu S, Porcu M, Primerano I et al (2022) Correction to: analysing the determinants of Italian university student mobility pathways. Genus 78(1):1–1
Coronnello C, Tumminello M, Micciche S et al (2009) Networks in biological systems: an investigation of the gene ontology as an evolving network. Il nuovo cimento C 32(2):157–160
Dey P, Goel K, Agrawal R (2020) P-simrank: Extending simrank to scale-free bipartite networks. Proc Web Conf 2020:3084–3090
Dickison ME, Magnani M, Rossi L (2016) Multilayer Social Networks. Cambridge University Press, Cambridge
Dotti NF, Fratesi U, Lenzi C et al (2014) Local labour market conditions and the spatial mobility of science and technology university students: evidence from italy. Rev Regional Res: Jahrbuch für Regionalwissenschaft 34(2):119–137
Edler D, Bohlin L, Rosvall M (2017) Mapping higher-order network flows in memory and multilayer networks with infomap. Algorithms 10(4):112
European Commission (2022) Community Research and Development Information Service. Retrieved from Cordis website: http://cordis.europa.eu
Everett MG, Borgatti SP (2019) Partitioning multimode networks. Advances in network clustering and blockmodeling pp 251–265
Fararo TJ, Doreian P (1984) Tripartite structural analysis: generalizing the breiger-wilson formalism. Social Netw 6(2):141–175
Findlay A, Packwood H, McCollum D et al (2018) Fees, flows and imaginaries: exploring the destination choices arising from intra-national student mobility. Glob Soc Educ 16(2):162–175
Foti NJ, Hughes JM, Rockmore DN (2011) Nonparametric sparsification of complex multiscale networks. PLoS ONE 6(2):1–10
Garas A, Argyrakis P (2009) A network approach for the scientific collaboration in the European framework programs. Europhys Lett 84(6):68005
Genova VG, Tumminello M, Enea M et al (2019) Student mobility in higher education: Sicilian outflow network and chain migrations. Electr J Appl Stat Anal 12(4):774–800
Genova VG, Tumminello M, Aiello F et al (2021) A network analysis of student mobility patterns from high school to master's. Stat Methods & Appl 30(5):1445–1464
Genova VG, Giordano G, Ragozini G, et al (2022) Clustering student mobility data in 3-way networks. In: Book of Abstracts IFCS 2022, 17th Conference of the International Federation of Classification Societies "Classification and Data Science in the Digital Age", Instituto Nacional de Estatística, pp 56
Giordano G, Primerano I (2018) The use of network analysis to handle semantic differential data. Quality & Quantity 52(3):1173–1192
Giordano G, Vitale MP (2011) On the use of external information in social network analysis. Adv Data Anal Classif 5(2):95–112
Giordano G, Ragozini G, Vitale MP (2019) Analyzing multiplex networks using factorial methods. Social Netw 59:154–170
Giordano G, Primerano I, Vitale P (2021) A network-based indicator of travelers performativity on instagram. Soc Indic Res 156(2):631–649
Interdonato R, Magnani M, Perna D et al (2020) Multilayer network simplification: approaches, models and methods. Comput Sci Rev 36(100):246
Kivelä M, Arenas A, Barthelemy M et al (2014) Multilayer networks. J Complex Netw 2(3):203–271
Kosztyán ZT, Fehérvölgyi B, Csizmadia T et al (2021) Investigating collaborative and mobility networks: reflections on the core missions of universities. Scientometrics 126(4):3551–3564
Maggioni MA, Breschi S, Panzarasa P (2013) Multiplexity, growth mechanisms and structural variety in scientific collaboration networks. Ind Innov 20(3):185–194
Magnani M, Wasserman S (2017) Introduction to the special issue on multilayer networks. Netw Sci 5(2):141–143

Meliciani V, Di Cagno D, Fabrizi A, et al (2022) Knowledge networks in joint research projects, innovation and economic growth across European regions. The Annals of Regional Science pp 1–38
Menichetti G, Remondini D, Panzarasa P et al (2014) Weighted multiplex networks. PloS One 9(6):e97857
MOBYSU.IT (2016) Database MOBYSU.IT, *Mobilità degli studi universitari italiani*, Protocollo di ricerca MIUR-Università degli Studi di Cagliari, Palermo, Siena, Torino, Sassari, Firenze e Napoli Federico II, Fonte dei dati ANS-MIUR/CINECA
Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69(2):026113
Prazeres L (2013) International and intra-national student mobility: trends, motivations and identity. Geogr Compass 7(11):804–820
Primerano I, Santelli F, Usala C (2021) A multiplex network approach to study Italian students' mobility. Book of short Papers SIS 2021:473–478
Rizzi L, Grassetti L, Attanasio M (2021) Moving from north to north: how are the students' university flows? Genus 77(1):1–22
Santelli F, Scolorato C, Ragozini G (2019) On the determinants of student mobility in an interregional perspective: A focus on campania region. Statistica Applicata-Italian J of Appl Stat 1:119–142
Santelli F, Ragozini G, Vitale MP (2022) Assessing the effects of local contexts on the mobility choices of university students in campania region in Italy. Genus 78(1):1–25
Saoud Z, Platoš J (2018) Community detection in bibsonomy using data clustering. In: Information Systems Architecture and Technology: Proceedings of 38th International Conference on Information Systems Architecture and Technology–ISAT 2017: Part I, Springer, pp 149–158
Slater PB (2009) Multiscale network reduction methodologies: Bistochastic and disparity filtering of human migration flows between 3,000+ us counties. arXiv preprint arXiv:0907.2393
Yongwan C, Griffith DA (2011) Modeling network autocorrelation in space-time migration flow data: an eigenvector spatial filtering approach. Ann Assoc Am Geogr 101(3):523–536
Zachary N (2014) The backbone of bipartite projections: inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. Social Netw 39:84–97
Zhu M, Kuskova V, Wasserman S et al (2016) Correspondence Analysis of Multirelational Multilevel Networks. Springer International Publishing, Cham, pp 145–172