

RESEARCH ARTICLE

Multilink communities of multiplex networks

Raul J. Mondragon¹*, Jacopo Iacovacci², Ginestra Bianconi²

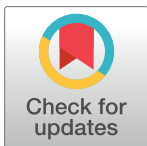
1 School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom, **2** School of Mathematical Sciences, Queen Mary University of London, London, United Kingdom

✉ These authors contributed equally to this work.

* r.j.mondragon@qmul.ac.uk

Abstract

Multiplex networks describe a large number of complex social, biological and transportation networks where a set of nodes is connected by links of different nature and connotation. Here we uncover the rich community structure of multiplex networks by associating a community to each multilink where the multilinks characterize the connections existing between any two nodes of the multiplex network. Our community detection method reveals the rich interplay between the mesoscale structure of the multiplex networks and their multiplexity. For instance some nodes can belong to many layers and few communities while others can belong to few layers but many communities. Moreover the multilink communities can be formed by a different number of relevant layers. These results point out that mesoscopically there can be large differences in the compressibility of multiplex networks.



OPEN ACCESS

Citation: Mondragon RJ, Iacovacci J, Bianconi G (2018) Multilink communities of multiplex networks. *PLoS ONE* 13(3): e0193821. <https://doi.org/10.1371/journal.pone.0193821>

Editor: Naoki Masuda, University of Bristol, UNITED KINGDOM

Received: September 10, 2017

Accepted: February 20, 2018

Published: March 20, 2018

Copyright: © 2018 Mondragon et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was supported by QMUL Research-IT and funded by EPSRC grant EP/K000128/1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The current Big Data explosion requires the development of new algorithms and theoretical methods to extract information from large datasets. Often in this context, it is advantageous to combine information coming from different sources and to represent the data by a multiplex network [1–5]. A multiplex network is formed by a set of nodes connected in different layers by links indicating interactions of different types. Multiplex networks are ubiquitous spanning from complex infrastructure networks [4, 6, 7], to social [8–11] biological [8, 12] and transportation networks [13, 14]. For instance, individuals can be related by different type of social ties, neurons can interact through chemical synapses and electrical gap junctions, and two locations can be connected by different means of transportation.

A multiplex network tends to have a richer structure than single networks and this richness is reflected in its communities [9, 10, 15–17]. The communities of a multiplex network cannot be obtained by considering its layers individually. Some communities might exist only in one layer, other communities can overlap on many layers and finally there are communities that only exist when considering the whole structure of the multiplex network. Several algorithms [9, 18–22] have been recently proposed to detect multilayer communities. These include methods based on multilayer modularity optimization [9, 20], diffusion properties on multilayer networks [18, 21] and consensus clustering [22]. All these techniques are node-based

community detection methods where each node or each replica-node (realization of a node in a given layer) is classified in one community. Interestingly in the framework of single-layer community detection [23, 24] it has been observed that link-based community detection methods [25, 26] can be very fruitful to describe the mesoscale organization of networks when nodes belong to several communities at the same time [27]. The need to extend the link communities to multiplex network is rather pressing. For instance if we consider individuals interacting through different on-line social network platforms, say Twitter and Facebook, it might be misleading to think that an individual or an account (a Twitter or Facebook account) might belong just to a single community. In fact, influential Twitter or Facebook accounts tend to reach more than one community of the same online platform.

In simple networks any two nodes can be either connected or not connected by a link, in multiplex network any two nodes can be connected in multiple ways. We say that two nodes are connected via a *multilink* [3, 11], where the multilink describes the pattern of connections between two nodes. In this work we propose a multilink community detection method for multiplex networks which extends link communities to the multiplex network framework. Our community detection method is based on the similarity of incident multilinks. In order to reduce unnecessary layer-information, the similarity between two multilinks is measured by comparing the local structure of the multiplex against a local, maximum entropy null model. To avoid introducing bias via the null model, the null model describes our state of knowledge of the multiplex in a way that is maximally noncommittal to the layered structure.

Here we show that using the proposed multilink community detection method not only we are able to extract relevant information on the mesoscale structure of multiplex networks, but also we can contribute to the scientific debate about the compressibility of the multiplex network structures. Recent research on multiplex networks questions whether it is opportune to aggregate or disaggregate their layers. Aggregation of layers could be useful for removing redundant information. De Domenico et al. [28], have shown that for the vast majority of multiplex networks there is trade-off between the information content and the minimization of their total number of layers. The case of disaggregating a single network to a multi-layer network has been considered by Vales-Catala et al. in Ref. [29]. According to their results some single networks are better represented as multiplex networks because they are effectively the result of a blind multiplex network aggregation procedure. Finally, Peixoto [30], using a statistical inference approach, has revealed that there is no clear answer, the benefits of the aggregation or disaggregation of the layers are dependent on the system under study.

Here we show that actually the optimal answer to the question whether it is more appropriate to aggregate or disaggregate a general multiplex network might not be global but meso-scale. Our analysis of social, biological and transportation networks reveals that in multiplex networks there is a very rich interplay between their mesoscale organization and their multiplexity. Multilinks communities can include connections of only one layer or of multiple layers. Additionally we observe that not always the layer activity (in how many layers a node is connected) correlates with the community activity (in how many communities a node can be found). For example there can be nodes that are connected in many layers (high layer activity) but belong only to few multilink communities (low community activity) and nodes belonging to few layers (low layer activity) but belonging to many multilink communities (high community activity). The first possibility suggests that mesoscopically the network could be compressed while the second possibility suggests that mesoscopically the network could be expanded into many layers making a case for a definition of a mesoscale compressibility of the multiplex network.

Materials and methods

Multiplex network

Let us consider a multiplex network formed by N nodes and M layers $\alpha = 1, 2, \dots, M$. The multiplex network is the set of M networks $\vec{G} = (G^{[1]}, G^{[2]}, \dots, G^{[\alpha]}, \dots, G^{[M]})$ where each network $G^{[\alpha]} = (V, E^{[\alpha]})$ is formed by the same set of N nodes $V = \{i; i = 1, 2, \dots, N\}$ and by the set of links $E^{[\alpha]}$ which describe the connections in layer α . We assume that all these networks are undirected and we represent each layer $\alpha = 1, 2, \dots, M$ by the adjacency matrix $\mathbf{a}^{[\alpha]}$. The whole multiplex network can be expressed via its multilinks [3, 11]. Every pair of nodes (i, j) is connected by a multilink

$$\vec{m}_{ij} = (m_{ij}^{[1]}, m_{ij}^{[2]}, \dots, m_{ij}^{[\alpha]}, \dots, m_{ij}^{[M]}), \tag{1}$$

with $m_{ij}^{[\alpha]} = a_{ij}^{[\alpha]}$ indicating in which layers of the multiplex network the two nodes are connected. The vector \vec{m}_{ij} specify what we call the *multilink composition*, i.e. in which layers node i and node j are connected. Whenever node i and node j are connected at least in one layer, i.e. $\vec{m} \neq \vec{0}$, we say that they are connected by a non-trivial multilink. To decide if a non-trivial multilink exist, it is convenient to construct the aggregated network \hat{G} formed by the N nodes of the multiplex. The adjacency matrix \mathbf{A} of the aggregated network \hat{G} has elements

$$A_{ij} = \theta \left(\sum_{\alpha=1}^M a_{ij}^{[\alpha]} \right), \tag{2}$$

where $\theta(x)$ is the step function $\theta(x) = 1$ if $x > 0$ and $\theta(x) = 0$ if $x \leq 0$. We indicate with $L = \sum_{i < j} A_{ij}$ the total number of links of the aggregated network, or equivalently the number of non-trivial multilinks.

In a multiplex network the nodes might not be connected in each layer. The number of layers in which a node is connected (or active) is called the *node activity* [8, 31] and reveals relevant coarse grained information about the node.

Multilink similarity

In the context of single networks several community detection methods use hierarchical clustering applied either to a similarity matrix between nodes [32] or between links [25, 26]. Here we construct a hierarchical clustering of multiplex networks based on a measure of similarity between incident multilinks. By defining the similarity between multilinks here we generalize the link communities previously defined for single layers [25, 26] to multiplex networks.

In a similar spirit to the use of the modularity function for detecting node communities [33], the similarity between incident multilinks is evaluated by comparing simultaneously the cohesiveness and the multiplexity of their neighborhood to a maximum entropy null model.

To every pair of multilinks connecting nodes i and k and nodes j and s we assign the similarity $S_{ik,js}$. The similarity $S_{ik,js}$ is non-zero only between incident multilinks (i.e. for $s = k$) and is a function of two parameters: ϵ and z . The parameter $\epsilon \in (0, 1)$ can be tuned depending on the role that we want to assign to the composition of the two incident multilinks with respect to their local neighborhood. The additional parameter $z \in (0, 1)$ evaluates the role of multiplexity and represent the cost we want to attribute to incident multilinks of different composition.

Specifically the non-zero similarities $S_{ik,jk}$ are given by

$$S_{ik,jk} = \epsilon \sigma_{ijk} + (1 - \epsilon) \sigma_{ij \setminus k}. \tag{3}$$

where σ_{ijk} evaluates the contribution of the two incident multilinks while $\sigma_{i\setminus k}$, evaluates instead the contribution due to the existence of other multilinks, joining node i and node j directly or by paths of length two excluding node k . The parameter $\epsilon \in (0, 1)$ tunes the relative importance between these two contributions.

The term σ_{ijk} is expressed as

$$\sigma_{ijk} = z^{\beta_{ik,jk}}, \tag{4}$$

with

$$\beta_{ij,rs} = 1 - \frac{\sum_{\alpha=1}^M m_{ij}^{[\alpha]} m_{rs}^{[\alpha]}}{M}. \tag{5}$$

The smaller is z the larger is the “penalty” imposed to the similarity $S_{ik,jk}$ between non-zero incident multilinks if node i and node j are connected in layers different from the one connecting node j and k . Therefore z tunes the layer dependance of the multilink communities. A large value of z favors communities existing only in one layer or overlapping in different layers, while a smaller value of z allows for multilink communities of multilinks with different composition. If the multilinks connecting nodes (i, k) and (j, k) have not even a link in a common layer then $\beta_{ik,jk} = 1$ and $z^{\beta_{ik,jk}} = z$, indicating the maximum cost attributed to multiplexity. If, on the contrary the two multilinks have the same layer composition, that is $m_{ij}^{[\alpha]} = m_{jk}^{[\alpha]} = 1$ for all α , then $\beta_{ik,jk} = 0$ and $z^{\beta_{ik,jk}} = 1$ indicating that we attribute no cost penalty to this configuration.

The term $\sigma_{i\setminus k}$ includes contributions from paths of length one (\mathcal{M}_{ij}) and two ($\hat{\mathcal{M}}_{ijr}$) between node i and node j that pass through node r with $r \neq k$, i.e.

$$\sigma_{i\setminus k} = \frac{1}{\mu} \left[\mathcal{M}_{ij} + \sum_{r \neq k} \hat{\mathcal{M}}_{ijr} \right], \tag{6}$$

where μ is a normalization constant with $\mu = \max(1, \nu)$ with

$$\nu = \min \left(\sum_{r \neq k} A_{ir}, \sum_{r \neq k} A_{jr} \right). \tag{7}$$

Similarly to the modularity measure [33], term \mathcal{M}_{ij} evaluates the significance of the observed multilink \vec{m}_{ij} against its expectation and, $\hat{\mathcal{M}}_{ijr}$ evaluates the significance of two non-trivial multilinks $\vec{m}_{ir}, \vec{m}_{jr}$ connecting respectively node i and node j to a common node $r \neq k$ against their expectations. These terms are

$$\mathcal{M}_{ij} = (A_{ij} - p_{ij}^{\vec{m}_{ij}}) z^{\beta_{ij,ij}} \delta(A_{ij}, 1) \tag{8}$$

and

$$\hat{\mathcal{M}}_{ijr} = (A_{ir}A_{jr} - p_{ir}^{\vec{m}_{ir}} p_{jr}^{\vec{m}_{jr}}) z^{\beta_{ir,jr}} \delta(A_{ir}A_{jr}, 1), \tag{9}$$

where $\beta_{ij,rs}$ is given by Eq (5), and $\delta(x, y)$ is the Kronecker delta (i.e. $\delta(x, y) = 1$ for $x = y$ and $\delta(x, y) = 0$ otherwise). The term $z^{\beta_{ij,rs}}$ puts a cost to the paths that are created using different layers. The expectation of multilink \vec{m}_{rs} is given by the probability $p_{rs}^{\vec{m}_{rs}}$, which is evaluated using maximum entropy ensembles preserving the degree of node i and node j in each layer α , and the multilinks $\vec{m}_{ik}, \vec{m}_{jk}$ (see S1 Appendix for further details).

Summing up, the parameter ϵ can be used to tune the contribution to the similarity $S_{ij,jk}$ coming from the composition of the two incident multilinks and the contribution coming from the local clustering of the multiplex network in proximity of the wedge ijk . In particular the smaller is ϵ the larger is the contribution due to the local clustering of the multiplex network.

From numerical experimentation, we observed that the time to evaluate the similarity matrix $S_{ik,jk}$ using our MATLAB code grows as N^3 for multiplex networks with the same number of layers and increases linearly with the number of layers M making our method suitable for small to medium size networks (see [S1 Fig](#)).

Multilink communities

From the $L \times L$ similarity matrix $S_{ik,jk}$ we construct a dendrogram via single linkage hierarchical clustering. The dendrogram contains information about the multiplex structure which cannot be obtained from the aggregated network. Finally the multilink communities are determined by cutting the dendrogram at a height that correspond to an optimal value of a appropriate score function.

To obtain the multilink communities we desire to use a score function that does not use any a priori assumptions about the multilink composition. To this end we have considered a score function used on single-layer link-community detection methods, i.e. the link modularity Q [26] (see [S1 Appendix](#) for its definition). An alternative choice could be to choose the partition density D used in [25]. The optimal partition is defined by the maximum value of Q obtained when considering all the heights in the dendrogram (see [S4 Fig](#) for typical profiles of this link modularity on real datasets).

Once every multilink is associated to a given multilink community we can assign to each node a *community activity* given by the number of communities to which its incident multilinks belong.

Results and discussion

A simple example

The community activity of a node resulting from the multilink community detection method is independent on its layer activity. To illustrate this property we consider the multilayer network shown in [Fig 1A](#) decomposed in three multilink communities [Fig 1C](#)) detected using the parameters $\epsilon = 0.4$ and $z = 0.6$. Node d is active in a single layer but belongs to two multilink communities. On the contrary node g is active in two layers but belongs to just one community.

Additionally the communities can be formed by interactions existing only in one layer or in multiple layers. For instance the community formed by the nodes $\{e, d, f\}$ of the multiplex network shown in [Fig 1A](#), only exist due to the combination of different layers in the multiplex. On the contrary the community formed by the nodes $\{a, b, c, d\}$ include only links of a single layer.

The dendrogram in [Fig 1B](#) shows the hierarchical structure of the link communities of the multiplex network in [Fig 1A](#) and reveals the multilayer nature of the network also in the case of this very symmetrical and clustered topology. In fact, the left and right communities of [Fig 1C](#), although they play the same role in the aggregated network, have a different decomposition into multilink sub-communities. There are two factors that contribute to this difference. The right community has a multilink formed by two layers (multilink 14) which are no present in the other community. The second factor is more subtle and it would generate differences in

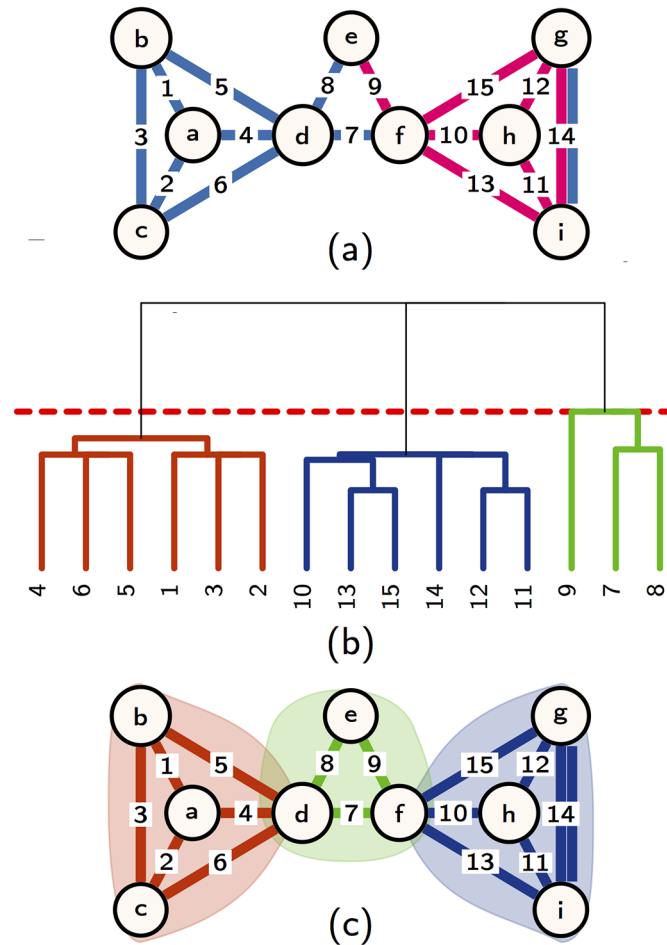


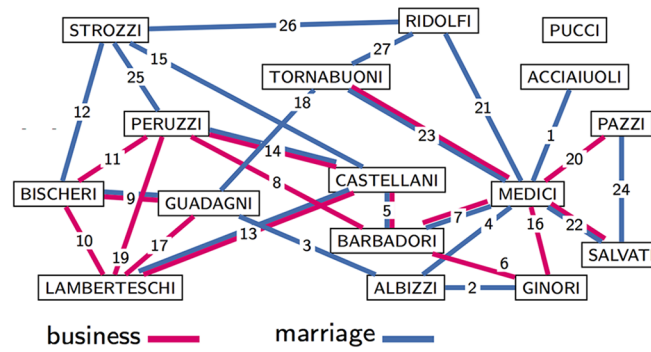
Fig 1. A two layer multiplex and its partition into multilink communities. (A) A simple two layer multiplex networks where the links in blue correspond to one layer and the dark pink ones to another layer. (B) Dendrogram of the multiplex network obtained from the multilink similarity. The dashed red line shows the maximum link modularity used to define the link communities. (C) Panel C shows the partition of the multiplex network into three communities revealing that communities can be formed by a single (community $\{a, b, c, d\}$) or multiple layers (community $\{d, e, f\}$) and that the nodes communities are independent on the node activity (node d belongs to two community and is active in one layer, node g is belongs to one community and is active in one layer). The colors of the dendrogram branches in panel B and of the color of the multilinks of the multiplex network in panel C indicate the different multilink communities. The multilink communities were detected using $\epsilon = 0.4, z = 0.6$.

<https://doi.org/10.1371/journal.pone.0193821.g001>

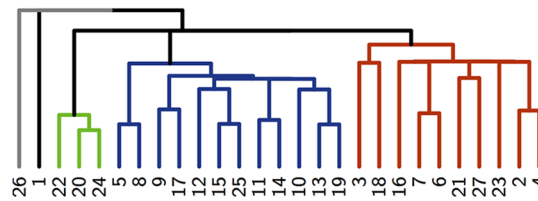
the hierarchical structure even if the community on the right included only links existing in a single layer (see [S1 Appendix](#) and [S2 Fig](#) for details).

Florentine families

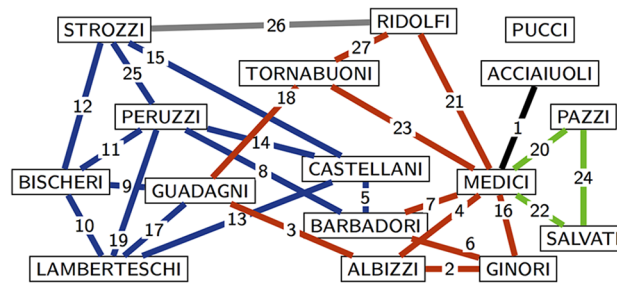
The Florentine Families Multiplex Network [34] consist of $M = 2$ two layers, one layer describes the business dealings between $N = 16$ florentine families in the XV century, the other layer their alliances due to marriages. [Fig 2A](#) shows these relationships between the families. [Fig 2B](#) shows the dendrogram describing the multilink communities for $\epsilon = 0.5, z = 0.6$ (see [S3 Fig](#) for the dependence of the number of clusters on ϵ and z).



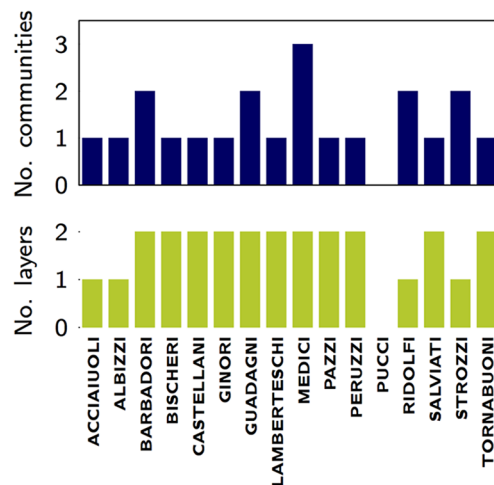
(a)



(b)



(c)



(d)

Fig 2. Multilink communities for the florentine families. (A) The Florentine Families Multiplex Network describing the business and marriage alliances of the XV century florentine families. (B) Heat map displaying the multilink similarity matrix and its relative dendrogram. (C) Partition of the Florentine Families multiplex network into five multilink communities. (D) Layer and community activity of the different families. The Medici family is characterized by achieving the maximum of the community activity. The multilink communities are detected using $\epsilon = 0.4$, $z = 0.6$.

<https://doi.org/10.1371/journal.pone.0193821.g002>

The two detected single multilink communities correspond to two different scenarios (Fig 2C). The multilink between the Strozzi and the Ridolfi family establish an interaction between two families which have connections between different clusters; the multilink between the Acciaiuoli and the Medici family is a leaf of the multiplex network, being the only multilink connecting the Acciaiuoli family to the rest of the multiplex network.

For each family we compare their *layer activity* and their *community activity* (Fig 2D). We observe that families with high community activity are powerful brokers between different communities. Most relevantly, the Medici play a pivotal role as they are brokers between three different communities. The Barbadori and the Guadagni family have the same community activity as the Ridolfi and the Strozzi family but while the first two are connected in both layers the latter two are connected to the other families exclusively in one layer (the marriage alliances).

Multiplex connectome of *C. elegans*

The Multiplex Connectome of *C. elegans* [35, 36] has two layers $M = 2$, the chemical synapses and the gap junctions describing the interactions between $N = 279$ neurons. As an example, we obtained the multilink communities for $\epsilon = 0.4$ and $z = 0.6$. The multiplex has 845 multilink communities of which 652 (about 77%) are made of single multilinks. The distribution of the sizes of the communities is broad. (Fig 3A). The largest community is formed by 878 multilinks followed communities including 67 links and 51 links. Although there is a large dominant community in the multiplex network, the internal structure of this community can be investigated via the dendrogram. We noticed that the ADAL and ADAR are the neurons that cluster first with some of their neighboring neurons (Fig 3B) for all values of z .

This multiplex network has neurons which have large community activity (Fig 3C). By ranking the neurons according to their community activity we find in the first two positions the RIBR and RIBL neurons, which are head interneurons connected via gap junctions to multiple other neuron classes, suggesting that these neurons play a role in brokering between different communities (Fig 3D).

European multiplex air transport network

The European Multiplex Air Transport Network [14] comprises of $N = 417$ European airports and $M = 37$ layers corresponding to the airlines that have flight connections between these airports. The total number of multilink describing these connections is 2953. For the case that $\epsilon = 0.4$ and $z = 0.6$, our algorithm obtains 1790 multilink communities. The largest community includes 723 nodes, about 24% of the total number of multilinks. The smallest communities are made of single multilinks and there are 1696 of them, about 57% of the multilinks.

We observe that the main communities have very different composition in term of single layers. Fig 4A and 4B shows the two largest communities. All the airlines (layers) contribute to the structure of the largest community (Fig 4A). The second largest community has a very different structure, only few airlines contribute to this community.

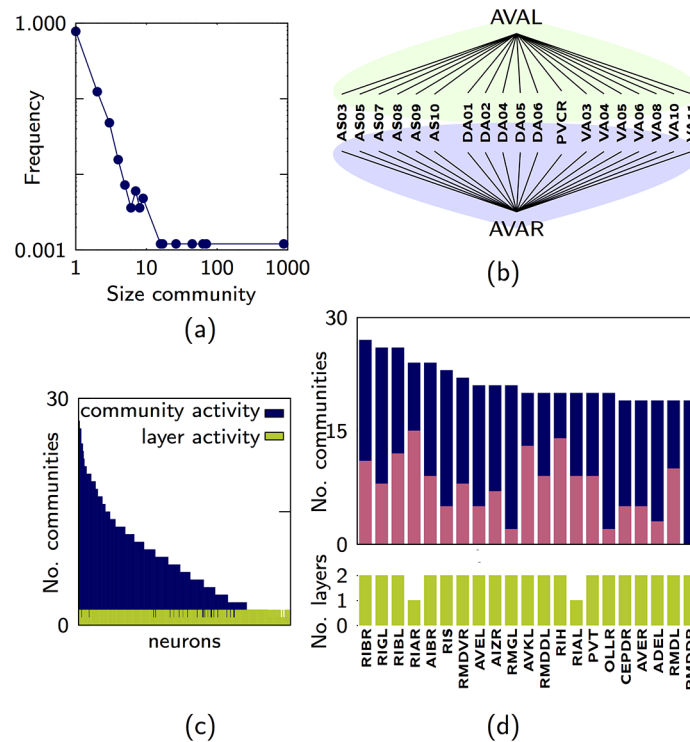


Fig 3. Properties of the multilink communities of the *C. elegans*. (A) Distribution of the communities sizes for the Multiplex Connectome of *C. elegans*. (B) The two most similar sub-communities contained in the largest multilink community. (C) Neurons ranked in decreasing order of their community activity. (D) Layer and community activity for the top ranked neurons. (the contribution of communities with single multilinks is in pink while the contributions of communities with more than one multilink is in blue). The multilink communities are detected using $\epsilon = 0.4$, $z = 0.6$.

<https://doi.org/10.1371/journal.pone.0193821.g003>

When comparing the airports and their community activity, we observe (Fig 4C) that while large layer activity, an airport serving multiple airline companies, seems to be correlated to high community activity, there is a significant variability in the community of airports that are active in many layers. For example Vienna (VIE) and Amsterdam (AMS) have a comparable layer activity but very different community activity. Similarly there are airports with small layer activity but significant community activity, for example Luton (LTN) and Bergamo (BGY) airports. This indicates that the airports might adopt different strategies to broker between different communities. These strategies might involve serving flights of many airline companies or serving flights of relatively fewer airline companies.

The parameter ϵ and z can be tuned to change the number of multilinks communities with larger values of z penalizing communities where multilinks have different composition \vec{m} , while larger values of ϵ allows to weight more the contribution to the multilink similarity coming from the local neighborhood of the wedge ikj . The dependence of the number of communities with the parameters z and ϵ for the three data sets is investigated in the Supporting Information (S1 Appendix). Our choice of the parameters $\epsilon = 0.4$ and $z = 0.6$ is dictated by the desire of having multilink communities that might span several layers (excluding very large values of z) extracted from a similarity matrix significantly affected by the local neighborhood of the wedges (excluding very low values of ϵ). Additionally these parameter values for the

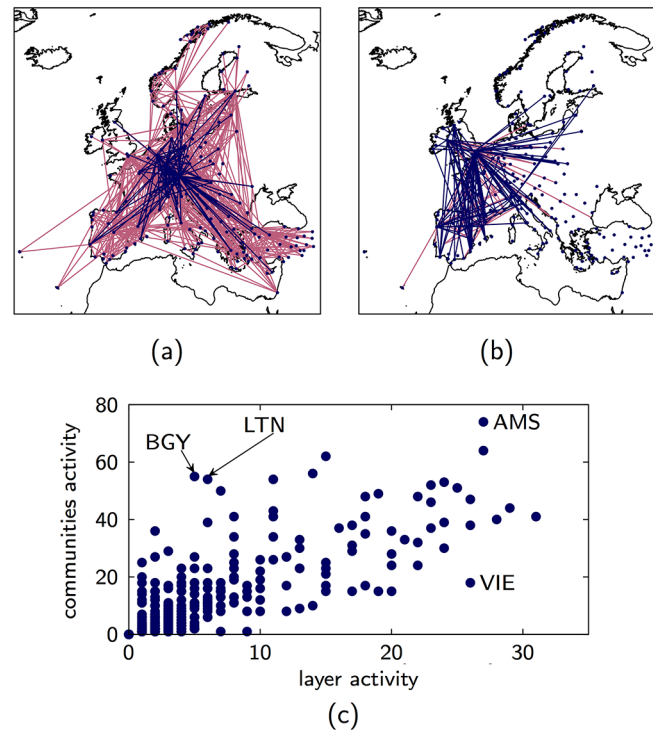


Fig 4. Community activity vs layer activity for the EU Airports multiplex. (A) Largest link community of the European Multiplex Air Transportation Network. Lufthansa’s flights are shown in blue, the other airlines in pink. (B) Second largest link community. Ryanair’s flights are shown in blue, the other airlines in pink. (C) Community vs. layer activity of the EU airports. While the layer activity appears to have a positive correlation with the communities activity, there are large differences in the communities activity between airports with large layer activity (compare for instance Amsterdam (AMS) and Vienna (VIE)). The multilink communities are detected using $\epsilon = 0.4, z = 0.6$.

<https://doi.org/10.1371/journal.pone.0193821.g004>

analysed real datasets lead to score function profiles with a well defined maximum (see S3 and S4 Figs).

However we point out that ϵ and z remain tunable parameters that can be chosen accordingly to the dataset and the scientific problem under study as it is done for instance with the resolution parameter in modularity optimization algorithms.

Composition of the multilink communities

To investigate whether the communities are formed exclusively by links of a single layer or include links of several layers we introduce the *layer specificity* $x_c^{[\alpha]}$ which is the fraction of multilinks in a community c which include a link in layer α . Therefore $x_c^{[\alpha]} = 1$ indicates that all the multilinks of a community include a link in layer α , while $x_c^{[\alpha]} = 0$ indicates that the community does not include any link in layer α . Note that since a single multilink can include links of different layers, the sum of the layer specificity $x_c^{[\alpha]}$ for community c in general do not add to one.

In the Multiplex Connectome of *C. elegans* we observe that many communities are exclusively formed by one type of multilink, however, the three largest communities have a multiplex nature as they include different types of multilinks (see Fig 5A where the larger communities are indicated by the labels 1, 2, 3 in order of decreasing size).

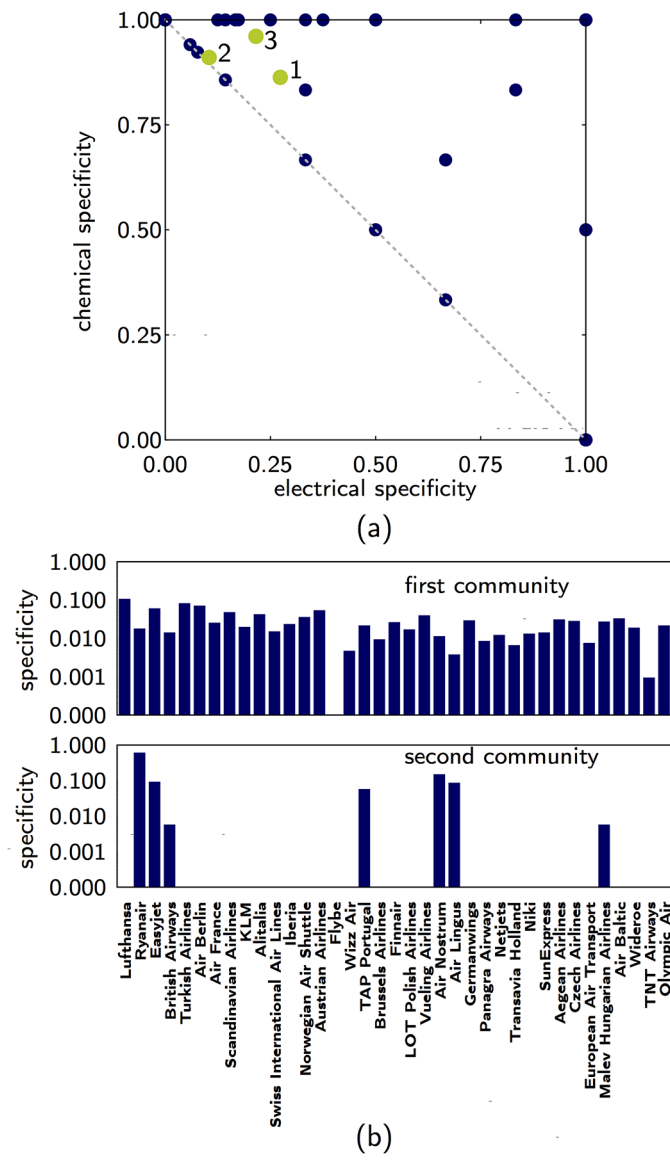


Fig 5. Specificity of the *C. elegans* and the EU Airlines. (A) Specificity for the communities in the Multiplex Connectome of *C. elegans*. (B) Specificity for the first and second largest communities for the European Air Transportation Network. Both panels indicate a large variability in the layer composition of different communities. The multilink communities are detected using $\epsilon = 0.4$, $z = 0.6$.

<https://doi.org/10.1371/journal.pone.0193821.g005>

In the European Multiplex Air Transportation Network, the largest community, apart from Flybe, contains flights from all other airlines (Fig 5B). The largest contribution comes from Lufthansa with an specificity of 0.10 followed by Turkish Airlines with 0.07 specificity. The second largest community has a different structure, in this case only seven airlines contribute to the community, the largest contribution is from Ryanair with a specificity of 0.60. In this multiplex, low-cost airlines like Ryanair, Easyjet and Wideroe have high specificity (often equal to 1) in many communities. However these airlines rarely have high specificity in the same

community. This is a consequence of the competition between low-cost airline companies as they tend to differentiate each other by having unique flights to some destinations.

Conclusion

Our method reveals the richness of multiplex networks at their mesoscale structure. This is achieved by associating to each pair of incident multilinks a similarity measure based on the comparison of the local connectivity of two multilinks against a null model. Our intrinsically multiplex community detection method allows us to associate to each node multiple communities independently on its layer activity. Specifically we can have nodes active exclusively in one layer and belonging to multiple communities or active in many layers but belonging only to few communities. The proposed method is here applied to several real datasets revealing that the mesoscale structure of a multiplex can be organized via communities containing links in many different layers and, at the same time, communities having one predominant layer. This suggests that the mesoscale organization of multiplex networks has a rich mesoscale structure that is not captured by methods that aim at compressing the information on few single layers.

Supporting information

S1 Appendix. Description of the multilink community method. Detailed description of the Multilink Community detection algorithm and additional results obtained by applying the Multilink Community detection algorithm to benchmark multiplex networks models and real multiplex networks datasets.

(PDF)

S1 Fig. Computational complexity as a function of the number of nodes and number of layers. (A) Average CPU time elapsed to evaluate all the elements of the matrix $S_{jk,jk}$ as a function of the number of nodes for two families of multiplexes. Bottom line the multiplexes are defined by $\langle k^{[\alpha]} \rangle = 3$ (bottom line) and $\langle k^{[\alpha]} \rangle = 3$, the dashed lines show the best fit via $f(N) = a + bN^z$. (B) Average CPU time as a function of M . The horizontal lines correspond to multiplexes with, from bottom to top, $N = 100, 500, 1000, 1500, 2000, 2500$.

(PDF)

S2 Fig. Example of a two layer multilink community. (A) A simple two layer multiplex network (purple and ochre links) and its multilink communities (shaded areas) and (B) its dendrogram obtained from the multilink similarity. The dashed red line shows the maximum link modularity used to define the link communities.

(PDF)

S3 Fig. Variation of the cluster size against the z and ϵ parameters. Variation of the number of clusters as a function of z with given ϵ (right panels) and as a function of ϵ with given z (left panels). The top panels show the Florentine families, the middle panels the *C. elegans* and the bottom panels the EU-airports multiplexes. In the left panels data are shown for $\epsilon = 0.4$ (blue triangles), $\epsilon = 0.5$ (pink circles) and $\epsilon = 0.6$ (green squares). In the right panels data are shown for $z = 0.4$ (blue triangles), $z = 0.5$ (pink circles) and $z = 0.6$ (green squares). The red circles show the values used in the main manuscript.

(PDF)

S4 Fig. Number of clusters against the score function Q (link-modularity). (A) Florentine Families Multiplex Network ($\epsilon = 0.5, z = 0.6$). (B) The Multiplex Connectome of *C. elegans* ($\epsilon =$

0.4, $z = 0.6$) and (C) for the European Multiplex Air Transportation Network ($\epsilon = 0.4, z = 0.6$). The maximum of Q determines the number of clusters which define the multilink communities of the multiplex network.

(PDF)

S5 Fig. Aggregated degree vs. community activity. (A) Multiplex Connectome of *C. elegans* and (B) for the European Multiplex Air Transportation Network.

(PDF)

S1 Data. Florentine family multiplex dataset.

(ZIP)

S2 Data. EU air transportation multiplex dataset.

(ZIP)

Acknowledgments

This research utilised Queen Mary's MidPlus computational facilities, supported by QMUL Research-IT and funded by EPSRC grant EP/K000128/1.

Author Contributions

Conceptualization: Raul J. Mondragon, Ginestra Bianconi.

Investigation: Jacopo Iacovacci, Ginestra Bianconi.

Methodology: Raul J. Mondragon, Jacopo Iacovacci, Ginestra Bianconi.

Software: Ginestra Bianconi.

Writing – original draft: Raul J. Mondragon, Jacopo Iacovacci, Ginestra Bianconi.

Writing – review & editing: Ginestra Bianconi.

References

1. Boccaletti S, Bianconi G, Criado R, Del Genio CI, Gómez-Gardenes J, Romance M, et al. The structure and dynamics of multilayer networks. *Physics Reports*. 2014; 544(1):1–122. <https://doi.org/10.1016/j.physrep.2014.07.001>
2. Kivela M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. *Journal of complex networks*. 2014; 2(3):203–271. <https://doi.org/10.1093/comnet/cnu016>
3. Bianconi G. Statistical mechanics of multiplex networks: Entropy and overlap. *Physical Review E*. 2013; 87(6):062806. <https://doi.org/10.1103/PhysRevE.87.062806>
4. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S. Catastrophic cascade of failures in interdependent networks. *Nature*. 2010; 464(7291):1025–1028. <https://doi.org/10.1038/nature08932> PMID: 20393559
5. Gomez S, Diaz-Guilera A, Gomez-Gardenes J, Perez-Vicente CJ, Moreno Y, Arenas A. Diffusion dynamics on multiplex networks. *Physical review letters*. 2013; 110(2):028701. <https://doi.org/10.1103/PhysRevLett.110.028701> PMID: 23383947
6. Shao J, Buldyrev SV, Havlin S, Stanley HE. Cascade of failures in coupled network systems with multiple support-dependence relations. *Physical Review E*. 2011; 83(3):036116. <https://doi.org/10.1103/PhysRevE.83.036116>
7. Huang X, Gao J, Buldyrev SV, Havlin S, Stanley HE. Robustness of interdependent networks under targeted attack. *Physical Review E*. 2011; 83(6):065101. <https://doi.org/10.1103/PhysRevE.83.065101>
8. Nicosia V, Latora V. Measuring and modeling correlations in multiplex networks. *Physical Review E*. 2015; 92(3):032805. <https://doi.org/10.1103/PhysRevE.92.032805>

9. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP. Community structure in time-dependent, multiscale, and multiplex networks. *Science*. 2010; 328(5980):876–878. <https://doi.org/10.1126/science.1184819> PMID: 20466926
10. Szell M, Lambiotte R, Thurner S. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*. 2010; 107(31):13636–13641. <https://doi.org/10.1073/pnas.1004008107>
11. Menichetti G, Remondini D, Panzarasa P, Mondragón RJ, Bianconi G. Weighted multiplex networks. *PloS one*. 2014; 9(6):e97857. <https://doi.org/10.1371/journal.pone.0097857> PMID: 24906003
12. Bentley B, Branicky R, Barnes CL, Chew YL, Yemini E, Bullmore ET, et al. The multilayer connectome of *Caenorhabditis elegans*. *PLoS Computational Biology*. 2016; 12(12):e1005283. <https://doi.org/10.1371/journal.pcbi.1005283> PMID: 27984591
13. De Domenico M, Solé-Ribalta A, Gómez S, Arenas A. Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences*. 2014; 111(23):8351–8356. <https://doi.org/10.1073/pnas.1318469111>
14. Cardillo A, Gómez-Gardenes J, Zanin M, Romance M, Papo D, Del Pozo F, et al. Emergence of network features from multiplexity. *Scientific reports*. 2013; 3:1344. <https://doi.org/10.1038/srep01344> PMID: 23446838
15. Battiston F, Iacovacci J, Nicosia V, Bianconi G, Latora V. Emergence of multiplex communities in collaboration networks. *PloS one*. 2016; 11(1):e0147451. <https://doi.org/10.1371/journal.pone.0147451> PMID: 26815700
16. Iacovacci J, Wu Z, Bianconi G. Mesoscopic structures reveal the network between the layers of multiplex data sets. *Physical Review E*. 2015; 92(4):042806. <https://doi.org/10.1103/PhysRevE.92.042806>
17. Kao TC, Porter MA. Layer Communities in Multiplex Networks. *arXiv preprint arXiv:170604147*. 2017;.
18. De Domenico M, Lancichinetti A, Arenas A, Rosvall M. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X*. 2015; 5(1):011027. <https://doi.org/10.1103/PhysRevX.5.011027>
19. Jeub LG, Mahoney MW, Mucha PJ, Porter MA. A local perspective on community structure in multilayer networks. *Network Science*. 2017; p. 1–20.
20. Bennett L, Kittas A, Muirhead G, Papageorgiou LG, Tsoka S. Detection of composite communities in multiplex biological networks. *Scientific reports*. 2015; 5:10345. <https://doi.org/10.1038/srep10345> PMID: 26012716
21. Kuncheva Z, Montana G. Community detection in multiplex networks using locally adaptive random walks. In: *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM; 2015. p. 1308–1315.
22. Lancichinetti A, Fortunato S. Consensus clustering in complex networks. *Scientific reports*. 2012;2.
23. Fortunato S, Hric D. Community detection in networks: A user guide. *Physics Reports*. 2016; 659:1–44. <https://doi.org/10.1016/j.physrep.2016.09.002>
24. Schaub MT, Delvenne JC, Rosvall M, Lambiotte R. The many facets of community detection in complex networks. *arXiv preprint arXiv:161107769*. 2016;.
25. Ahn YY, Bagrow JP, Lehmann S. Link communities reveal multiscale complexity in networks. *Nature*. 2010; 466(7307):761–764. <https://doi.org/10.1038/nature09182> PMID: 20562860
26. Evans T, Lambiotte R. Line graphs, link partitions, and overlapping communities. *Physical Review E*. 2009; 80(1):016105. <https://doi.org/10.1103/PhysRevE.80.016105>
27. Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*. 2005; 435(7043):814–818. <https://doi.org/10.1038/nature03607> PMID: 15944704
28. De Domenico M, Nicosia V, Arenas A, Latora V. Structural reducibility of multilayer networks. *Nature communications*. 2015; 6. <https://doi.org/10.1038/ncomms7864>
29. Valles-Catala T, Massucci FA, Guimera R, Sales-Pardo M. Multilayer stochastic block models reveal the multilayer structure of complex networks. *Physical Review X*. 2016; 6(1):011036. <https://doi.org/10.1103/PhysRevX.6.011036>
30. Peixoto TP. Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. *Physical Review E*. 2015; 92(4):042807. <https://doi.org/10.1103/PhysRevE.92.042807>
31. Cellai D, Bianconi G. Multiplex networks with heterogeneous activities of the nodes. *Physical Review E*. 2016; 93(3):032302. <https://doi.org/10.1103/PhysRevE.93.032302> PMID: 27078361
32. Ravasz E, Barabási AL. Hierarchical organization in complex networks. *Physical Review E*. 2003; 67(2):026112. <https://doi.org/10.1103/PhysRevE.67.026112>

33. Newman ME. Modularity and community structure in networks. *Proceedings of the national academy of sciences*. 2006; 103(23):8577–8582. <https://doi.org/10.1073/pnas.0601602103>
34. Padgett JF, Ansell CK. Robust Action and the Rise of the Medici, 1400-1434. *American journal of sociology*. 1993; 98(6):1259–1319. <https://doi.org/10.1086/230190>
35. Chen BL, Hall DH, Chklovskii DB. Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103(12):4723–4728. <https://doi.org/10.1073/pnas.0506806103> PMID: 16537428
36. De Domenico M, Porter MA, Arenas A. MuxViz: a tool for multilayer analysis and visualization of networks. *Journal of Complex Networks*. 2014; p. cnu038.