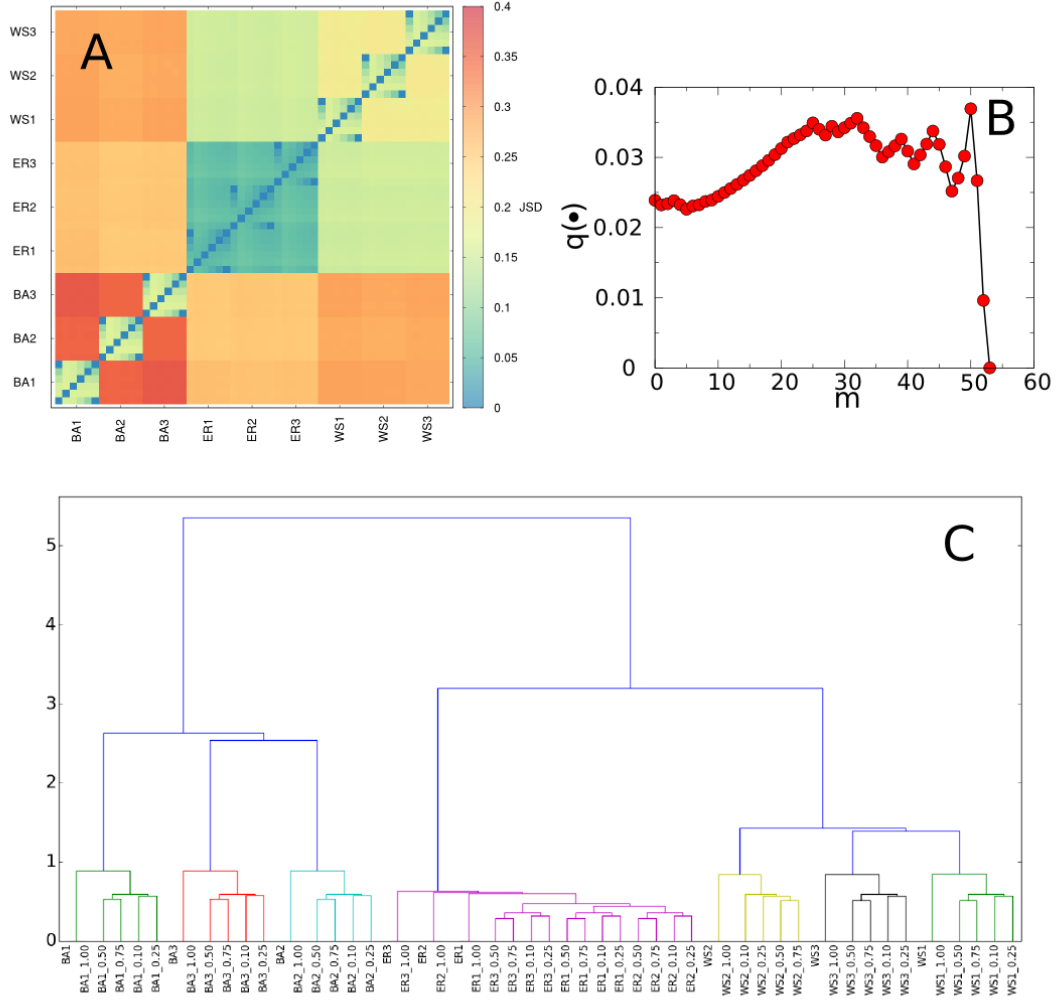
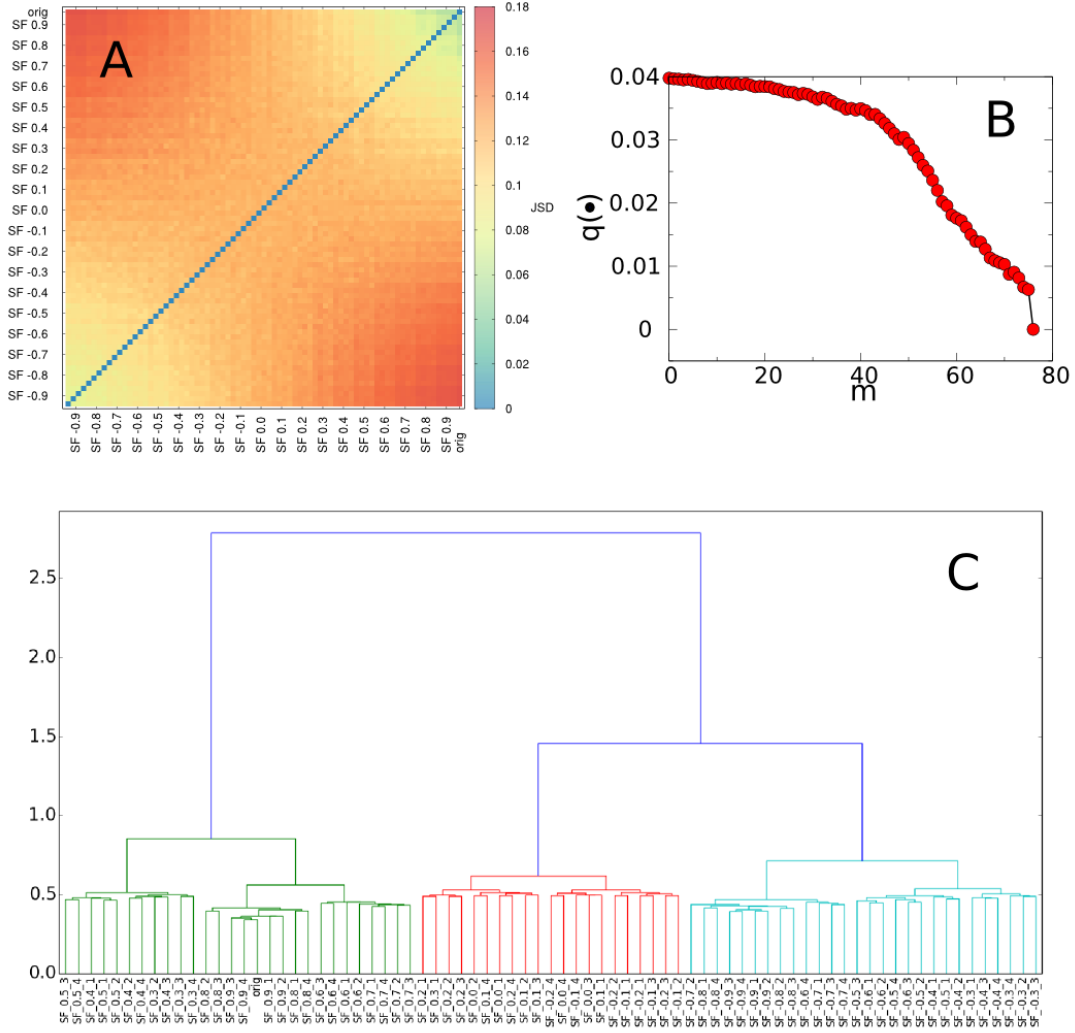


Supplementary Figure 1: **Variation of the Von-Neumann entropy due to layer aggregation.** Distribution of  $\Delta H = h_{A[1]} + h_{A[2]} - h_{A[1]+A[2]}$  in a multilayer configuration with two layers. *Left panel:* system consisting of an Erdos-Renyi layer and a layer with only one link. *Right panel:* system consisting of two different Erdos-Renyi layers. The distribution is computed over 4950 realisations.

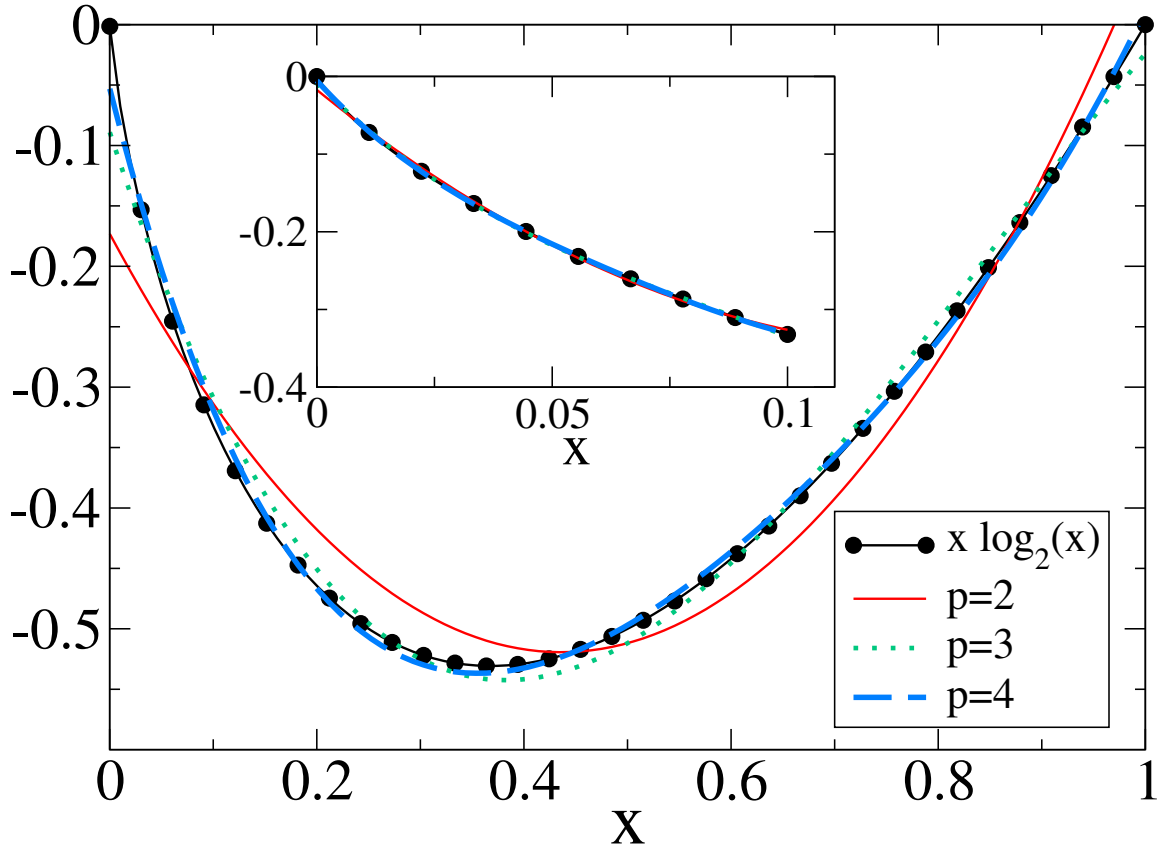


Supplementary Figure 2: **Edge intersection benchmark.** (A) Jensen–Shannon Distance matrix of the benchmark multilayer network with several realisations of three network models with different amount of edge intersection (see Supplementary Methods). Each group of layers is labelled  $MODEL_n$ , where  $MODEL$  can be one of BA (Barabási–Albert), ER (Erdős–Renyi) and WS (Watts–Strogatz), while  $n$  is the corresponding realisation number. Each group of layers consists (from left to right) of a master layer and 5 layers having an increasing amount of edge overlap with the master layer, respectively 10%, 25%, 50%, 75%, 100%. (B) The quality function  $q(\bullet)$  has a global maximum for  $m = 50$ , which corresponds to the higher-level partition of the layers in groups constructed from the same model. The dendrogram is reported in panel (C). Interestingly, groups of layers associated to the same master layer are consistently merged first.

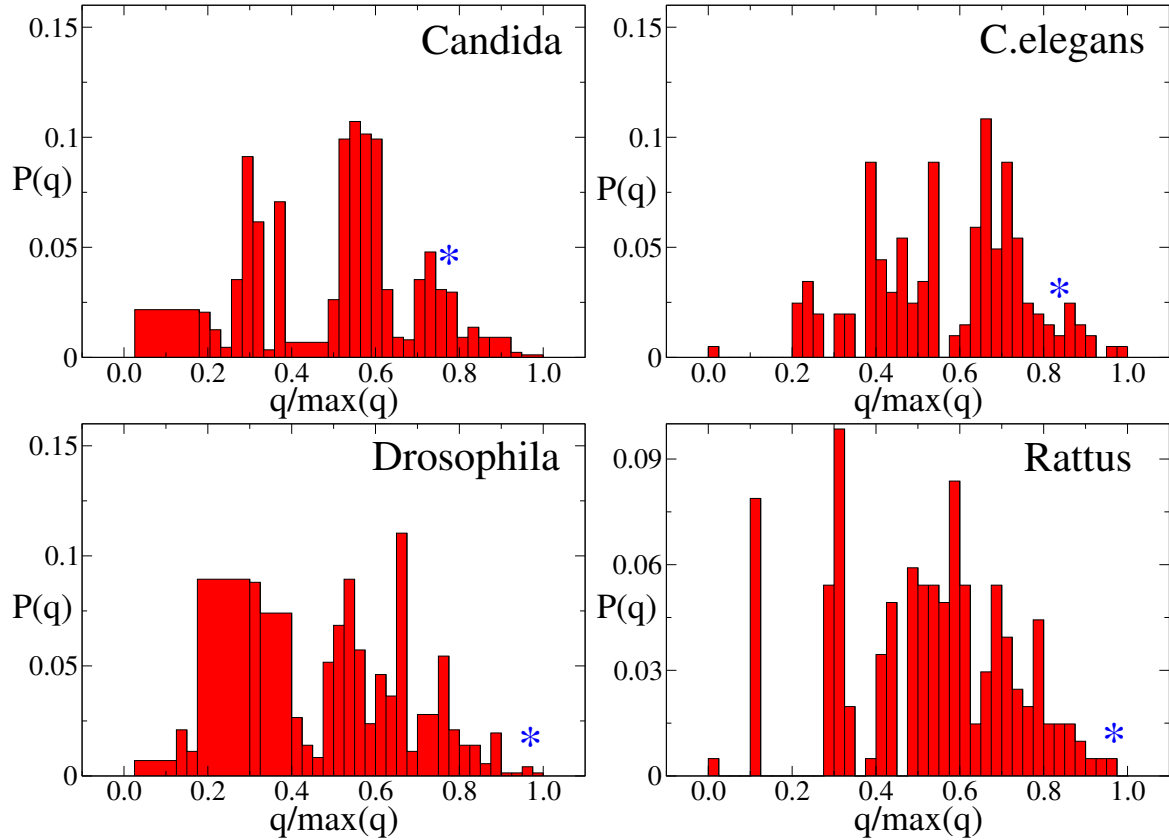


Supplementary Figure 3: **Inter-layer degree correlations benchmark.** (A) The Jensen-Shannon Distance between all the pairs of layers of the multilayer benchmark with prescribed inter-layer degree correlations. For each value of the inter-layer degree correlation coefficient  $\rho$  we constructed 4 independent relabellings of a reference layer (indicated as “orig”). Notice that each group of layers corresponding to the same value of  $\rho$  is indicated by a single label in the form “SF- $\rho$ ”. (B) Despite all the layers of this benchmark are isospectral (and, consequently, have the same value of Von Neumann entropy) the global maximum of the quality function  $q(\bullet)$  is located at  $m = 0$ , indicating that, due to the relabelling, each layer is indeed carrying a substantial amount of new information, which would be lost in aggregating them. Nevertheless, the value of  $q(\bullet)$  remains very close to the maximum at least for  $m < 40$ . (C) The visual inspection of the corresponding dendrogram reveals that the layer aggregation algorithm correctly identifies the existence of structural similarities among the layers characterised by identical or close values of  $\rho$ . The three large groups correspond, respectively, to large positive (left side, green), close to zero (center, red) and large negative values of  $\rho$  (right side, cyan). Notice that layers corresponding to the same value of  $\rho$  are consistently aggregated first.

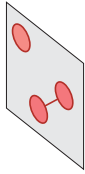
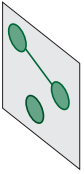
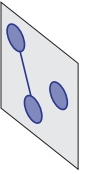
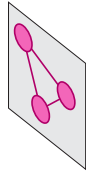


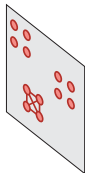
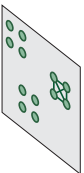
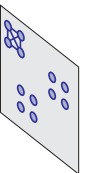
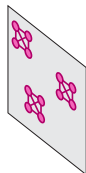


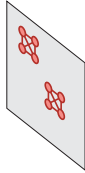
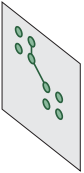
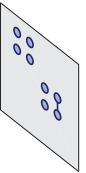
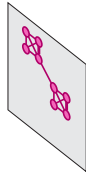
Supplementary Figure 5: **Approximating the Von Neumann entropy.** Approximation of the function  $f(x) = x \log(x)$  (solid black line with dots) in the interval  $[0, 1]$  through polynomials of increasing degree  $p$ , namely  $p = 2$  (solid red line),  $p = 3$  (dotted green line) and  $p = 4$  (dashed blue line). Notice that a pretty good accuracy is already obtained for  $p = 4$ . In the inset we focus on the approximation of  $f(x)$  in the interval  $[0, 0.1]$ , which more closely resembles the real scenario for the computation of the Von Neumann entropy of a graph (recall that due to Eq. (2) the maximum eigenvalue of the rescaled Laplacian is much smaller than 1). In this case a very good accuracy can be already obtained for  $p = 2$  (the three lines are collapsed in a single one, which is practically indistinguishable from  $f(x)$ ).

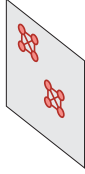
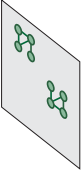
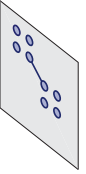
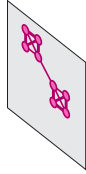


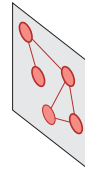

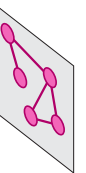
Supplementary Figure 6: **Performance of the greedy optimisation algorithm.** The four panels report, as an example, the distributions of  $q(\bullet)/\max q(\bullet)$  over all the possible layer partitions of the multi-layer protein-gene interaction networks of *Candida*, *C.elegans*, *Drosophila* and *Rattus*. In each panel the asterisk (\*) indicates the position of the optimal value of  $q(\bullet)$  found by the hierarchical clustering procedure. Notice that the greedy exploration of the partition space is always able to find sub-optimal solutions yielding a quite high value of  $q(\bullet)$  and in some cases, e.g. for *Rattus*, the solution provided by the hierarchical clustering algorithm corresponds with the overall optimum.

MULTIPLEX TRIANGLE				Merge Seq.	$h_{C[1]} + h_{C[2]}$	$h_{C[1]+C[2]}$	$\bar{H}(C)$	$q(\bullet)$	desirable merge
				Original	-	-	0	1	-
				Layers 1 & 2	0	0.811	0.405	0.594	NO
				Layers 1+2 & 3	0.811	1	1	0	NO
Layer 1	Layer 2	Layer 3	Aggregate						

DISCONNECTED CLIQUES				Merge Seq.	$h_{C[1]} + h_{C[2]}$	$h_{C[1]+C[2]}$	$\bar{H}(C)$	$q(\bullet)$	desirable merge
				Original	-	-	1.585	0.5	-
				Layers 1 & 2	3.170	2.585	2.085	0.342	NO
				Layers 1+2 & 3	4.167	3.170	3.170	0	NO
Layer 1	Layer 2	Layer 3	Aggregate						

BRIDGED CLIQUES				Merge Seq.	$h_{C[1]} + h_{C[2]}$	$h_{C[1]+C[2]}$	$\bar{H}(C)$	$q(\bullet)$	desirable merge
				Original	-	-	1.309	0.499	-
				Layers 1 & 2	3.396	2.618	2.613	0	NO
				Layers 1+2 & 3	2.618	2.613	2.613	0	NO
Layer 1	Layer 2	Layer 3	Aggregate						

BRIDGED CLIQUES & QUASI-REDUNDANT LAYER				Merge Seq.	$h_{C[1]} + h_{C[2]}$	$h_{C[1]+C[2]}$	$\bar{H}(C)$	$q(\bullet)$	desirable merge
				Original	-	-	1.784	0.361	-
				Layers 1 & 2	5.352	2.784	1.392	0.501	YES
				Layers 1+2 & 3	2.784	2.792	2.792	0	NO
Layer 1	Layer 2	Layer 3	Aggregate						

REDUNDANT LAYERS				Merge Seq.	$h_{C[1]} + h_{C[2]}$	$h_{C[1]+C[2]}$	$\bar{H}(C)$	$q(\bullet)$	desirable merge
			Original	-	-	1.757	0	-	
			Layers 1 & 2	3.514	1.757	1.757	0	YES	
Layer 1	Layer 2	Aggregate							

Supplementary Table 1: Representative illustrations of multilayer configurations where it is desirable to merge (or not) some of the layers.

Network	$q(\mathcal{C}_{\text{opt}})$	$\mathcal{C}_{\text{opt}}$	$q(\mathcal{C}_{\text{max}})$	$\mathcal{C}_{\text{max}}$	$q(\mathcal{C}_{\text{opt}})/q(\mathcal{C}_{\text{max}})$
Arabidopsis	0.4367	(1,6); (2); (3); (4,5); (7)	0.4746	(1); (2); (3,4,5); (6); (7)	0.92
Bos	0.4942	(1); (2); (3,4)	0.4942	(1); (2); (3,4)	1.0
Candida	0.5276	(1); (2,4,5); (3,6); (7)	0.7115	(1); (2,4,5,7); (3); (6)	0.74
C. elegans	0.4252	(1); (2); (3); (4); (5); (6)	0.4965	(1,4,5,6); (2); (3)	0.86
Drosophila	0.4256	(1,6); (2); (3); (4,5); (7)	0.4427	(1,3,4,5,6); (2); (7)	0.96
Gallus	0.5150	(1); (2); (3); (4); (5); (6)	0.5771	(1); (2); (3); (4,5); (6)	0.89
Mus	0.3763	(1); (2); (3); (4,5); (6); (7)	0.3913	(1); (2); (3,4,5); (6); (7);	0.96
Plasmodium	0.5000	(1,2); (3);	0.6111	(1); (2); (3)	0.82
Rattus	0.5038	(1); (2); (3,4,5); (6)	0.5038	(1); (2); (3,4,5); (6)	1.00
S.cerevisiae	0.1150	(1,6,7); (2); (3); (4,5)	0.1346	(1,4,5,6,7); (2); (3)	0.85
S.pombe	0.197	(1,6); (2); (3); (4,5,7)	0.2296	(1,4,5,6,7); (2); (3)	0.86
Xenopus	0.4237	(1); (2,3,4); (5);	0.4237	(1); (2,3,4); (5)	1.00

Supplementary Table 2: For each network in the BIOGRID data set we report, respectively, the (sub-)optimal value  $q(\mathcal{C}_{\text{opt}})$  of the quality function obtained through hierarchical clustering, the corresponding layer partition, the overall maximum value of the quality function  $q(\mathcal{C}_{\text{max}})$  computed over all the possible partitions, the corresponding layer partition, and the ratio  $q(\mathcal{C}_{\text{opt}})/q(\mathcal{C}_{\text{max}})$ . In all the cases the optimal value of  $q(\bullet)$  yielded by the greedy procedure is between 74% and 100% of the overall maximum, while the partition corresponding to the optimal cut of the dendrogram is very similar to the best one.



## Supplementary Note 1

In this section we show the results obtained by using the layer aggregation method proposed in the main text on three instructive synthetic multilayer benchmarks. We show that the procedure is able to detect similarities in the structural properties of the layers at a microscopic, mesoscopic and macroscopic scale. In particular, the aggregation algorithm is sensitive to the existence of edge overlap among the layers, to the presence of inter-layer degree correlations, and to the existence of communities spanning over several layers.

**Edge intersection.** The first benchmark we considered consists of synthetic networks with  $N = 200$  nodes constructed through three different models, namely, Barabási-Albert linear preferential attachment graphs (BA,  $m = 4$ ), Erdős-Renyi graphs (ER,  $p = 0.5$ ) and Watts-Strogatz small-world network (WS,  $m = 5$ ,  $p = 0.2$ ). In particular, we considered three realisations for each model, named “master layers” and respectively called BA\_1, BA\_2, BA\_3, ER\_1, ER\_2, ER\_3, WS\_1, WS\_2, WS\_3. For each realisation, we constructed 5 more layers, each characterised by an increasing amount of edge intersection (see Supplementary Methods) with the corresponding master layer, i.e 10%, 25%, 50%, 75%, 100%.

The algorithm to generate such multilayer networks works as follows. Given a master layer, a copy is created and a biased configuration model is employed. In this surrogate layer (with 100% of edge intersection with the master layer, by construction) we choose two distinct nodes  $i$  and  $j$  with two distinct neighbors,  $h$  and  $l$  respectively, requiring also that  $i$  ( $h$ ) is not connected to  $j$  ( $l$ ) on both layers to avoid creating multiple edges. Then, in the surrogate layer, we rewire the link connecting the pairs  $(i, h)$  and  $(j, l)$  to links connecting the pairs  $(i, l)$  and  $(j, h)$  to decrease the current edge intersection. The current amount of edge intersection is calculated and the procedure is repeated iteratively until the desired edge intersection is reached. When the desired value of the intersection is small, for instance smaller than 20%, it is faster to start from a different configuration, where the surrogate layer is built in order to have zero overlap with the master layer. In this case, the procedure will rewire links in order to increase the amount of edge intersection.

In conclusion, the benchmark multilayer network consists of 9 groups of 6 layers each, for a total of  $M = 54$  layers. In principle, we expect that layers with large values of edge intersection should be aggregated earlier, since each overlapping edge is just redundant.

The results of layer aggregation on this benchmark are reported in Supplementary Figure 2. Notice that, in general, the Jensen-Shannon distance between layers having high edge overlap with the same master layer is pretty small (see Supplementary Figure 2A). Interestingly, the quality function  $q(\bullet)$  has a global maximum for  $m = 50$  (see Supplementary Figure 2B), which corresponds to a partition of the layers in three groups, respectively BA, ER and WS networks (see the dendrogram reported in Supplementary Figure 2C). A visual inspection of the dendrogram reveals the internal structure of each of these groups of layers, which matches perfectly the presence of edge intersection with different master layers.

**Inter-layer degree correlations.** This second benchmark is a multilayer network consisting of identical scale-free networks ( $P(k) \sim k^{-2.5}$ ,  $k_{\min} = 3$ ) with  $N = 1000$  nodes. The nodes on each layer were independently relabelled in order to guarantee a prescribed value of the Spearman’s inter-layer degree correlation coefficient  $\rho$  with the first layer, kept as reference. The algorithm used to set the pairwise inter-layer degree correlations is based on simulated annealing and is described in Supplementary Ref. [1]. We considered 19 values of  $\rho$  in the interval  $[-0.9 : 0.9]$ , (namely,  $\rho = -0.9, -0.8, -0.7, \dots, 0.8, 0.9$ ) and we constructed 4 layer relabellings for each value of  $\rho$ , obtaining a multilayer network with  $M = 77$  layers ( $19 \times 4$  layers with prescribed values of  $\rho$  plus the reference layer). The naming convention for each layer is “SF- $\rho$ - $r$ ”, where  $\rho$  is the value of the inter-layer degree correlation coefficient with respect to the reference layer and  $r$  is the number of the realisation for that value of  $\rho$  ( $r = 1, 2, 3, 4$ ).

The results of hierarchical layer aggregation on this benchmark multilayer network are reported in Supplementary Figure 3, where the reference layer is indicated as “orig”. It is interesting to note that all the layers of the graph are isospectral, meaning that their associated Laplacians have exactly the same set of eigenvalues, and are thus characterised by exactly the same value of Von Neumann

entropy. This is trivially due to the fact that all the layers are obtained by relabelling the same graph (the reference layer), and the spectrum of a graph is preserved under node relabelling. Nevertheless, as shown in Supplementary Figure 3B, the algorithm correctly recognises that each layer adds a certain amount of information to the multilayer network, since the relabelling procedure shuffles the degree sequence and destroys any edge overlap, so that the global maximum of the quality function  $q(\bullet)$  is obtained for  $m = 0$ , i.e. when all the layers of the network are kept separated.

However, by looking at the corresponding dendrogram, shown in Supplementary Figure 3C, it is still possible to visually identify three major groups of layers, namely *i*) those having relatively large positive values of  $\rho$  (left side of Supplementary Figure 3C,  $0.3 \leq \rho \leq 0.9$ , including the reference layer, marked as “orig”); *ii*) those having relatively large negative values of  $\rho$  (right side of Supplementary Figure 3C,  $-0.9 \leq \rho \leq -0.3$ ); and *iii*) those having a value of  $\rho$  close to zero (central region in Supplementary Figure 3C,  $-0.2 \leq \rho \leq 0.3$ ). Each of this high-level groups is internally structured in a few sub-groups which include all the layers characterised by relatively close values of  $\rho$ . Interestingly, the original network (labelled as “orig”) is put in the same cluster of the layers having  $\rho = 0.9$ , i.e. a high positive value of inter-layer degree correlations.

**Community structure.** The third benchmark is a multilayer network whose layers have a strong community structure and communities on different layers have a certain percentage of nodes in common [2]. The benchmark consists of 26 groups of layers with different size, respectively 12 groups of 2 layers, 8 groups of 3 layers and 6 groups of 4 layers, for a total of  $M = 72$  layers. Each layer is a network with  $N = 1024$  nodes, formed by 64 cliques of 16 nodes, where an edge connects two nodes in different cliques with probability  $p_{out} = 0.05$ . The communities of the layers belonging to the same group share a certain percentage of nodes, in this case 25% (which corresponds to 4 nodes, on average). For instance, if we consider a group of three layers, then 4 nodes in each clique will appear in the same community on all the three layers. The algorithm adopted to build such multilayer networks is based on simulated annealing. Each layer is denoted as  $L_{n\_r\_l}$ , where  $n$  is the size of the group to which the layer belongs,  $r$  is the number identifying the group and  $l$  is the number identifying the layer inside the group. Using this notation, “L4\_r3\_l2” denotes the second layer in the third 4-layer group.

The results are reported in Supplementary Figure 4. Notice that due to the existence of overlap among the communities of the same group, the Jensen-Shannon Distance among layers belonging to the same group is usually pretty small, compared to that between layers in different groups (see Supplementary Figure 4A). Therefore, it comes as no surprise that the algorithm aggregates first the layers belonging to the same group and then the groups having similar size, as shown by the dendrogram reported in Supplementary Figure 4C. Interestingly, the quality function  $q(\bullet)$  has a maximum for  $m = 0$ , indicating that the best representation of the system is obtained when all the layers are kept separated.

## Supplementary Note 2

In this section we comment on an exotic property of the Von Neumann entropy of a graph and we present some synthetic examples to discuss the behaviour of the function  $q(\bullet)$  used to evaluate the quality of a given configuration of the layers of a multi-layer network.

We start by noticing that, according to the interpretation recently proposed in Ref. [3], the rescaled Laplacian  $\mathcal{L}_G$  of a graph  $G$  can be considered as the quantum density operator associated to a mixed state. Such a quantum state (which in practice corresponds to the graph  $G$ ) consists of the uniform mixing of pure states on the space of the vertices of  $G$ , where each pure state corresponds to exactly one edge of the graph. The Von Neumann entropy of  $G$  measures the mixedness of such a state, where  $\mathcal{H}_{VN}(G) = 0$  if and only if  $G$  is a pure state, i.e. if  $G$  is a graph consisting of exactly one edge, while in general  $\mathcal{H}_{VN}(G) > 0$  when  $G$  has more than one edge. As a consequence, the Von Neumann entropy of a graph is expected to increase if the graph becomes denser.

However, this is not always the case, in general. The reason lies in the definition of this entropy in terms of the eigenvalues of the density matrix constructed from the adjacency matrix. For instance, starting from a graph with a certain adjacency matrix, one can build a denser graph just by summing up a new adjacency matrix with only one link. It is a well known problem to predict the magnitude of the eigenvalues of the resulting matrix by starting from the knowledge of the eigenvalues of each matrix separately, and only bounds can be set (as, for instance, those ones given by Weyl inequalities). If we interpret each graph of a multilayer network as a quantum state, the aggregation of two layers (given by the sum of their corresponding adjacency matrices) is equivalent to the superposition of the states in the same space, rather than a mixed state in a space given by the tensorial product of the original spaces. The aggregation itself is a pure network state, whereas it represents a new mixed (entangled) state of the system made by vertices and edges. However, it is not always guaranteed the sub-additivity property of the Von Neumann entropy of the aggregation, i.e.  $h_{A^{[1]}+A^{[2]}} \leq h_{A^{[1]}} + h_{A^{[2]}}$ , making more difficult the interpretation of this measure in terms of information in the context of complex networks.

Here, we show a simple case where sub-additivity can be violated by the superposition of two different network states. The first layer  $A^{[1]}$  we consider is a mixed network state, i.e. an Erdos-Renyi network with 100 nodes and connection probability  $p = 0.5$ ; the second layer  $A^{[2]}$  is a pure network state, i.e a network with only one (undirected) link. We keep constant  $A^{[1]}$  while we vary the position of the link in  $A^{[2]}$ , considering all possible pairs of distinct nodes. For each configuration, we calculate the Von Neumann entropy corresponding to each layer, namely  $h_{A^{[1]}}$  and  $h_{A^{[2]}}$ , the entropy  $h_{A^{[1]}+A^{[2]}}$  of their superposition (i.e. the resulting aggregated network) and the variation  $\Delta H = h_{A^{[1]}} + h_{A^{[2]}} - h_{A^{[1]}+A^{[2]}}$ . In practice,  $h_{A^{[2]}} = 0$  in any configuration, since  $A^{[2]}$  consists of exactly one edge and is thus a pure network state. In the left panel of Supplementary Figure 1 we show the distribution of  $\Delta H$  over all the possible configurations of  $A^{[2]}$ . It is evident that there exist superpositions of states (i.e. aggregation of layers) whose entropy is larger than the sum of the entropy of the two states considered separately, and other superpositions for which instead sub-additivity is preserved. However, it is worth noticing that a violation of the sub-additivity of the Von Neumann entropy is observed only when the two layers have very different edge densities, whereas for networks with the same order of links it is more likely that sub-additivity is instead preserved, as shown in the right panel of Supplementary Figure 1 where we report the distribution of  $\Delta H$  for many realizations of pairs of Erdos-Renyi networks with the same average degree.

This peculiar property of the Von Neumann entropy of a graph is responsible, among other factors, for the increasing (or decreasing) behavior of the entropy per layer of the multilayer network and, as a consequence, of the corresponding behavior of the quality function  $q(\bullet)$ .

In Supplementary Table 1 we report some illustrative examples of the reduction procedure applied to small multilayer networks. In each case we report the entropy of each of the two layers to be merged, the corresponding value of  $\Delta H$  and  $q(\bullet)$  and we indicate if the given aggregation is desirable or not. It is evident that our procedure tends to avoid creating patterns that are not already present in the multilayer network. Moreover, it tends to keep separated the structures observed in each layer even if they are disconnected in the fully aggregated network. Nevertheless, the method correctly

suggests the merge of partially or completely redundant layers.

### Supplementary Note 3

In order to check whether the proposed hierarchical clustering procedure is able to find a good approximation of the configuration of layers yielding the highest quality function, one can compare the optimal value of  $q(\bullet)$  given by the greedy algorithm with the real optimal value of the quality function  $q(\bullet)$  over all the possible partitions of  $M$  layers. We note that, given a set of  $M$  elements (in our case, the layers of the multiplex), the total number of partitions of this set is given by the  $M^{\text{th}}$  Bell number, which is recursively defined as:

$$B_M = \sum_{k=0}^{M-1} \binom{M-1}{k} B_k, \quad B_0 = B_1 = 1. \quad (1)$$

Unfortunately, this number increases super-exponentially with  $M$  (the  $M^{\text{th}}$  Bell number is known to be  $O((M/\log M)^M)$ ), making the enumeration of all the partitions of a set practically unfeasible already for small values of  $M$  (in fact, we have  $B_{10} \sim 10^5$ ,  $B_{15} \sim 1.3 \times 10^9$  and  $B_{20} \sim 5 \times 10^{13}$ ). For this reason, we limited our exploration to all the multiplex networks in the BIOGRID data set, which all have between 3 and 7 layers. In particular, we compared the maximum value of  $q(\bullet)$  obtained by means of hierarchical clustering with the global maximum yielded by the overall best configuration of layers. In Supplementary Table 2 we report the results. Notice that in most of the cases the maximum value of  $q(\bullet)$  obtained through the hierarchical clustering procedure is very close to the actual maximum, and in general the corresponding partition is very similar to the one associated to  $\max q(\bullet)$ . This confirms that the greedy agglomerative procedure, although not perfect, still provides a good trade-off between the efficient exploration of the configuration space and the total number of operations required.

In Supplementary Figure 6 we also report, as an example, the distribution of  $q(\bullet)/\max q(\bullet)$  over all the possible layer partitions of four multi-layer networks, where we indicate with an asterisk the position of the sub-optimal value of  $q(\bullet)$  found through hierarchical clustering.

## Supplementary Methods

### Approximation of the Von Neumann Entropy

The calculation of the Von Neumann entropy requires a considerable computational effort in the case of large multilayer networks, since it involves the computation of the spectrum of the rescaled Laplacian matrix associated to each of the  $M$  layers, which has time complexity  $O(MN^{2.37})$  where  $N$  is the number of nodes in the network. Here we show that it is possible to obtain a good approximation of the Von Neumann entropy by employing an efficient procedure based on the shape of the function  $f(x) = x \log(x)$  and on the fact that all the eigenvalues of the rescaled Laplacian are non-negative and smaller than 1.

Given a graph  $G(V, E)$  with  $N = |V|$  nodes and  $K = |E|$  edges, represented by the adjacency matrix  $A = \{a_{ij}\}$  where  $a_{ij} = 1$  if node  $i$  and node  $j$  are connected through an edge, it is possible to show [4] that the largest eigenvalue  $\ell_{\max}$  of the Laplacian matrix associated to  $A$  satisfies the inequality

$$\ell_{\max} \leq \max_{i,j} \{k_i + k_j; (i, j) \in E\}, \quad (2)$$

where  $k_i = \sum_j a_{ij}$  is the degree of node  $i$ . We recall that the Von Neumann entropy of the graph is defined as

$$h_G = - \sum_{i=1}^N \lambda_i \log_2(\lambda_i).$$

where  $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$  are the eigenvalues of the density matrix  $\hat{\mathcal{L}}$  (the so called rescaled Laplacian), which is obtained by dividing the Laplacian matrix by the quantity

$$\left( \sum_{i \in V} k_i \right) = 2K. \quad (3)$$

The eigenvalues of the rescaled Laplacian satisfy the conditions  $0 \leq \lambda_i \leq 1$  ( $i = 1, 2, \dots, N$ ) and  $\sum_{i \in V} \lambda_i = 1$ . In particular, as a consequence of Eq. (2), we obtain

$$\lambda_{\max} = \lambda_N \leq \frac{\max_{i,j} \{k_i + k_j; (i, j) \in E\}}{\sum_{i \in V} k_i} \leq \frac{k_{\max}}{K}. \quad (4)$$

where  $k_{\max}$  is the largest node degree in the graph. In practice, the calculation of  $h_G$  involves the computation of the function  $f(x) = x \log_2(x)$  in the points  $x = \lambda_i$ , which all lie in the range  $0 \leq x \leq \lambda_{\max} \leq 1$ . Due to the smooth character of  $f(x)$  in  $[0, 1]$ , it is possible to approximate the function  $f(x)$  through an appropriate polynomial of a chosen order  $p$  for  $x \in [0, 1]$ , obtaining the values of the coefficients  $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_p$  such that

$$f(x) \approx \sum_{n=0}^p \alpha_n x^n \quad (5)$$

by means of any standard fitting procedure. Our numerical experiments confirmed that a quite accurate approximation can be obtained already for  $p = 4$  (See Supplementary Figure 5). By inserting the polynomial approximation of  $f(x)$  in the definition of the Von Neumann entropy we obtain

$$h \approx - \sum_{i=1}^N \sum_{n=0}^p \alpha_n \lambda_i^n, \quad (6)$$

and noting that

$$\sum_{i=1}^N \lambda_i^n = \text{Tr}[\hat{\mathcal{L}}^n], \quad n = 1, 2, \dots, \quad (7)$$

we can rewrite the Von Neumann entropy in terms of the traces of the first  $p$  powers of the rescaled Laplacian matrix:

$$h \approx -\alpha_0 N - \sum_{n=1}^p \alpha_n \text{Tr}[\hat{\mathcal{L}}^n]. \quad (8)$$

The time complexity of this calculation is again  $\mathcal{O}(N^{2.37})$ , as in the case of the full spectrum of the Laplacian, but in practice the computation of Eq. (8) is much faster, since it requires exactly  $p$  matrix multiplications.

It is also worth noting that the approximation of  $f(x) = x \log(x)$  can be effectively restricted to the interval  $[0, \lambda_{\max}] \subseteq [0, 1]$ , allowing to obtain a relatively greater accuracy for the same order  $p$  of the fitting polynomial (See inset of Supplementary Figure 5). Actually, the computation of the leading eigenvalue  $\lambda_{\max}$  can be avoided by restricting the fit to the interval  $[0, k_{\max}/K]$  since  $\lambda_{\max} \leq \frac{k_{\max}}{K}$ .

Our numerical experiments showed that by using Eq. (8) one obtains a speed-up of order  $> 10$  over the computation of the full spectrum already for  $N = 5000$  and  $p = 4$ , with an error of less than 1% on the estimated entropy. An error  $< 0.1\%$  is easily obtained in most cases by using  $p = 10$  and restricting the fit to  $[0, k_{\max}/K]$ . We have checked that the impact of this approximated calculation on the layer aggregation method proposed in the main text is negligible.

### Edge overlap and edge intersection in multilayer networks.

In real-world multilayer networks two nodes  $i$  and  $j$  can be connected on more than one of the  $M$  layers. Given a pair of nodes  $(i, j)$  the corresponding *edge overlap* is the fraction of layers in which the edge  $(i, j)$  exists [5]:

$$o_{ij} = \frac{1}{M} \sum_{\alpha=1}^M a_{ij}^{[\alpha]} \quad (9)$$

The aggregated topological network associated to a  $M$ -layer multilayer network is the single-layer graph defined by the adjacency matrix:

$$a_{ij} = \begin{cases} 1 & \text{if } o_{ij} \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

so that an edge exists in the resulting graph between  $i$  and  $j$  if there exists at least one layer  $\alpha$  for which  $a_{ij}^{[\alpha]} \neq 0$ . It is worth remarking that this particular aggregation should not be confused with the aggregation performed to reduce the multilayer network, where the aggregated layers are obtained by summing up the corresponding adjacency matrices. The edge overlap of a multilayer network is defined as the average of  $o_{ij}$  over all the node pairs for which  $a_{ij} = 1$ :

$$\langle o \rangle = \frac{1}{M \sum_{i,j} a_{ij}} \sum_{i,j} o_{ij} = \frac{1}{2MK} \sum_{i,j} o_{ij} \quad (11)$$

Notice that the average edge overlap  $\langle o \rangle$  is equal to 1 only if all the  $M$  layers are identical.

The intersection of two layers of a multilayer network with  $N$  nodes is a single-layer network in which an edge exists between node  $i$  and node  $j$  if and only if  $i$  and  $j$  are connected through an edge on each of the two layers. The *edge-intersection index* is defined as the fraction of edges existing on both layers with respect to the total number of connections in the multilayer. This concept can be straightforwardly extended to more than two layers and can be formalised as

$$I = \frac{M \sum_{i,j=1}^N \min \{a_{ij}^{[1]}, a_{ij}^{[2]}, \dots, a_{ij}^{[M]}\}}{\sum_{\alpha=1}^M \sum_{i,j=1}^N a_{ij}^{[\alpha]}}. \quad (12)$$

The edge overlap and the edge intersection index provide complementary information about the edge persistence pattern across the layers of a multilayer network.

## Supplementary References

- [1] Nicosia, V. & Latora, V. Measuring and modelling correlations in multiplex networks. *Preprint at <http://arxiv.org/abs/1403.1546>* (2014).
- [2] De Domenico, M., Lancichinetti, A., Arenas, A. & Rosvall, M. Identifying modular flows on multi-layer networks reveals highly overlapping organization in social systems. *Phys. Rev. X. (In Press)* (2014).
- [3] De Beaudrap, N., Giovannetti, V., Severini, S. & Wilson, R. Interpreting the von neumann entropy of graph laplacians, and coentropic graphs. *Preprint at <http://arxiv.org/abs/1304.7946>* (2013).
- [4] Anderson, W. & Morley, T. Eigenvalues of the laplacian of a graph. *Lin. Multilin. Algebra* **18**, 141–145 (1985).
- [5] Battiston, F., Nicosia, V. & Latora, V. Structural measures for multiplex networks. *Phys. Rev. E* **89**, 032804 (2014).