ARTICLE

OPEN

Check for updates

Higher-order motif analysis in hypergraphs

Quintino Francesco Lotito [1], Federico Musciotto [2], Alberto Montresor[1] & Federico Battiston [3]✉

A deluge of new data on real-world networks suggests that interactions among system units are not limited to pairs, but often involve a higher number of nodes. To properly encode higher-order interactions, richer mathematical frameworks such as hypergraphs are needed, where hyperedges describe interactions among an arbitrary number of nodes. Here we systematically investigate higher-order motifs, defined as small connected subgraphs in which vertices may be linked by interactions of any order, and propose an efficient algorithm to extract complete higher-order motif profiles from empirical data. We identify different families of hypergraphs, characterized by distinct higher-order connectivity patterns at the local scale. We also propose a set of measures to study the nested structure of hyperedges and provide evidences of structural reinforcement, a mechanism that associates higher strengths of higher-order interactions for the nodes that interact more at the pairwise level. Our work highlights the informative power of higher-order motifs, providing a principled way to extract higher-order fingerprints in hypergraphs at the network microscale.

---

[1] Department of Information Engineering and Computer Science, University of Trento, via Sommarive 9, I-38123 Trento, Italy. [2] Dipartimento di Fisica e Chimica Emilio Segrè, Università di Palermo, Viale delle Scienze, Ed. 18, I-90128 Palermo, Italy. [3] Department of Network and Data Science, Central European University, 1100 Vienna, Austria. ✉email: battistonf@ceu.edu

1

Over the last two decades, networks have emerged as a powerful tool to analyze the complex topology of interacting systems[1]. From social networks to the brain, several systems have been represented as a collection of nodes and links, encoding dyadic interactions among pairs of units. Yet, growing empirical evidence is now suggesting that a large number of such interactions are not limited to pairs, but rather occur in larger groups[2,3]. Examples include collaboration networks[4], human face-to-face interactions[5], species interactions in complex ecosystems[6], cellular networks[7] and structural and functional brain networks[8,9].

To properly encode such higher-order interactions[2,3], richer mathematical frameworks such as hypergraphs[10] are needed, where hyperedges describe interactions taking place among an arbitrary number of nodes. To characterize these higher-order systems[2], computational tools from algebraic topology have been proposed[11,12], as well as generalization of common network concepts, including centrality measures[13,14], directedness[15], clustering[16,17] and assortativity[18]. An explicit treatment of higher-order interactions, including their inference and reconstruction[19], is necessary to understand network formation mechanisms[20–23], fully capture the real community structure of higher-order systems[24–26] and extract their statistically validated higher-order backbone[27]. Noticeably, taking into account higher-order interactions might be crucial to understand the emergent behavior of complex systems, as they have been found to profoundly impact diffusion[28,29], synchronization[30–34], social[35–37] and evolutionary[38] processes.

Networked systems may be differentiated by their preferential patterns of connectivity at the microscale, encoding a characteristic fingerprint often relevant for system functions. This may be quantified by measuring network motifs, small connected subgraphs that appear in an observed network at a frequency that is significantly higher than in a random-graph null model[39]. The analysis of the motifs of a network revealed the emergence of "superfamilies" of networks, i.e., clusters of networks that display similar local structure. These clusters tend to group networks from similar domains or networks that have evolved via similar evolutionary processes[40]. In fact, motifs can be interpreted as elementary computational circuits, with specific functionalities that can be shared by similar networks. For example, transportation networks are designed to simplify the traffic flow, whereas gene regulation and neuron networks are often thought to be evolved to process information. These functional differences in such systems are reflected in the emergence of different significant motifs in the networks that describe them. In this regard, studying motifs can also give new insights into the dynamics and resilience of classes of networks[40,41]. To explicitly uncover the relation between the dynamical processes that unfold on a network and its structural decomposition at the local scale, recently a refined notion of process motifs has been proposed[42], introducing a framework to assess the contribution of each motif to the overall dynamical behavior of the system.

Network motifs have been used in a wide range of applications. In biology, motifs have been extensively studied for the analysis of transcription regulation networks (i.e., networks that control gene expression). Studies show that diverse organisms from bacteria to humans exhibit common regulation patterns, each with its very own function in determining gene expression[43–47]. Similarly, motif analysis has been applied to show how complex and flexible neural functions emerge from the composition of fundamental circuits in brain networks[48]. Moreover, motifs have also been used as a feature for the identification of cancer[49]. Eventually, the need to analyze biological datasets of ever-increasing size has been a strong motivation for the development of more efficient algorithms[50]. Besides biology, motifs have also been applied to

provide fingerprints of the local structures of social networks[51,52], for the early detection of crisis-leading structural changes in financial networks[53] and to study the networks of direct and indirect interactions across species in ecology[54,55].

The interest of the research community in extracting fingerprints at the network microscale of real-world systems has led to considering richer frameworks for motif analysis[56], including extensions to more general network models such as weighted[57], temporal[58] and multilayer[59] networks. Weighted networks can be characterized in terms of the intensity and coherence of the link weights of their subgraphs[60]. Temporal networks can be studied at both topological and temporal micro- and mesoscale by considering time-restricted patterns of interactions[61,62]. Statistically over-expressed small multilayer subgraphs[63] highlight the local structure of multilayer networks such as the human brain[64]. Nevertheless, the methods, algorithms and tools proposed in literature so far mostly consider only patterns of pairwise interactions, thus limiting our capabilities of characterizing the local structure of systems that involve group interactions. Recently, Lee et al.[65] made the first contribution to close this gap: at difference with traditional motif analysis that focuses on patterns of interactions among small sets of nodes, they investigated patterns associated with connected hyperedges, in particular the 26 possible ways in which 3 connected hyperedges can overlap, allowing to extract information on the design principles of hypergraphs.

In order to systematically study the local structure of higher-order networks, here we investigate higher-order network motifs by providing a general and scalable methodology that naturally generalize to hypergraphs the seminal notion and analysis of network motifs proposed by Milo et al.[39] for traditional graphs. Higher-order network motifs are defined as statistically over-expressed connected subgraphs of a given number of nodes, which can be connected by higher-order interactions of arbitrary order. We propose a combinatorial characterization of these new mathematical objects and develop an efficient algorithm to evaluate the statistical significance of each higher-order motif on empirical data. We show that we are able to extract fingerprints at the network microscale of higher-order real-world systems, and highlight the emergence of families of systems that show a similar higher-order local structure. Finally, we propose a set of measures to investigate the nested structure of hyperedges (i.e., the collection of lower-order hyperedges defined on a subset of the nodes of a hyperedge) and provide evidence of the phenomenon of structural reinforcement, for which real-world group interactions are stronger if they are supported by a rich nested structure of pairwise interactions.

Results

Motif analysis has established itself as a fundamental tool in network science to extract fingerprints of networks at the microscale and to identify their structural and functional building blocks. By directly extending the traditional definition of network motifs, we can define higher-order network motifs as small connected patterns of higher-order interactions that appear in an observed hypergraph at a frequency that is significantly higher than a suitably randomized system.

Similarly to what happens with traditional motifs, the steps required to perform a higher-order motif analysis are (i) counting the frequency of each higher-order motif in a network, (ii) comparing the frequency of each motif with that observed in a null model, and (iii) evaluating their over- or under-expression using a statistical measure. Algorithms for counting traditional motifs fail to capture information about group interactions, since they do not consider patterns of hyperedges. A detailed description of our proposal for algorithms and tools able to extract and evaluate higher-order motifs is reported in the Methods section.

For our motif analysis of real-world higher-order systems, we collected a number of freely available networked datasets. The datasets[16,66–82] come from a variety of domains: sociology (proximity contacts, votes), technology (e-mails), biology (gene/disease, drugs) and co-authorship. Each dataset has been manually tagged and associated to a specific domain. The description of each dataset is reported in Supplementary Note 1. In some datasets, higher-order structures are naturally encoded as hyperedges (e.g., three authors collaborating on the same paper), in others we infer higher-order structures from pairwise interactions (e.g., for face-to-face interactions recorded over time, we promote cliques of size k to hyperedges of order k if the corresponding three dyadic encounters happened at the same time. We note that the choice of the specific time-window for aggregation does not affect our results, as presented in Supplementary Note 2.

Combinatorial analysis of higher-order motifs. The number of possible patterns of pairwise undirected interactions involving three connected nodes is only two; however, it grows to six when considering also higher-order interactions (Fig. 1a). Finding an analytical form encoding the dependence of the number of higher-order motifs on the motif order k is a challenging task due to the constraints related to the computation of all possible combinations of higher-order interactions among k nodes. However, we are able to compute upper and lower bounds for this number. We denote with m the number of all the possible non-isomorphic connected hypergraphs of k vertices (we recall that two hypergraphs are isomorphic if they are identical modulo relabeling of the vertices). To compute an upper bound on m, we can count the number of labeled hypergraphs ignoring the constraint on being non-isomorphic and connected. There are $\binom{k}{i}$ possible hyperedges of size i over k vertices. We are interested only in the hyperedges with cardinality at least 2; therefore, there are $\sum_{i=2}^{k} \binom{k}{i} = 2^k - k - 1$ possible hyperedges. When creating a labeled hypergraph we can either include each hyperedge or not, this yields a total number of possible labeled hypergraphs equal to 2^{2^k-k-1}. To compute the lower bound of m, we construct connected hypergraphs on k vertices as follows. First, we pick any chain of edges and put all the edges in the hypergraph. This uses $k - 1$ edges and makes sure the hypergraph is connected. There are $(2^k - k - 1) - (k - 1) = 2^k - 2k$ potential edges left over. Each of those edges can be added or not to the

hypergraph, yielding at least 2^{2^k-2k} connected hypergraphs. However, we have to count only non-isomorphic copies, and have so far counted labeled graphs. For each unlabeled graph, there are at most $k!$ ways of labeling the vertices. So the number of non-isomorphic connected hypergraph is at least $\frac{2^{2^k-2k}}{k!}$. Figure 1b shows the upper and lower bounds on the growth of the possible higher-order motifs as a function of the order, as well as the exact count for small orders, showing that this function has a super-exponential growth. The combinatorial explosion of higher-order motifs makes their storing and indexing in memory (required steps for counting their occurrences in empirical hypergraphs and evaluating their over- or under-expression) intractable for high orders. Given these combinatorial difficulties, in the following, we focus on the analysis of the higher-order motifs of order 3 and 4.

Motifs of order 3. The over- and under-expression measures of each higher-order motif (abundance with respect to a null model, see Methods) in a hypergraph are concatenated in a significance profile (SP, see Methods) that constitutes a fingerprint of the local structure of the network. In this section, we characterize the local connectivity of empirical networks at the smallest scale, with higher-order motifs of order 3.

After having computed the SPs of all the datasets, a first question one could ask is how hypergraphs from different domains differ on average in their SPs. We compute the SPs of a domain by grouping and averaging the SPs of all networks that belong to it (more information about the disaggregated SPs can be found in Supplementary Note 4). The analysis of the higher-order profiles of order 3 of each domain highlights the relative structural importance of certain patterns of higher-order interactions (Fig. 2a). The pairwise triangle II appears to be a highly over-expressed motif in all the domains, whereas the greatest differences across domains emerge from motifs that involve a 3-hyperedge and at least one dyadic edge. In the social and technological domains, the motif VI made by a 3-hyperedge and a triangle of dyadic edges is highly over-expressed, suggesting that entities interacting in groups also tend to interact individually. In co-authorship networks, the most over-expressed motifs are IV and V, which involve a 3-hyperedge and one or two dyadic edges, indicating that in these domains there might be a hierarchical structure that prevents all nodes from interacting equally in pairs, as in the case of a research leader that co-authors papers with students and postdocs while the latter do not co-author papers
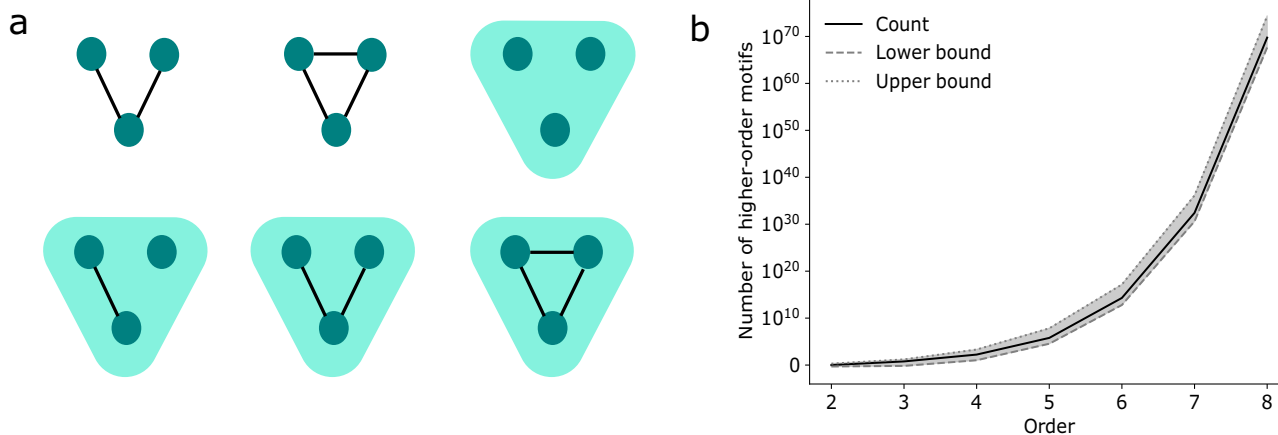
Fig. 1 Combinatorics of higher-order motifs. a Enumeration of all the six possible patterns of higher-order interactions involving three nodes. Green shaded triangles represent higher-order interactions, whereas black lines represent pairwise interactions. **b** Upper and lower bounds on the number of higher-order motifs as a function of the order (gray shaded area). The black line represents the exact count for small orders.
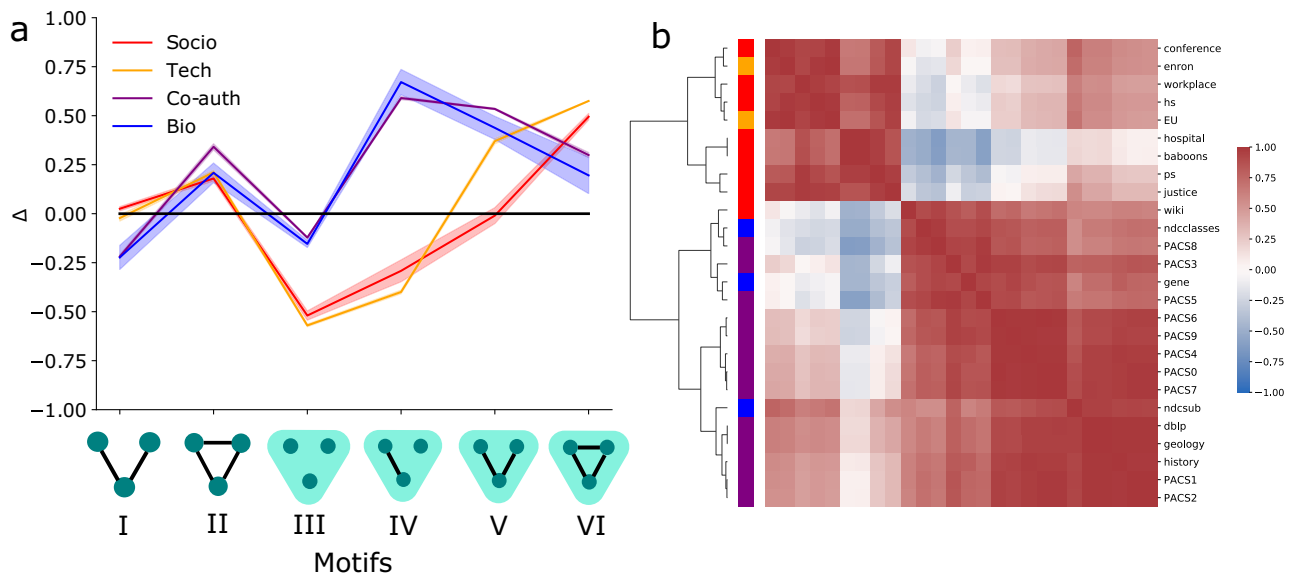
Fig. 2 A higher-order fingerprint for hypergraphs at the network microscale. a Significance profiles (SP) of hypergraphs from higher-order motifs of order 3 (labeled I–VI). Δ is the abundance of each motif relative to random networks. Over-expressed higher-order motifs are associated to specific functionalities of the system. To simplify the plot, we averaged and grouped higher-order motif profiles of networks from the same domain. For each domain, we represent the mean of the respective higher-order motif profiles with a solid line and the standard error of the mean with a shaded area. **b** Correlation matrix of the investigated datasets computed on the SPs. SPs of networks from similar domains display a positive correlation. We identify two large higher-order families of hypergraphs, characterized by distinct higher-order connectivity patterns at the local scale. Each row of the correlation matrix is labeled with different colors depending on the domain of the respective dataset: red for the social domain, orange for e-mails, purple for the co-authorship domain and blue for the biological domain. Moreover, we show the clustering tree computed by applying a hierarchical clustering algorithm on the significance profiles, considering correlation as a measure of similarity. The clustering tree highlights the hierarchical organization of the emerged clusters. In the correlation matrix, red squares represent high positive correlation while blue squares represent high negative correlation.

without the former. A similar motif is also found to be over-expressed in biological systems. Moreover, SPs allow also to analyze anti-motifs, i.e., motifs that are highly under-expressed. An anti-motif in the social and technological domains is III, the 3-hyperedge without any dyadic interaction, indicating that it is unlikely that an interaction in the group is not followed or preceded by any pairwise interaction. The biological and co-authorship domain do not display any anti-motif.

Another interesting question is whether the domain categorization naturally emerges from individually clustering the SPs of all the empirical hypergraphs. We perform a hierarchical cluster analysis considering the pairwise correlation between the distributions of the occurrences of the higher-order motifs for each dataset as (the inverse of) a distance (Fig. 2b). The analysis shows the emergence of two main clusters, i.e., families of higher-order networks that share similar patterns of higher-order interactions at the microscale. The clusters, here inferred in a purely data-driven manner, reproduce the partitions of domains displayed in Fig. 2a (social and technological datasets in a cluster, biological and co-authorship ones in the other), offering a more nuanced view on the similarity across datasets.

Motifs of order 4. In the previous section, we have systematically investigated the smallest higher-order motifs. The number of possible patterns of higher-order interactions involving 4 nodes is significantly higher than the corresponding with 3 nodes, as it grows from 6 to 171. Despite the difficulties associated to this increase, analyzing higher-order motifs of order 4 provides more nuanced information about the local structure of networks compared to 3-motifs.

In Fig. 3a, we group together similar domains based on the analysis in the previous section showing the average of their SPs

with the higher-order motifs of order 4. The order of motifs along the x-axis maximizes the visual difference in SPs across clusters. On the left end of the x-axis, we find motifs that are highly over-expressed in the Bio/Co-auth domain, while they are under-expressed in the Socio/Tech domain. Conversely, on the right end of the x-axis, we find motifs that are over-expressed in the Socio/Tech domain, while not characteristic for the other domain. This observation suggests that both the extremes of the x-axis carry information about the structural differences among the clusters.

The richer structural information captured by the higher-order motifs of order 4 compared to their counterparts of order 3 is highlighted in the clustering analysis (Fig. 3b). When focusing on the two main clusters, the results are comparable with the previous cluster analysis. However, a richer hierarchical intra-cluster organization naturally emerges, as well as a better separation between the two clusters (See Supplementary Note 3).

Finally, we characterize the Socio/Tech and the Bio/Co-auth clusters by means of their most over-expressed, and therefore most representative, higher-order motifs of order 4 (Fig. 3c). The Socio/Tech domain shows an over-expression of structures involving more lower-order nested relations (e.g., dyadic links), while the Bio/Co-auth domain displays a preference toward less relations but of higher-order. This pattern might be caused by the fact that people interacting in groups are likely to interact also in single pairs, therefore it is plausible that group interactions in the Socio/Tech domain are supported by a large number of lower-level interactions. On the other hand, people tend to write papers in large groups and tend to maintain the same research group over time, with few additions or removals. Therefore, patterns involving only dyadic relations are penalized. For a more in-depth description of the most over- and under-expressed higher-order motifs of order 4, we refer to Supplementary Note 5.
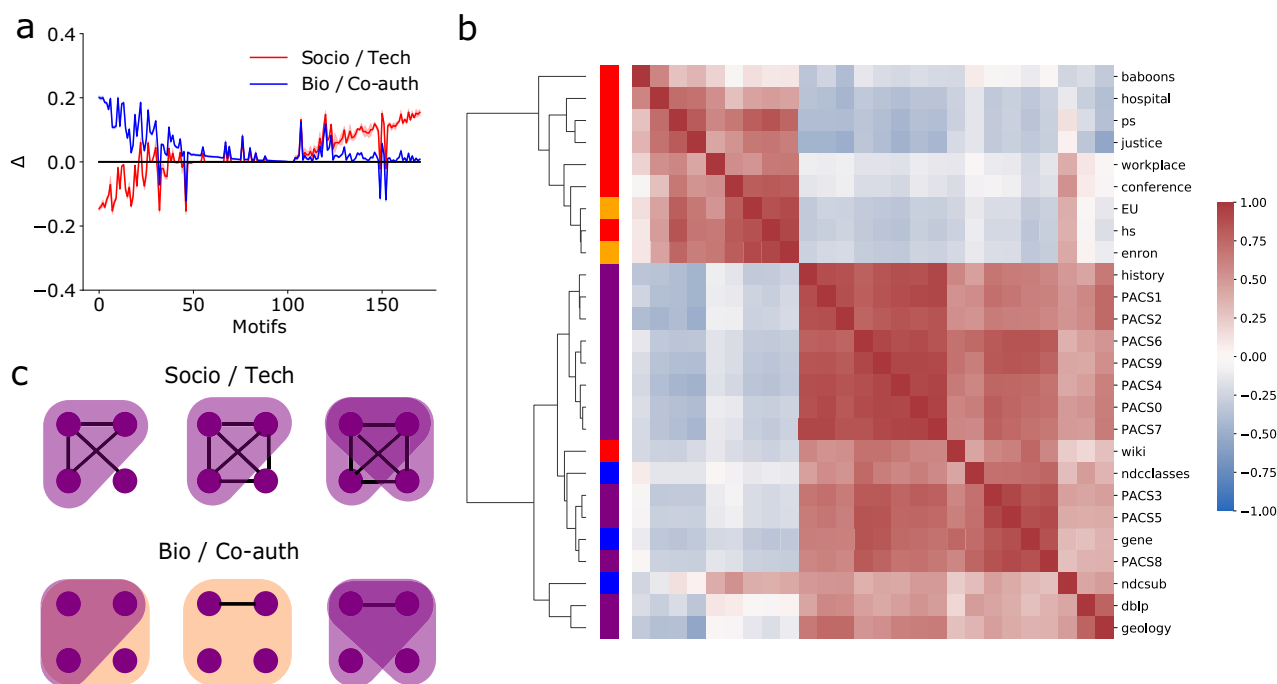
Fig. 3 Analyzing the local structure of hypergraphs via higher-order motifs of order 4. a Significance Profiles (SP) of hypergraphs from higher-order motifs of order 4. Δ is the abundance of each motif relative to random networks. SPs are much more complex due to the increase in the number of considered patterns of higher-order interactions. We group and average the SPs of networks from the same higher-order family (i.e., Socio/Tech and bio/ Co-auth) and sort the motifs on the x-axis based on their ability to discriminate the two higher-order families. Distinct characteristic higher-order motifs of order 4 are associated with the two classes of networks. The shaded area represents the standard error of the mean. If the shaded area is not visible, it is of the same size as the line thickness. **b** Correlation matrix of the investigated datasets computed on SPs of order 4. The matrix provides richer information than its equivalent at order 3 on the local structure of networks: the two big clusters emerge again but are better separated, and display a richer intra-cluster hierarchical structure. Each row of the correlation matrix is labeled with different colors depending on the domain of the respective dataset: red for the social domain, orange for e-mails, purple for the co-authorship domain and blue for the biological domain. Moreover, we show the clustering tree computed by applying a hierarchical clustering algorithm on the significance profiles, considering correlation as a measure of similarity. With respect to the analysis with higher-order motifs of order 3, the clustering tree highlights a better separation between the two big clusters, as well as a richer intra-cluster hierarchical organization. In the correlation matrix, red squares represent a high correlation while blue squares represent a low correlation. **c** The six most representative higher-order motifs from the two clusters. Purple shaded triangles and orange shaded squares represent respectively higher-order interactions of size 3 and 4, whereas black lines represent pairwise interactions.

Nested organization of higher-order interactions. We now turn our attention to characterize the nested structure of large higher-order hyperedges. We define the nested structure of a large hyperedge h as the collection of hyperedges existing on a subset of the nodes of h, and extract statistics on the nested structure of hyperedges of any size. The advantage of this approach is that it still provides information about the local structure of sub-modules of a network, while its computational complexity is only linear in the number of hyperedges in the hypergraph.

First, we consider the average number of edges in the nested structures of hyperedges of different sizes (Fig. 4a). The networks are grouped according to their domain. While biological and co-authorship networks do not display evident differences in the number of nested edges with the growth of the hyperedge size, social and technological networks show a clear growing trend with a change of slope after orders 5 and 6.

In order to complement this information, we looked at how the mean size of the nested edges changes with the growth of the size of the analyzed hyperedges (Fig. 4b). In this case, all the domains show a growing trend, with biological and co-authorship networks displaying a faster growth. Thus, while social and technological networks tend to have more edges in the nested structure of their large hyperedges, they tend to be of small size. Biological and co-authorship networks, instead, shows an opposite behavior. All in all, this suggests that, in agreement with our previous findings, also at higher scales Socio/Tech network motifs are systematically more nested.

Higher-order motifs and reinforcement. In order to understand if and how the occurrence of nested dyadic interactions affects the strength of group interactions, we investigate how much the weight of each hyperedge (i.e., the number of times each group interaction occurs) is correlated with the number of nested pairwise links. We find that a positive trend emerges, indicating the existence of a correlation between a rich nested pairwise structure and the weight of a hyperedge (Fig. 5a). We dubbed this phenomenon, similar to the one highlighted in ref. [59] for multi-layer networks, as higher-order structural reinforcement.

Moreover, we used the metadata about personal relationships between students recorded in the High School dataset from SocioPatterns to understand if similar reinforcing behavior is observed in the presence of friendship interactions between individuals. Friendship data have been collected in two ways, from Facebook accounts and through a questionnaire. In the first case, two students are always reciprocally friends, while in the second case a friendship can be unreciprocated. In Fig. 5b, we analyze the relationship between the average number of friends (both on Facebook and by questionnaire) and the topology of the different motifs in the proximity hypergraph. Our results show
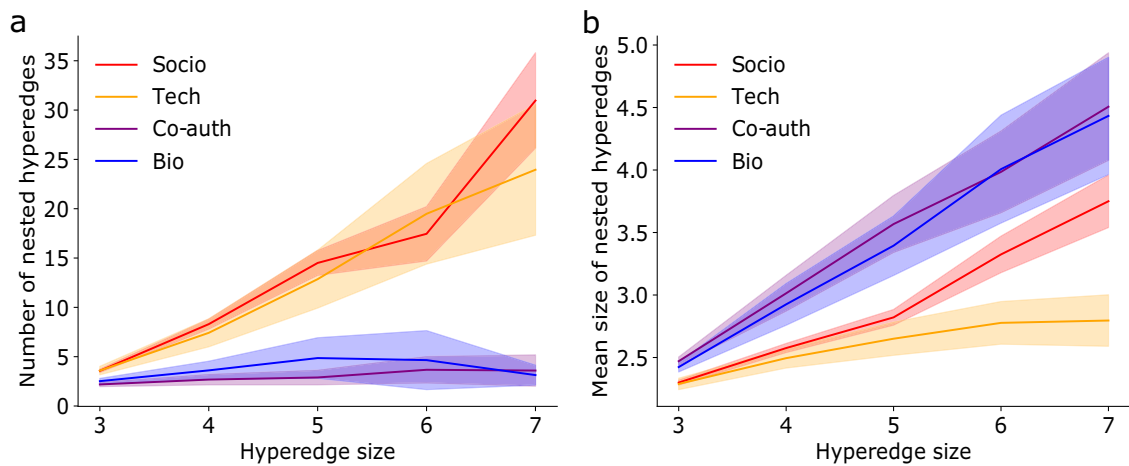
Fig. 4 Nested organization of group interactions. Different higher-order families of hypergraphs can display very different hierarchical organization of their higher-order interactions. **a** Mean number of hyperedges in the nested structure of large hyperedges as a function of their size. Biological and co-authorship networks display a static behavior, while social and technological networks show a clear increasing richness of the nested hierarchical structures of the hyperedges. **b** Mean average size of the hyperedges in the nested structure of large hyperedges as a function of their size. All the domains show a linear growing trend; however, biological and co-authorship networks grow faster. All in all, Socio/Tech networks tend to have a lot of small-size edges in the nested structure of their hyperedges. The Bio/Co-auth domain, instead, tend to prefer few large-size edges. In both panels, the shaded area represents the standard deviation.
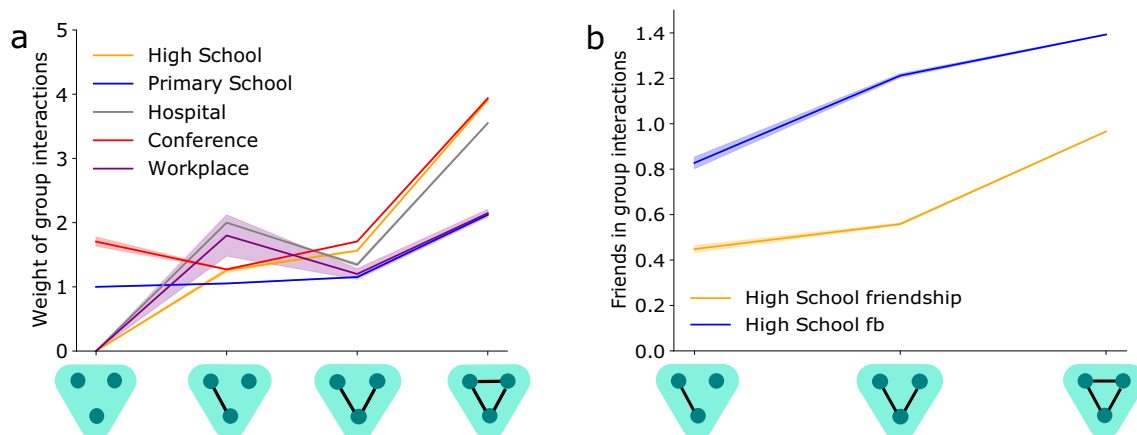
Fig. 5 Structural reinforcement. A rich supporting nested structure of pairwise links makes group interactions stronger. In both panels, stronger levels of connectivity are observed when the number of dyadic interactions increases. **a** Mean weight of each group interaction (i.e., the number of times each group interaction occurs) as a function of the number of its nested pairwise links. **b** Mean number of friends (certified by a Facebook friendship or by a questionnaire) in group interactions as a function of the number of their nested pairwise links. In both panels, the shaded area represents the standard error of the mean.

that the higher the number of pairwise interactions between students that interact in hyperedges of size three, the higher will be the number of friends in the group, further suggesting the existence of reinforcement mechanisms.

Discussion
The framework of network motifs is widely recognized as a fundamental tool for the analysis of complex networks. Able to highlight local structural characteristics of networks and influence their dynamics, motifs can be considered the fundamental building blocks of networks, and have produced applications in a number of fields such as biology and social network analysis.

Modeling complex systems by means of hypergraphs have recently emerged as a fundamental tool in Network Science, prompting the question of how to identify and assess network motifs in the presence of higher-order interactions. With the aim of extracting the local fingerprint of hypergraphs, in this work we introduced the notion of higher-order network motifs, which are

small, possibly overlapping patterns of higher-order interactions that are statistically over-expressed with respect to a null model. We proposed a combinatorial characterization of higher-order network motifs, as well as an efficient algorithm to evaluate their statistical significance on empirical data. These tools allowed us to extract fingerprints of a variety of real-world systems by focusing on their characteristic patterns of higher-order interactions among small groups of nodes, showing the emergence of families of hypergraphs characterized by similar local structures. Moreover, we proposed a set of measures to study the nested structure of hyperedges and provided evidence of a structural reinforcement mechanism that associates stronger weights of higher-order interactions to groups of nodes that interact more at the pairwise level.

Similarly to the case of traditional pairwise network motifs, we believe that higher-order network motifs can pave the way to applications in a number of domains, pushed by the growing awareness of the relevance of the higher-order nature of interactions in many real-world systems. Given the possible applications

of this framework in data-intensive domains, a limitation of our proposed approach is its scalability. In this work, indeed we proposed an algorithm that allows us to perform an exhaustive search, and for this reason, focuses on higher-order network motifs of size 3 and 4. However, we believe that there is room for different approaches, which sacrifices exhaustiveness but could allow us to gain deep insights on motifs of greater size. As the first step in this direction, we looked at the nested structure of patterns of hyperedges of larger orders. In addition to this, we believe that the development of sampling methods for the statistical evaluation of higher-order network motifs will be critical for more widespread real-world applications. All in all, our work highlights the informative power of higher-order motifs, providing an initial approach to extract higher-order fingerprints in hypergraphs at the network microscale.

Methods

A higher-order motif analysis involves three steps: (i) counting the frequency of each target higher-order motif in an observed network, (ii) comparing them with those of a null model, and (iii) establish the over- or under-expression of certain sub-hypergraph patterns.

Here, we propose an exact algorithm to count the frequency of each higher-order motif of order k in a hypergraph. The first fundamental sub-task to solve efficiently is the hypergraph isomorphism problem (i.e., establishing the equivalence under relabeling of two hypergraphs). In fact, for each occurrence of a connected sub-hypergraph with k nodes, we need to update the frequency of the respective higher-order motif of order k. This problem can be solved efficiently by enumerating and indexing all the higher-order motifs of order k with all the respective relabelings, allowing to update and count occurrences of patterns of sub-hypergraphs in constant time via a hash map. Since we are interested only in patterns of size 3 and 4, this is doable. In fact, the number of possible non-isomorphic patterns of higher-order interactions involving 4 nodes is 171, a number that makes all the relabelings storable in memory.

To enumerate sub-hypergraphs of size k we use an algorithm that proceeds in a hierarchical way. It first iterates over all the hyperedges of size k, which are able to directly induce a motif, i.e., a hyperedge of size k gives all the nodes to construct a motif of order k. Then it iteratively considers hyperedges of lower orders until it reaches the traditional dyadic links. Since hyperedges of order lower than k are not able to directly induce a motif, the algorithm proceeds in a way similar to[83] and selects the remaining nodes by considering the neighborhood of the sub-hypergraph. Once selected k nodes, to efficiently construct their induced sub-hypergraph, we iterate over the power set of the k nodes (which corresponds to 2^k possible hyperedges) and keep only the hyperedges that exist in the original hypergraph.

As a null model, we use the configuration model proposed by Chodrow[21]. We sample from the configuration model $n = 100$ times and compute the frequencies of the higher-order motifs in each sample. To validate the over- and under-expression of certain patterns, we use the abundance Δ_i of each motif i relative to random networks proposed in[40],

$$\Delta_i = \frac{N\mathrm{real}_i - \langle N\mathrm{rand}_i \rangle}{N\mathrm{real}_i + \langle N\mathrm{rand}_i \rangle + \epsilon} \tag{1}$$

Following[40], we set $\epsilon = 4$.

We define the SP of a network as the vector of Δ_i normalized to length 1,

$$\mathrm{SP} = \frac{\Delta_i}{\sqrt{\sum \Delta_i^2}} \tag{2}$$

Data availability

The datasets analyzed in this paper are available at https://github.com/FraLotito/higher-order-motifs.

Code availability

The code for higher-order motif analysis is available at https://github.com/FraLotito/higher-order-motifs.

References

1. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006).

2. Battiston, F. et al. Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.* **874**, 1–92 (2020).

3. Battiston, F. et al. The physics of higher-order interactions in complex systems. *Nat. Phys.* **17**, 1093–1098 (2021).

4. Patania, A., Petri, G. & Vaccarino, F. The shape of collaborations. *EPJ Data Sci.* **6**, 18 (2017).

5. Cencetti, G., Battiston, F., Lepri, B. & Karsai, M. Temporal properties of higher-order interactions in social networks. *Sci. Rep.* **11**, 7028 (2021).

6. Grilli, J., Barabás, G., Michalska-Smith, M. J. & Allesina, S. Higher-order interactions stabilize dynamics in competitive network models. *Nature* **548**, 210 (2017).

7. Klamt, S., Haus, U.-U. & Theis, F. Hypergraphs and cellular networks. *PLOS Comput. Biol.* **5**, e1000385 (2009).

8. Petri, G. et al. Homological scaffolds of brain functional networks. *J. R. Soc. Interface* **11**, 20140873 (2014).

9. Giusti, C., Ghrist, R. & Bassett, D. S. Two's company, three (or more) is a simplex. *J. Comput. Neurosci.* **41**, 1–14 (2016).

10. Berge, C. *Graphs and Hypergraphs* (North-Holland Pub. Co., 1973).

11. Patania, A., Vaccarino, F. & Petri, G. Topological analysis of data. *EPJ Data Sci.* **6**, 7 (2017).

12. Sizemore, A. E., Phillips-Cremins, J. E., Ghrist, R. & Bassett, D. S. The importance of the whole: topological data analysis for the network neuroscientist. *Netw. Neurosci.* **3**, 656–673 (2019).

13. Estrada, E. & Rodríguez-Velázquez, J. A. Subgraph centrality and clustering in complex networks. *Phys. A* **364**, 581–594 (2006).

14. Benson, A. R. Three hypergraph eigenvector centralities. *SIAM J. Math. Data Sci.* **1**, 293–312 (2019).

15. Chodrow, P. & Mellor, A. Annotated hypergraphs: models and applications. *Appl. Netw. Sci.* **5**, 9 (2020).

16. Benson, A. R., Abebe, R., Schaub, M. T., Jadbabaie, A. & Kleinberg, J. Simplicial closure and higher-order link prediction. *Proc. Natl Acad. Sci. USA* **115**, E11221–E11230 (2018).

17. Yin, H., Benson, A. R. & Leskovec, J. Higher-order clustering in networks. *Phys. Rev. E* **97**, 052306 (2018).

18. Veldt, N., Benson, A. R. & Kleinberg, J. Higher-order homophily is combinatorially impossible. *arXiv*. Preprint at arXiv:2103.11818 (2021).

19. Young, J.-G., Petri, G. & Peixoto, T. P. Hypergraph reconstruction from network data. *Commun. Phys.* **4**, 1–11 (2021).

20. Courtney, O. T. & Bianconi, G. Generalized network structures: the configuration model and the canonical ensemble of simplicial complexes. *Phys. Rev. E* **93**, 062311 (2016).

21. Chodrow, P. S. Configuration models of random hypergraphs. *J. Complex Netw.* **8**, cnaa018 (2020).

22. Kovalenko, K. et al. Growing scale-free simplices. *Commun. Phys.* **4**, 1–9 (2021).

23. Millán, A. P., Ghorbanchian, R., Defenu, N., Battiston, F. & Bianconi, G. Local topological moves determine global diffusion properties of hyperbolic higher-order networks. *Phys. Rev. E* **104**, 054302 (2021).

24. Carletti, T., Fanelli, D. & Lambiotte, R. Random walks and community detection in hypergraphs. *J. Phys.: Complex.* **2**, 015011 (2021).

25. Eriksson, A., Edler, D., Rojas, A., de Domenico, M. & Rosvall, M. How choosing random-walk model and network representation matters for flow-based community detection in hypergraphs. *Commun. Phys.* **4**, 1–12 (2021).

26. Chodrow, P. S., Veldt, N. & Benson, A. R. Generative hypergraph clustering: from blockmodels to modularity. *Sci. Adv.* **7**, eabh1303 (2021).

27. Musciotto, F., Battiston, F. & Mantegna, R. N. Detecting informative higher-order interactions in statistically validated hypergraphs. *Commun. Phys.* **4**, 218 (2021).

28. Schaub, M. T., Benson, A. R., Horn, P., Lippner, G. & Jadbabaie, A. Random walks on simplicial complexes and the normalized Hodge Laplacian. *SIAM Rev.* **62**, 353–391 (2020).

29. Carletti, T., Battiston, F., Cencetti, G. & Fanelli, D. Random walks on hypergraphs. *Phys. Rev. E* **101**, 022308 (2020).

30. Bick, C., Ashwin, P. & Rodrigues, A. Chaos in generically coupled phase oscillator networks with nonpairwise interactions. *Chaos* **26**, 094814 (2016).

31. Skardal, P. S. & Arenas, A. Higher-order interactions in complex networks of phase oscillators promote abrupt synchronization switching. *Commun. Phys.* **3**, 218 (2020).

32. Millán, A. P., Torres, J. J. & Bianconi, G. Explosive higher-order kuramoto dynamics on simplicial complexes. *Phys. Rev. Lett.* **124**, 218301 (2020).

33. Lucas, M., Cencetti, G. & Battiston, F. Multiorder laplacian for synchronization in higher-order networks. *Phys. Rev. Res.* **2**, 033410 (2020).

34. Gambuzza, L. et al. Stability of synchronization in simplicial complexes. *Nat. Commun.* **12**, 1–13 (2021).

35. Iacopini, I., Petri, G., Barrat, A. & Latora, V. Simplicial models of social contagion. *Nat. Commun.* **10**, 2485 (2019).

36. Chowdhary, S., Kumar, A., Cencetti, G., Iacopini, I. & Battiston, F. Simplicial contagion in temporal higher-order networks. *J. Phys.: Complexity* **2** (2021).

37. Neuhäuser, L., Schaub, M. T., Mellor, A. & Lambiotte, R. Opinion Dynamics with Multi-body Interactions. In *NetGCooP* (2020).

38. Alvarez-Rodriguez, U. et al. Evolutionary dynamics of higher-order interactions in social networks. *Nat. Hum. Behav.* **5**, 586–595 (2021).

39. Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).

40. Milo, R. et al. Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542 (2004).

41. Dey, A. K., Gel, Y. R. & Poor, H. V. What network motifs tell us about resilience and reliability of complex networks. *Proc. Natl Acad. Sci. USA* **116**, 19368–19373 (2019).

42. Schwarze, A. C. & Porter, M. A. Motifs for Processes on Networks. *SIAM J. Appl. Dyn. Syst.* **20**, 2516–2557 (2021).

43. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450 (2007).

44. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64 (2002).

45. Mazurie, A., Bottani, S. & Vergassola, M. An evolutionary and functional assessment of regulatory network motifs. *Genome Biol.* **6**, 1–12 (2005).

46. Dobrin, R., Beg, Q. K., Barabási, A.-L. & Oltvai, Z. N. Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network. *BMC Bioinforma.* **5**, 1–8 (2004).

47. Yeger-Lotem, E. et al. Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proc. Natl Acad. Sci. USA* **101**, 5934–5939 (2004).

48. Sporns, O. & Kötter, R. Motifs in brain networks. *PLoS Biol.* **2**, e369 (2004).

49. Chen, L. et al. Identification of breast cancer patients based on human signaling network motifs. *Sci. Rep.* **3**, 1–7 (2013).

50. Patra, S. & Mohapatra, A. Review of tools and algorithms for network motif discovery in biological networks. *IET Syst. Biol.* **14**, 171–189 (2020).

51. Hong-lin, X., Han-bing, Y., Cui-fang, G. & Ping, Z., Social network analysis based on network motifs. *J. Appl. Math.* **2014**, 874708 (2014).

52. Juszczyszyn, K., Kazienko, P. & Musiał, K. Local topology of social network based on motif analysis. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 97–105 (Springer, 2008).

53. Saracco, F., Di Clemente, R., Gabrielli, A. & Squartini, T. Detecting early signs of the 2007–2008 crisis in the world trade. *Sci. Rep.* **6**, 1–11 (2016).

54. Bascompte, J. & Stouffer, D. B. The assembly and disassembly of ecological networks. *Philos. Trans. R. Soc. B: Biol. Sci.* **364**, 1781–1787 (2009).

55. Simmons, B. I. et al. Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos* **128**, 154–170 (2019).

56. Benson, A. R., Gleich, D. F. & Leskovec, J. Higher-order organization of complex networks. *Science* **353**, 163–166 (2016).

57. Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proc. Natl Acad. Sci. USA* **101**, 3747–3752 (2004).

58. Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519**, 97–125 (2012).

59. Battiston, F., Nicosia, V. & Latora, V. Structural measures for multiplex networks. *Phys. Rev. E* **89**, 032804 (2014).

60. Onnela, J.-P., Saramäki, J., Kertész, J. & Kaski, K. Intensity and coherence of motifs in weighted complex networks. *Phys. Rev. E* **71**, 065103 (2005).

61. Kovanen, L., Karsai, M., Kaski, K., Kertész, J. & Saramäki, J. Temporal motifs in time-dependent networks. *J. Stat. Mech. Theory Exp.* **2011**, P11005 (2011).

62. Paranjape, A., Benson, A. R. & Leskovec, J. Motifs in temporal networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 601–610 (ACM, 2017).

63. Kivelä, M. & Porter, M. A. Isomorphisms in multilayer networks. *IEEE Trans. Netw. Sci. Eng.* **5**, 198–211 (2018).

64. Battiston, F., Nicosia, V., Chavez, M. & Latora, V. Multilayer motif analysis of brain networks. *Chaos: Interdisc. J. Nonlinear Sci.* **27**, 047404 (2017).

65. Lee, G., Ko, J. & Shin, K. Hypergraph motifs: concepts, algorithms, and discoveries. *Proc. VLDB Endow.* **13**, 2256–2269 (2020).

66. Epstein, L., Walker, T. G., Hendrickson, N. S. & Roberts, J. *The U.S. Supreme Court Justices Database* (2019).

67. Physical Review Journals. APS Data Sets for Research. https://journals.aps.org/datasets (2021).

68. Mastrandrea, R., Fournet, J. & Barrat, A. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE* **10**, 1–26 (2015).

69. Chaintreau, A. et al. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Trans. Mob. Comput.* **6**, 606–620 (2007).

70. Kunegis, J. KONECT: The Koblenz Network Collection. In *Proceedings of the 22nd International Conference on World Wide Web* 1343–1350 (Association for Computing Machinery, 2013).

71. Génois, M. & Barrat, A. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Sci.* **7**, 11 (2018).

72. Gelardi, V., Godard, J., Paleressompoulle, D., Claidière, N. & Barrat, A. Measuring social networks in primates: wearable sensors versus direct observations. *Proc. R. Soc. A: Math., Phys. Eng. Sci.* **476**, 20190737 (2020).

73. Sinha, A. et al. An overview of microsoft academic service (MAS) and applications. In *Proceedings of the 24th International Conference on World Wide Web* (ACM Press, 2015).

74. Leskovec, J., Huttenlocher, D. & Kleinberg, J. *Signed Networks in Social Media*, 1361–1370 (Association for Computing Machinery, New York, NY, USA, 2010).

75. Leskovec, J., Huttenlocher, D. & Kleinberg, J. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, 641–650 (Association for Computing Machinery, New York, NY, USA, 2010).

76. Piñero, J. et al. The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **48**, D845–D855 (2019).

77. Piñero, J. et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2016).

78. Piñero, J. et al. Disgenet: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* **2015**, bav028 (2015).

79. Queralt-Rosinach, N., Piñero, J., Serrano, A., Sanz, F. & Furlong, L. I. DisGeNET-RDF: harnessing the innovative power of the semantic web to explore the genetic basis of diseases. *Bioinformatics* **32**, 2236–2238 (2016).

80. Bauer-Mehren, A. et al. Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. *PLoS ONE* **6**, 1–13 (2011).

81. Génois, M. et al. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. *Netw. Sci.* **3**, 326–347 (2015).

82. Vanhems, P. et al. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS ONE* **8**, 1–9 (2013).

83. Wernicke, S. Efficient detection of network motifs. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **3**, 347–359 (2006).

Author contributions
Q.F.L., F.M., A.M., and F.B. designed research; Q.F.L. analyzed data, developed the algorithms and performed the computations; Q.F.L., F.M., A.M., and F.B. analyzed results and wrote the paper.

Competing interests
The authors declare no competing interest. F.B. is an Editorial Board Member for *Communications Physics*, but was not involved in the editorial review of, or the decision to publish this article.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-022-00858-7.

Correspondence and requests for materials should be addressed to Federico Battiston.

Peer review information *Communications Physics* thanks Anton Eriksson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.