

Detecting informative higher-order interactions in statistically validated hypergraphs

Federico Musciotto ^{1✉}, Federico Battiston ² & Rosario N. Mantegna ^{1,3}

Recent empirical evidence has shown that in many real-world systems, successfully represented as networks, interactions are not limited to dyads, but often involve three or more agents at a time. These data are better described by hypergraphs, where hyperlinks encode higher-order interactions among a group of nodes. In spite of the extensive literature on networks, detecting informative hyperlinks in real world hypergraphs is still an open problem. Here we propose an analytic approach to filter hypergraphs by identifying those hyperlinks that are over-expressed with respect to a random null hypothesis, and represent the most relevant higher-order connections. We apply our method to a class of synthetic benchmarks and to several datasets, showing that the method highlights hyperlinks that are more informative than those extracted with pairwise approaches. Our method provides a first way, to the best of our knowledge, to obtain statistically validated hypergraphs, separating informative connections from noisy ones.

¹Dipartimento di Fisica e Chimica Emilio Segrè, Università di Palermo, Viale delle Scienze, Ed. 18, I-90128 Palermo, Italy. ²Department of Network and Data Science, Central European University, 1100 Vienna, Austria. ³Complexity Science Hub Vienna, Josefstädter Strasse 39, 1080 Vienna, Austria.
✉email: musci8@gmail.com

Over the last years, advances in technology have made available a deluge of new data on biological and socio-technical systems, which have helped scientists build more efficient and precise data-informed models of the world we live in. Most of these data sources, from online social networks to the world trade web and the human brain, have been fruitfully represented as graphs, where nodes describe the units of the systems, and links encode their pairwise interactions¹. Yet, prompted by new empirical evidence, it is now clear that in most real-world systems interactions are not limited to pairs, but often involve three or more agents at the same time². In our brain, neurons communicate through complex signals involving multiple partners at the same time^{3,4}. In nature, species co-exist and compete following an intricate web of relationships that can not be understood by considering pairwise interactions only⁵. In science, most advances are achieved by combining the expertise of multiple individuals in the same team⁶.

To fully keep into account the higher-order organization of real networks, new mathematical frameworks have been proposed, rapidly becoming widespread in the last few years. Computational techniques from algebraic topology have made possible to extract the “shape” of the data, investigating the topological features associated with the existence of higher-order interactions from social networks to the brain^{7,8}. In parallel, traditional network measures have been generalized to account for the existence of non-pairwise interactions. This includes new proposals for centrality measures^{9,10}, community structure¹¹, and simplicial closure, which is a generalization of clustering coefficient to higher-order interactions¹². The temporal evolution of higher-order social networks has been investigated, showing the presence of nontrivial correlations and burstiness at all orders of interactions¹³. Besides, explicitly considering the higher-order structure of real-world systems has led to the discovery of new collective phenomena and dynamical behavior, from social contagion^{14–16} and human cooperation¹⁷ to models of diffusion^{18,19} and synchronization^{20–23}.

Among the several frameworks, hypergraphs, collections of nodes, and hyperlinks encoding interactions among any number of units represent the most natural generalization of traditional networked structures to explicitly consider systems beyond pairwise interactions^{2,24}. However, mapping data and mathematical frameworks present us with some new challenges. For instance, for some systems, higher-order interactions might be difficult to observe, or only be recorded as a collection of pairwise data. To overcome this limitation, recent work has developed a Bayesian framework to reconstruct higher-order connections from simple pairwise interactions following a principle of parsimony²⁵.

In spite of the explosion of new methods to analyze systems interacting at higher orders, a filtering technique working for hypergraphs is not yet available. Filtering techniques are a relatively recent addition to network analysis. Extracting the filtered elements of a network allows to focus on relevant connections that are highly representative of the system, discarding all the redundant and/or noisy information carried by those nodes and connections that can be described by an appropriate statistical null hypothesis (e.g., the configuration model of the system). Different names have been proposed so far to address this approach. The first name used was *backbone* of a network²⁶. In this case, the stress was on the links of nodes that were not compatible with a null hypothesis of equally distributed strength. Another proposal was a statistically validated network²⁷. A statistically validated network is a subgraph of an original graph where the selected links are those associated with a pair node activity that is not compatible with the one estimated under a random null hypothesis taking into account the heterogeneity of

activity of nodes. Statistically based filtering of real networks has been investigated in studies focusing on classic examples of networks, such as airports²⁶ and actor/movies²⁷ networks, trading decisions of investors^{28–31}, mobile phone calls of a large set of users^{32,33}, financial credit transactions occurring in an Interbank market³⁴, intraday lead-lag relationships of returns of financial assets traded in major financial markets³⁵, the Japanese credit market³⁶, international trade networks³⁷, social networks of news consumption³⁸, and rating networks of e-commerce platforms³⁹. The procedure of filtering nodes and links in a real network is not unique both in terms of methodology and in terms of the null hypothesis. Examples of different approaches have been recently proposed in the literature^{40–44}. Several of these techniques have been reviewed and discussed in^{45–48}. To the best of our knowledge, all works available so far have performed network filtering at the level of pair of nodes.

In this work, we introduce a filtering methodology for complex systems where interactions can be of various possible orders. Our approach explicitly takes into account the heterogeneity of the system and therefore it is able to highlight overexpression of hyperlinks of different sizes and weights. In particular, by mapping each layer of hyperlinks of a specific size in a bipartite system, our method identifies those hyperlinks that are overexpressed with respect to a random null hypothesis. To show the informativeness of our filtering method, through a synthetic benchmark we show that our approach detects real hyperlinks with higher sensitivity and accuracy than other traditional filtering techniques. We then apply our method to three different empirical social datasets. We show that the results obtained with our analysis are able to highlight information that is not obtainable neither from the unfiltered hypergraphs nor with a pairwise statistically validated analysis of the same system.

Results and discussion

Traditional network filtering approaches, such as the disparity filter²⁶ or the SVN approach²⁷, are not suited for higher-order data, since by design they mistreat all the information on connections beyond pairwise interactions. This implies that cliques of size n highlighted with a pairwise approach might not correspond to genuinely statistically validated hyperlinks, possibly producing both false positive and false negative. Consider, for example, the collaboration network between three authors that have strongly interacted in pairs in their research but have never published a paper altogether. The pairwise analysis might detect a clique of three nodes whereas the hyperlink with three nodes would not exist, thus generating a false positive. Similarly, false negatives might emerge in the case of an overexpressed hyperlink of size n that is not matched by a clique of validated pairwise links. Figure 1a illustrates the different possibilities of pairwise validation and hyperlink validation for $n = 3$.

Here, we propose an analytical filtering technique designed to detect overexpressed hyperlinks of various sizes. Our method works with weighted hypergraphs, where groups of nodes are connected through interactions (hyperlinks) of any size that are not limited to pairwise links. In particular, we model the weight of a hyperlink composed of n nodes as the intersection of n sets, where each set represents all the hyperlinks in which each node is active (see Fig. 1b for a schematic illustration of the method for $n = 3$). We are interested in evaluating if the weight of a hyperlink is compatible with a null model in which all nodes are selecting randomly their partners. This problem can be solved analytically (see “Methods”), associating a p value to each hyperlink. Our null model preserves the heterogeneity of the degree of all n nodes. This is particularly relevant in the case of hypergraphs whose higher-order degree distribution is strongly heterogeneous. This

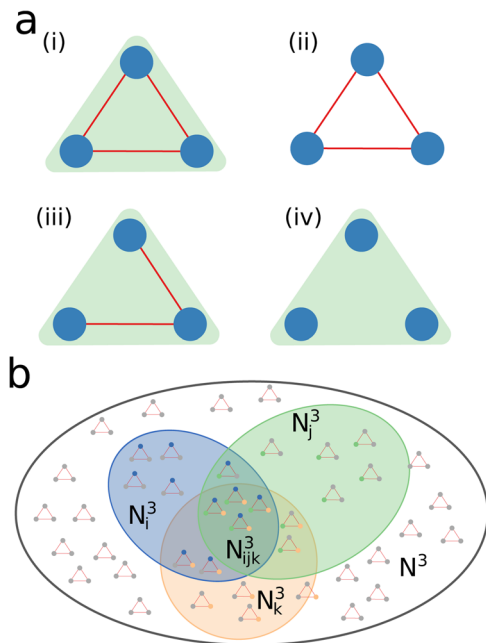


Fig. 1 A higher-order filter for hypergraphs. **a** Schematic illustration of false positives and false negatives in the investigation of statistically validated hyperlinks of size $n = 3$ detected in the Statistically Validated Hypergraph (SVH) compared with an approach based on statistically validated pairwise interactions. Green shaded triangles represent statistically validated hyperlinks with SVH, whereas red lines depict statistically validated pairwise interactions. Considering the various possibilities: (i) both the hyperlink and the 3-clique of pairwise interactions are validated; (ii) the SVH does not statistically validate the hyperlink whereas a 3-clique among the three nodes emerges from the validation of pairwise interactions. Taking the statistically validated pairwise overexpression as an indication of full interaction between nodes provides a conclusion on the overexpression of the hyperlink that is a false positive; (iii) and (iv) the hyperlink is statistically validated but the statistically validated pairwise links do not form a clique (i.e., detecting hyperlinks by evaluating the statistical validation of pairwise interactions would produce a false negative). **b** Schematic description of the SVH method for hyperlinks of size $n = 3$. In this example, node i is blue, node j is green, and node k is orange. The shaded areas represent the sets of hyperlinks of each node. The hyperlink $\{i, j, k\}$ is occurring four times. By knowing the number of all N^3 hyperlinks of size 3 in the system, and N_i^3 , N_j^3 , and N_k^3 number of hyperlinks including node i , j , and k respectively, our methodology allows us to compute the probability of observing $N_{i,j,k}^3$ occurrences (see Eq. (1) in “Methods”).

is a ubiquitous characteristic of real systems. Finally, the Statistically Validated Hypergraph (SVH) is obtained by putting together all hyperlinks at different sizes that are validated against our null hypothesis.

Benchmark. We generate hypergraphs that we use then as benchmark in the following way. We select N nodes and a set of sizes for hyperlinks, $\{n\} = \{2, 3, \dots, n\}$. For each size n , we select a fraction f of the N nodes, split them in non-overlapping groups of size n and connect each group with m hyperlinks. Our noiseless benchmark can be perturbed by adding some noise, that simulates random fluctuations or errors in the collection of the data. Specifically, we include an additional parameter p_n that represents the probability that the m interactions that define a hyperlink are assigned to a different, randomly selected group of the same size. Thus, our benchmark is defined by the parameters $(N, m, \{n\}, f, p_n)$. Here, we set $N = 500$, $m = 5$, $\{n\} = \{2, 3, 4, 5, 6, 7, 8\}$, but the

following results are not affected by the specific values of N and m and hold also for larger values of n . On the benchmark, we look at the groups of different sizes that are identified by the SVH (i.e., by our methodology) and SVN (i.e., inferred by pairwise statistically validated links) approaches. We choose the SVN as a pairwise approach because it is specifically tailored to work with bipartite networks, that represent the lowest order representation of hypergraphs. In this representation, a hyperlink of size n is mapped as n links between n nodes of set A and a node of set B. In our comparison, for SVH we select the groups of size n defined by validated hyperlinks, whereas for SVN we extract the maximal cliques of size n from the validated projected network. We first set $f = 0.5$ and study the performance of the two methods when p_n changes.

To show the performances of the two methods we compute the true-positive rate, $\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$ of validated hyperlinks at different sizes n as a function of the parameter p_n (Fig. 2a). The TPR quantifies the fraction of true groups that the two methods are able to identify. For each value of p_n and each size n , we generate 1000 realizations of the benchmark and take the median of TPR on all realizations. While the SVH approach is very robust in the presence of noise, with a moderate decrease of its TPR when p_n grows, the SVN significantly fails to detect the right hyperlinks already at moderate values of p_n . In fact, the detection of validated cliques on a pairwise network is strongly sensitive to noise, as missing only one of the $\binom{n}{2}$ links that compose a clique of size n is enough to fail its detection.

To complement this result, we also look at the false discovery rate (FDR), defined as $\text{FDR} = \frac{\text{FP}}{\text{FP} + \text{TP}}$, that quantifies the fraction of false positives on the total number of detected groups (see Supplementary Fig. 1 and Supplementary Note 1). By investigating the benchmark, we find that SVH never detects a false positive, while the SVN has different performances depending on the size of the hyperlink and the value of p_n . Indeed, already at low values of p_n , the FDR is significantly high for hyperlinks of small sizes, as it is more likely that cliques at such sizes are detected by the SVN due to spurious combinations of pairwise links. The higher the size, the less likely it is to obtain spurious cliques, which is reflected in lower values of FDR. For large values of p_n , the FDR associated with SVN worsens at almost all sizes.

We then looked at the performance of the two methods when both p_n and f vary, and we report the results in Supplementary Figs. 2 and 3. Color spots in the matrices of the Figures represent the median of the difference in TPR (Supplementary Fig. 2) and FDR (Supplementary Fig. 3) between SVH and SVN as a function of the two parameters. The first row of each matrix represents the difference of the curves plotted in Fig. 2a and Supplementary Fig. 1 respectively. Although the TPR difference pattern is similar for all rows, for high values of f the difference in performance between SVH and SVN becomes larger already for intermediate values of p_n . In fact, the parameter f affects the probability that a node participates in interactions at different sizes (Fig. 2b), thus the higher f the more challenging it is to correctly identify all overexpressed hyperlinks. Indeed, for large f each node is active in groups of different sizes and a filtering method that works only at the pairwise level is likely to produce overexpressed cliques in the SVN that overestimate the real size of an overexpressed hyperlink because the pairwise analysis can merge groups of nodes of different size.

In order to further check the consistency of the SVH approach, we generated ensembles of random hypergraphs, where hyperlinks are assigned to nodes compatibly with our null model. In Supplementary Tables 1 and 2, we report the result of this analysis. The SVH is never detecting spurious hyperlinks at any value of density of the random hypergraphs when adopting the

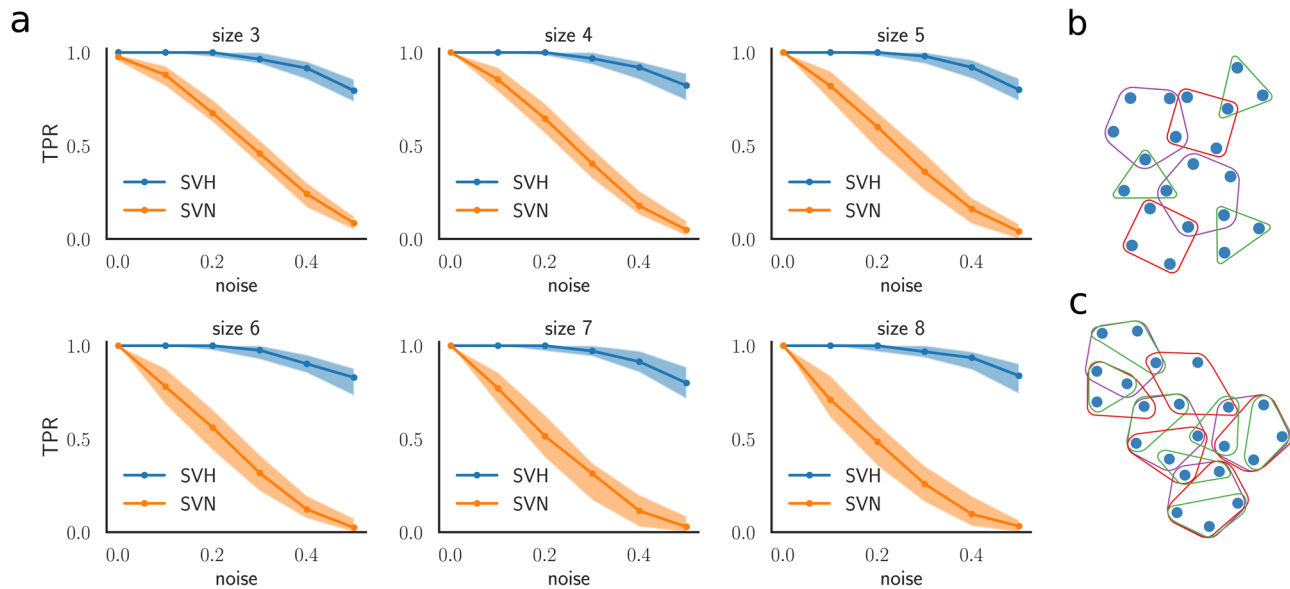


Fig. 2 Hypergraph benchmark and performance analysis. **a** Numerical simulations of the benchmark characterized by the presence of hyperlinks of size n ranging from 3 to 8. Median of the true positive rate of the detected hyperlinks obtained by using the Statistically Validated Hypergraph (SVH) methodology (blue line) and the detection of the pairwise overexpression in the Statistically Validated Network (SVN) (orange line) as a function of the noise parameter. The shaded area around the lines represents the interval between 10th and 90th percentile observed in 1000 realizations. Each panel is related to hyperlinks of different sizes n . **b** Benchmark realization with 22 nodes and fraction $f = 0.4$. The sizes of hyperlinks are 3 (green lines), 4 (red lines), and 5 (purple lines). **c** Benchmark realization with 22 nodes and $f = 0.9$.

control for the false discovery rate as a correction for multiple hypothesis testing. Furthermore, we checked if the performance of the SVN increases when we select all the cliques of validated links instead of only the maximal ones. We find that this different approach does not improve the TPR of SVN (Supplementary Fig. 4), while it is strongly worsening its FDR (Supplementary Fig. 5). The reason for this behavior is that by selecting all cliques, for each clique of size n we are validating all nested cliques of smaller size, no matter if spurious or not. On one side this approach may help in improving the TPR (but only in the case of nested hyperlinks, which are not strongly present in our benchmark), but it dramatically worsens the FDR.

US supreme court. In this section, we apply the SVH methodology to a dataset that records all votes expressed by the justices of the Supreme Court in the US from 1946 to 2019 case by case⁴⁹. This dataset has been extensively investigated in political science to understand and try to predict the patterns of justices' decision by looking at their political alignment during the period in which they were active^{50,51}. Similar research ideas have started to percolate also the complex systems' community, as shown by a recent work that proposes a link prediction model to forecast the evolution of the citation network spanned by cases ruled by the twin European institution, the Court of Justice⁵².

We start noting that such a system naturally fits the framework of hypergraphs, with hyperlinks of size n representing groups of n justices that voted in a case in the same way. As the Supreme Court is made of 9 justices, n can vary from 1 to 9 (in the case of unanimous decisions). In the investigated period, we observe 38 different justices judging 8915 cases. We find that the most frequent decisions are the unanimous ones (~ 2600), while all the other possible grouping of justices are present with at least 1000 entries (see Supplementary Table 3 and Supplementary Fig. 6a of Supplementary Note 2). Moreover, we find that the median of the number of decisions that justice has taken in a group of size n increases with the size of the group (Supplementary Fig. 6b), signaling that justices are more likely to vote as part of a large

majority than in a small minority. This evidence suggests that an approach that does not take into account interactions beyond the pairwise level is suboptimal to identify groups of justices that show overexpression of voting together since each justice typically voted in groups of different sizes. In fact, this observed behavior is analog to the behavior seen in the benchmark when we set a large value of f . Indeed, in this system when we use both SVH and SVN to detect overexpressed groups of different sizes (Supplementary Fig. 6c), we find that SVN is unable to find groups at smaller sizes because it is impossible to discriminate groups different from the majority and minority when a pairwise analysis is performed. Moreover, justices vote in the same way in a large number of cases (all the unanimous or almost unanimous ones). Even if we remove the unanimous votes, the situation does not change (see Supplementary Fig. 6d). Conversely, the SVH detects a much larger number of overexpressed groups at all possible sizes of interaction. A summary of the numbers of validated groups at different values of n is given in Supplementary Table 3.

An analysis of groups detected by the SVH provides informative insights on the activity of justices. Indeed, we can characterize each justice with the Segal–Cover (SC) score⁵⁰, which represents the level of judicial liberalism of each justice throughout her activity in the Supreme Court. There is a general SC score and several other scores focusing on specialized categories of legal decisions. When we compute the standard deviation of the SC score for the groups highlighted by the SVH and we compare it with that computed (i) on all the groups of justices observed to vote together at least once and (ii) on all possible groups of justices (to extract the latter we only consider justices that were contemporarily active in the Supreme Court), we find that the groups of justices detected by the SVH have the lowest diversity in liberalism SC score (Fig. 3a). This means that, with respect to their level of liberalism, the groups of justices of size n that present an overexpressed number of joint votes were more similar among them than the set of possible groups of justices. It is worth to note that the SC score is computed exclusively looking at the individual activity of justices case by

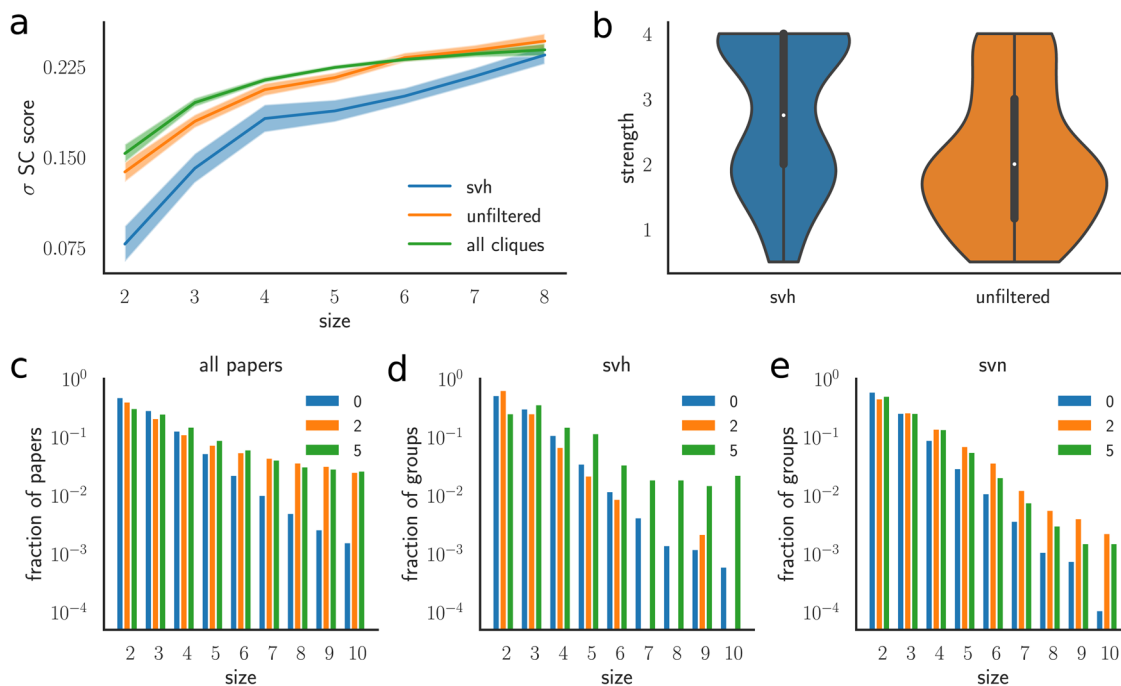


Fig. 3 Filtering real-world hypergraphs. **a** Average standard deviation of the Segal–Cover score as a function of a group size of justices of the US Supreme Court for the hyperlinks detected in the Statistically Validated Hypergraph (blue line), for the hyperlinks of the unfiltered hypergraph (orange line), and all possible groups formed by justices that were jointly acting at the Supreme Court during the same time period (green line). Lines represent mean values and shaded areas represent the standard error of the mean. **b** Violin plots of the average strength (as reported in the survey diaries) of (i) the hyperlinks statistically validated by the SVH (blue box) and (ii) the hyperlinks of the unfiltered hypergraph (orange box). **c** Fractions of papers written by groups of authors of different sizes for General Physics (blue bars), Nuclear Physics (orange bars), and Physics of Gases and Plasma PACS categories. **d, e** Fractions of overexpressed groups of authors as a function of size n in the overexpressed hyperlinks validated by SVH (**d**) and of groups of authors that are cliques of size n in the SVN (**e**) for 0-General Physics (blue bars), 2-Nuclear Physics (orange bars), and 5-Physics of Gases and Plasma (green bars) PACS categories.

case, while with the SVH method we are validating groups exclusively looking at their common decisions.

However, the SC liberalism score reduces to a mono-dimensional quantity a piece of information that can be more nuanced. Indeed, the Supreme Court has jurisdiction on cases of different legal areas, ranging from civil rights and criminal procedures to economic decisions. For this reason, the SC score is specialized in a number of distinct scores that capture the attitude of a justice in the different areas⁴⁹. Supplementary Table 4 reports the justices’ scores for the three main areas of criminal procedures, civil rights, and economic. As the area of each case is reported in our data, we are able to separately validate groups of justices for the different legal areas. This gives additional insights into the activity of justices. On one side, we find cases as that of Justice Antonin Scalia, that was consistently voting with a conservative attitude in all areas, and all the validated groups of size < 5 in which he is validated are always composed by other conservative justices. On the other side, we find more nuanced cases as that of justice Byron White, that was appointed by US President John F. Kennedy. White was progressive on economic issues, and indeed he is present in validated groups of size 3 with two other progressive justices, Thurgood Marshall and William Brennan. Conversely, he had a much more conservative attitude on issues related to civil rights and criminal procedures, and this is detected by our approach: in cases related to these areas he is validated in groups of size 3 with the conservative justices Warren Burger, William Rehnquist, and John Marshall Harlan. A summary of the hyperlinks validated in the SVHs of the different areas is reported in Supplementary Tables 5–7.

High school data. In this section, we detect the overexpressed hyperlinks of social interactions observed between students

during their stay at a French high school⁵³. These data are part of the SocioPatterns project that aimed at integrating social network analysis traditionally performed on surveys with actual contact data tracked through radio-frequency identification sensors. The contacts are detected and stored when pairs of students are physically located nearby at a given time t . The recording was occurring with a temporal resolution of the 20s. The data contain also information about self-reported contacts and friendships and Facebook networks that were present among students. These data have already been analyzed to understand the overlap between network and contact data⁵³. Here, we focus on the detection of higher-order interactions from the tracked contacts. In fact, it is important to track the presence of higher-order interactions in a social system as it has an impact on the dynamical processes that can occur on top of it^{14,17,22}. In order to build a hypergraph, we extract the higher-order interactions from the raw data. To do so, for each time step t we build the graph of interactions occurring at time t and we extract the maximal cliques of any size. Indeed, if n students are tracked in a fully connected clique at time t , it means that they had a collective interaction at that time. We find that on top of pairwise interactions the network contains hyperlinks that involve up to 5 students interacting at the same time (Supplementary Fig. 7a and Supplementary Note 3). After extracting the overexpressed hyperlinks with both SVH and SVN, we find that, as in the case of US justices, the two methods have a different distribution of validated groups (Supplementary Fig. 7b). Specifically, SVH detects much more groups than SVN at lower size but significantly less at the higher size. We verify that most of the groups detected by SVN at higher size are spurious (Supplementary Fig. 7c), which means that with SVN we validate groups of size n even if the corresponding students were never simultaneously interacting altogether in a group of that size. In

this system, the cliques detected by the SVN do not represent a reliable proxy of the overexpressed hyperlinks.

In order to understand the nature of the hyperlinks validated by the SVH, we analyze the validated groups using the available metadata. We use the contact diaries that were filled by the students at the end of a specific day of data collection. This additional information stores contacts that were self-reported by students themselves and can be seen as robust information about the system, since it contains interactions that were strong enough to be remembered by the students. We use the diaries to extract cliques at different sizes of interacting students, and we compare this information with the hyperlinks present in the unfiltered dataset and with those validated with the SVH approach. To maintain consistency across the two datasets, we drop contact data that do not contain students that filled the diaries surveys and drop from the diaries contacts that were not tracked by the sensors. Furthermore, we limit contact data to that recorded in the same day of the diaries survey. We find that the SVH is very precise in retrieving the self-reported cliques (it contains less “spurious” hyperlinks that do not correspond to diaries cliques) but it is not highly accurate (i.e., it contains only a smaller fraction of diaries cliques) (Supplementary Table 8).

The reported contacts come with a discrete weight provided by the students themselves that represent the duration of each reported interaction ((i) at most 5 min if $w = 1$, (ii) between 5 and 15 min if $w = 2$, (iii) between 15 min and 1 h if $w = 3$, (iv) more than 1 h if $w = 4$). For each clique in the contact diaries that is also detected in the unfiltered hypergraph or in the SVH, we compute the overall strength by averaging the strengths of the links that constitute it. We find that the distribution of the average strength of SVH hyperlinks is higher than the one of unfiltered groups (Fig. 3b), showing that the SVH hyperlinks detect the most persistent groups. The difference in the distribution of average strength is statistically significant according to a non-parametric Kruskal-Wallis test with score ~ 18 and $P < 0.0001$. It is worth to stress that the hyperlinks present in the SVH do not necessarily correspond exclusively to the interactions with the highest weight. Indeed, we find that hyperlinks of the same weight can be present or absent in the SVH, depending on the heterogenous activity of the involved nodes (Supplementary Fig. 8). In fact, for less active nodes hyperlinks with a small weight are more likely to be validated, while hyperlinks that involve more active nodes need a higher weight in order to be validated.

Physics authors. In this section, we analyze the hypergraph of scientific collaborations among Physics authors. To do so, we investigate the APS dataset, which contains authorship data on papers published in journals of the APS group from 1893 to 2015. This dataset has already been extensively investigated to characterize structural and dynamical properties of scientific collaborations, with respect to both authors’ careers and topics’ evolution^{54–56}. Here, we match the papers present in the APS dataset with the Web of Science database using the doi and identify the authors with the ID curated by WoS to maximize disambiguation, which is a well-known issue that affects the accuracy of the dataset. Since we are interested in interactions among authors, we limit our investigation to papers with a maximum number of ten authors to avoid larger collaborations for which direct interaction between all authors is less likely. From the APS dataset, we retrieve the PACS of each paper. This allows us to split the set of papers in ten subfields of physics by using the highest hierarchical level of PACS classification. We focus on the papers published from 1985 onwards as from this year reporting one of more PACS per paper became compulsory. This leaves us with 269,887 papers and 114,856 authors.

First, we look at the distribution of papers of different sizes for each subfield (Fig. 3c). Here, we focus on the categories of General Physics (PACS hierarchical integer number 0), Nuclear Physics (PACS number 2) and Physics of Gases and Plasma (PACS number 5) but the CDFs for all PACS categories are shown in Supplementary Fig. 9 (Supplementary Note 4). As expected, we find that PACS have different distributions of team size that highlight different publication habits of the researchers publishing in different PACS categories. In our selection, we find that the subfield with higher percentages of smaller groups is General Physics. On the other side, in Nuclear Physics and Physics of Gases and Plasma there are higher percentages of larger research groups (in Fig. 3c, all percentages sum to 1 because we are cutting from the distributions all papers that are written by groups larger than ten).

We then apply to the dataset both SVH and SVN methodologies and extract the distribution of the validated groups with the two methods (Fig. 3d, e, respectively). At first sight, we find that the distributions of the groups validated with the SVH show relevant differences with the original ones, while with the SVN we obtain similar trends. Specifically, we find that for Nuclear Physics the fraction of overexpressed groups in the SVH goes rapidly to 0 when the size increases, showing an opposite trend if compared to the distribution of the number of authors per paper. This means that the size of most of the SVH validated groups for this PACS is relatively small, in spite of the fact that there are many papers written by larger numbers of authors. In the case of overexpressed groups of the SVN, the distribution for this category of PACS is similar to that of Fig. 3c. Conversely, PACS 5 (Physics of Gases and Plasma) maintains a similar profile across the different distributions.

In order to understand this finding we looked, for each size, at the relationship across PACS between the fraction of validated groups and the average number of papers written by a group of that size (Supplementary Fig. 10 for SVH and 11 for SVN). We find that in the case of SVH these two quantities are strongly correlated, with Pearson coefficient ranging from 0.84 to 0.99 and being always statistically significant. This means that with the SVH, we validate more groups of authors when these groups are writing on average more papers together. This is a result showing the reliability of SVH results. It is interesting to note that PACS 5 has among the highest average numbers of papers per group at higher sizes, making it clear why the SVH validates more groups at these sizes. On the other hand, in Nuclear Physics most (for some sizes even all) of the detected groups at higher sizes wrote only one paper altogether, so the number of validated groups is much lower. Conversely, with the SVN approach, the fraction of detected groups is not related to the average activity of groups of that size (in the scatter plots of Supplementary Fig. 11, all correlations between the fraction of detected groups and the average number of papers per group are not statistically different from zero). Due to the fact that with the pairwise SVN approach all groups are obtained through the aggregation of pairwise overexpressed links, the details of higher-order interactions are missed.

Summing up, the SVH approach gives us an insight that is not evident in the raw data or with methods that are limited to the characterization of pairwise interactions: we find that research areas like Nuclear Physics and Physics of Gases and Plasma are similar with respect to the distribution of papers that are written by research groups of different sizes, but the research groups in the two PACS have different publication habits. In Physics of Gases and Plasma, it is more likely that exactly the same group publishes more papers together (and with SVH we identify the groups that do so in a significant way), while in Nuclear Physics

most of the medium size collaborations produce a paper just in a single occurrence.

Conclusion. In the last decade, a deluge of new data on biological and sociotechnical systems has become available, showing the importance of filtering techniques able to highlight potentially informative network structures. Recently, hypergraphs have emerged as a fundamental tool to map real-world interacting systems. Yet, extracting the relevant interactions from higher-order data is still an open problem. In this work we proposed Statistically Validated Hypergraphs (SVH) as a method to identify the most meaningful relations between entities of a higher-order system, reducing the complexity carried by noisy and/or spurious interactions.

Our method is able to quantify the probability that an observed hyperlink is compatible with a process in which all involved nodes were randomly selecting their counterparts, reducing the detection of false negatives and false positives occurring with basic pairwise filters. Besides, the null model that we developed naturally reproduces the heterogeneous activity of each node, a crucial feature that overcomes the limitations of a threshold-based filtering approach. We have showcased the application of our method to three different systems: the US Supreme Court, the social connections of students interacting in a French high school and the scientific collaborations of Physics authors publishing in the Journal of the American Physical Society. In all cases, statistically validated groups carry more coherent information than that observed in the unfiltered hypergraphs. For the US Supreme Court, groups of justices with more similar SC profiles are highlighted. For the students of a French high school, groups of students characterized by an intense social interaction are detected, and for the authors of physics papers, the analysis of the SVH unveils a difference in publication habits across subfields that is not evident when looking at the complete system.

A foreseeable development of our methodology is a generalization capable to take into account the temporal dynamics leading to the emergence of a hypergraph, similarly to what was proposed in ref. 43 for pairwise interactions only. Taken together, we believe that our method, separating meaningful connections from less informative node interactions, is a powerful tool capable to capture the different nuances of higher-order interacting systems.

Methods

In a SVH, each hyperlink of size n represents a group of n nodes that is over-expressed by comparing its occurrence with that of a null hypothesis that reproduces random group interactions. To extract the p value of a hyperlink of size n , we select a subset of the hypergraph considering only hyperlinks of size n , and we compute the weighted degree of each node with respect to this subgraph, $N_{x_1}^n, N_{x_2}^n, \dots, N_{x_n}^n$. We then extract the weight of the hyperlink connecting all n nodes, $N_{x_1 \dots x_n}^n$ and the total number of hyperlinks of size n , N^n . We can then assess the probability of $N_{x_1 \dots x_n}^n$ being compatible with a null model where each of the nodes in the group randomly selects its hyperlinks from the whole set of hyperlinks of size n . In fact, evaluating the probability of observing a hyperlink of weight $N_{x_1 \dots x_n}^n$ is equivalent to evaluating the probability of having an intersection of size $N_{x_1 \dots x_n}^n$ among n sets⁵⁷. To illustrate the method, we start with the simplest case, $n = 3$, with three nodes i, j , and k being active respectively in N_i^3, N_j^3 , and N_k^3 interactions of size 3 (Fig. 1b, c). The probability of having the three nodes interacting together N_{ijk}^3 times under a random null model is written as

$$p(N_{ijk}^3) = \sum_X H(X|N^3, N_i^3, N_j^3) \times H(N_{ijk}^3|N^3, X, N_k^3) = \frac{1}{\binom{N^3}{N_j^3} \binom{N^3}{N_i^3}} \sum_X \binom{N_i^3}{X} \binom{N^3 - N_i^3}{N_j^3 - X} \binom{N^3 - X}{N_k^3} \binom{N^3 - N_k^3}{N_{ijk}^3} \quad (1)$$

where $H(N_{AB}|N, N_A, N_B)$ is the hypergeometric distribution that computes the probability of having an intersection of size N_{AB} between two sets A and B of size N_A and N_B given N total elements.

The probability $p(N_{ijk}^3)$ in Eq. (1) is obtained through the convolution of two instances of the hypergeometric distribution. Indeed, to compute $p(N_{ijk}^3)$ we start from the probability of having an intersection of size X between nodes i and j and multiply it with the probability of having an intersection of size N_{ijk}^3 between node k and the intersection set of size X between i and j . This product is then summed over all possible values of X , i.e., all possible intersections between i and j which are compatible with the observed number of interactions between all the three nodes. Starting from Eq. (1), we then compute a p value for the hyperlink connecting i, j , and k through the survival function,

$$p(x \geq N_{ijk}^3) = 1 - \sum_{x=0}^{N_{ijk}^3-1} p(x). \quad (2)$$

The p value provides the probability of observing N_{ijk}^3 or more occurrences of the hyperlink composed by i, j, k . Once all the p values for all observed hyperlinks are computed, they are tested against a threshold of statistical significance α . In all, the results presented in this paper we use $\alpha = 0.01$. The statistical test is performed by using the control for the false discovery rate⁵⁸ as a multiple hypothesis test correction. The total number of test considered is $N_t = \binom{N_{nodes}}{3}$, which is the number of all possible triplets of the N_{nodes}^3 elements that are active in hyperlinks of size 3. Thus, when applying the control for the False Discovery Rate method, we start from a Bonferroni threshold computed as $\alpha_B = \alpha/N_t$.

For an hyperlink of generic size n , Eq. (1) becomes

$$p(N_{x_1 \dots x_n}^n) = \sum_{X_{x_1 x_2}} H(X_{x_1 x_2}|N^n, N_{x_1}^n, N_{x_2}^n) \times \sum_{X_{x_1 x_2 x_3}} H(X_{x_1 x_2 x_3}|N^n, X_{x_1 x_2}, N_{x_3}^n) \times \dots \times \sum_{X_{x_1 x_2 \dots x_{n-1}}} H(X_{x_1 x_2 \dots x_{n-1}}|N^n, X_{x_1 x_2 \dots x_{n-2}}, N_{x_{n-1}}^n) \times H(N_{x_1 x_2 \dots x_n}^n|N^n, X_{x_1 x_2 \dots x_{n-1}}, N_{x_n}^n). \quad (3)$$

As in the case with three nodes, the main idea of Eq. (3) is to write the overall probability of having an intersection of size $N_{x_1 x_2 \dots x_n}^n$ between the activity of n nodes as the product of multiple probabilities of hierarchical pairwise intersections, summed over all the possible configurations compatible with $N_{x_1 x_2 \dots x_n}^n$. The specific order of the nodes in the hierarchical intersections does not affect the value of $p(N_{x_1 \dots x_n}^n)$. From Eq. (3), we extract a p value as in Eq. (2). The number of tests to consider when correcting for multiple testing is $N_t = \binom{N_{nodes}}{n}$. In our numerical computation, we use the approach developed in ref. 57. The approach is analytic but requires heavy combinatorial computation, and it might be of the difficult application when the hyperlinks are of size larger than about fifteen nodes (Supplementary Note 5). Supplementary Fig. 12 shows the computational time as a function of the hyperlink size, and Supplementary Fig. 13 plots the dependence of computational time on the weight of a hyperlink.

Data availability

The data analyzed in this paper are available at <https://github.com/musci8/SVH>.

Code availability

The code to extract Statistically Validated Hypergraphs is available at <https://github.com/musci8/SVH>.

Received: 13 April 2021; Accepted: 9 August 2021;

Published online: 24 September 2021

References

- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006).
- Battiston, F. et al. Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.* **874**, 1–92 (2020).
- Petri, G. et al. Homological scaffolds of brain functional networks. *J. R. Soc. Interface* **11**, 20140873 (2014).
- Giusti, C., Ghrist, R. & Bassett, D. S. Two’s company, three (or more) is a simplex. *J. Comput. Neurosci.* **41**, 1–14 (2016).
- Grilli, J., Barabás, G., Michalska-Smith, M. J. & Allesina, S. Higher-order interactions stabilize dynamics in competitive network models. *Nature* **548**, 210 (2017).
- Patania, A., Petri, G. & Vaccarino, F. The shape of collaborations. *EPJ Data Sci.* **6**, 18 (2017).
- Patania, A., Vaccarino, F. & Petri, G. Topological analysis of data. *EPJ Data Sci.* **6**, 7 (2017).
- 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).
- Estrada, E. & Rodríguez-Velázquez, J. A. Subgraph centrality and clustering in complex hyper-networks. *Phys. A* **364**, 581–594 (2006).

10. Benson, A. R. Three hypergraph eigenvector centralities. *SIAM J. Math. Data Sci.* **1**, 293–312 (2019).
11. Carletti, T., Fanelli, D. & Lambiotte, R. Random walks and community detection in hypergraphs. *J. Phys.: Complexity* **2**, 015011 (2021).
12. 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).
13. Cencetti, G., Battiston, F., Lepri, B. & Karsai, M. Temporal properties of higher-order interactions in social networks. *Sci. Rep.* **11**, 7028 (2021).
14. Iacopini, I., Petri, G., Barrat, A. & Latora, V. Simplicial models of social contagion. *Nat. Commun.* **10**, 2485 (2019).
15. Chowdhary, S., Kumar, A., Cencetti, G., Iacopini, I. & Battiston, F. Simplicial contagion in temporal higher-order networks. *Journal of Physics: Complexity* **2** 035019 (2021).
16. Neuhäuser, L., Schaub, M. T., Mellor, A. & Lambiotte, R. Opinion dynamics with multi-body interactions. Preprint at <https://arxiv.org/abs/2004.00901> (2020).
17. Alvarez-Rodriguez, U. et al. Evolutionary dynamics of higher-order interactions in social networks. *Nat. Hum. Behav.* **5**, 586–595 (2021).
18. 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).
19. Carletti, T., Battiston, F., Cencetti, G. & Fanelli, D. Random walks on hypergraphs. *Phys. Rev. E* **101**, 022308 (2020).
20. Skardal, P. S. & Arenas, A. Higher-order interactions in complex networks of phase oscillators promote abrupt synchronization switching. *Commun. Phys.* **3**, 1–6 (2020).
21. Millán, A. P., Torres, J. J. & Bianconi, G. Explosive higher-order Kuramoto dynamics on simplicial complexes. *Phys. Rev. Lett.* **124**, 218301 (2020).
22. Lucas, M., Cencetti, G. & Battiston, F. A multi-order Laplacian for synchronization in higher-order networks. *Phys. Rev. Res.* **2**, 033410 (2020).
23. Gambuzza, L. V. et al. Stability of synchronization in simplicial complexes. *Nat. Commun.* **12**, 1–13 (2021).
24. Berge, C. *Graphs and hypergraphs*. North-Holland Publishing Company (1973).
25. Young, J.-G., Petri, G. & Peixoto, T. P. Hypergraph reconstruction from network data. *Commun. Phys.* **4**, 135 (2021).
26. Serrano, M. Á., Boguná, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci. USA* **106**, 6483–6488 (2009).
27. Tumminello, M., Micciche, S., Lillo, F., Piilo, J. & Mantegna, R. N. Statistically validated networks in bipartite complex systems. *PLoS ONE* **6**, e17994 (2011).
28. Tumminello, M., Lillo, F., Piilo, J. & Mantegna, R. N. Identification of clusters of investors from their real trading activity in a financial market. *New J. Phys.* **14**, 013041 (2012).
29. Musciotto, F., Marotta, L., Micciché, S., Piilo, J. & Mantegna, R. N. Patterns of trading profiles at the nordic stock exchange. a correlation-based approach. *Chaos, Solitons, Fractals* **88**, 267–278 (2016).
30. Musciotto, F., Marotta, L., Piilo, J. & Mantegna, R. N. Long-term ecology of investors in a financial market. *Palgrave Commun.* **4**, 1–12 (2018).
31. Challet, D., Chicheportiche, R., Lallouache, M. & Kassibrakis, S. Statistically validated lead-lag networks and inventory prediction in the foreign exchange market. *Adv. Complex Syst.* **21**, 1850019 (2018).
32. Li, M.-X. et al. Statistically validated mobile communication networks: the evolution of motifs in European and Chinese data. *New J. Phys.* **16**, 083038 (2014).
33. Li, M.-X. et al. A comparative analysis of the statistical properties of large mobile phone calling networks. *Sci. Rep.* **4**, 1–12 (2014).
34. Hatzopoulos, V., Iori, G., Mantegna, R. N., Micciché, S. & Tumminello, M. Quantifying preferential trading in the e-mid interbank market. *Quantitative Finance* **15**, 693–710 (2015).
35. Curme, C., Tumminello, M., Mantegna, R. N., Stanley, H. E. & Kenett, D. Y. Emergence of statistically validated financial intraday lead-lag relationships. *Quantitative Finance* **15**, 1375–1386 (2015).
36. Marotta, L. et al. Backbone of credit relationships in the Japanese credit market. *EPJ Data Sci.* **5**, 1–14 (2016).
37. Straka, M. J., Caldarelli, G. & Saracco, F. Grand canonical validation of the bipartite international trade network. *Phys. Rev. E* **96**, 022306 (2017).
38. Becatti, C., Caldarelli, G., Lambiotte, R. & Saracco, F. Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections. *Palgrave Commun.* **5**, 1–16 (2019).
39. Becatti, C., Caldarelli, G. & Saracco, F. Entropy-based randomization of rating networks. *Phys. Rev. E* **99**, 022306 (2019).
40. Gemmetto, V., Cardillo, A. & Garlaschelli, D. Irreducible network backbones: unbiased graph filtering via maximum entropy. Preprint at <https://arxiv.org/abs/1706.00230> (2017).
41. Coscia, M. & Neffke, F. M. H. Network backbone with noisy data. in *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, 425–436 (IEEE, 2017).
42. Saracco, F. et al. Inferring monopartite projections of bipartite networks: an entropy-based approach. *New J. Phys.* **19**, 053022 (2017).
43. Kobayashi, T., Takaguchi, T. & Barrat, A. The structured backbone of temporal social ties. *Nat. Commun.* **10**, 1–11 (2019).
44. Marcaccioli, R. & Livan, G. A pólya urn approach to information filtering in complex networks. *Nat. Commun.* **10**, 1–10 (2019).
45. Iori, G. & Mantegna, R. N. Empirical analyses of networks in finance. in *Handbook of Computational Economics*, (eds Hommes, C. & Le Baron, B.) Vol. 4, 637–685 (Elsevier, 2018).
46. Straka, M. J., Caldarelli, G., Squartini, T. & Saracco, F. From ecology to finance (and back?): a review on entropy-based null models for the analysis of bipartite networks. *J. Stat. Phys.* **173**, 1252–1285 (2018).
47. Micciché, S. & Mantegna, R. N. A primer on statistically validated networks. *Computat. Soc. Sci. Complex Syst.* **203**, 91 (2019).
48. Cimini, G. et al. The statistical physics of real-world networks. *Nat. Rev. Phys.* **1**, 58–71 (2019).
49. Epstein, L., Walker, T. G., Hendrickson, N. S. S. & Roberts, J. The U.S. Supreme Court Justices Database. <http://epstein.wustl.edu/research/justicesdata.html> (2019).
50. Segal, J. A. & Cover, A. D. Ideological values and the votes of u.s. supreme court justices. *Am. Political Sci. Rev.* **83**, 557–565 (1989).
51. Epstein, L. The supreme court compendium: data, decisions and development. *Congressional Quarterly Inc.* (1994).
52. Mones, E., Sapieżyński, P., Thordal, S., Olsen, H. P. & Lehmann, S. Emergence of network effects and predictability in the judicial system. *Sci. Rep.* **11**, 2045–2322 (2021).
53. 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).
54. Radicchi, F. & Castellano, C. Rescaling citations of publications in physics. *Phys. Rev. E* **83**, 046116 (2011).
55. Battiston, F. et al. Taking census of physics. *Nat. Rev. Phys.* **1**, 89–97 (2019).
56. Chinazzi, M., Gonçalves, B., Zhang, Q. & Vespignani, A. Mapping the physics research space: a machine learning approach. *EPJ Data Sci.* **8**, 33 (2019).
57. Wang, M., Zhao, Y. & Zhang, B. Efficient test and visualization of multi-set intersections. *Sci. Rep.* **5**, 16923 (2015).
58. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, 289–300 (1995).

Acknowledgements

F.M. and R.N.M. acknowledge financial support by the MIUR PRIN project 2017WZFTZP, Stochastic forecasting in complex systems. F.B. acknowledges partial support from ERC synergy grant 810115 (DYNASNET).

Author contributions

F.M., F.B., and R.N.M. designed research; F.M. analyzed the data; F.M. performed numerical simulations; F.M., F.B., and R.N.M. 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/s42005-021-00710-4>.

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

Peer review information *Communications Physics* thanks Paolo Barucca 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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021