Supplementary Information
Detecting informative higher-order interactions in statistically validated hypergraphs

Federico Musciotto, Federico Battiston, and Rosario N. Mantegna

Supplementary Note 1. BENCHMARKS

A. Type I errors

In order to assess the rate of type I errors detected by the SVH on our benchmark we compute the False Discovery Rate (FDR), defined as $FDR = \frac{FP}{FP+TP}$. The behavior of FDR as a function of p_n for SVH and SVN with $f = 0.5$ is shown in Figure Supplementary Figure 1. Results are obtained by performing over 1000 realizations of the benchmark. The large dispersion of the FDR of SVN for large values of p_n is due to a digitization effect, as the number of true positives detected by SVN decreases significantly when p_n grows. We then studied the performance of our method when both f and p_n vary. Figure Supplementary Figure 2 and Supplementary Figure 3 show the median of the difference in TPR and FDR between SVH and SVN as a function of both parameters.

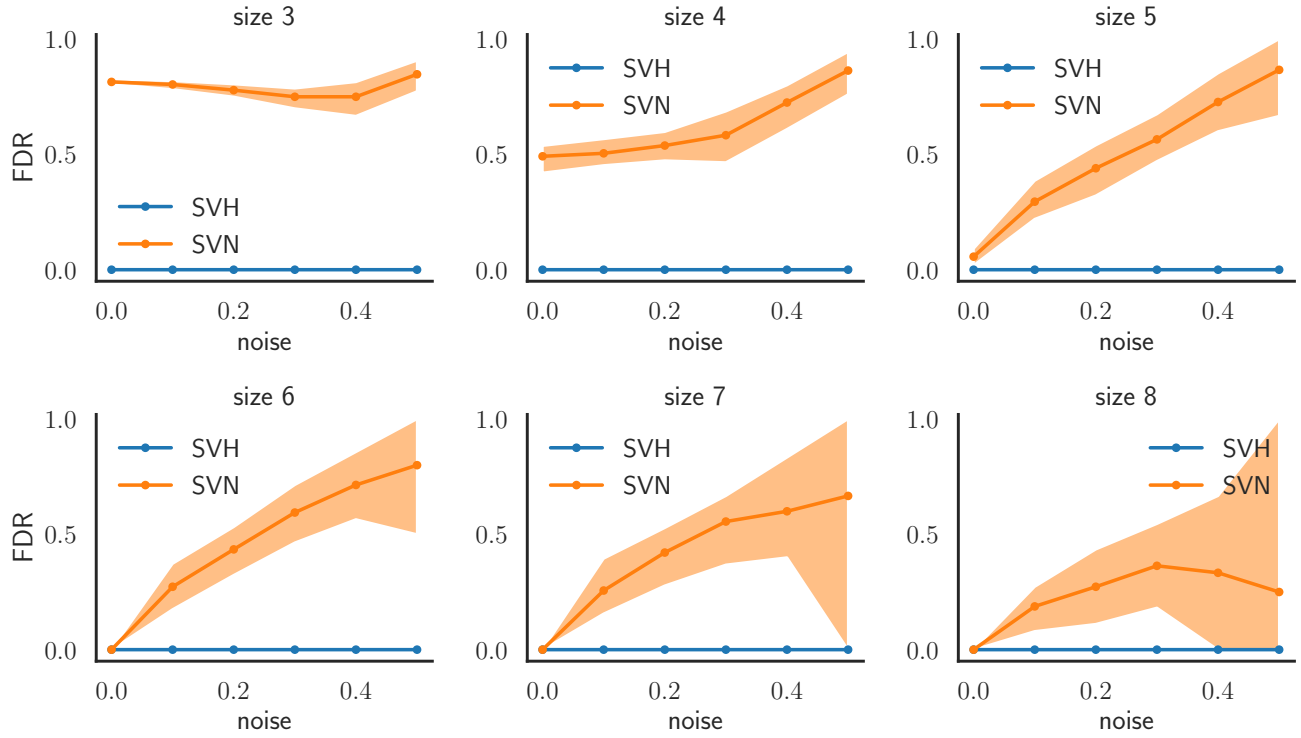

FIG. Supplementary Figure 1. Numerical simulations of the benchmark characterized by the presence of hyperlinks of size n ranging from 3 to 8 (see description in the main text). False Discovery Rate of the detected hyperlinks obtained by using the SVH methodology (blue line) and of the groups inferred by the pairwise over-expression of SVN (orange line) as a function of parameter p_n. The shaded area around the lines represents the interval between 10_{th} and 90_{th} percentile observed in 1000 realizations. Each panel is related to hyperlinks of different size n.

B. Control of False Discovery Rate

In order to check the robustness of the control of False Discovery Rate of our approach we generated ensembles of random hypergraph according to the following scheme. Given a set of N nodes and $\{n\}$ sizes of hyperlinks, we

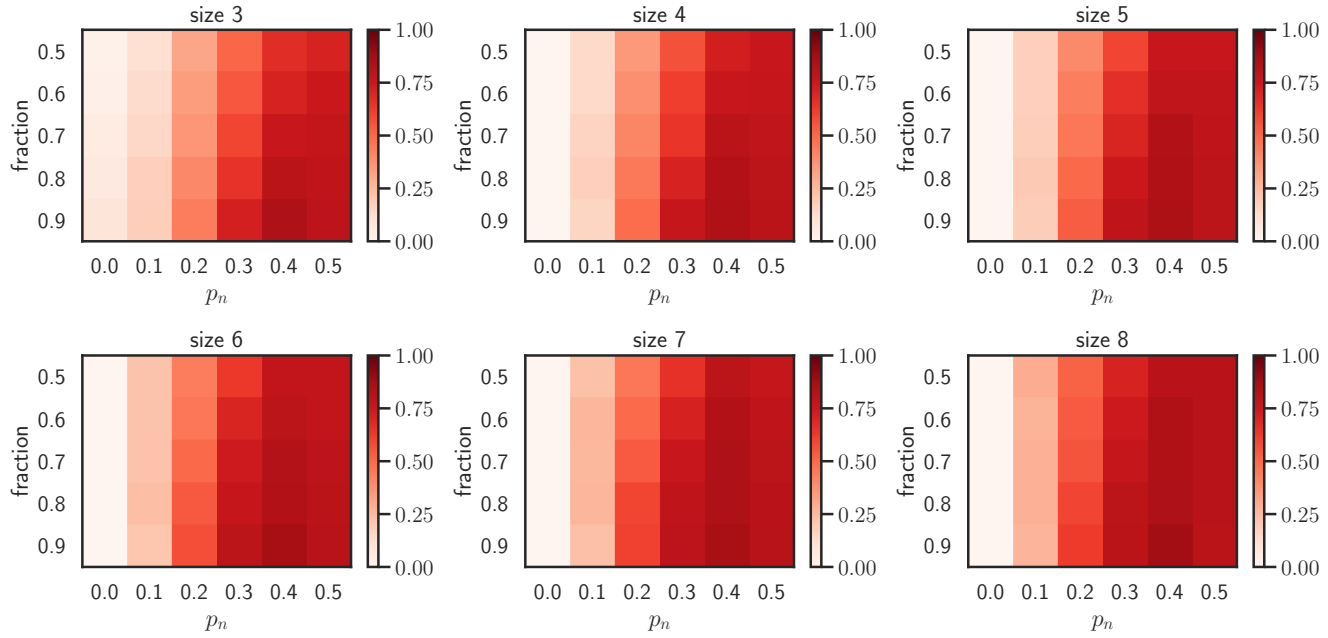

FIG. Supplementary Figure 2. Numerical simulations of the benchmark characterized by the presence of hyperlinks of size n ranging from 3 to 8 (see description in the main text). Median of the difference of TPR(SVH) - TPR(SVN) of the detected hyperlinks obtained by using the SVH methodology and the detection of the pairwise over-expression of SVN as a function of parameter f (x-axis) and p_n (y-axis) over 1000 realizations. Each panel is related to hyperlinks of different size n.

generate a random hypergraph of density d by randomly placing a fraction d of all available hyperedges of size n, whose number corresponds to the binomial coefficient $\binom{N}{n}$. We set $N = 50$ and $\{n\} = \{2, 3, 4\}$. We generated 100 random hypergraph for each value of d in $\{0.01, 0.05, 0.10, 0.30, 0.50\}$ and applied the SVH approach on each hypergraph, counting the detected hyperlinks at all sizes. When we use the control for the False Discovery Rate as a correction for multiple hypothesis testing, we never detect a hyperlink from the random ensemble (Table Supplementary Table 1), no matter the value of density and the distribution of weights of hyperlinks. Without the control for the FDR correction, we detect a certain amount of validated hyperlinks also in the random case (Table Supplementary Table 2), which is a confirmation of the need for a multiple hypothesis testing correction.

C. Maximal cliques versus all cliques

When using the SVN approach on hypergraph, we select the maximal cliques from the statistically validated network as a proxy of the validated hyperlinks. The only other viable alternative is to select as hyperlinks all the cliques at any size. If we adopt this alternative approach on our simulations of the investigated benchmarks, we find that the TPR does not improve (Figure Supplementary Figure 4), while the FDR significantly worsens (Figure Supplementary Figure 5). The patterns of Figures Supplementary Figure 4, Supplementary Figure 5 are related to the level of nestedness of hyperlinks of different sizes, i.e. to the occurrence of having hyperlinks of a size contained in hyperlinks of larger sizes. In fact, if the level of nestedness is high, taking all cliques is more convenient as it improves the TPR without increasing significantly the FDR. However, in most real applications the level nestedness is not equal for all hyperlinks, which means that the SVN may have unpredictable precision and accuracy with both approaches. This specific aspect will be investigated in a future study.

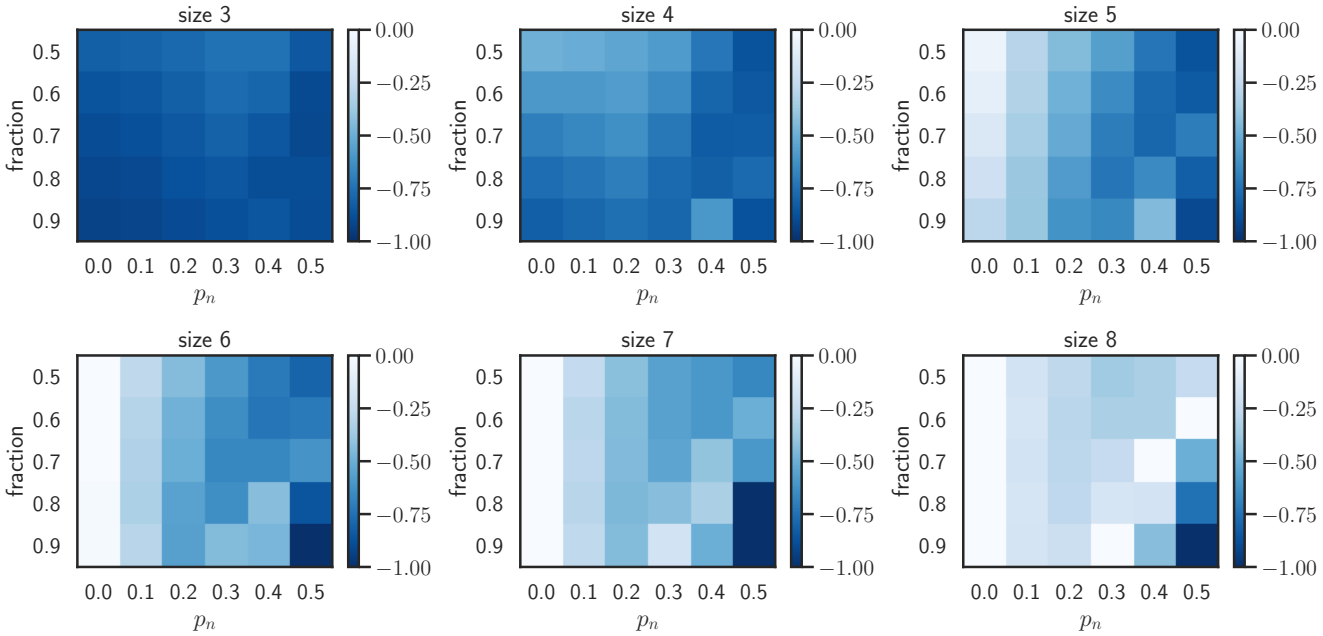

FIG. Supplementary Figure 3. Numerical simulations of the benchmark characterized by the presence of hyperlinks of size n ranging from 3 to 8 (see description in the main text). Median of the difference of FDR(SVH) - FDR(SVN) of the detected hyperlinks obtained by using the SVH methodology and the detection of the pairwise over-expression of SVN as a function of parameter p_n (x-axis) and f (y-axis) over 1000 realizations. Each panel is related to hyperlinks of different size n.

Density	Avg val 2	Avg val 3	Avg val 4	Max val 2	Max val 3	Max val 4	Avg weight 2	Avg weight 3	Avg weight 4	Max weight 2	Max weight 3	Max weight 4
0.01	0	0	0	0	0	0	1.00	1.01	1.00	1	2	3
0.05	0	0	0	0	0	0	1.02	1.03	1.03	2	3	4
0.10	0	0	0	0	0	0	1.05	1.05	1.05	3	3	4
0.30	0	0	0	0	0	0	1.15	1.16	1.16	4	6	5
0.50	0	0	0	0	0	0	1.27	1.27	1.27	5	6	7

TABLE Supplementary Table 1. Summary of the results of SVH on random hypergraphs. Each row reports the density of the generated random hypergraphs, and for sizes 2, 3 and 4 (i) the average number of validated hyperlinks, (ii) the maximum number of validated hyperlinks, (iii) the average hyperlink weight and (iv) the maximum hyperlink weight across 100 realizations. When validating, the significance threshold is corrected with the control for the False Discovery Rate procedure.

Density	Avg val 2	Avg val 3	Avg val 4	Max val 2	Max val 3	Max val 4	Avg weight 2	Avg weight 3	Avg weight 4	Max weight 2	Max weight 3	Max weight 4
0.01	0.0	1.2	12.4	0	3	14	1.00	1.01	1.01	1	2	2
0.05	0.4	1.5	0.1	2	3	1	1.02	1.03	1.03	2	3	3
0.10	0.9	1.3	0.0	2	3	0	1.05	1.05	1.05	3	4	4
0.30	0.5	0.2	0.0	2	1	0	1.18	1.16	1.16	4	6	6
0.50	0.1	0.0	0.0	1	0	0	1.28	1.27	1.27	6	6	7

TABLE Supplementary Table 2. Summary of the results of SVH on random hypergraphs. Each row reports the density of the generated random hypergraphs, and for sizes 2, 3 and 4 (i) the average number of validated hyperlinks, (ii) the maximum number of validated hyperlinks, (iii) the average hyperlink weight and (iv) the maximum hyperlink weight across 100 realizations. When validating, the significance threshold is not corrected with any multiple hypothesis testing procedure.

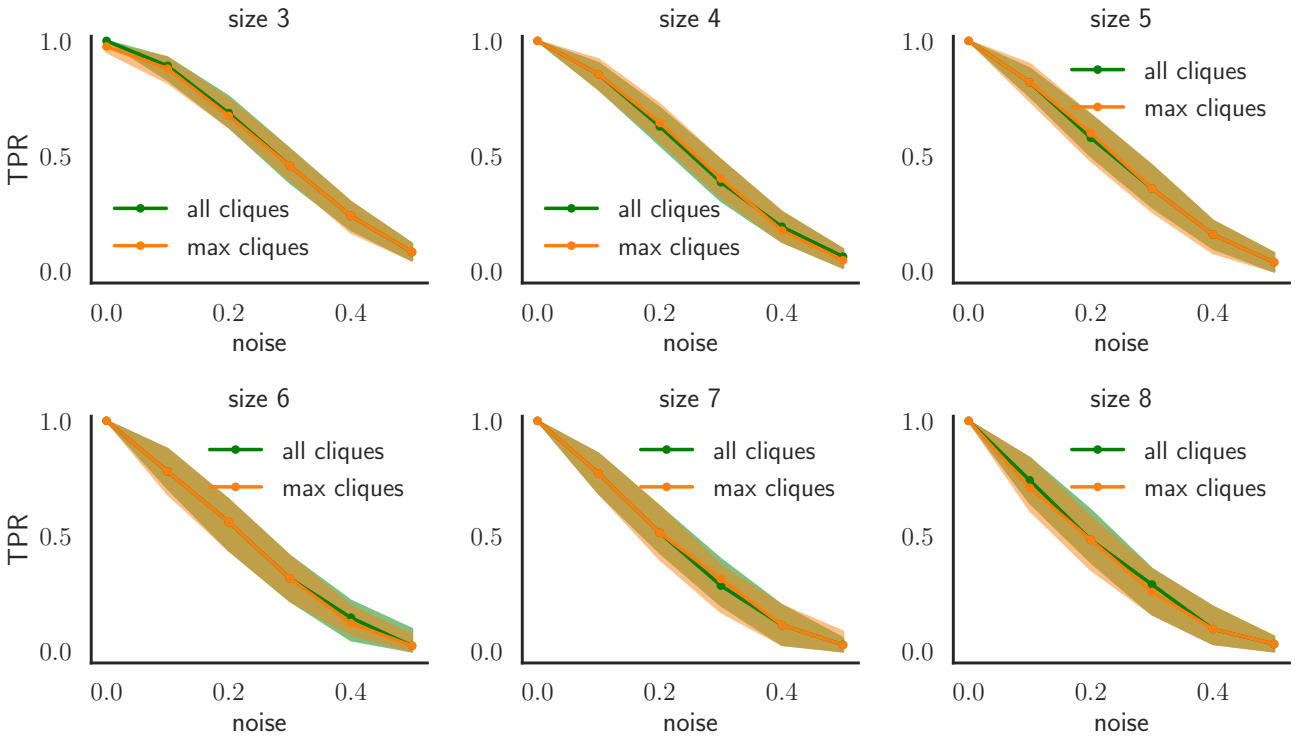

FIG. Supplementary Figure 4. Numerical simulations of the benchmark characterized by the presence of hyperlinks of size n ranging from 3 to 8 (see description in the main text). True Positive Rate of the hyperlinks obtained by detecting all cliques (green line) and the maximal cliques (orange line) from the SVN as a function of parameter p_n. The shaded area around the lines represents the interval between 10_{th} and 90_{th} percentile observed in 1000 realizations. Each panel is related to hyperlinks of different size n.

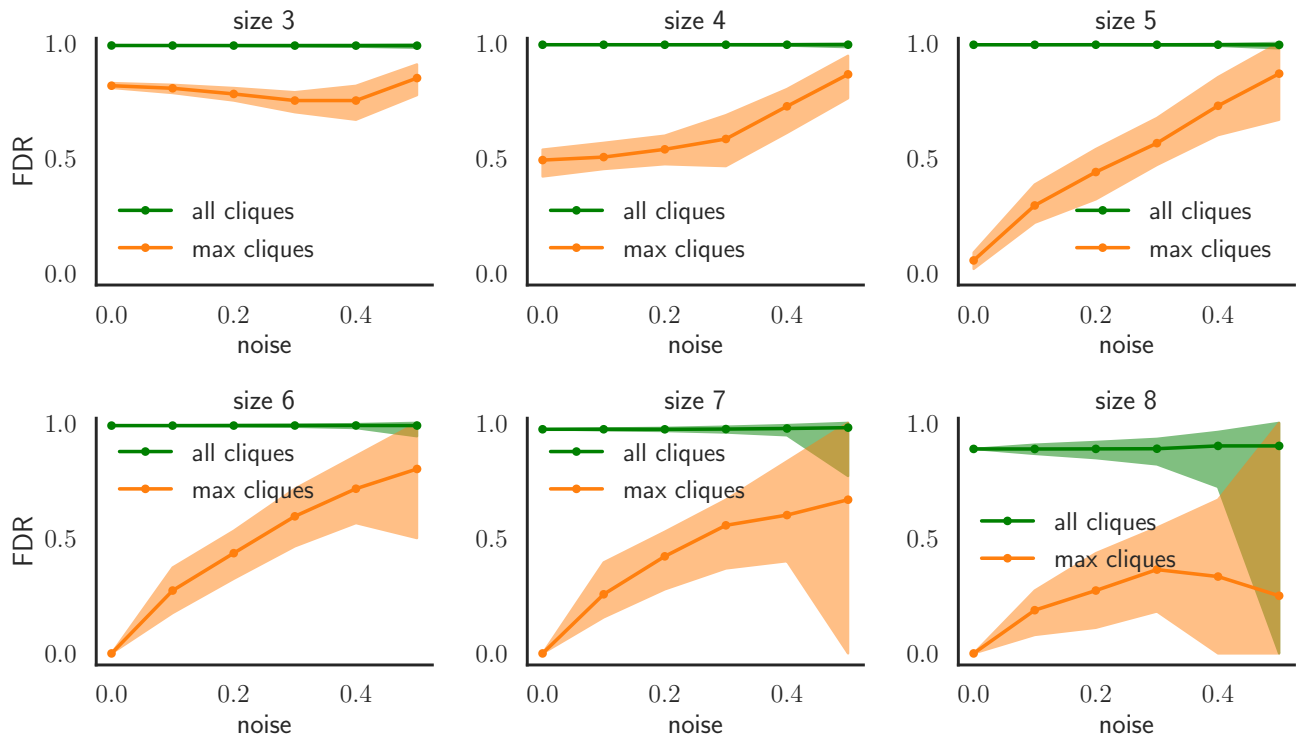

FIG. Supplementary Figure 5. Numerical simulations of the benchmark characterized by the presence of hyperlinks of size n ranging from 3 to 8 (see description in the main text). False Discovery Rate of the hyperlinks obtained by detecting all cliques (green line) and the maximal cliques (orange line) from the SVN as a function of parameter p_n. The shaded area around the lines represents the interval between 10_{th} and 90_{th} percentile observed in 1000 realizations. Each panel is related to hyperlinks of different size n.

Size	Justices	Votes	Possible groups	Observed groups	Validated groups
1	34	1089	-	-	-
2	37	1390	264	180	21
3	38	1621	868	354	49
4	38	1447	1666	358	49
5	38	1764	2016	465	60
6	38	1668	1568	484	91
7	38	1336	764	360	84
8	38	1474	213	167	86
9	38	2601	26	26	26

TABLE Supplementary Table 3. Summary of the co-decisions (i.e. same type of vote) of justices in groups of different size. For all sizes we report the number of justices active at least once and the number of votations in which groups of that size were observed. For all sizes larger than 1 we report the number of possible groups of co-voting justices (taking into account as a contrain the period of co-activity of justices), the number of observed groups and the number of groups validated in the SVH.

Supplementary Note 2. US SUPREME COURT

The US Supreme Court is a system characterized by a high density of interactions, with a relatively small number of justices (38) interacting over 8915 sentences. Table Supplementary Table 3 offers a breakdown of the activity of justices by reporting the number of justices and groups active in co-decisions of different sizes, together with the number of hyperlinks validated in the SVH. Other evidences of this high density are given by Figure Supplementary Figure 6a, that plots the number of sentences as a function of majority size and Figure Supplementary Figure 6b, that shows the median number of voting decisions in groups of different sizes made by the justices. This high density results in a poor performance of the SVN method in detecting validated hyperlinks, as shown in Figure Supplementary Figure 6c. Figure Supplementary Figure 6d shows how the performance of SVN remains poor also after dropping all unanimous sentences from the data.

In Figure 3a of the main text we show how the groups of justices validated by the SVH approach have a higher similarity with respect to the Seagal Cover score. In Table Supplementary Table 4 we report the SC score of all the 38 justices active in the dataset. Together with the SC score we also report the scores on three separate areas, (i) Criminal procedures, (ii) Civil rights and (iii) Economic issues. A summary of the activity of justices and the outcome of SVH for these three areas is presented in Tables Supplementary Table 5, Supplementary Table 6 and Supplementary Table 7.

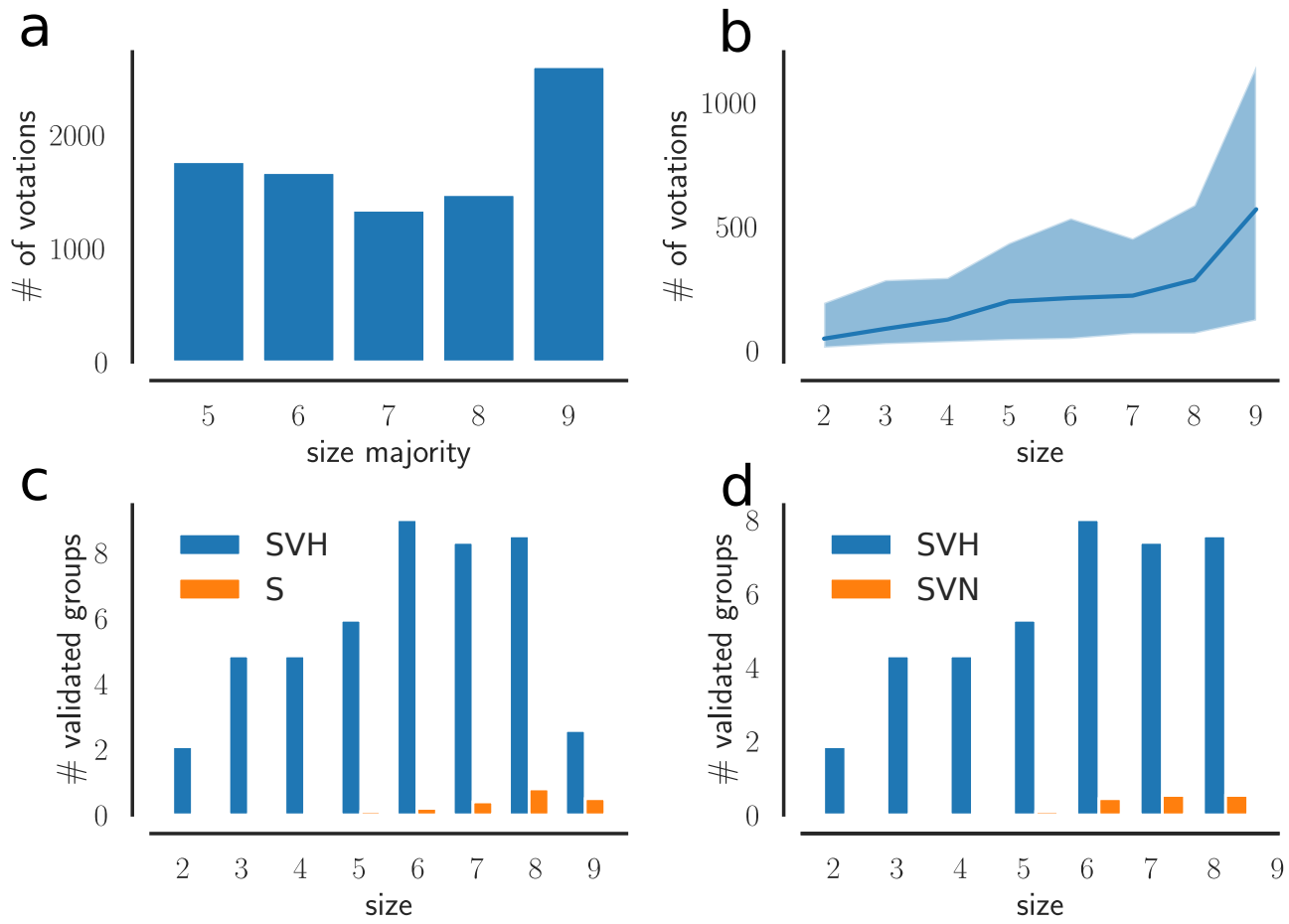

FIG. Supplementary Figure 6. **a)** Number of sentences as a function of the size of majority in the US Supreme Court. **b)** Median of the number of sentences made by a justice as a function of the size of the co-voting group (blue line). The shaded area around the line represents the interval between 10_{th} and 90_{th} percentile of the distribution of all justices. **c)** Number of validated groups in the SVH (blue bars) and in the SVN (orange bars) as a function of group size. **d)** Same of c) after excluding all unanimous votations when extracting the groups with the SVN approach.

name	SC score	Criminal procedure	Civil rights	Economic
HLBlack	0.875	0.668	0.701	0.811
SFReed	0.725	0.318	0.471	0.658
FFrankfurter	0.665	0.5	0.622	0.54
WODouglas	0.730	0.804	0.872	0.796
FMurphy	1.000	0.805	0.82	0.759
RHJackson	1.000	0.378	0.519	0.513
WBRutledge	1.000	0.738	0.723	0.748
HHBurton	0.280	0.284	0.433	0.533
FMVinson	0.750	0.27	0.453	0.599
TCClark	0.500	0.321	0.535	0.725
SMinton	0.720	0.231	0.362	0.62
EWarren	0.750	0.733	0.81	0.813
JHarlan2	0.875	0.378	0.42	0.367
WJBrennan	1.000	0.762	0.825	0.698
CEWhittaker	0.500	0.398	0.442	0.336
PStewart	0.750	0.449	0.484	0.45
BRWhite	0.500	0.329	0.552	0.593
AJGoldberg	0.750	0.75	0.98	0.632
AFortas	0.845	0.815	0.821	0.722
TMarshall	1.000	0.801	0.845	0.63
WEBurger	0.115	0.19	0.374	0.45
HABlackmun	0.115	0.423	0.621	0.55
LFPowell	0.165	0.284	0.403	0.456
WHRehnquist	0.045	0.196	0.306	0.494
JPStevens	0.250	0.669	0.643	0.562
SDOConnor	0.415	0.264	0.454	0.465
AScalia	0.000	0.274	0.306	0.462
AMKennedy	0.365	0.333	0.44	0.446
DHSouter	0.325	0.572	0.693	0.542
CThomas	0.160	0.224	0.25	0.462
RBGinsburg	0.680	0.623	0.7	0.566
SGBreyer	0.475	0.556	0.697	0.5
JGRoberts	0.120	0.311	0.424	0.445
SAAlito	0.100	0.186	0.388	0.424
SSotomayor	0.780	0.669	0.707	0.5
EKagan	0.730	0.625	0.719	0.5
NMGorsuch	0.110	0	0.5	0.333
BMKavanaugh	0.070	-	-	-

TABLE Supplementary Table 4. Seagal Cover score of all the 38 justices present in our dataset. The second, third and fourth columns report the scores related to votations concerning respectively (i) criminal procedures, (ii) civil rights and (iii) economic issues.

Supplementary Note 3. HIGH SCHOOL STUDENTS

In this section we present additional information on the analysis performed on the French high school dataset. Figure Supplementary Figure 7a plots the number of observed hyperlinks as a function of their size, showing the predominance of interactions at smaller sizes. The distribution of validated hyperlinks in the SVH (blue bars in Figure Supplementary Figure 7b) reproduces the same pattern of the unfiltered hypergraph, while the distribution of groups detected by the SVN (orange bars in Figure Supplementary Figure 7c) has a different shape. However, most of the groups detected by the SVN at higher sizes are spurious (i.e. they don't correspond to observed hyperlinks of that size in simultaneous interactions of students) as shown in Figure Supplementary Figure 7c.

To enrich the description of the hyperlinks validated in the SVH we compared both the SVH and the unfiltered hypergraph with the diaries data self reported by the student. Table Supplementary Table 8 reports the details of this comparison, showing how the SVH is more precise (i.e. it detect less spurious groups) but less accurate (i.e. it detects less true groups) of the unfiltered hypergraph.

At the end of the section in the main text we stress how the detection of validated hyperlinks in the SVH is more nuanced than a filtering procedure based on putting a threshold on weights. Figure Supplementary Figure 8 shows

Size	Justices	Votes	Possible groups	Observed groups	Validated groups
1	25	202	-	-	-
2	36	360	264	82	9
3	37	435	868	141	22
4	38	433	1666	130	24
5	38	505	2016	180	32
6	38	436	1568	180	25
7	38	321	764	134	23
8	38	273	213	89	21
9	38	481	26	25	24

TABLE Supplementary Table 5. Summary of the co-decisions of justices in groups of different size for the votations related to criminal procedures. For all sizes we report the number of justices active at least once and the number of votations in which groups of that size were observed. For all sizes larger than 1 we report the number of possible groups of co-voting justices (taking into account as a contrain the period of co-activity of justices), the number of observed groups and the number of groups validated in the SVH.

Size	Justices	Votes	Possible groups	Observed groups	Validated groups
1	24	173	-	-	-
2	33	218	264	75	8
3	38	289	868	113	11
4	37	232	1666	76	15
5	37	288	2016	101	22
6	38	276	1568	140	11
7	38	219	764	114	15
8	38	243	213	74	16
9	38	416	26	26	23

TABLE Supplementary Table 6. Summary of the co-decisions of justices in groups of different size for the votations related to civil rights. For all sizes we report the number of justices active at least once and the number of votations in which groups of that size were observed. For all sizes larger than 1 we report the number of possible groups of co-voting justices (taking into account as a contrain the period of co-activity of justices), the number of observed groups and the number of groups validated in the SVH.

the distributions of weights for the hyperlinks in the SVH and those that are not validated at different sizes. For all sizes there is an overlap in weights between the two groups. For these hyperlinks, the difference between being validated or not lies in the heterogeneous activity of its nodes. Indeed, 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.

Supplementary Note 4. PHYSICS AUTHORS

The APS dataset used in our analysis is the same of Ref: Battiston, Musciotto, Wang, Barabasi, Szell and Sinatra, Taking census of physics *Nature Reviews Physics*, 1(1):89-97, 2019. In the main text we focus on three PACS, General Physics (PACS hierarchical integer number 0), Nuclear Physics (PACS number 2) and Physics of Gases and Plasma (PACS number 5). Figure Supplementary Figure 9 plots the cumulative density functions of papers (panel a), groups validated in the SVH (panel b) and in the SVN (panel c) for all ten PACS.

In the main text we show that the correlation between the average number of papers written by research groups correlates with the fraction of groups validated with the SVH. Figure Supplementary Figure 10 shows the scatter plots for all sizes of hyperlinks from 2 to 10. Figure Supplementary Figure 11 shows the same plots for SVN. From the Figure we conclude that the two variables are not correlated.

Size	Justices	Votes	Possible groups	Observed groups	Validated groups
1	28	234	-	-	-
2	34	280	264	100	5
3	37	297	868	148	8
4	37	235	1666	135	12
5	38	295	2016	168	10
6	38	317	1568	188	19
7	38	278	764	156	19
8	38	302	213	109	28
9	38	525	26	26	23

TABLE Supplementary Table 7. Summary of the co-decisions of justices in groups of different size for the votations related to economic issues. For all sizes we report the number of justices active at least once and the number of votations in which groups of that size were observed. For all sizes larger than 1 we report the number of possible groups of co-voting justices (taking into account as a contrain the period of co-activity of justices), the number of observed groups and the number of groups validated in the SVH.

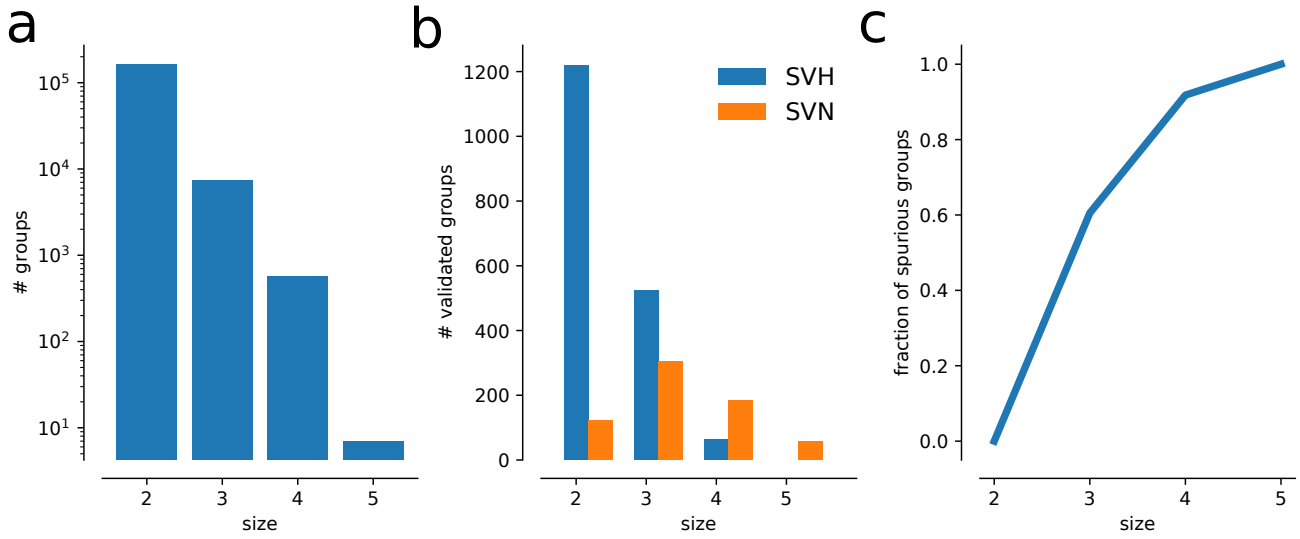

FIG. Supplementary Figure 7. **a)** Number of hyperlinks observed in the high school data as a function of size. **b)** Number of validated hyperlinks in the SVH (blue bars) and in the SVN (orange bars) as a function of size. **c)** Fraction of spurious hyperlinks (i.e. maximal pairwise cliques that do not correspond to observed hyperlinks) detected by the SVN as a function of size.

Supplementary Note 5. COMPUTATIONAL COMPLEXITY

The largest contribution to the computational complexity of the SVH approach is associated to the estimation of p-values, that are obtained through Eq. 3 of the main text as a convolution of hypergeometric functions. We investigated how the computational time grows as a function of the size of hyperlinks (Figure Supplementary Figure 12), finding an exponential growth after size 8. Another element that impacts the computational complexity is the weight of hyperlinks, as larger weights imply the evaluation of binomial coefficients with larger numbers. Figure Supplementary Figure 13, that plots the computational time as a function of weight, shows a linear dependence.

size	# diaries	# SVH	# unfiltered	SVH ∩ diaries	unfiltered ∩ diaries	spurious SVH	spurious unfiltered
2	202	351	1322	78	191	0.78	0.86
3	90	61	265	14	30	0.77	0.89
4	16	3	14	0	0	1.00	1.00

TABLE Supplementary Table 8. Summary of the comparison between the diaries cliques and the hyperlinks in the SVH and the unfiltered hypergraph. For each hyperlink size we report the number of groups in the diaries data, in the SVH and in the unfiltered hypergraph, followed by the intersection with the diaries and the fraction of spurious groups.

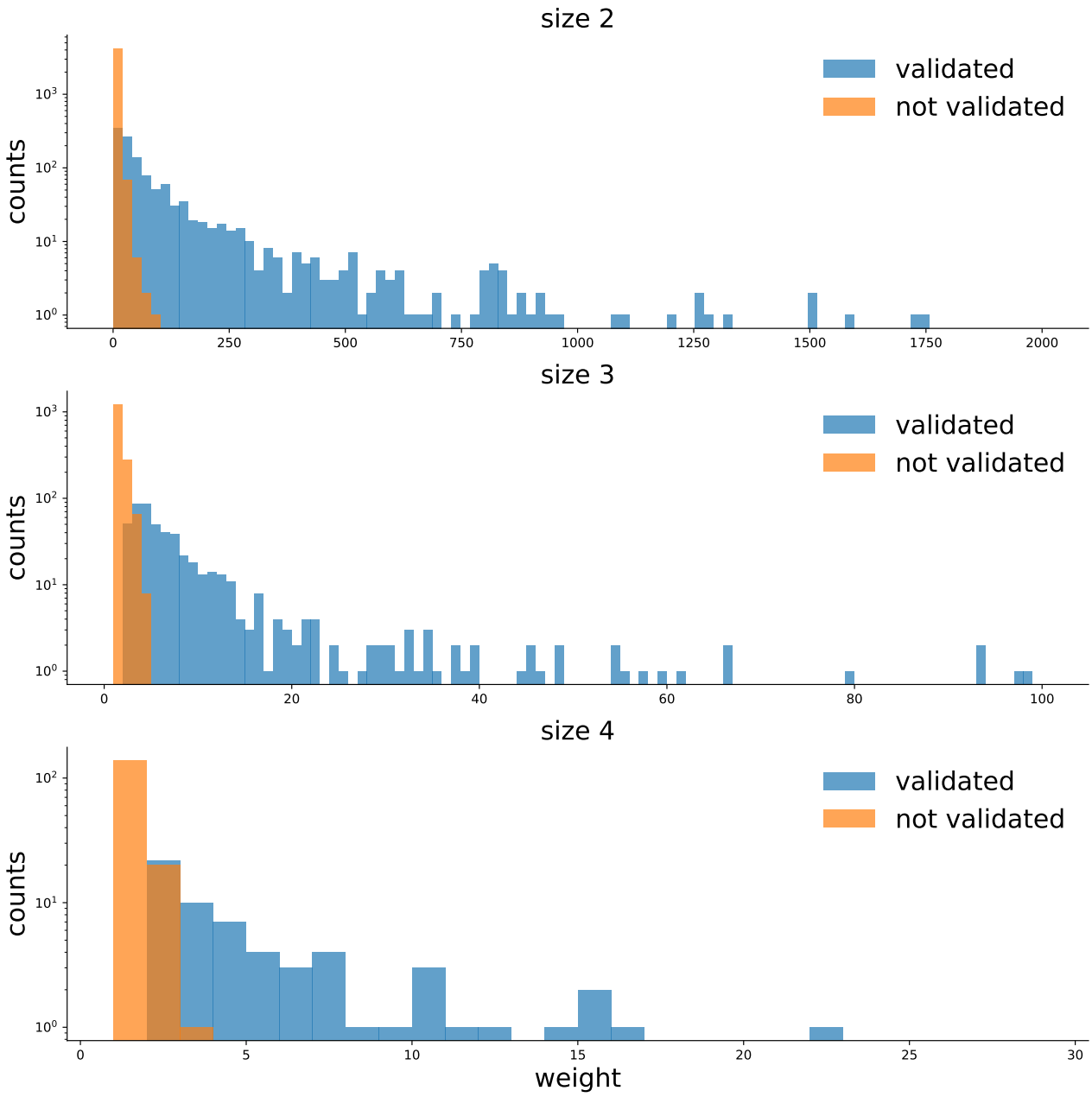

FIG. Supplementary Figure 8. Distribution of weights for the validated (blue bars) and the not validated (orange bars) hyperlinks at size 2 (top panel), size 3 (middle panel) and size 4 (bottom panel).

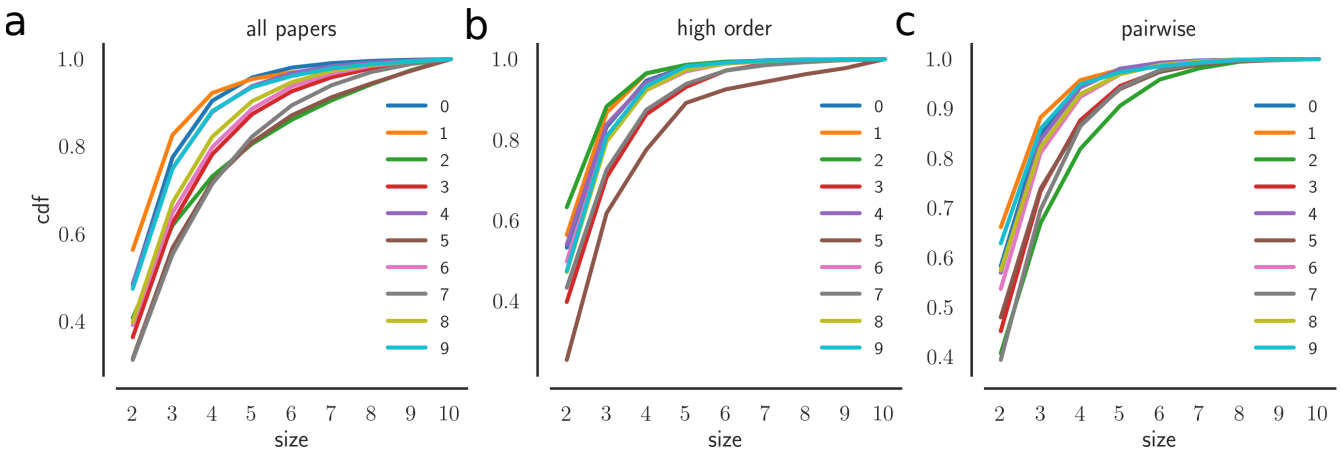

FIG. Supplementary Figure 9. **a)** Cumulative density function of papers as a function of team size for all ten PACS. **b)** Cumulative distribution function of hyperlinks in the SVH as a function of their size for all ten PACS. **c)** Cumulative distribution function of groups detected in the SVN as a function of their size for all ten PACS. For all panels, the legend reports the color code of the different PACS.

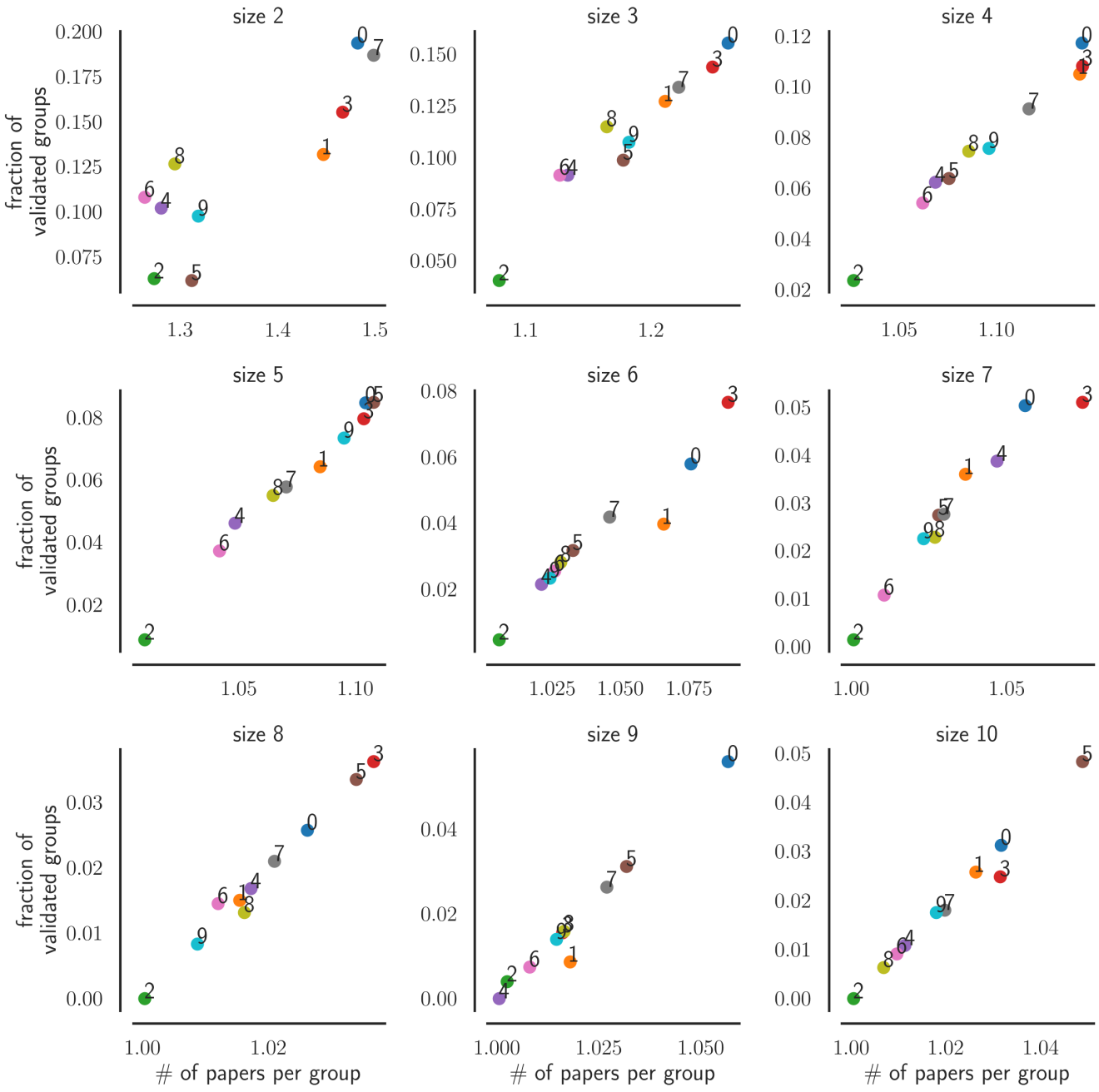

FIG. Supplementary Figure 10. Scatter plots of the average number of papers written by a group vs the fraction of validated hyperlinks in the SVH. Each dot represents a different PACS and each panel is related to a different group size. The color code for PACS is the same of Figure Supplementary Figure 9. The number on each dot is the PACS identifier.

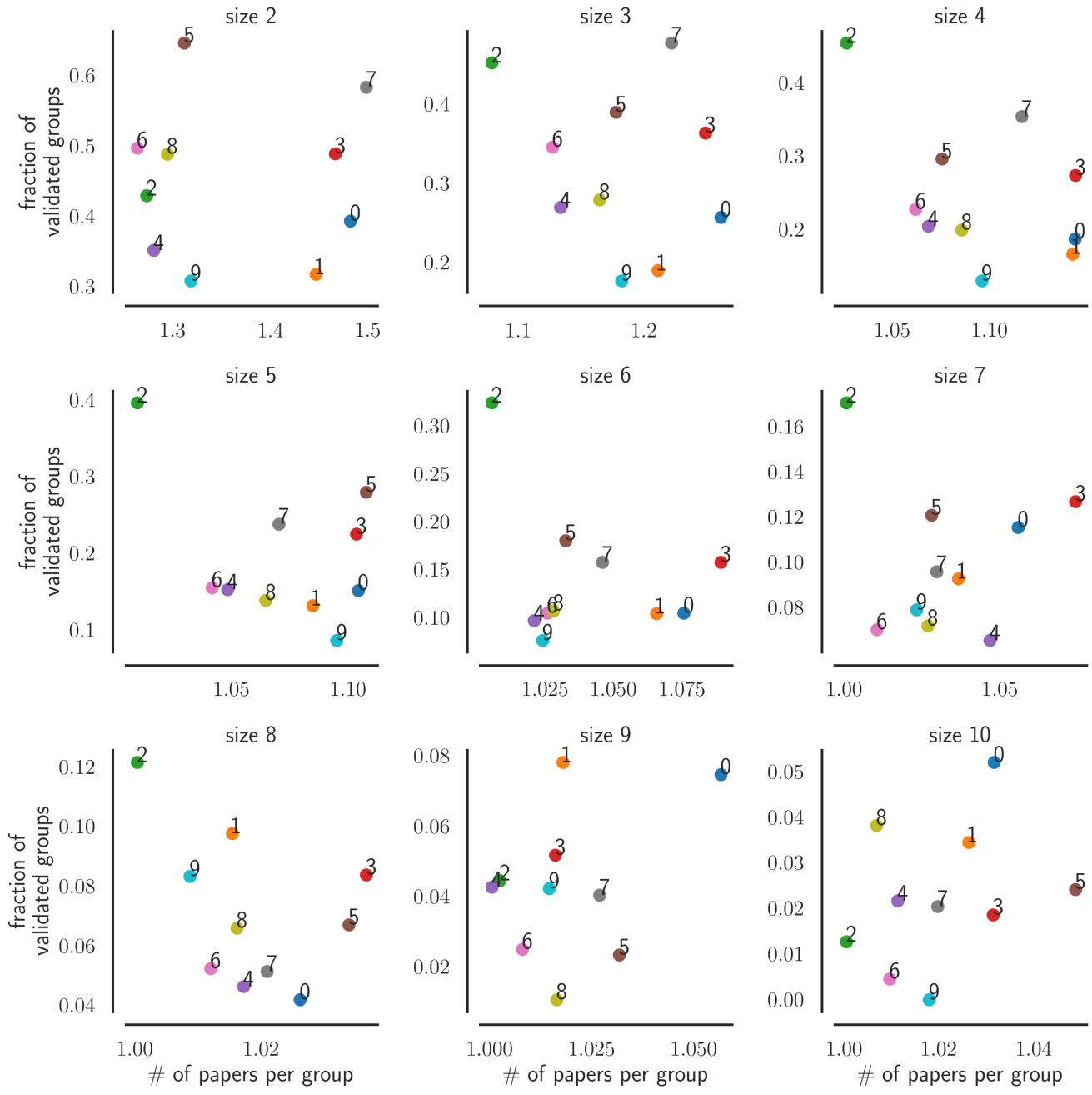

FIG. Supplementary Figure 11. Scatter plots of the average number of papers written by a group vs the fraction of validated hyperlinks in the SVH. Each dot represents a different PACS and each panel is related to a different group size. The color code for PACS is the same of Figure Supplementary Figure 9. The number on each dot is the PACS identifier.

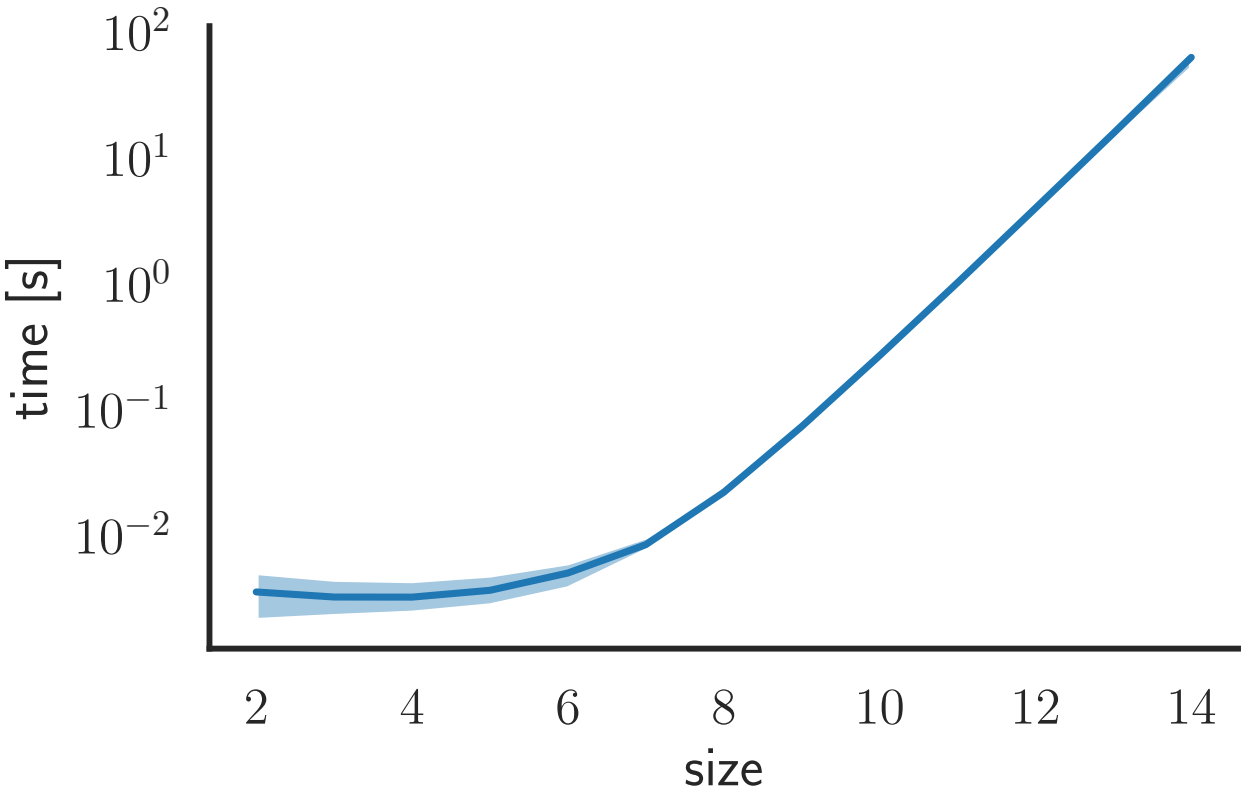

FIG. Supplementary Figure 12. Median over 1000 realizations of the computational time needed on a single core to validate a hyperlink as a function of its size (blue line). The shaded area represent the interval between the 10_{th} and the 90_{th} percentile.

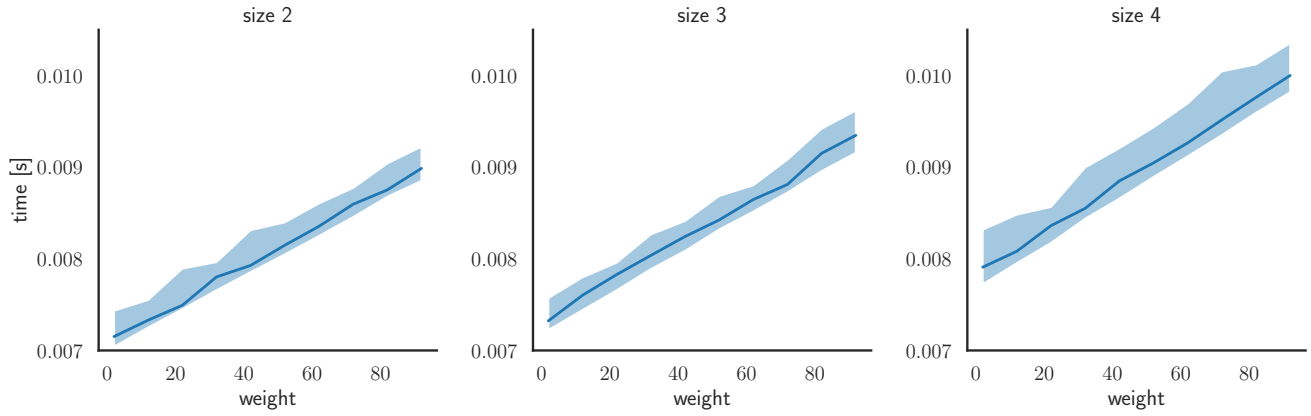

FIG. Supplementary Figure 13. Median over 1000 realizations of the computational time needed on a single core to validate a hyperlink as a function of its weight (blue line) for different hyperlinks sizes. The shaded area represent the interval between the 10_{th} and the 90_{th} percentile.