communications physics

ARTICLE

https://doi.org/10.1038/s42005-022-00890-7 OPEN

Detecting mesoscale structures by surprise

Emiliano Marchese [1✉], Guido Caldarelli[2] & Tiziano Squartini [1,3]

The importance of identifying mesoscale structures in complex networks can be hardly overestimated. So far, much attention has been devoted to detect modular and bimodular structures on binary networks. This effort has led to the definition of a framework based upon the score function called 'surprise', i.e. a p-value that can be assigned to any given partition of nodes. Hereby, we make a step further and extend the entire framework to the weighted case: six variants of surprise, induced by just as many variants of the hypergeometric distribution, are, thus, considered. As a result, a general, statistically grounded approach for detecting mesoscale network structures via a unified, suprise-based framework is presented. To illustrate its performances, both synthetic benchmarks and real-world configurations are considered. Moreover, we attach to the paper a Python code implementing all variants of surprise discussed in the present manuscript.

[1] IMT School for Advanced Studies, Lucca 55100, Italy. [2] Ca Foscari University of Venice, Venice 30123, Italy. [3] Institute for Advanced Study, University of Amsterdam, Amsterdam 1012 GC, The Netherlands. ✉email: emiliano.marchese@imtlucca.com

The importance of identifying the signature of some kind of mesoscopic organization in complex networks (be it due to the presence of communities or bipartite, core-periphery, bow-tie structures) can be hardly overestimated[1,2], the best example of complex systems whose behavior is deeply affected by their mesoscopic structural organization (e.g. resilience to the propagation of shocks, to the failure of nodes, etc.) being provided by financial networks[3–6].

So far, much attention has been devoted to the detection of binary mesoscale structures, i.e. communities and, to a far less extent, core-periphery structures: the efforts to solve these problems have led to a number of approaches that are briefly sketched below (for a detailed review of them, see refs. [7,8]).

Community detection has been initially approached by attempting a definition of communities based on the concepts of clustering, cliques, k-core; core-periphery structures have, instead, been defined in a purely top-down fashion, by imagining a fully connected subgraph (i.e. the core) surrounded by (peripherical) vertices exclusively linked to the first ones[3]. As stressed in[5], the deterministic character of these definitions makes their tout-court application to real-world systems extremely difficult. This is the reason why the intuitive requirements that 'the number of internal edges is larger than the number of external edges' and that 'the core portion of a network is densely connected, while its periphery is loosely connected'[8] are, now, interpreted in a purely probabilistic way: a community has, thus, become a subgraph whose vertices have a larger probability to be inter-connected than to be connected to any other vertex in the graph—and analogously for the core-periphery structure. In other words, the top-down approach defining a golden standard and looking for (deviations from) it has left the place to a bottom-up approach where structures are supposed to emerge as the result of non-trivial (i.e. non-casual) interactions between the nodes.

This change of perspective leads to a number of problems. The first one concerns the definition of models stating how edges are formed and has been solved by adopting the rich formalism defining the Exponential Random Graphs framework[9]; the second, and most important, one concerns the definition of a (statistically sound) procedure for selecting the best model among the ones providing competing descriptions of the data. This has led to the identification of three broad classes of algorithms—for an alternative classification see ref. [10]: according to our intuition, all of them implement some kind of statistical inference, the major difference lying in the way the corresponding test of hypothesis is implemented; from a practical point of view, instead, all these methods are designed for optimization, the functional form of the specific score function determining the class to which a given algorithm belongs.

The most representative algorithms among those belonging to the first class are the ones based on modularity. Although these methods compare the empirical network with a benchmark, they do not provide any indication of the statistical significance of the recovered partition, the reason being that none of them is designed as a proper statistical test. For instance, let us consider the definition of modularity, whose generic addendum is proportional to the term $(a_{ij} - p_{ij})$: while it embodies a comparison between the (empirical) adjacency matrix $\mathbf{A}$ and the matrix of probability coefficients $\mathbf{P}$ defining the benchmark, it does not implement any proper test of hypothesis. The algorithms prescribing to maximize a plain likelihood function belong to this group as well. Although popular, this way of proceeding is known to be affected by overfitting issues: an example is provided by the straight maximization of the likelihood defining the Stochastic Block Model (SBM), or its degree-corrected version, over the whole set of possible partitions, that outputs the trivial one where each vertex is a cluster on its own[11].

The aforementioned, major limitation is overcome by the algorithms belonging to the second class. They implement tests of hypothesis either à la Fisher or à la Neyman-Pearson, i.e. either defining a single benchmark (the null hypothesis of the first scenario) or two, alternative ones (the null and the alternative hypothesis of the second scenario): from a practical point of view, such a result is achieved by identifying the aforementioned benchmarks with proper probability distributions and the best partition of nodes with the one minimizing the corresponding p-value. Surprise-based algorithms belong to this second group: as it has been shown in ref. [12]—for a particular case; such a result will be generalized in what follows—optimizing (asymptotic) surprise amounts at carrying out a (sort of) Likelihood Ratio Test aimed at choosing between two alternative models.

Hypothesis testing can be further refined by allowing for more than two hypotheses to be tested at a time: results of the kind are particularly useful for model selection and, in fact, have produced a plethora of criteria (e.g. the Akaike Information Criterion, the Bayesian Information Criterion and the Minimum Description Length) for singling out the best statistical model out of a basket of competing ones. Generally speaking, optimization, here, seeks for the maximum of a corrected likelihood function embodying the trade-off between accuracy and parsimony of a description. An example of the algorithms belonging to this third class is represented by Infomap; other examples are represented by the recipes employing the SBM within a Bayesian framework (see ref. [13] and the references therein).

With this paper, we pose ourselves within the second research line and adopt a bottom-up approach that prescribes to compare any empirical network structure with the outcome of a properly-defined benchmark model. To this aim, we devise a unified framework for mesoscale structures detection based upon the score function called 'surprise', i.e. a p value that can be assigned to any given partition of nodes, on both undirected and directed networks (our function is named 'surprise' since it generalizes the function proposed in ref. [14] for community detection; in our case, however, it indicates a proper probability and not its logarithm, as in ref. [14]): while for binary community detection this is achieved by employing the binomial hypergeometric distribution[14,15], with bimodular structures like the bipartite and the core-periphery ones, one needs to consider its multinomial variant[12]. Here, we aim at making a step further, by extending the entire framework to the weighted case: as a result, we present a general, statistically grounded approach to the problem of detecting mesoscale structures on networks via a unified, surprise-based framework.

Other examples of the use of the hypergeometric distribution to carry out tests of hypothesis on networks can be found in the papers[16] (where the authors introduce a method to provide a statistically validated, monopartite projection of a bipartite network—the considered null hypothesis encoding the heterogeneity of the system),[17] (where the authors employ the same validation procedure to detect cores of communities within each set of nodes of a bipartite system) and[18] (where the authors extend the framework proposed in the aforementioned references to carry out a statistical validation of motifs observed in hypergraphs). For a review on the use of the hypergeometric distribution for network analyses see ref. [19] and the references therein.

Results

Surprise has recently received a lot of attention: the advantages of employing such a score function have been extensively discussed in[12,14,15,20–23] where researchers have tested and compared its performance from a purely numerical perspective. A characterization of the statistical properties of surprise is, however, still missing: for this reason, we will, first, make an effort to translate

the problem of detecting a given mesoscale network structure into a proper, exact significance test (to the best of our knowledge, the only other attempt of the kind—specifically, to detect communities—is the one in ref. [24]) and, then, show how the rich—yet, still underexplored—surprise-based formalism can properly answer such a question.

The basic equation underlying exact tests reads

$$\Pr(x \geq x^*) = \sum_{x \geq x^*} f(x) \tag{1}$$

and returns the probability of observing an outcome of the random variable X that is more extreme than the realized one, i.e. x^*. In the setting above, f represents the distribution encoding the null hypothesis and $\Pr(x \geq x^*)$—commonly known with the name of p value—answers the question 'is the realized value $X = x^*$ compatible with the (null) hypothesis that X is distributed according to f?'.

The p-value constitutes the basic quantity for carrying out any significance test: hence, in what follows we will tackle the problem of detecting the signature of a (statistically significant) mesoscale network structure by individuating a specific test (or, equivalently, a suitable functional form for f).

Community detection in binary networks. The topic of nodes partitioning into densely-connected groups has traditionally received a lot of attention, with applications ranging from the analysis of social networks to the definition of recommendation systems[1,2]. Within the surprise-based framework, binary community detection is carried out via the identification

$$\begin{aligned} f(l_\bullet) &\equiv \mathrm{H}(l_\bullet | V, V_\bullet, L) \\ &= \frac{\prod_{i=\bullet,\circ} \binom{V_i}{l_i}}{\binom{V}{L}} = \frac{\binom{V_\bullet}{l_\bullet}\binom{V_\circ}{l_\circ}}{\binom{V}{L}} = \frac{\binom{V_\bullet}{l_\bullet}\binom{V-V_\bullet}{L-l_\bullet}}{\binom{V}{L}} \end{aligned} \tag{2}$$

i.e. by calculating the p value

$$\mathscr{S} \equiv \sum_{l_\bullet \geq l_\bullet^*} f(l_\bullet) \tag{3}$$

of a binomial hypergeometric distribution whose parameters read as above. In this formalism, the $\bullet$ subscript will be meant to indicate quantities that are internal to communities, while the $\circ$ subscript will be meant to indicate quantities that are external to communities. More precisely, the binomial coefficient $\binom{V_\bullet}{l_\bullet}$ enumerates the number of ways $l_\bullet$ links can be redistributed within communities, i.e. over the available $V_\bullet$ node pairs, while the binomial coefficient $\binom{V_\circ}{l_\circ}$ enumerates the number of ways the remaining $l_\circ = L - l_\bullet$ links can be redistributed between communities, i.e. over the remaining $V_\circ = V - V_\bullet$ node pairs. Notice that, although $l_\bullet$ is 'naturally' bounded by the value $V_\bullet$, it cannot exceed L - whence the usual requirement $l_\bullet \in [l_\bullet^*, \min\{L, V_\bullet\}]$.

From a merely statistical point of view, surprise interprets a network as a population of V node pairs, L of which have been drawn; out of the L extracted ones, $l_\bullet$ node pairs have the desired feature of being internal to communities. Hence, for a given partition of nodes, $\mathscr{S}$ quantifies the probability of observing at least $l_\bullet^*$ successes (i.e. intra-cluster edges) out of L draws: the lower this probability, the 'more surprising' the observation of the corresponding partition, hence the better the partition itself.

Bimodular structures detection in binary networks. The surprise-based framework can be easily extended to detect what can be called bimodular structures, a term that will be used to compactly indicate core-periphery[25–27] and bipartite structures[28,29]. The reason for adopting such a terminology lies in the evidence that both kinds of structures are defined by bimodular partitions, i.e. partitions of nodes into two different groups.

As shown elsewhere[12], the issue of detecting binary bimodular structures can be addressed by considering a multivariate (or multinomial) hypergeometric distribution, i.e. by identifying

$$\begin{aligned} f(l_\bullet, l_\circ) &\equiv \mathrm{MH}(l_\bullet, l_\circ | V, V_\bullet, V_\circ, L) \\ &= \frac{\prod_{i=\bullet,\circ,\top} \binom{V_i}{l_i}}{\binom{V}{L}} = \frac{\binom{V_\bullet}{l_\bullet}\binom{V_\circ}{l_\circ}\binom{V_\top}{l_\top}}{\binom{V}{L}} \\ &= \frac{\binom{V_\bullet}{l_\bullet}\binom{V_\circ}{l_\circ}\binom{V-(V_\bullet+V_\circ)}{L-(l_\bullet+l_\circ)}}{\binom{V}{L}} \end{aligned} \tag{4}$$

where $V_\top \equiv V - (V_\bullet + V_\circ)$ indicates the number of node pairs between the modules $\bullet$ and $\circ$ and $l_\top \equiv L - (l_\bullet + l_\circ)$ indicates the number of links that must be assigned therein. While the binomial coefficient $\binom{V_\bullet}{l_\bullet}$ enumerates the number of ways $l_\bullet$ links can redistributed within the first module (e.g. the core portion) and the binomial coefficient $\binom{V_\circ}{l_\circ}$ enumerates the number of ways $l_\circ$ links can redistributed within the second module (e.g. the periphery portion), the third binomial coefficient $\binom{V-(V_\bullet+V_\circ)}{L-(l_\bullet+l_\circ)}$ enumerates the number of ways the remaining $L - (l_\bullet + l_\circ)$ links can be redistributed between the first and the second module, i.e. over the remaining $V - (V_\bullet + V_\circ)$ node pairs.

This choice induces the definition of the binary bimodular surprise

$$\mathscr{S}_{/\!/} \equiv \sum_{l_\bullet \geq l_\bullet^*} \sum_{l_\circ \geq l_\circ^*} f(l_\bullet, l_\circ); \tag{5}$$

analogously to the univariate case, $l_\bullet$ and $l_\circ$ are naturally bounded by the values $V_\bullet$ and $V_\circ$ — notice, however, that the sum $l_\bullet + l_\circ$ cannot exceed L (although it may not reach such a value, e.g. in case $V_\bullet + V_\circ < L$).

Community detection in weighted networks. Within the surprise-based framework, the problem of detecting binary communities has been rephrased as an aleatory experiment whose random variable is the number of links within communities. Interestingly enough, such an experiment can be easily mapped into a counting problem, allowing us to interpret $\mathscr{S}$ as indicating the number of configurations whose number of internal links (i.e. within communities) is larger than the observed one.

When dealing with weighted networks, we would like to proceed along similar guidelines and consider the total, internal weight as our new random variable, to be redistributed across the available node pairs. Adopting this approach has three major consequences: (1) weights must be considered as composed by an integer number of binary links; (2) each node pair must be allowed to be occupied by more than one link; (3) the total weight must be allowed to vary even beyond the network size (when handling real-world networks, the case $W \gg V$ is often encountered).

The proper setting to define an aleatory experiment satisfying the requests above is provided by the so-called stars and bars model, a combinatorial technique that has been introduced to handle the counting of configurations with multiple occupancies. Basically, the problem of counting in how many ways $w_\bullet$ particles (our links) can be redistributed among $V_\bullet$ boxes (our node pairs), while allowing more than one particle to occupy each box, can be tackled by allowing both the particles and the bars delimiting the boxes to be permuted[30]. Since $V_\bullet$ boxes are delimited by $V_\bullet - 1$ bars, a term like $\binom{V_\bullet + w_\bullet - 1}{w_\bullet}$ is needed.

In order to better grasp the meaning of such a term, let us make a simple example. Let us imagine to observe a network with three nodes and two links, carrying a weight of 1 and 2, respectively. Now, were we interested in a purely binary analysis,

we may ask ourselves in how many ways we could place the two links among the $\frac{N(N-1)}{2}=\frac{3(3-1)}{2}=3$ available pairs: the answer is provided by the binary binomial coefficient $\binom{V}{l_\bullet}=\binom{3}{2}=3$. The implicit assumption we make is that the three links must not occupy the same node pairs - otherwise the total number of connections wouldn't be preserved.

This perspective changes from the purely weighted point of view. Since we are now interested in preserving just the total weight of our network, irrespectively of the number of connections it is placed upon, the number of admissible configurations becomes precisely $\binom{V_\bullet+w_\bullet-1}{w_\bullet}=\binom{2+3}{3}=10$. Such a number is larger than before since, now, weights are disaggregated into binary links and multiple occupations of the latter ones are allowed.

The considerations above lead us to generalize the community detection problem to the weighted case by identifying

$$\begin{aligned} f(w_\bullet) &\equiv \mathrm{NH}(w_\bullet|V+W,W,V_\bullet)\\ &=\frac{\prod_{i=\bullet,\circ}\binom{V_i+w_i-1}{w_i}}{\binom{V+W-1}{W}}=\frac{\binom{V_\bullet+w_\bullet-1}{w_\bullet}\binom{V_\circ+w_\circ-1}{w_\circ}}{\binom{V+W-1}{W}}\\ &=\frac{\binom{V_\bullet+w_\bullet-1}{w_\bullet}\binom{(V-V_\bullet)+(W-w_\bullet)-1}{W-w_\bullet}}{\binom{V+W-1}{W}} \end{aligned} \tag{6}$$

i.e. by replacing the binomial hypergeometric distribution considered in the purely binary case with a negative hypergeometric distribution, a choice inducing the definition of the weighted surprise

$$\mathscr{W} \equiv \sum_{w_\bullet \geq w_\bullet^*} f(w_\bullet) \tag{7}$$

where the binomial coefficient $\binom{V_\bullet+w_\bullet-1}{w_\bullet}=\binom{V_\bullet+w_\bullet-1}{V_\bullet-1}$ enumerates the number of ways $w_\bullet$ links can be redistributed within communities, i.e. over the available $V_\bullet$ node pairs, and the binomial coefficient $\binom{V_\circ+w_\circ-1}{w_\circ}=\binom{V_\circ+w_\circ-1}{V_\circ-1}$ enumerates the number of ways the remaining $w_\circ=W-w_\bullet$ links can be redistributed between communities, i.e. over the remaining $V_\circ=V-V_\bullet$ node pairs. Differently from the binary case, the sum ranges up to the empirical weight of the network, i.e. $w_\bullet\in[w_\bullet^*,W]$.

Bimodular structures detection in weighted networks. Let us now introduce the third generalization of the surprise-based formalism: following the same line of reasoning that led us to approach the detection of binary bimodular structures by considering the multinomial analogue of the distribution introduced for binary community detection, we are now led to focus on the multinomial (or multivariate) negative hypergeometric distribution, i.e.

$$\begin{aligned} f(w_\bullet,w_\circ) &\equiv \mathrm{MNH}(w_\bullet,w_\circ|V+W,W,V_\bullet,V_\circ)\\ &=\frac{\prod_{i=\bullet,\circ,\top}\binom{V_i+w_i-1}{w_i}}{\binom{V+W-1}{W}}\\ &=\frac{\binom{V_\bullet+w_\bullet-1}{w_\bullet}\binom{V_\circ+w_\circ-1}{w_\circ}\binom{V_\top+w_\top-1}{w_\top}}{\binom{V+W-1}{W}}\\ &=\frac{\binom{V_\bullet+w_\bullet-1}{w_\bullet}\binom{V_\circ+w_\circ-1}{w_\circ}\binom{V-(V_\bullet+V_\circ)+W-(w_\bullet+w_\circ)-1}{W-(w_\bullet+w_\circ)}}{\binom{V+W-1}{W}} \end{aligned} \tag{8}$$

while the binomial coefficient $\binom{V_\bullet+w_\bullet-1}{w_\bullet}$ enumerates the number of ways $w_\bullet$ links can redistributed within the first module (e.g. the core portion), the binomial coefficient $\binom{V_\circ+w_\circ-1}{w_\circ}$enumerates the number of ways $w_\circ$ links can be redistributed within the second module (e.g. the periphery portion) and the binomial coefficient $\binom{V-(V_\bullet+V_\circ)+W-(w_\bullet+w_\circ)-1}{W-(w_\bullet+w_\circ)}$ enumerates the number of ways the remaining $w_\top\equiv W-(w_\bullet+w_\circ)$ links can be redistributed between the first and the second module, i.e. over the remaining $V_\top\equiv V-(V_\bullet+V_\circ)$ node pairs. Such a position induces the definition of the weighted bimodular surprise

$$\mathscr{W}_{//} \equiv \sum_{w_\bullet\geq w_\bullet^*}\sum_{w_\circ\geq w_\circ^*} f(w_\bullet,w_\circ); \tag{9}$$

as for the (weighted) community detection, weights are understood as integer numbers—equivalently, as composed by an integer number of binary links. For what concerns the limits of the summations, $w_\bullet$ and $w_\circ$ are naturally bounded by W; notice, however, that the sum $w_\bullet+w_\circ$ itself cannot exceed such a value.

Enhanced community detection. The recipe to detect communities on weighted networks can be further refined to account for the information encoded into the total number of links, beside the one provided by the total weight, by combining two of the distributions introduced above. To this aim, let us proceed in a two-step fashion: first, let us recall that the number of ways L links can be placed among V node pairs, in such a way that $l_\bullet$ connections are internal to the clusters - while the remaining $L-l_\bullet$ ones are, instead, external - is precisely

$$\mathrm{H}(l_\bullet|V,V_\bullet,L)=\frac{\binom{V_\bullet}{l_\bullet}\binom{V-V_\bullet}{L-l_\bullet}}{\binom{V}{L}}; \tag{10}$$

now, for each of the binary configurations listed above, $W-L$ links remain to be assigned: while $w_\bullet-l_\bullet$ of them must be placed within the clusters, on top of the $l_\bullet$ available internal links, the remaining $(W-L)-(w_\bullet-l_\bullet)$ ones must be placed between the clusters, on top of the $L-l_\bullet$ available inter-cluster connections. Hence, the conditional negative hypergeometric distribution reading

$$\begin{aligned} \mathrm{NH}(w_\bullet|W,W-L,l_\bullet)=&\frac{\binom{l_\bullet+(w_\bullet-l_\bullet)-1}{w_\bullet-l_\bullet}}{\binom{L+(W-L)-1}{W-L}}\cdot\\ &\cdot\frac{\binom{(L-l_\bullet)+(W-L)-(w_\bullet-l_\bullet)-1}{(W-L)-(w_\bullet-l_\bullet)}}{\binom{L+(W-L)-1}{W-L}}; \end{aligned} \tag{11}$$

remains naturally defined; now, combining the two distributions above, simplifying and re-arranging, the generic term of the enhanced hypergeometric distribution can be rewritten as

$$\begin{aligned} \mathrm{EH}(l_\bullet,w_\bullet|V,V_\bullet,L,W)=&\mathrm{H}(l_\bullet|V,V_\bullet,L)\cdot\\ &\cdot\mathrm{NH}(w_\bullet|W,W-L,l_\bullet)\\ =&\frac{\binom{V_\bullet}{l_\bullet}\binom{V_\circ}{l_\circ}}{\binom{V}{L}}\cdot\frac{\binom{w_\bullet-1}{w_\bullet-l_\bullet}\binom{w_\circ-1}{w_\circ-l_\circ}}{\binom{W-1}{W-L}}; \end{aligned} \tag{12}$$

with a clear meaning of the symbols. An analytical characterization of it is provided into the Supplementary Note 3: for the moment, let us simply notice that the definition provided above works for the values $0<l_\bullet<L$.

By posing $f(l_\bullet,w_\bullet)\equiv\mathrm{EH}(l_\bullet,w_\bullet|V,V_\bullet,L,W)$, our distribution induces the definition of the enhanced surprise, i.e.

$$\mathscr{E} \equiv \sum_{l_\bullet\geq l_\bullet^*}\sum_{w_\bullet\geq w_\bullet^*} f(l_\bullet,w_\bullet); \tag{13}$$

although $l_\bullet$ and $w_\bullet-l_\bullet$ are naturally bounded by $V_\bullet$ and $W-L$, respectively, the former one cannot exceed L. In order to better understand how the enhanced surprise works, let us consider again the aforementioned example: given a network with three nodes and two links, carrying a weight of 1 and 2, respectively, we observe $\binom{V}{l_\bullet}=\binom{3}{2}=3$ (purely) binary configurations with

exactly the same number of links and $\binom{V_\bullet+w_\bullet-1}{w_\bullet}=\binom{2+3}{3}=10$ (purely weighted) configurations with exactly the same total weight. If we, now, constrain both the total number of links and the total weight of the network, the number of admissible configurations becomes $\binom{V_\bullet}{l_\bullet}\binom{w_\bullet-1}{w_\bullet-l_\bullet}=\binom{3}{2}\binom{3-1}{3-2}=3\cdot 2=6$, as it can be easily verified upon explicitly listing them. Naturally, the configurations admissible by the enhanced surprise are a subset of the configurations admissible by the weighted surprise, i.e. precisely the ones with the desired number of links (see also Fig. 1).

Enhanced bimodular structures detection. The last generalization of surprise concerns its use for the detection of bimodular structures within the enhanced framework. This amounts at considering the following multinomial variant of the enhanced hypergeometric distribution

$$\begin{aligned}&\mathrm{MEH}(l_\bullet,l_\circ,w_\bullet,w_\circ|V,V_\bullet,V_\circ,L,W)\\&=\frac{\binom{V_\bullet}{l_\bullet}\binom{V_\circ}{l_\circ}\binom{V_\top}{l_\top}}{\binom{V}{L}}\cdot\frac{\binom{w_\bullet-1}{w_\bullet-l_\bullet}\binom{w_\circ-1}{w_\circ-l_\circ}\binom{w_\top-1}{w_\top-l_\top}}{\binom{W-1}{W-L}}\\&=\frac{\binom{V_\bullet}{l_\bullet}\binom{V_\circ}{l_\circ}\binom{V-(V_\bullet+V_\circ)}{L-(l_\bullet+l_\circ)}}{\binom{V}{L}}\cdot\frac{\binom{w_\bullet-1}{w_\bullet-l_\bullet}\binom{w_\circ-1}{w_\circ-l_\circ}\binom{W-(w_\bullet+w_\circ)-1}{(W-L)-((w_\bullet+w_\circ)-(l_\bullet+l_\circ))}}{\binom{W-1}{L-1}}\end{aligned}\tag{14}$$

where $V_\top\equiv V-(V_\bullet+V_\circ)$ indicates the number of node pairs between the modules $\bullet$ and $\circ$ and $l_\top\equiv L-(l_\bullet+l_\circ)$ indicates the number of links that must be assigned therein. An analytical characterization of it is provided into the Supplementary Note 3: for the moment, let us simply notice that the definition provided above works for the values $0<l_\bullet,l_\circ<L$. The position $f(l_\bullet,l_\circ,w_\bullet,w_\circ)\equiv\mathrm{MEH}(l_\bullet,l_\circ,w_\bullet,w_\circ|V,V_\bullet,V_\circ,L,W)$ induces the definition of the enhanced bimodular surprise

$$\mathscr{E}_{/\!/}\sum_{l_\bullet\geq l_\bullet^*}\sum_{l_\circ\geq l_\circ^*}\sum_{w_\bullet\geq w_\bullet^*}\sum_{w_\circ\geq w_\circ^*}f(l_\bullet,l_\circ,w_\bullet,w_\circ).\tag{15}$$

Notice that $l_\bullet$ and $l_\circ$ are naturally bounded by $V_\bullet$ and $V_\circ$: still, their sum cannot exceed L; analogously, $w_\bullet-l_\bullet$ and $w_\circ-l_\circ$ are naturally bounded by $W-L$: still, their sum cannot exceed $W-L$.

As for its binomial counterpart, the expression of the MEH can be rearranged in a term-by-term fashion, in such a way that the module-specific binomial coefficients can be grouped together. Upon doing so, it becomes clearer that the MEH counts the number of ways $w_\bullet-l_\bullet$ links can be placed on top of the $l_\bullet$ binary links characterizing the connectance of the $\bullet$ module, times the number of ways $w_\circ-l_\circ$ links can be placed on top of the $l_\circ$ binary links characterizing the connectance of the $\circ$ module, times the number of ways the remaining $W-(w_\bullet+w_\circ)-(L-(l_\bullet+l_\circ))$ links can be placed on top of the $L-(l_\bullet+l_\circ)$ binary links characterizing the connectance of the third module.

Table 1 in the Methods section gathers all the variants of the surprise-based formalism, illustrating both the full and the asymptotic expression for each of them. To sum up, detecting a weighted mesoscale structure implies considering the negative version of the probability mass function working in the corresponding binary case (e.g. moving from the hypergeometric to the negative hypergeometric one); detecting a bimodular structure, instead, implies considering the multinomial version of the probability mass function working in the corresponding binary case (e.g. moving from the binomial hypergeometric to the multinomial one).

Comparing methods for the detection of mesoscale structures. Let us now carry out a comparison among some of the methods designed to detect mesoscale structures (for a consistency check of our surprise-based formalism and a theoretical comparison between modularity and surprise, we redirect the interested reader to the Supplementary Notes 4 and 5). To this aim, let us consider two popular algorithms for mesoscale structures detection, i.e. modularity maximization (Q has been considered in its full definition, e.g. $\langle a_{ij}\rangle=p_{ij}=\frac{k_ik_j}{2L},\forall\, i<j$ for binary undirected configurations) and Infomap[31], Upon doing so, we are able to compare one algorithm per class, i.e. modularity for the first class, surprise for the second class and Infomap for the third class.

To this aim, we have focused on different kinds of benchmarks, i.e. classes of synthetic networks with well-defined planted partitions, the aim being that of inspecting the goodness of a

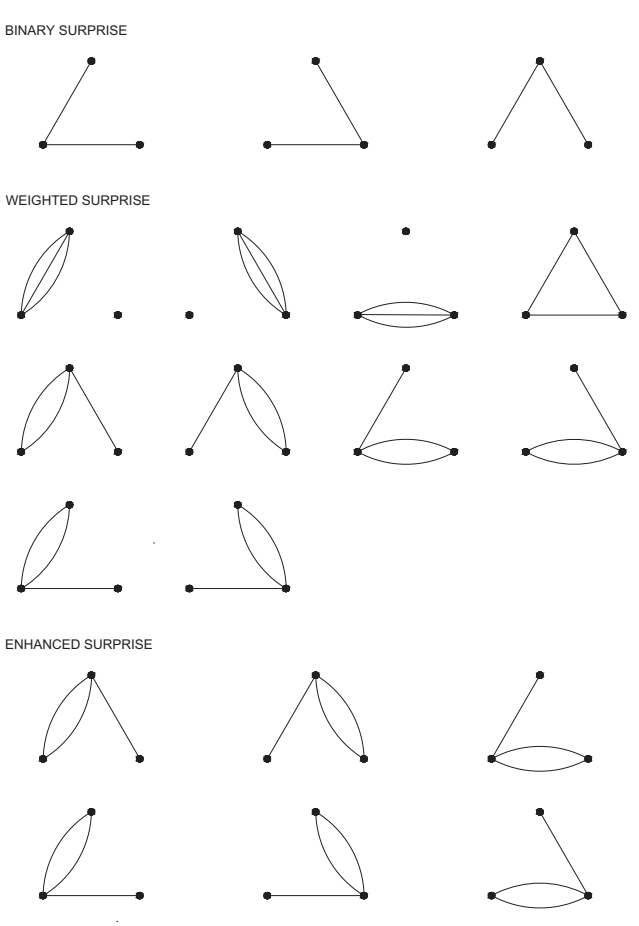

Fig. 1 Graphical comparison of different ways of counting admissible configurations when dealing with surprise. Let us imagine to observe a network with three nodes and two links, carrying a weight of 1 and 2, respectively. Were we interested in a purely binary analysis, we may ask ourselves in how many ways we could place the two links among the three available pairs: the answer is provided by the 'binary' binomial coefficient $\binom{V_\bullet}{l_\bullet}=\binom{3}{2}=3$. Were we interested in a purely weighted analysis, we may ask ourselves in how many ways we could place the two links among the three available pairs while preserving the total weight of our network, irrespectively of the number of connections it is placed upon; the number of admissible configurations becomes $\binom{V_\bullet+w_\bullet-1}{w_\bullet}=\binom{2+3}{3}=10$. Such a number is larger than before since, now, weights are 'disaggregated' into binary links and multiple occupations of the latter ones are allowed. Were we interested in the enhanced analysis, we may ask ourselves in how many ways we could place the two links among the three available pairs while preserving both the total number of links and the total weight of the network; the number of admissible configurations becomes $\binom{V_\bullet}{l_\bullet}\binom{w_\bullet-1}{w_\bullet-l_\bullet}=\binom{3}{2}\binom{3-1}{3-2}=3\cdot 2=6$, as can be easily verified upon explicitly listing them.

Table 1 Table illustrating all the generalizations of the surprise-based formalism proposed in the present paper.

	Community detection	Bimodular structures detection
Full expression		
Binary case	$f=\frac{\prod_{i=\bullet,\circ}\binom{V_i}{l_i}}{\binom{V}{L}}$	$f=\frac{\prod_{i=\bullet,\circ,\top}\binom{V_i}{l_i}}{\binom{V}{L}}$
Weighted case	$f=\frac{\prod_{i=\bullet,\circ}\binom{V_i+w_i-1}{w_i}}{\binom{V+W-1}{W}}=\frac{\prod_{i=\bullet,\circ}\left(\binom{V_i}{w_i}\right)}{\left(\binom{V}{W}\right)}$	$f=\frac{\prod_{i=\bullet,\circ,\top}\binom{V_i+w_i-1}{w_i}}{\binom{V+W-1}{W}}=\frac{\prod_{i=\bullet,\circ,\top}\left(\binom{V_i}{w_i}\right)}{\left(\binom{V}{W}\right)}$
Enhanced case	$f=\frac{\prod_{i=\bullet,\circ}\binom{V_i}{l_i}\binom{w_i-1}{l_i-1}}{\binom{V}{L}\binom{W-1}{W-L}}$	$f=\frac{\prod_{i=\bullet,\circ,\top}\binom{V_i}{l_i}\binom{w_i-1}{l_i-1}}{\binom{V}{L}\binom{W-1}{W-L}}$
Asymptotic expression		
Binary case	$f\propto\frac{\mathrm{Ber}(V,L,p)}{\prod_{i=\bullet,\circ}\mathrm{Ber}(V_i,l_i,p_i)}$	$f\propto\frac{\mathrm{Ber}(V,L,p)}{\prod_{i=\bullet,\circ,\top}\mathrm{Ber}(V_i,l_i,p_i)}$
Weighted case	$f\propto\frac{\mathrm{Geo}(V,W,q)}{\prod_{i=\bullet,\circ}\mathrm{Geo}(V_i,w_i,q_i)}$	$f\propto\frac{\mathrm{Geo}(V,W,q)}{\prod_{i=\bullet,\circ,\top}\mathrm{Geo}(V_i,w_i,q_i)}$
Enhanced case	$f\propto\frac{\mathrm{BF}(V,L,W,p,r)}{\prod_{i=\bullet,\circ}\mathrm{BF}(V_i,l_i,w_i,p_i,r_i)}$	$f\propto\frac{\mathrm{BF}(V,L,W,p,r)}{\prod_{i=\bullet,\circ,\top}\mathrm{BF}(V_i,l_i,w_i,p_i,r_i)}$

Table shows both the full and the asymptotic expression of each probability mass function—with the only exception of the numerical coefficient characterizing each asymptotic expression. While detecting a weighted mesoscale structure implies considering the negative version of the probability mass function working in the corresponding binary case (e.g. moving from the hypergeometric to the negative hypergeometric one), detecting a 'bimodular' structure, instead, implies considering the multinomial version of the probability mass function working in the corresponding binary case (e.g. moving from the binomial hypergeometric to the multinomial one). Upon employing the multiset notation, a nice formal symmetry can be recovered between the purely binary and the purely weighted cases

given algorithm in recovering the imposed partition. As an indicator of the goodness of the partition retrieved by each algorithm, we have followed[17] and employed three different indices (see the Methods section for their definition): the normalized mutual information (NMI), the adjusted Rand index (ARI) and the adjusted Wallace index (AWI).

First, let us inspect the performance of modularity, surprise and Infomap to detect cliques arranged in a ring. Specifically, we have considered seven different ring-like configurations, each one linking twenty binary cliques (i.e. K_3, K_4, K_5, K_8, K_{10}, K_{15}, K_{20}). As Fig. 2 reveals, surprise always recovers the planted partition; on the other hand, modularity maximization leads to miss the partitions with K_3 and K_4 and Infomap misses the partition with K_3, a result that may be a consequence of the resolution limit, affecting both the aforementioned algorithms.

Let us now ask us if the presence of weights affects the detection of mesoscale structures. Generally speaking, the answer is yes, as the comparison between the ring of binary cliques and the ring of weighted cliques cases shows. In particular, the result according to which surprise minimization is able to discriminate the inter-linked cliques changes once the weight of the links connecting any two cliques is risen: in fact, this leads the algorithm to reveal as communities two tightly-connected pairs of cliques, now. We explicitly notice that the results shown in Fig. 3 also depend on the relative magnitude of the weights within and between cliques: however, as long as the inter-cliques weight is up to two orders of magnitude larger than the intra-cliques one, it holds true.

In order to expand the set of comparisons, we have focused on two different kinds of well-established benchmarks, i.e. the Lancichinetti–Fortunato–Radicchi (LFR) one and the Aldecoa's relaxed-caveman (RC) one[32] (see the Supplementary Note 6 for their definition and a pictorial example of them).

Results on specific implementations of the LFR benchmark are shown in Fig. 4. Infomap is, generally speaking, a strong performer; as evident upon looking at the first row, however, its performance decreases abruptly as the mixing parameter exceeds a threshold value that depends on the particular setting of the LFR benchmark. Modularity, instead, seems to be more robust (i.e. its performance degrades less rapidly as the parameters μ_t increases) although the resolution limit manifests itself when configurations with small communities are considered. Overall, the performance of surprise seems to constitute a good compromise between the robustness of modularity and the steadily high accuracy of Infomap. We would also like to stress that surprise competes with modularity although it employs much less information than the latter: in fact, while the benchmark employed by modularity coincides with the (sparse version of the) Configuration Model—hence, encodes the information on the entire degree sequence—surprise compares the RGM with the SBM, hence employing the information on the link density, both in a global and in a block-wise fashion. Surprise becomes the best performer when binary, directed configurations are considered—see the second row of Fig. 4: while the performance of modularity starts decreasing as soon as the value of μ_t is risen and $\mathrm{NMI}_{\mathrm{Infomap}}\simeq 0$ when μ_t crosses the value of 0.6, the performance of surprise degrades much more slowly - in fact, for some instances of the LFR benchmark, it achieves a large value of NMI even for values $\mu_t\geq 0.8$.

Let us now comment on the performance of our algorithms when weighted configurations are considered. The results are, again, shown in Fig. 4 (to be noticed that we have kept one of the two parameters fixed and studied the dependence of the NMI on the other: specifically, we have frozen the topological mixing parameter and studied the dependence of the results on μ_w, thus inspecting the performance of our algorithms as the weights are redistributed on a fixed topology). Infomap is, again, a strong performer although its performance keeps decreasing abruptly as μ_w exceeds a threshold value depending on the particular setting of the LFR benchmark; modularity, instead, performs worse than in the binary case although it is still more robust than Infomap. Although degrading less sharply than Infomap, the performance of the purely weighted surprise seems to be the worst, here; on the other hand, the enhanced surprise outperforms the competing algorithms for intermediate values of the topological mixing parameter, irrespectively from the size of the communities: in fact, $\mathrm{NMI}_{\mathrm{surprise}}=1$ even for $\mu_w=0.8$. Similar considerations hold true when weighted, directed configurations are considered (with the only difference that, now, modularity steadily performs worse than the other algorithms, except for the largest values of the mixing parameter). As for the binary cases, surprise competes with modularity although it employs much less information than the latter: in fact, while the benchmark employed by modularity

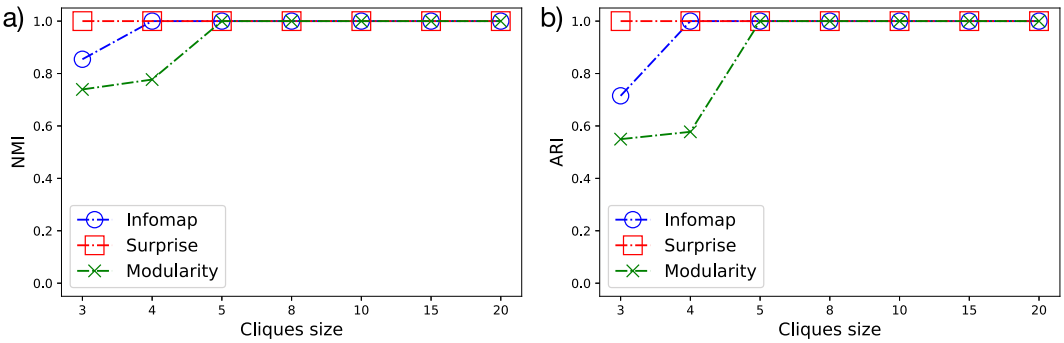

Fig. 2 Performance of community detection algorithms on 'homogeneous' rings of cliques. For each configuration, twenty cliques of the same size (i.e. K_3, K_4, K_5, K_8, K_{10}, K_{15}, K_{20}) have been considered. Panels **a**, **b** respectively show the Normalized Mutual Information (NMI) and the Adjusted Rand Index (ARI) as a function of the number of cliques. While surprise always recovers the planted partition, modularity misses the partitions induced by K_3 and K_4 and Infomap misses the partition induced by K_3—a result that may be a consequence of the resolution limit, known to affect the last two algorithms.

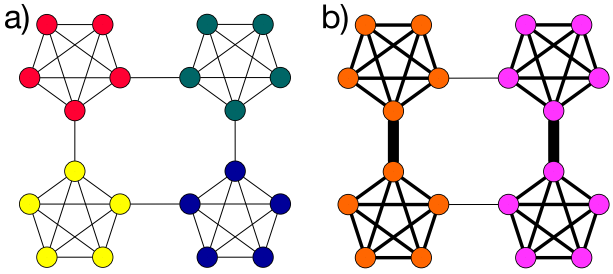

Fig. 3 Comparison between the 'ring of binary cliques' and the 'ring of weighted cliques' cases. Panel **a** depicts the four communities detected by surprise on a ring of binary cliques. As panel **b** shows, the result according to which surprise minimization is able to discriminate the cliques linked in a ring-like fashion changes once the weights come into play: in fact, as the weight of the links connecting any two cliques is risen, the algorithm reveals pairs of tightly-connected cliques as communities.

now coincides with the (sparse version of the) Weighted Configuration Model—hence, encodes the information on the entire strength sequence— surprise compares the WRGM with the WSBM, hence employing the information on the magnitude of the total weight, both in a global and in a block-wise fashion.

Let us now consider the RC benchmark. Results on specific implementations of this are shown in Fig. 5: surprise outperforms both competing algorithms across the entire domain of the degradation parameter p. More specifically, while modularity degrades slowly as the value of p is risen, Infomap degrades abruptly as $p \geq 0.4$. Hence, for small values of such a parameter, Infomap outperforms modularity; on the other hand, for large values of p, modularity outperforms Infomap (although both $\mathrm{NMI}_{\mathrm{modularity}}$ and $\mathrm{ARI}_{\mathrm{modularity}}$ achieve a value which is around 0.6, i.e. already far from the maximum). Interestingly, for small values of p, the performance of Infomap and that of surprise overlap, both achieving NMI and ARI values which are very close to 1: as p crosses the value of 0.4, however, the two trends become increasingly different with Infomap being outperformed by modularity which, in turn, is outperformed by surprise. From a more general perspective, these results confirm what has been already observed elsewhere[14], i.e. that the best-performing algorithms on the LFR benchmarks often perform poorly on the RC benchmarks and vice versa.

Let us now inspect the performance of surprise in recovering binary bimodular structures. To this aim, we have defined a benchmark mimicking the philosophy of the RC one, i.e. progressively degrading an initial, well-defined configuration:

- let us consider N_c core nodes and N_p periphery nodes. The core is completely connected (i.e. the link density of the $N_c \times N_c$ block is 1) and the periphery is empty (i.e. the link density of the $N_p \times N_p$ block is 0). So far, our benchmark is reminiscent of a core-periphery structure à la Borgatti–Everett;
- let us now focus on the topology of the $N_c \times N_p$ bipartite network embodying the connections between the core and the periphery: in particular, let us consider each entry of such an adjacency matrix and pose $a_{cp} = 1$ with probability p_{cp}. Upon doing so, such a subgraph will have a link density amounting precisely at p_{cp};
- let us now 'degrade' such an initial configuration, by progressively filling the periphery and emptying the core. This can be achieved by (1) considering all peripherical node pairs and link them with probability q_p; (2) considering all core node pairs and keep them linked with probability $1 - q_c$ (or, equivalently, disconnect them with probability q_c). Upon doing so, we end up with a core whose link density is precisely $1 - q_c$ and with a periphery whose link density is precisely q_p. Now, varying q_p in the interval $[0, p_{cp}]$ and q_c in the interval $[0, 1 - p_{cp}]$ allows us to span a range of configurations starting with the Borgatti–Everett one and ending with an Erdös-Rényi one.

Specifically, here we have considered $N_c = 100$, $N_p = 300$ and $p_{cp} = 0.5$. The result of our exercise is shown in Fig. 6: as expected, the performance of the surprise worsens as the degradation parameter becomes closer to $p_{cp} = 0.5$; however, both the NMI and the ARI indices steadily remain very close to 1—a result meaning that surprise optimization is not only able to correctly classify true positives (i.e. to keep the nodes originally in the same communities together) but also the other, possible kinds of node pairs.

As with the exercise on community detection, let us now ask us how the presence of weights impacts on the detection of bimodular structures. Let us consider a toy core-periphery network: rising the weight of any two links connecting the core with the periphery allows the two nodes originally part of the periphery to be detected as belonging to the core (see Fig. 7c, d). Analogously, if a bipartite topology is modified by adding weights between some of the nodes belonging to the same layer, $\mathscr{W}_{/\!/}$ will detect a core-periphery structure as significant, the core nodes being the ones linked by the heaviest connections (see Fig. 7a, b).

Testing surprise on real-world networks. When coming to study real systems, particularly insightful examples are provided by

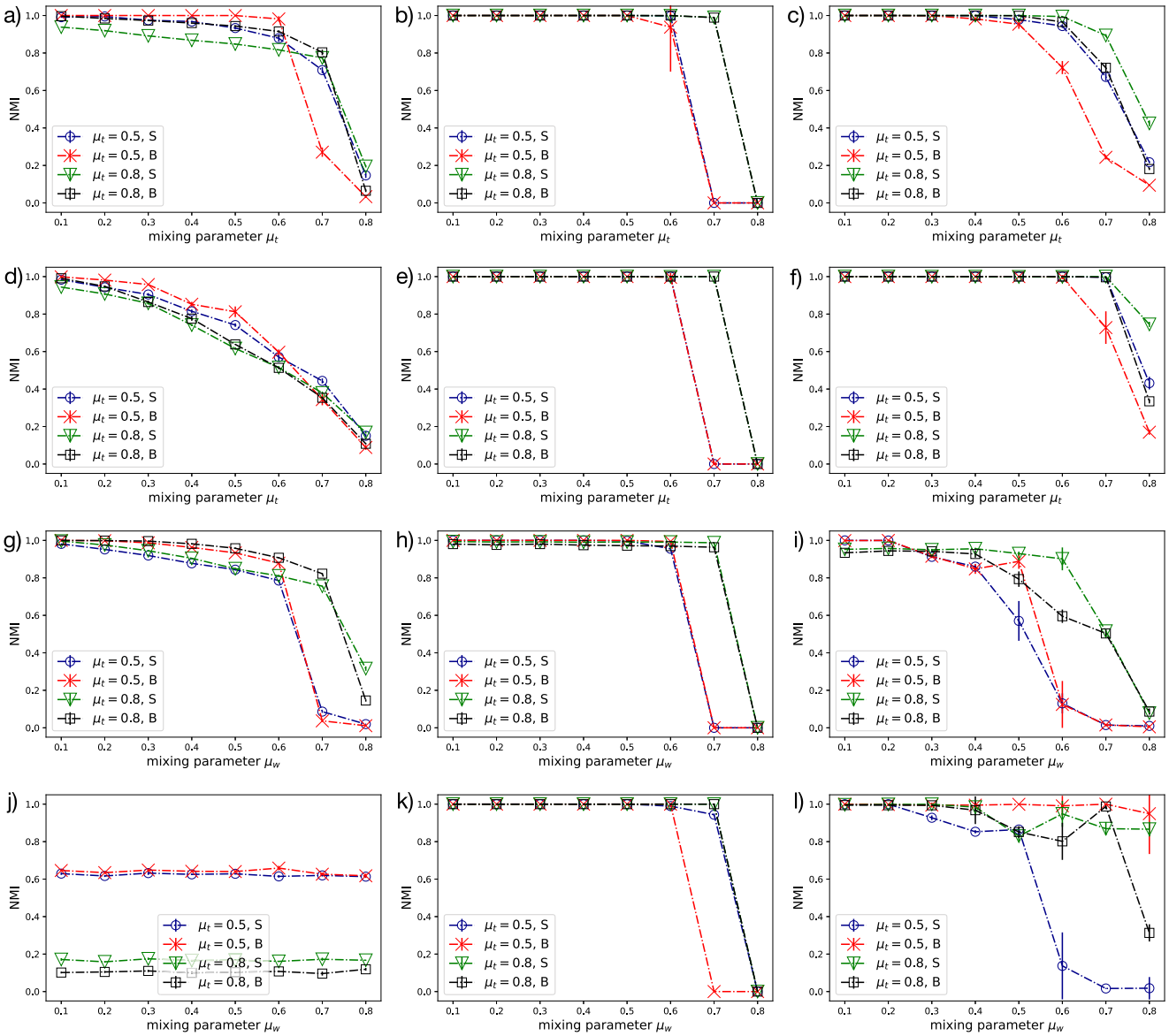

Fig. 4 Comparison of different community detection algorithms on the Lancichinetti–Fortunato–Radicchi (LFR) benchmark. Evolution of the Normalized Mutual Information (NMI) on binary undirected (**a**–**c**), binary directed (**d**–**f**), weighted undirected (**g**–**i**) and weighted directed benchmarks (**j**–**l**). The algorithms considered here are based upon modularity (**a**, **d**, **g** and **j**), Infomap (**b**, **e**, **h** and **k**) and surprise (**c**, **f**, **i** and **l**). In every panel four curves are shown: they correspond to two different network sizes (blue and red: 1000 nodes; green and black: 5000 nodes) and, for a given size, to two different ranges for the size of communities, indicated by the letters S (small, communities have between 10 and 50 nodes) and B (big, communities have between 20 and 100 nodes). In the case of weighted networks, all configurations have the same size (5000 nodes) the difference lying in the value of the binary mixing parameter (blue and red: $u_t = 0.5$; green and black: $u_t = 0.8$). When considering binary undirected networks, we have set $\tau_1 = -2$ and $\tau_2 = -1$ (the average degree is 20 and the maximum degree is 50). When considering binary directed networks, μ_t refers to in-degrees, that are distributed according to a power law while the out-degrees are kept constant for all nodes; the other input parameters, instead, are the same ones used for the undirected configurations. When weighted undirected configurations are considered, an additional mixing parameter is needed, i.e. μ_w, accounting for the percentage of a node strength to be distributed on the links that connect it to the nodes outside its own community; the exponent of the strength distribution has been set to 1.5 for all realizations considered here. When weighted directed configurations are considered, μ_w refers to in-strengths. The trend of the NMI, plotted as a function of the mixing parameters, reveals Infomap to be a strong performer even if its performance decreases abruptly as the mixing parameters exceed a threshold value that depends on the particular setting of the benchmark; the performance of modularity, instead, degrades less rapidly. The performance of surprise seems to constitute a good compromise between the robustness of modularity and the steadily high accuracy of Infomap. The error bars quantify the variability (standard deviation) of the NMI values over the set of one hundred networks generated for each setting of the parameters.

social networks. To this aim, let us consider the one induced by the co-occurrences of characters within the 'Star Wars' saga (i.e. the three trilogies)[33]. As shown in Fig. 8 we have both considered the binary and the weighted version of it. In both cases, two major clusters are visible. For what concerns the binary version of such a network, the optimization of $\mathscr{S}$ reveals the presence of two major clusters: those clusters are induced by the characters of Episodes I–III (e.g. Yoda, Qui-Gon, Obi-Wan, Anakin, Padme, the Emperor, Count Dooku, etc.) and by the characters of Episodes IV-IX (e.g. C-3PO, Leia, Han, Lando, Poe, Finn); a third cluster,

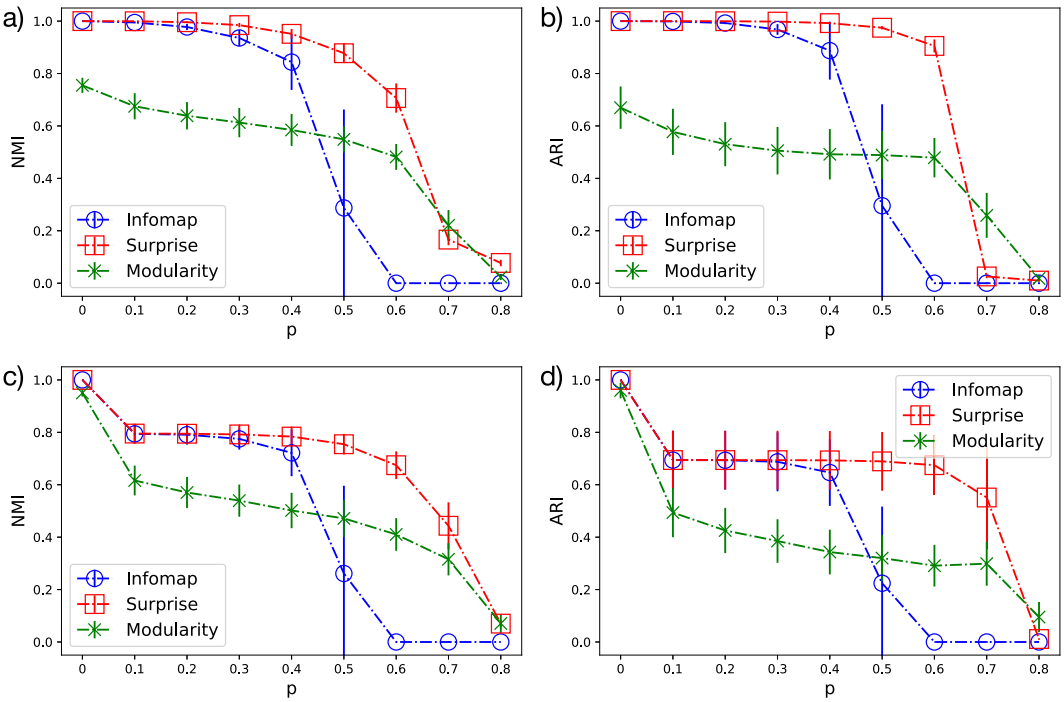

Fig. 5 Comparison of different community detection algorithms on the Relaxed-Caveman (RC) benchmark. Evolution of the Normalized Mutual Information (NMI) and of the Adjusted Rand Index (ARI) on binary undirected (**a**, **b**) and binary directed (**c**, **d**) benchmarks. The chosen configurations consist of 512 nodes, grouped in 16 communities arranged in a ring-like fashion and whose size obeys a power-law with exponent 1.8; the smallest community is composed by 3 nodes. The initial configurations are progressively degraded according to the following mechanism: first, a percentage p of links is randomly selected and removed; afterward, a percentage p of links is randomly selected and rewired. For each value of the degradation parameter, one hundred (undirected and directed) networks are generated: the error bars quantify the variability (standard deviation) of the NMI and the ARI values. The trend of the NMI, plotted as a function of the (only) degradation parameter p that drives the evolution of the initial ring of cliques towards a progressively less-defined clustered configuration, reveals surprise to outperform both modularity and Infomap. From a more general perspective, these results confirm what has been already observed elsewhere, i.e. that the best-performing algorithms on the Lancichinetti–Fortunato–Radicchi (LFR) benchmarks often perform poorly on the Relaxed Caveman (RC) benchmarks and vice versa.

instead, concerns the villains of Episodes VII-IX (i.e. Snoke, Kylo Ren, Phasma, Hux). Interestingly, Rey, BB-8, Maz-Kanata and other characters living on Jakku are clustered together; moreover, the interactions between the characters of Episodes IV-VI and those of Episodes VII-IX causes the former ones and the latter ones to be recovered within the same cluster. This picture is further refined once weights are taken into account: in fact, two of the aforementioned clusters are now merged, giving origin to the cluster of heroes of Episodes IV-IX (e.g. C-3PO, Leia, Han, Lando, Poe, Finn, Rey, Maz-Kanata).

Let us now inspect the effectiveness of our framework in revealing weighted communities by considering the friendship network among the terrorists involved in the train bombing of Madrid in 2004[34] and the one among the residents living in an Australian University Campus[34] (see Fig. 8). As the optimization of $\mathscr{W}$ reveals, while fully connected subsets of nodes are considered as communities in case links have unitary weights, sparser subgraphs can be considered as communities as well whenever their inner connections are heavy enough. On the other hand, both bottom panels of Fig. 8 seem to confirm that one of the main limitations of surprise-like functionals is that of recovering a large number of small cluster of nodes.

Let us now compare the performance of $\mathscr{S}_{//}$ and $\mathscr{W}_{//}$ in order to see if, and how, the presence of weights affects the bimodular mesoscale organization of networks. To this aim, let us focus on the network of co-occurrences of the characters of the novel 'Les Miserables'[34]. As shown in Fig. 9, link weights indeed modify the picture provided by just considering the simple presence of links (see also ref. [12]): the core of the weighted network is, in fact, constituted by the nodes connected by the heavier links, irrespectively from the link density of the former one.

After having applied our framework to the analysis of social networks, let us move to consider financial networks. One of the most popular examples of the kind is provided by the electronic Italian Interbank Money Market (e-MID)[35], depicted in Fig. 9. Notice that, for such a network, the vast majority of core links are also the heavier ones, an evidence confirming a tendency that is ubiquitous in financial and economic systems, i.e. binary and weighted quantities—even at the mesoscale—are closely related.

The surprise-based formalism presented in this paper can be also employed in a hierarchical fashion to highlight either nested communities or nested bimodular structures. To clarify this point, let us consider the World Trade Web (WTW) in the year 2000 as a case study[36]. First, let us run $\mathscr{W}_{//}$ to highlight the core portion of the weighted version of such a network; as Fig. 10 shows, the bipartition distinguishes countries with a large strength from those whose trade volume is low (basically, a bunch of African, Asian and South-American countries). Repeating our analysis within the core portion of the network allows us to discover the presence of a (statistically significant) nested core: in fact, the secound-round optimization reveals that the core-inside-the-core is composed by countries such as Canada, USA, the richest European countries, China and Russia.

Let us now compare it with $\mathscr{E}_{//}$, run in a hierarchical fashion as well. The results of our exercise are shown in Fig. 10. As evident from looking at it, the enhanced surprise is more restrictive than

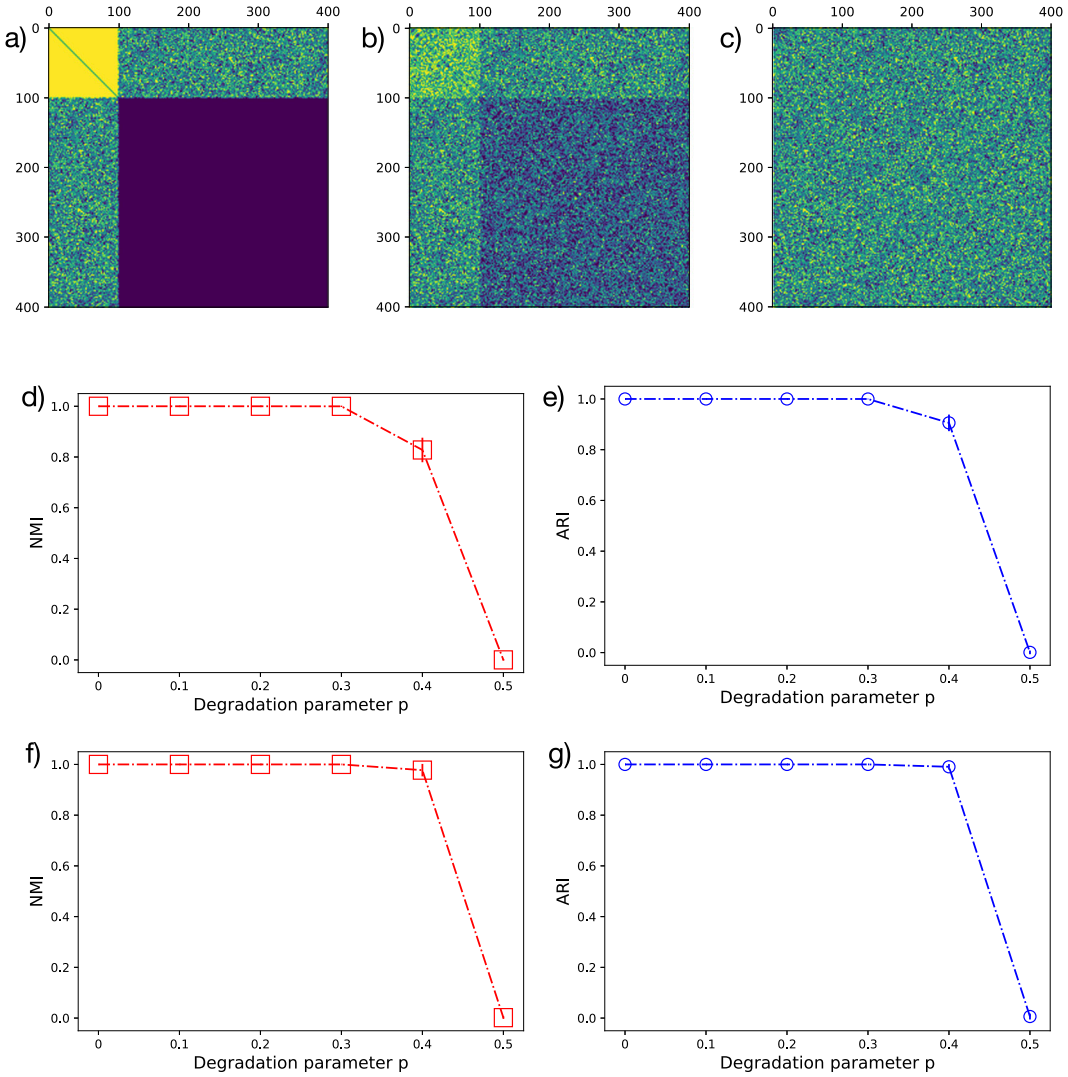

Fig. 6 Benchmark for testing surprise on recovering core-periphery structures. The benchmark defined here progressively 'degrades' the initial configuration represented in panel **a** and defined by (1) a completely connected core, (2) an empty periphery, (3) an intermediate part whose link density amounts at $p_{cp} = 0.5$. Such a configuration is 'degraded' by progressively filling the periphery and emptying the core. This is achieved by (1) considering all peripherical node pairs and link them with probability q_p; (2) considering all core node pairs and keep them linked with probability $1 - q_c$: varying q_p in the interval $[0, p_{cp}]$ and q_c in the interval $[0, 1 - p_{cp}]$ allows us to span a range of configurations starting with the Borgatti-Everett one and ending with an Erdös-Renyi one. Panels **b** and **c** represent two configurations for which $q_c = q_p = 0.3$ and $q_c = q_p = 0.5$. In panels **d** and **e** we show the evolution of the Normalized Mutual information (NMI) and of the Adjusted Rand Index (ARI) as a function of the degradation parameter $q_c = q_p$, for the partitions recovered by surprise on undirected benchmarks; analogously, in panels **f** and **g**, for the directed benchmarks. As expected, the performance of the surprise worsens as $q_c = q_p \simeq p_{cp} = 0.5$; however, both the NMI and the ARI steadily remain very close to 1.

the purely weighted one, as a consequence of constraining the degrees beside the strengths. Hence, while the first run excludes the countries with both a small degree and a small strength, the second run excludes the Russia, a result seemingly indicating that while its strength is large enough to allow it to be a member of the core, its degree is not. In a sense, the optimization of $\mathscr{E}_{//}$ corrects the picture provided by the optimization of $\mathscr{W}_{//}$ as the core becomes less populated by low degree nodes—an effect which is likely to become more evident on systems that are neither financial nor economic in nature.

Let us now run, and compare, modularity, Infomap and surprise on the bunch of real-world networks above. Table 2 sums up the results. A first observation concerns the number of detected communities: while Infomap is the algorithm producing the smallest number of clusters, surprise is the one producing the largest number of clusters—more precisely, surprise outputs more and smaller clusters than the other two methods. As a consequence, our three algorithms produce partitions with an overall small overlap, as indicated by the NMI; the ARI confirms such an observation—although indicating that the pictures provided by Infomap and surprise are (overall) more similar than those provided by modularity and surprise (and, as a consequence, by modularity and Infomap). Interestingly enough, the values of the AWI are quite large—and larger than the corresponding NMI and ARI values: since it just focuses on the percentage of true positives, a good performance under such an index indicates that the two tested algorithms agree on the nodes to be clustered together (although they may not—and, in general, will not—agree on the number of communities). Hence, the discrepancy between the ARI and the AWI may be explained by the presence of statistical noise (i.e. misclassified pairs of nodes, although the word may not be correct as the information about

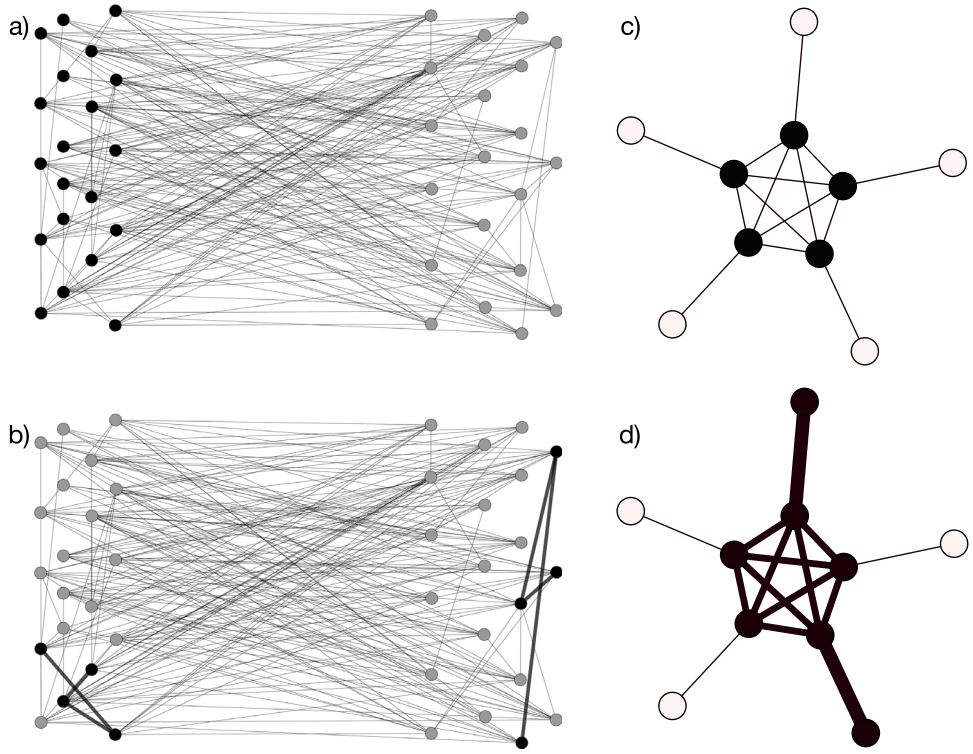

Fig. 7 Impact of weights on the detection of bimodular structures. The presence of weights affects the detection of bimodular mesoscale structures as well. In fact, rising the weight of any two links connecting the core with the periphery of a toy network allows the two nodes originally part of the periphery (**a**) to be detected as belonging to the core (**b**); analogously, if a bipartite binary topology (**c**) is modified by adding weights between some of the nodes belonging to the same layer (**d**), a core-periphery structure will be detected as significant.

the true partition is not available) around the bulk of nodes to be put together.

Let us stress once more that, whenever real-world networks are considered, information about the existence of a true partition is rarely available; for this reason, exercises as the one we have carried out here may be useful to gain insight on the system under study: instead of trusting just one algorithm, combining pairs of them—e.g. by considering as communities the subsets of nodes output by both—may be the right solution to overcome the limitations affecting each single method.

Discussion

The hypergeometric distribution—together with its many variants—has recently revived the interest of researchers who have employed it to define novel network ensembles[37], recipes for projecting bipartite networks[38], etc.

The distributions related to it allow for a wide variety of benchmarks to be defined, each one embodying a different set of constraints. In the present paper we have explored the power of the hypergeometric formalism to carry out the detection of mesoscale structures: it allows proper statistical tests to be definable for revealing the presence of modular, core-periphery and bipartite structures on any kind of network, be it binary or weighted, undirected or directed. According to the classification proposed in the introductory section of the paper, we believe surprise to belong to the second class of algorithms—its asymptotic expression embodying a sort of LRT aimed at choosing between alternative hypotheses (the RGM and the SBM, the WRG and the WSBM, etc.).

More in general, our approach reveals the superiority of the algorithms for mesoscale structures detection belonging to the second and to the third class with respect to those belonging to the first one: still, the two classes of statistically grounded approaches compete on some specific benchmarks, as the comparisons carried out on the LFR and the RC ones clearly show (specifically, methods performing well on LFR benchmarks do not on RC benchmarks and vice versa). This also suggests a strategy to handle mesoscale structures detection on real-world networks: as the information on the presence of a possible, true partition is rarely available, a good strategy may be that of running different algorithms on the same empirical networks, check the consistency of their output and combine them, e.g. by taking the overlap—in a way that is reminiscent of multi-model inference.

Although the surprise-based approach is powerful and versatile, its downside is represented by the specific kinds of tests that are induced by the optimization of surprise-like score functions: comparisons between benchmarks ignoring the local structure of nodes (i.e. their degree) are, in fact, carried out. While this seems to be perfectly reasonable when considering core-periphery structures—see also the contribution[27] whose authors claim that a core-periphery structure is always compatible with a network degree sequence—this is no longer true for the community detection task[39]—indeed, as it has been noticed elsewhere, ignoring degrees may be at the origin of the large number of singletons output by surprise as assigning nodes with few neighbors to larger clusters may be disfavored from a statistical point of view.

The observations above call for the extension of the hypergeometric formalism we have studied here to include more refined benchmarks as the ones constraining the entire degree sequence.

Methods

Community detection in binary networks. In order to gain more insight into the surprise-based formalism, let us derive an asymptotic expression for $\mathcal{S}$[40]. To this

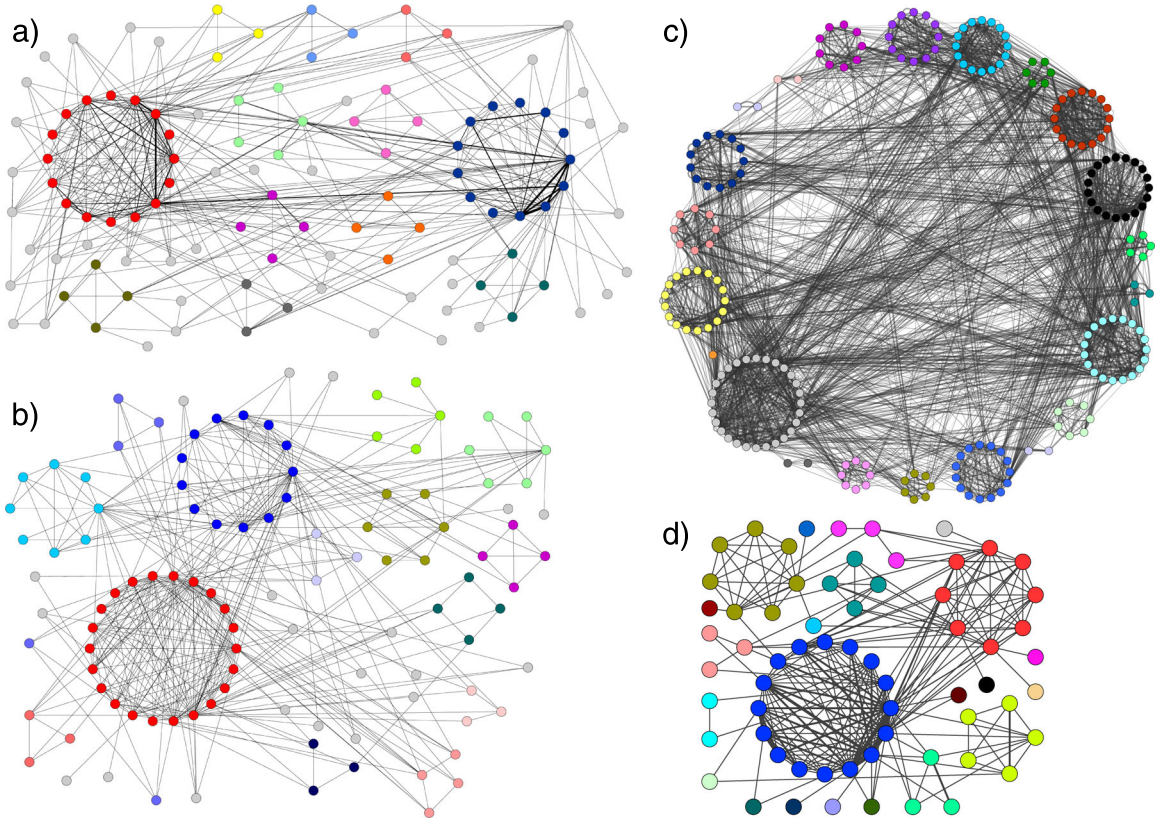

Fig. 8 Application of the surprise-based formalism for the detection of communities on real-world networks. Panels **a** and **b** show, respectively, the result of surprise minimization on the binary and the weighted version of the network of co-occurrences of Star Wars characters. Overall, link weights refine the picture provided by just considering the presence of links: in the binary case, in fact, surprise detects two major clusters, induced by the characters of Episodes I-III and by the characters of Episodes IV-IX; once weights are taken into account, these clusters merge and give origin to the cluster of heroes of Episodes IV-IX. Panel **c** depicts the communities detected by minimizing surprise on the friendship network among the terrorists involved in the train bombing of Madrid in 2004[34]; panel **d** depicts the communities detected by minimizing surprise on the friendship network among the residents living in an Australian University Campus[34].

aim, let us consider that it can be simplified by Stirling-approximating the binomial coefficients that appear within it. By exploiting the recipe $n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$, $\mathscr{S}$ can be rewritten as

$$\mathscr{S} \simeq \sum_{l_\bullet \geq l_\bullet^*} A(l_\bullet)\left[\frac{\mathrm{Ber}(V,L,p)}{\prod_{i=\bullet,\circ}\mathrm{Ber}(V_i,l_i,p_i)}\right] \tag{16}$$

(see the Supplementary Note 1 for more details) where the expression

$$\mathrm{Ber}(x,y,z) = z^y(1-z)^{x-y} \tag{17}$$

defines a Bernoulli probability mass function. The parameters appearing in eq. (16) read $p = \frac{L}{V}$ and $p_i = \frac{l_i}{V_i}$ and the coefficient in front of the sum is

$$A(l_\bullet) = \sqrt{\frac{\sigma^2}{2\pi\prod_{i=\bullet,\circ}\sigma_i^2}} \tag{18}$$

with $\sigma^2 = Vp(1-p)$ and $\sigma_i^2 = V_i p_i(1-p_i)$. Equation (16) makes it explicit that employing $\mathscr{S}$ for binary community detection ultimately amounts at comparing the description of a networked configuration provided by the Random Graph Model (RGM), and encoded into the expression

$$\mathrm{Ber}(V,L,p) = p^L(1-p)^{V-L}, \tag{19}$$

with the description of the same configuration provided by the Stochastic Block Model (SBM)[39], and encoded into the expression

$$\prod_{i=\bullet,\circ}\mathrm{Ber}(V_i,l_i,p_i) = p_\bullet^{l_\bullet}(1-p_\bullet)^{V_\bullet-l_\bullet}\cdot p_\circ^{l_\circ}(1-p_\circ)^{V_\circ-l_\circ} \tag{20}$$

where $p_\bullet = \frac{l_\bullet}{V_\bullet}$ and $p_\circ = \frac{l_\circ}{V_\circ}$. Naturally, the SBM takes as input the two sets indexed by $\bullet$ and $\circ$ and distinguish the connections found within the clusters - contributing to the probability of the whole configuration with the term $\mathrm{Ber}(V_\bullet, l_\bullet, p_\bullet) = p_\bullet^{l_\bullet}(1-p_\bullet)^{V_\bullet-l_\bullet}$ - from the ones found between the clusters - contributing to the probability of the whole configuration with the term $\mathrm{Ber}(V_\circ, l_\circ, p_\circ) = p_\circ^{l_\circ}(1-p_\circ)^{V_\circ-l_\circ}$.

Notice also that the asymptotic expression of the surprise guarantees that the parameters of the null models defining it are tuned according to the maximum-of-the-likelihood principle. To see this explicitly, let us consider $\mathrm{Ber}(x,y,z)$ whose log-likelihood reads

$$\mathscr{L} = y\ln z + (x-y)\ln(1-z); \tag{21}$$

upon maximizing it with respect to z, one finds $z = \frac{y}{x}$.

The way $\mathscr{S}$ works, i.e. by comparing two different null models, is reminiscent of more traditional likelihood ratio tests, where a null hypothesis H_0 (in our case, a given partition is compatible with the RGM) is tested against an alternative hypothesis H_1 (in our case, the given partition is compatible with the SBM): as the asymptotic expression of the surprise clarifies, minimizing it amounts at finding the partition least likely to occour under the RGM than under the SBM.

Bimodular structures detection in binary networks. Analogously to the univariate case, the asymptotic expression for $\mathscr{S}_{//}$ can be derived upon Stirling-approximating the binomial coefficients appearing within it. This time, the recipe $n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ leads to

$$\mathscr{S}_{//} \simeq \sum_{l_\bullet \geq l_\bullet^*}\sum_{l_\circ \geq l_\circ^*} B(l_\bullet, l_\circ)\left[\frac{\mathrm{Ber}(V,L,p)}{\prod_{i=\bullet,\circ,\top}\mathrm{Ber}(V_i,l_i,p_i)}\right] \tag{22}$$

where, as before, $\mathrm{Ber}(x,y,z) = z^y(1-z)^{x-y}$ defines a Bernoulli probability mass function (see the Supplementary Note 1 for more details) and the parameters read $p = \frac{L}{V}$ and $p_i = \frac{l_i}{V_i}$; the numerical coefficient appearing in front of the whole expression, now, reads

$$B(l_\bullet, l_\circ) = \frac{1}{2\pi}\sqrt{\frac{\sigma^2}{\prod_{i=\bullet,\circ,\top}\sigma_i^2}} \tag{23}$$

with $\sigma^2 = Vp(1-p)$ and $\sigma_i^2 = V_i p_i(1-p_i)$. The quantity $\mathscr{S}_{//}$ compares the description of a networked configuration provided by the RGM, and encoded into the expression $\mathrm{Ber}(V,L,p) = p^L(1-p)^{V-L}$, with the description of the same

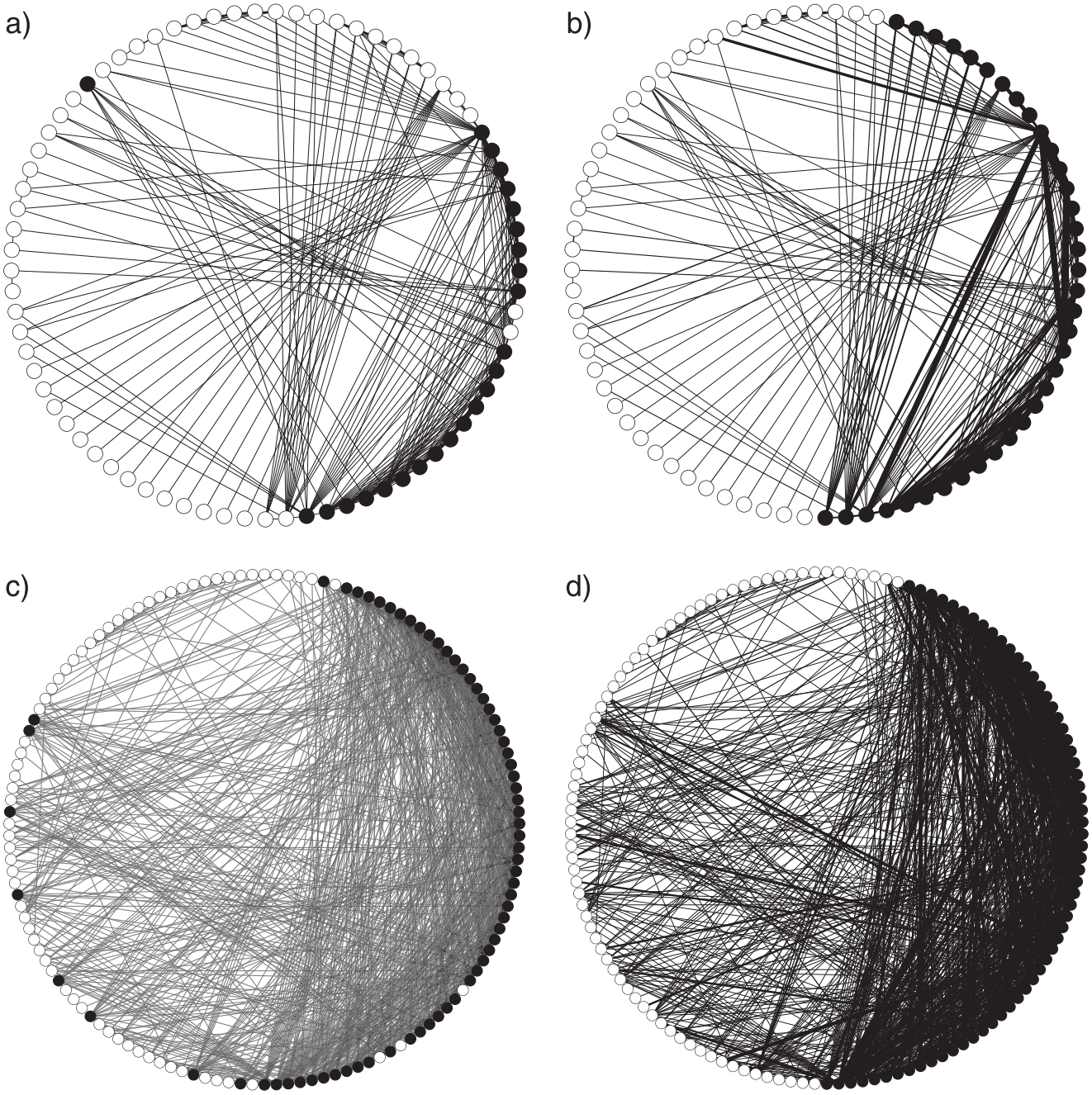

Fig. 9 Bimodular structures detection on real-world social and financial networks. In panels **a** and **b** we show the results of the application of our framework for the detection of bimodular structures on the network of co-occurrences of the characters of 'Les Miserables'[34] (binary case in panel **a**; weighted case in panel **b**). Link weights refine the picture provided by just considering the presence of links: the core (in black) is, in fact, constituted by the nodes connected by the 'heavier' links, irrespectively from the link density of the former one. Similarly, panels **c** and **d** depict the results of the application of our framework for the detection of bimodular structures on the electronic Italian Interbank Money Market (e-MID)[35], during the maintenance period approximately corresponding to May 2009 (binary case in panel **c**; weighted case in panel **d**). As the picture shows, the (purely binary) core links are also the 'heavier' ones—an evidence confirming an ubiquitous tendency in economic and financial systems, i.e. binary and weighted quantities are closely related.

configuration provided by the SBM (now, defined by three - instead of two - different blocks), represented by the denominator of the expression defined in eq. (22), i.e.

$$\prod_{i=\bullet,\circ,\top}\mathrm{Ber}(V_i,l_i,p_i)=p_\bullet^{l_\bullet}(1-p_\bullet)^{V_\bullet-l_\bullet}\cdot p_\circ^{l_\circ}(1-p_\circ)^{V_\circ-l_\circ}\cdot p_\top^{l_\top}(1-p_\top)^{V_\top-l_\top}. \tag{24}$$

Community detection in weighted networks. The asymptotic expression for $\mathscr{W}$ can be deduced by following the same reasoning that has allowed us to derive the asymptotic expression for $\mathscr{S}$. Stirling-approximating the binomial coefficients entering into the definition of $\mathscr{W}$ leads to the writing

$$\mathscr{W}\simeq\sum_{w_\bullet\geq w_\bullet^*}C(w_\bullet)\left[\frac{\mathrm{Geo}(V,W,q)}{\prod_{i=\bullet,\circ}\mathrm{Geo}(V_i,w_i,q_i)}\right] \tag{25}$$

(see the Supplementary Note 2 for more details) where the expression

$$\mathrm{Geo}(x,y,z)=z^y(1-z)^x \tag{26}$$

defines a geometric probability mass function and the parameters appearing in eq. (25) read $q=\frac{W}{V+W-1}$ and $q_i=\frac{w_i}{V_i+w_i-1}$. In the weighted case, the Bernoulli

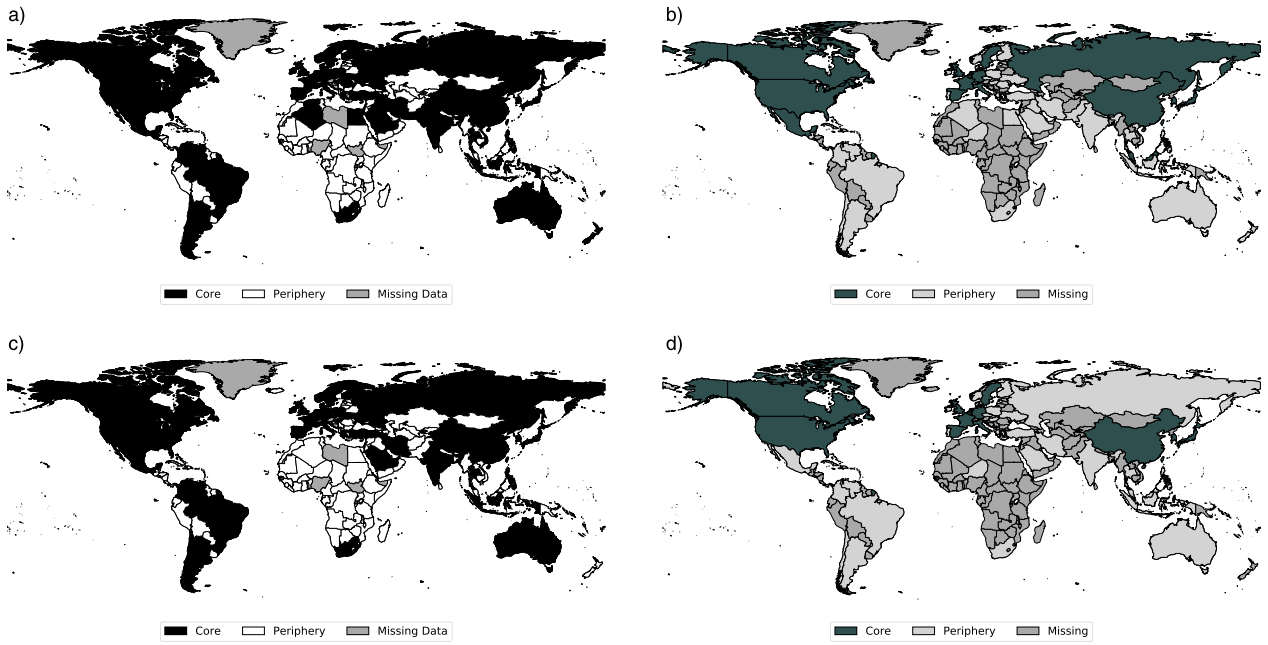

Fig. 10 Detection of nested bimodular structures. Result of the application of our framework for the detection of weighted bimodular structures on the world trade web (WTW) in the year 2000[36]. **a**, **b** upon running $\mathscr{W}^{\frown}_{//}$ in a hierarchical fashion, a core-within-the-core is detected, revealing that the richest world countries (i.e. Canada, USA, the richest European countries, China and Russia—right panel, in dark green) constitute an even more tightly linked cluster of nations among the ones with the largest trading activity (**a**). **c**, **d** $\mathscr{E}_{//}$ penalizes the countries with both a small degree and a small strength; in fact, the second run excludes the Russia, a result seemingly indicating that while its strength is 'large enough' to be a member of the core, its degree is not.

Table 2 Table comparing the performance of modularity, Infomap and surprise on the real-world networks considered in the present paper.

	# Communities								
	Modularity	**Infomap**	**Surprise**	**NMI_{SM}**	**NMI_{SI}**	**ARI_{SM}**	**ARI_{SI}**	**AWI_{SM}**	**AWI_{SI}**
Binary networks									
Madrid bombing terrorists	6	5	26	0.50	0.44	0.33	0.27	0.83	0.73
'Star Wars' characters	9	5	35	0.37	0.53	0.15	0.28	0.39	0.94
Australian	13	6	30	0.47	0.72	0.20	0.46	0.76	0.93
University Campus									
'Les Miserables' characters	8	6	33	0.56	0.56	0.46	0.50	0.92	0.83
Weighted networks									
Madrid bombing terrorists	8	5	22	0.50	0.67	0.37	0.62	0.65	0.94
'Star Wars' characters	10	6	52	0.36	0.43	0.18	0.24	0.79	0.79
Australian	14	6	25	0.47	0.74	0.22	0.46	0.72	0.91
University Campus									
'Les Miserables' characters	10	6	32	0.51	0.61	0.37	0.49	0.72	0.74

Interestingly enough, the values of the Adjusted Wallace Index (AWI) are quite large: since it just focuses on the percentage of true positives, a good performance under such an index indicates that the tested algorithms agree on the nodes to be clustered together: hence, the discrepancy between the adjusted rand index (ARI) and the AWI may be explained by the presence of statistical noise (i.e. pairs of nodes on whose classification the algorithms disagree) around the bulk of nodes to be put together - on which the algorithms agree. The values of the Normalized Mutual Information (NMI) fall in between the ARI and the AWI but still confirm that the partitions detected by surprise are more similar to the ones revealed by Infomap than to those detected by modularity.
While Infomap is the algorithm producing the smallest number of clusters, surprise is the one producing the largest number of clusters; as a consequence, the three partitions output by our algorithms have an overall small overlap.

probability mass function appearing in the asymptotic expression of $\mathscr{S}$ is replaced by a geometric probability mass function: this implies that (asymptotically) the comparison is, now, carried out between the description of a networked configuration provided by the Weighted Random Graph Model (WRGM), and encoded into the expression

$$\mathrm{Geo}(V, W, q) = q^{W}(1-q)^{V}, \tag{27}$$

with the description of the same configuration provided by the Weighted Stochastic Block Model (WSBM) and encoded into the expression

$$\prod_{i=\bullet,\circ} \mathrm{Geo}(V_i, w_i, q_i) = q_{\bullet}^{w_{\bullet}}(1-q_{\bullet})^{V_{\bullet}} \cdot q_{\circ}^{w_{\circ}}(1-q_{\circ})^{V_{\circ}} \tag{28}$$

where $q_{\bullet} = \frac{w_{\bullet}}{V_{\bullet}+w_{\bullet}}$ and $q_{\circ} = \frac{w_{\circ}}{V_{\circ}+w_{\circ}}$. As for the binary case, the asymptotic expression of the weighted surprise clarifies that minimizing it amounts at finding the partition least likely to occour under the WRGM than under the WSBM.

As in the binary case, the parameters characterizing the geometric probability mass functions defining the asymptotic weighted surprise can be estimated via the maximum-of-the-likelihood principle, according to which the log-likelihood of the expression $\mathrm{Geo}(x, y, z)$ reads

$$\mathscr{L} = y \ln z + x \ln(1-z); \tag{29}$$

upon maximizing it with respect to z, one finds $z = \frac{y}{x+y}$: hence, one can pose $q \simeq \frac{W}{V+W}$ and $q_i \simeq \frac{w_i}{V_i+w_i}$.

The numerical coefficient appearing in front of the whole expression, instead, reads

$$C(w_\bullet) = \sqrt{\frac{\mu}{2\pi\prod_{i=\bullet,\circ}\mu_i}} \tag{30}$$

with $\mu \simeq Vq$ and $\mu_i \simeq V_i q_i$.

Bimodular structures detection in weighted networks. Let us now derive the asymptotic expression for $\mathscr{W}_{//}$. As usual, let us Stirling-approximate the binomial coefficients entering into the definition of $\mathscr{W}_{//}$; such a simplification leads to the expression

$$\mathscr{W}_{//} \simeq \sum_{w_\bullet \geq w_\bullet^*}\sum_{w_\circ \geq w_\circ^*} D(w_\bullet, w_\circ)\left[\frac{\mathrm{Geo}(V, W, q)}{\prod_{i=\bullet,\circ,\top}\mathrm{Geo}(V_i, w_i, q_i)}\right] \tag{31}$$

where, as before, $\mathrm{Geo}(x, y, z) = z^y(1-z)^x$ defines a geometric probability mass function (see the Supplementary Note 2 for more details) and the parameters read $q = \frac{W}{V+W-1}$ and $q_i = \frac{w_i}{V_i+w_i-1}$ but can be approximated, according to the maximum-of-the-likelihood principle, as $q \simeq \frac{W}{V+W}$ and $q_i \simeq \frac{w_i}{V_i+w_i}$; the numerical coefficient multiplying the whole expression reads

$$D(w_\bullet, w_\circ) = \frac{1}{2\pi}\sqrt{\frac{\mu}{\prod_{i=\bullet,\circ,\top}\mu_i}} \tag{32}$$

with $\mu \simeq Vq$ and $\mu_i \simeq V_i q_i$. The analysis of $\mathscr{W}_{//}$ in the asymptotic regime reveals that it compares the description of a networked configuration provided by the WRGM, and encoded into the expression $\mathrm{Geo}(V, W, q) = q^W(1-q)^V$, with the description of the same configuration provided by the WSBM (now, defined by three - instead of two - different blocks), represented by the denominator of the expression defined in eq. (31), i.e.

$$\begin{aligned}\prod_{i=\bullet,\circ,\top}\mathrm{Geo}(V_i, w_i, q_i) = {} & q_\bullet^{w_\bullet}(1-q_\bullet)^{V_\bullet}\cdot q_\circ^{w_\circ}(1-q_\circ)^{V_\circ} \\ & \cdot q_\top^{w_\top}(1-q_\top)^{V_\top}.\end{aligned} \tag{33}$$

Enhanced community detection. Analogously to the other functionals, the asymptotic expression of $\mathscr{E}$ can be derived by Stirling-approximating the binomial coefficients entering into its definition as well:

$$\mathscr{E} \simeq \sum_{l_\bullet \geq l_\bullet^*}\sum_{w_\bullet \geq w_\bullet^*} E(l_\bullet, w_\bullet)\left[\frac{\mathrm{BF}(V, L, W, p, r)}{\prod_{i=\bullet,\circ}\mathrm{BF}(V_i, l_i, w_i, p_i, r_i)}\right]; \tag{34}$$

in the formula above (see the Supplementary Note 3 for more details), the expression

$$\begin{aligned}\mathrm{BF}(x, y, u, z, t) &= z^y(1-z)^{x-y}\cdot t^{u-y}(1-t)^y \\ &= \mathrm{Ber}(x, y, z)\cdot\mathrm{Geo}(y, u-y, t)\end{aligned} \tag{35}$$

defines a Bose-Fermi probability mass function[41]. While a Bernoulli probability mass function characterizes the asymptotic behavior of the purely binary surprise and a geometric probability mass function characterizes the asymptotic behavior of the purely weighted surprise, the enhanced surprise is asymptotically characterized by a distribution whose functional form is halfway between the two previous ones: while its Bernoulli-like portion controls for the presence of the links, its conditional geometric-like portion controls for the magnitude of the remaining weights. To stress its mixed character, such a distribution has been named 'Bose-Fermi' since it can be retrieved within the Exponential Random Graphs framework as a consequence of Shannon entropy maximization, constrained to simultaneously reproduce both the total number of links and the total weight of a network[41,42].

The parameters appearing in eq. (34) read $p = \frac{L}{V}$, $p_i = \frac{l_i}{V_i}$ and $r = \frac{W-L}{W-1}$, $r_i = \frac{w_i-l_i}{w_i-1}$; while the first class of parameters can be tuned according to the maximum-of-the-likelihood principle, the ones belonging to the second class can be approximated according to the same recipe. To see this explicitly, let us consider $\mathrm{BF}(x, y, u, z, t)$, whose log-likelihood reads

$$\mathscr{L} = y\ln z + (x-y)\ln(1-z) + (u-y)\ln t + y\ln(1-t); \tag{36}$$

maximizing it with respect to z leads to $z = \frac{y}{x}$ while maximizing it with respect to t leads to $t = \frac{u-y}{u}$: hence, one can pose $r \simeq \frac{W-L}{W}$, $r_i \simeq \frac{w_i-l_i}{w_i}$.

The numerical coefficient multiplying the whole expression is defined as

$$E(l_\bullet, w_\bullet) = \frac{1}{2\pi}\sqrt{\frac{\sigma^2\mu}{\prod_{i=\bullet,\circ}\sigma_i^2\mu_i}} \tag{37}$$

where $\sigma^2 = Vp(1-p)$, $\sigma_i^2 = V_i p_i(1-p_i)$, $\mu \simeq Lr$ and $\mu_i \simeq l_i r_i$.

Similarly to the other cases, the asymptotic expression of the enhanced surprise compares the description of a networked configuration provided by the Enhanced Random Graph Model (ERGM), and encoded into the expression $\mathrm{BF}(V, L, W, p, r) = p^L(1-p)^{V-L}\cdot r^{W-L}(1-r)^L$, with the description of the same configuration provided by its block-wise counterpart, i.e. the Enhanced Stochastic Block Model (ESBM) and encoded into the expression $\prod_{i=\bullet,\circ}\mathrm{BF}(V_i, l_i, w_i, p_i, r_i)$.

Enhanced bimodular structures detection. The enhanced bimodular surprise admits an asymptotic expression as well, i.e.

$$\begin{aligned}\mathscr{E}_{//} \simeq {} & \sum_{l_\bullet \geq l_\bullet^*}\sum_{l_\circ \geq l_\circ^*}\sum_{w_\bullet \geq w_\bullet^*}\sum_{w_\circ \geq w_\circ^*} F(l_\bullet, l_\circ, w_\bullet, w_\circ) \\ & \cdot\left[\frac{\mathrm{BF}(V, L, W, p, r)}{\prod_{i=\bullet,\circ,\top}\mathrm{BF}(V_i, l_i, w_i, p_i, r_i)}\right]\end{aligned} \tag{38}$$

that can be recovered by Stirling-approximating the binomial coefficients entering into the definition of $\mathscr{E}_{//}$. While the expression BF denotes a Bose-Fermi probability mass function[41], the parameters appearing in eq. (38) read $p = \frac{L}{V}$, $p_i = \frac{l_i}{V_i}$ and $r = \frac{W-L}{W-1}$, $r_i = \frac{w_i-l_i}{w_i-1}$; as for $\mathscr{E}$, the maximum-of-the-likelihood principle determines (approximates) the first (second) class of parameters.

The numerical coefficient multiplying the whole expression is defined as

$$F(l_\bullet, l_\circ, w_\bullet, w_\circ) = \frac{1}{(2\pi)^2}\sqrt{\frac{\sigma^2\mu}{\prod_{i=\bullet,\circ,\top}\sigma_i^2\mu_i}} \tag{39}$$

where $\sigma^2 = Vp(1-p)$, $\sigma_i^2 = V_i p_i(1-p_i)$, $\mu \simeq Lr$ and $\mu_i \simeq l_i r_i$. As observed for the other functionals, the asymptotic expression of the enhanced bimodular suprise compares the description of a networked configuration provided by the ERGM, and encoded into the expression $\mathrm{BF}(V, L, W, p, r) = p^L(1-p)^{V-L}\cdot r^{W-L}(1-r)^L$, with the description of the same configuration provided by its block-wise counterpart, i.e. the ESBM (now, defined by three - instead of two - different blocks), represented by the denominator of the expression defined in eq. (38), i.e. $\prod_{i=\bullet,\circ,\top}\mathrm{BF}(V_i, l_i, w_i, p_i, r_i)$.

Quantifying the goodness of a partition. As indicators of the goodness of the partition retrieved by each algorithm we have employed three different indices. The first one is the normalized mutual information (NMI), defined as

$$\bar{I}(X, Y) = \frac{2I(X, Y)}{H(X)+H(Y)} \tag{40}$$

and comparing the partitions X and Y; $H(X) = -\sum_x f_x \ln f_x$ and $f_x = \frac{n_x}{n}$ is the fraction of nodes assigned to the cluster labeled with x; analogously, for the partition Y. The term $I(X, Y) = \sum_x\sum_y f_{xy}\ln\left(\frac{f_{xy}}{f_x f_y}\right)$ is the proper mutual information and $f_{xy} = \frac{n_{xy}}{n}$ is the fraction of nodes assigned to cluster x in partition X and to cluster y in partition Y. Naturally, $\bar{I}(X, Y)$ equals 1 if the partitions are identical and 0 if the partitions are independent.

The second index we have considered is the adjusted Rand index (ARI) that reads

$$\mathrm{ARI} = \frac{TP + TN - \langle TP + TN\rangle}{TP + FP + TN + FN - \langle TP + TN\rangle} \tag{41}$$

and represents a sort of accuracy corrected by a term that quantifies the agreement between the reference partition and a random partition - the term random referring to the Permutation Model. The number of true positives (TP) is the number of pairs of nodes being in the same community both in the considered and in the reference partition; the number of false positives (FP) is the number of pairs of nodes being in the same community in the considered partition but in different communities in the reference partition; the number of true negatives (TN) is the number of pairs of nodes being in the same community neither in the considered nor in the reference partition; the number of false negatives (FN) is the number of pairs of nodes being in the same community in the reference partition but not in the considered partition. Naturally, the closer the ARI to 0, the more random the provided partition.

The third index we have considered is the adjusted Wallace index (AWI), defined as

$$\mathrm{AWI} = \frac{TP - \langle TP\rangle}{TP + FP - \langle TP\rangle} \tag{42}$$

and representing a sort of corrected positive predicted value. Again, the closer the AWI to 0, the more random the provided partition.

Data availability

The data that support the findings of this study are freely downloadable at the following URLs:

- https://github.com/evelinag/StarWars-social-network;
- http://konect.cc/networks;
- http://ksglediteh.com/exptradegdp.html.

Code availability

As an additional result, we release a comprehensive package, coded in Python, that implements all the aforementioned surprise-like variants (see the Supplementary Note 7 for more details about the algorithms). Its name is 'SurpriseMeMore' - the name recalls the package released in 2014[43] - and is freely available on GitHub at the following link https://github.com/EmilianoMarchese/SurpriseMeMore.

Received: 13 September 2021; Accepted: 14 April 2022;
Published online: 30 May 2022

References

1. Fortunato, S. & Hric, D. Community detection in networks: a user guide. *Phys. Rep.* **659**, 1–44 (2016).
2. Khan, B. S. & Niazi, M. A. Network community detection: a review and visual survey. Preprint at https://arxiv.org/abs/1708.00977 (2017).
3. Borgatti, S. P. & Everett, M. G. Models of core/periphery structures. *Soc. Netw.* **21**, 375–395 (2000).
4. Craig, B. & Von Peter, G. Interbank tiering and money center banks. *J. Financ. Intermed.* **23**, 322–347 (2014).
5. Van Lelyveld, I. et al. Finding the core: Network structure in interbank markets. *J. Bank. Financ.* **49**, 27–40 (2014).
6. Luu, D. T., Napoletano, M., Barucca, P. & Battiston, S. Collateral unchained: rehypothecation networks, concentration and systemic effects. *J. Financ. Stab.* **52**, 100811 (2021).
7. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).
8. Cimini, G., Mastrandrea, R. & Squartini, T. *Reconstructing Networks* (Cambridge University Press, 2021).
9. Fronczak, A. Exponential random graph models (2012).
10. Peixoto, T. P. Descriptive vs. inferential community detection: pitfalls, myths and half-truths. Preprint at https://arxiv.org/abs/2112.00183 (2021).
11. Karrer, B. & Newman, M. E. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 (2011).
12. de Jeude, Jv. L., Caldarelli, G. & Squartini, T. Detecting core-periphery structures by surprise. *EPL* **125**, 68001 (2019).
13. Peixoto, T. P. Bayesian stochastic blockmodeling. in *Advances in Network Clustering and Blockmodeling* (eds Doreian, P., Batagelj, V. & Ferligoj, A.) 289–332 (Wiley, New York, 2019).
14. Aldecoa, R. & Marin, I. Surprise maximization reveals the community structure of complex networks. *Sci. Rep.* **3**, 1–9 (2013).
15. Nicolini, C. & Bifone, A. Modular structure of brain functional networks: breaking the resolution limit by surprise. *Sci. Rep.* **6**, 1–13 (2016).
16. Tumminello, M., Micciche, S., Lillo, F., Piilo, J. & Mantegna, R. N. Statistically validated networks in bipartite complex systems. *PLoS ONE* **6**, e17994 (2011).
17. Bongiorno, C., London, A., Miccichè, S. & Mantegna, R. N. Core of communities in bipartite networks. *Phys. Rev. E* **96**, 022321 (2017).
18. Musciotto, F., Battiston, F. & Mantegna, R. N. Detecting informative higher-order interactions in statistically validated hypergraphs. Preprint at https://arxiv.org/abs/2103.16484 (2021).
19. Miccichè, S. & Mantegna, R. N. A primer on statistically validated networks. *Comput. Soc. Sci. Complex Syst.* **203**, 91 (2019).
20. Jiang, Y., Jia, C. & Yu, J. An efficient community detection algorithm using greedy surprise maximization. *J. Phys. A Math. Theor.* **47**, 165101 (2014).
21. Del Ser, J., Lobo, J. L., Villar-Rodriguez, E., Bilbao, M. N. & Perfecto, C. Community detection in graphs based on surprise maximization using firefly heuristics. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, 2233–2239 (IEEE, 2016).
22. Tang, Y.-N. et al. An effective algorithm for optimizing surprise in network community detection. *IEEE Access* **7**, 148814–148827 (2019).
23. Gamermann, D. & Pellizaro, J. A. An algorithm for network community structure determination by surprise. *Physica A: Statistical Mechanics and its Applications* **595**, 127063 (2022).
24. Kojaku, S. & Masuda, N. A generalised significance test for individual communities in networks. *Scientific Reports* **8**, 1–10 (2018).
25. Zhang, X., Martin, T. & Newman, M. E. Identification of core-periphery structure in networks. *Phys. Rev. E* **91**, 032803 (2015).
26. Barucca, P. & Lillo, F. Disentangling bipartite and core-periphery structure in financial networks. *Chaos Solitons Fractals* **88**, 244–253 (2016).
27. Kojaku, S. & Masuda, N. Core-periphery structure requires something else in the network. *New J. Phys.* **20**, 043012 (2018).
28. Holme, P., Liljeros, F., Edling, C. R. & Kim, B. J. Network bipartivity. *Phys. Rev. E* **68**, 056107 (2003).
29. Estrada, E. & Rodriguez-Velazquez, J. A. Spectral measures of bipartivity in complex networks. *Phys. Rev. E* **72**, 046105 (2005).
30. Prokhorov, Y. V., & Feller, W. An introduction to probability theory and its applications. *Teoriya Veroyatnostei i ee Primeneniya* **10**, 204–206 (1965).
31. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl Acad. Sci. USA* **105**, 1118–1123 (2008).
32. Lancichinetti, A. & Fortunato, S. Community detection algorithms: a comparative analysis. *Phys. Rev. E* **80**, 056117 (2009).
33. Star wars characters network. https://github.com/evelinag/StarWars-social-network.
34. The konect project. http://konect.cc/.
35. De Masi, G., Iori, G. & Caldarelli, G. Fitness model for the italian interbank money market. *Phys. Rev. E* **74**, 066112 (2006).
36. Gleditsch, K. S. Expanded trade and gdp data. *J. Confl. Resolut.* **46**, 712–724 (2002).
37. Casiraghi, G. & Nanumyan, V. Configuration models as an urn problem. *Sci. Rep.* **11**, 1–10 (2021).
38. Cimini, G., Carra, A., Didomenicantonio, L. & Zaccaria, A. Meta-validation of bipartite network projections. *Commun. Phys.* **5**, 76 (2022).
39. Karrer, B. & Newman, M. E. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 (2011).
40. Traag, V. A., Aldecoa, R. & Delvenne, J.-C. Detecting communities using asymptotical surprise. *Phys. Rev. E* **92**, 022816 (2015).
41. Garlaschelli, D. & Loffredo, M. I. Generalized bose-fermi statistics and structural correlations in weighted networks. *Phys. Rev. Lett.* **102**, 038701 (2009).
42. Mastrandrea, R., Squartini, T., Fagiolo, G. & Garlaschelli, D. Enhanced reconstruction of weighted networks from strengths and degrees. *New J. Phys.* **16**, 043022 (2014).
43. Aldecoa, R. & Marin, I. Surpriseme: an integrated tool for network community structure characterization using surprise maximization. *Bioinformatics* **30**, 1041–1042 (2014).

Acknowledgements

The authors acknowledge support from the EU project SoBigData-PlusPlus (grant no. 871042).

Author contributions

E.M. and T.S. developed the methods. E.M. performed the analyses. E.M., G.C., and T.S. wrote, reviewed, and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-022-00890-7.

Correspondence and requests for materials should be addressed to Emiliano Marchese.

Peer review information *Communications Physics* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

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.