Entropy-based randomization of rating networks

Carolina Becatti,[1,*] Guido Caldarelli,[1,2,3] and Fabio Saracco[1]
[1]*IMT School for Advanced Studies, Piazza S.Francesco 19, 55100 Lucca, Italy*
[2]*Istituto dei Sistemi Complessi (ISC)-CNR UoS Università "Sapienza", Piazzale Aldo Moro 5, 00185 Roma, Italy*
[3]*ECLT San Marco 2940, 30124 Venezia, Italy*

(Received 16 April 2018; revised manuscript received 30 November 2018; published 7 February 2019)

In recent years, due to the great diffusion of e-commerce, online rating platforms quickly became a common tool for purchase recommendations. However, instruments for their analysis did not evolve at the same speed. Indeed, interesting information about users' habits and tastes can be recovered just considering the bipartite network of users and products, in which links represent products' purchases and have different weights due to the score assigned to the item in users' reviews. With respect to other weighted bipartite networks, in these systems we observe a maximum possible weight per link, that limits the variability of the outcomes. In the present article we propose an entropy-based randomization method for this type of networks (i.e., bipartite rating networks) by extending the configuration model framework: the randomized network satisfies the constraints of the degree per rating, i.e., the number of given ratings received by the specified product or assigned by the single user. We first show that such a null model is able to reproduce several nontrivial features of the real network better than other null models. Then, using our model as benchmark, we project the information contained in the real system on one of the layers: To provide an interpretation of the projection obtained, we run the Louvain community detection on the obtained network and discuss the observed division in clusters. We are able to detect groups of music albums due to the consumers' taste or communities of movies due to their audience. Finally, we show that our method is also able to handle the special case of categorical bipartite networks: we consider the bipartite categorical network of scientific journals recognized for the scientific qualification in economics and statistics. In the end, from the outcome of our method, the probability that each user appreciate every product can be easily recovered. Therefore, this information may be employed in future applications to implement a more detailed recommendation system that also takes into account information regarding the topology of the observed network.

DOI: 10.1103/PhysRevE.99.022306

I. INTRODUCTION

Network theory [1,2] proved successful [3] in the description and modeling of a wide variety of systems, ranging from the obvious cases of the internet [4,5], the world wide web [6], and social networks [7]. In these settings it formed the evidence on which computational social science is based [8], to cell properties in biology [9], and fMRI imaging in brain analysis [10,11], contributing to the new field of network medicine [12,13] and to banks in financial systems [14,15]. Networks come in various shapes, from the simplest case of similar vertices connected by binary edges, to weighted and/or directed networks, to multigraphs where more than one edge can connect two vertices, to bipartite graphs where two distinct sets of vertices are present. Simple examples of the latter case are bipartite graphs in which a connection is drawn if an individual, on one set, performs or not a given task, on the other set. A lot of work has been developed so far to analyze this kind of data, a great part of it being focused on different methods to identify the structure of the network (see, for example, Ref. [16], where the authors discuss the drawbacks of finding communities in bipartite networks and then propose a new solution based on bipartite stochastic block models to address this topic).

This work deals with the specific case of bipartite *rating* networks, where the two sets of nodes are individuals and items' purchases while the edges represent reviews of products given by consumers and are weighted by the numerical score received, as for example, in the well-known Amazon review system. These kinds of graphs have been mostly studied from a machine learning and computer science perspective, to train models able to recommend items to people, based on their taste and preferences. Different methodologies are employed for this purpose; see Ref. [17] for an thorough review of the literature on the topic and the recorded progress.

In this paper we focus on a different approach, providing an analytical tool that could reveal useful in the development of a recommendation system based on network topology. We follow the stream of literature introduced by Refs. [18–20], which defines an appropriate method to construct benchmark models for the observed networks. Therefore, our attention can be focused on the assessment of the significance of several topological quantities measured on the graph. More specifically, we compare the real system with its randomization, represented by an ensemble of graphs with the same number of nodes and all possibile link configurations. To have an effective filter, we constrain the average over the ensemble of

*carolina.becatti@imtlucca.it

 ©2019 American Physical Society

some topological quantities—in this case the degree sequence per rating—and check if other nontrivial measures of the actual network are reproduced. If not, then there is a signal of a behavior that it is not captured by the constraints only.

The method works as follows: It first prescribes to define an appropriate ensemble of graphs, with constant number of nodes; second, it defines a probability distribution over the ensemble through a constrained entropy maximization procedure; then, the maximization of the related likelihood function provides the probability that any possible pair of nodes in the network of interest is connected. The constraints introduced in the first maximization procedure are the topological quantities of the real network, i.e., for binary and undirected networks the degree of each node is used as a constraint. Once the theoretical framework is complete, we can state if the real values of some topological quantities substantially deviate from the theoretical distribution, by comparing the actual observations with the expectations of the null model.

Rating networks may be interpreted as classical weighted networks, whose edges are weighted by a finite set of discrete scores. In this context, appropriate constraints are represented by the specification of nodes' strengths only (weighted configuration model, in Ref. [20]). Because of the extremely poor predictive power of vertices' strengths, an enriched version of the previous model has also been introduced (enhanced configuration model) in Ref. [21]; this method adds the topology as additional information. However, the presence (in our framework) of a finite number of discrete weights complicates the problem formulation and increases the required computational effort. For these reasons, a preliminar "binarization" procedure is often employed (it is the approach of Ref. [22], but it is also common in recommendation systems, like in Ref. [23]), by thresholding the edges' weights. In this way, the resulting network is binary and can be easily randomized with the Bipartite Configuration Model in Ref. [24]. In this paper we propose an alternative approach, constraining not only on the presence of positive reviews, but on the exact ratings. Due to its application, we indicate it in the following as bipartite score configuration model (BiSCM). The peculiarity of our approach is that we avoid the scores-related problems by specifying a *multidegree* for each node in the network, i.e., by *specifying the entire distribution of scores received by a node*. We will show that adding more constraints allows us to define a more restrictive null model, thus to reproduce with higher accuracy the features of the original network.

Let us highlight that our approach is general enough to permit to randomize bipartite signed networks (as a subset of rating networks) and bipartite categorical networks, the latter being a subject to which, to the best of our knowledge, there is a substantial scarcity of theoretical tools for their analysis. For instance, while there are different proposals for measuring the similarity among items in categorical datasets [25], nontrivial null-models and benchmarks are practically absent. The present methodology tries to fill this gap.

The rest of this paper is structured as follows. In the Sec. II we explain the ensemble construction procedure and show how the model can be employed in two different kinds of analysis: evaluation of the significance of some topological quantities and projection of the bipartite network on one of its layers. In Sec. III we briefly review the datasets used to test the methods. The main results regarding the motifs analysis are reported in Sec. IV, where we describe all the performed analyses, while we characterize the communities found in the projected and validated networks in Sec. V. Finally, we discuss possible future developments of the method in Sec. VI.

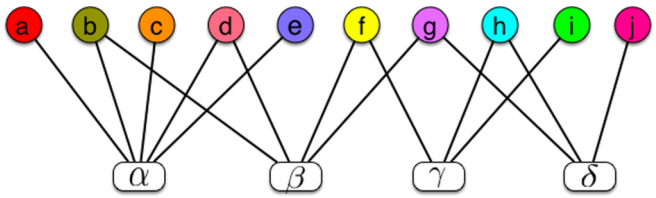

FIG. 1. A simple bipartite graph. In the following, Latin letters will indicate goods, while Greek letters will denote users.

II. METHODS

In this section, we briefly introduce the used notation and explain the necessary steps to construct the null model ensemble. A bipartite network is a network that can be partitioned in two sets of nodes, such that only edges between nodes belonging to different sets can be observed (see Fig. 1). This kind of structure naturally arises whenever considering collaboration networks (i.e., actors in the first set, movies on the second set), export of products, consumers, goods, etc. To distinguish between the two sets, the index running on one set L is typically indicated by Latin letters, while the index running on the second set Γ is indicated by Greek letters. The number of nodes belonging to the two sets L and Γ will be denoted with the symbols N_L and N_Γ, respectively. A bipartite *rating* network with $N = N_L + N_\Gamma$ vertices and E edges can be entirely specified by its $N_L \times N_\Gamma$ adjacency matrix $\mathbf{M}$ with entries $m_{i,\alpha} = \beta$ whenever product i has been reviewed and assigned score β by user α and $m_{i,\alpha} = 0$, otherwise. In what follows, we only deal with the case in which users are required to assign discrete numerical scores and the number of possible scores is known, denoted from now on as $\beta_{\max}$. Therefore, $\beta \in \{1, \ldots, \beta_{\max}\}$. All members of the benchmark ensemble will have a constant number of vertices per layer, respectively, equal to N_L and N_Γ. For the sake of simplicity, a binary representation of the adjacency matrix entries will be considered, defining $m_{i,\alpha,\beta} = \delta(m_{i,\alpha}, \beta)$ for all β, where δ is the Kronecker δ function. By doing so, the variable $m_{i,\alpha,\beta}$ will be equal to 1 if node α has reviewed node i with the numerical score β and $m_{i,\alpha,\beta} = 0$ otherwise. We use the notation

$$k_{i,\beta}(\mathbf{M}) = \sum_{\alpha} m_{i,\alpha,\beta} \quad i = 1, \ldots, N_L, \tag{1}$$

$$k_{\alpha,\beta}(\mathbf{M}) = \sum_{i} m_{i,\alpha,\beta} \quad \alpha = 1, \ldots, N_\Gamma \tag{2}$$

to indicate the number of reviews with score β, respectively, received by a generic product i in Eq. (1) and assigned by a generic user α in Eq. (2). The specification of Eqs. (1) and (2) for all scores β defines the distribution of scores received by each node and constitute the fundamental constraints of our problem.

At this point we look for the probability distribution maximizing the (Shannon's) entropy,

$$S = -\sum_{\mathbf{M}} P(\mathbf{M}) \ln P(\mathbf{M}), \tag{3}$$

under the constraints $\langle k_{i,\beta}\rangle = k_{i,\beta}$ and $\langle k_{\alpha,\beta}\rangle = k_{\alpha,\beta}$ for all i, α and for all scores β. In other words, we consider the probability distribution over the ensemble such that the expected degree of each node, for every possible rating, equals on average its observed value, while keeping all the rest maximally random. The solution to this bipartite maximization problem gives the following probability distribution over the ensemble,

$$P(\mathbf{M}|\vec{x},\vec{y}) = \prod_{i,\alpha} q_{i,\alpha}(m_{i,\alpha,\beta}|\vec{x},\vec{y}), \tag{4}$$

where $\vec{x}$ is a $N_L\beta_{\max}$ dimensioned vector of Lagrangian multipliers that controls the expected degrees for each possible rating for the set of products, while $\vec{y}$ is the analogous $N_\Gamma\beta_{\max}$ dimensioned vector of Lagrangian multipliers for the users. The quantity

$$q_{i,\alpha}(m_{i,\alpha,\beta}|\vec{x},\vec{y}) = \frac{\prod_\beta (x_{i,\beta}\, y_{\alpha,\beta})^{m_{i,\alpha,\beta}}}{1+\sum_\beta x_{i,\beta}\, y_{\alpha,\beta}} \tag{5}$$

for all scores determines the probability to observe one of the entries between nodes i and α (refer to the Appendix for further details). Notice that each node has been assigned a vectorial Lagrangian multiplier ($\vec{x}_i$ if it belongs to the layer L, $\vec{y}_\alpha$ if it belongs to the layer Γ) of dimension $\beta_{\max}$. Thus, the probability to observe a positive outcome, i.e., a link with rating β, can be expressed as

$$p_{i,\alpha,\beta} = \frac{x_{i,\beta}\, y_{\alpha,\beta}}{1+\sum_\beta x_{i,\beta}\, y_{\alpha,\beta}} \tag{6}$$

for all i, α, and β. Therefore, the outcome of our method allows to easily recover the probability that each user assigns a given score to all items, for all observed rating levels in the network.

To determine the numerical values for our Lagrangian multipliers, let us consider a specific real-world rating network $\mathbf{M}^*$, for which the degree sequence $\{k_{i,\beta}(\mathbf{M}^*), k_{\alpha,\beta}(\mathbf{M}^*)\}$ is known for all i, α and for each rating level β. The log-likelihood defined by Eq. (4) is given by

$$\mathcal{L}(\vec{x},\vec{y}\,|\mathbf{M}^*) = \sum_{i,\beta} k_{i,\beta}(\mathbf{M}^*)\ln x_{i,\beta} + \sum_{\alpha,\beta} k_{\alpha,\beta}(\mathbf{M}^*)\ln y_{\alpha,\beta} - \sum_{i,\alpha}\ln\left(1+\sum_\beta x_{i,\beta}\, y_{\alpha,\beta}\right). \tag{7}$$

Then, the maximization procedure consists in finding the specific parameter values $(\vec{x}^*,\vec{y}^*)$ that maximize the probability to observe the network of interest $\mathbf{M}^*$. Thus, the benchmark model for the real-world network $\mathbf{M}^*$ is completely specified and it is possible to compare its observed topological properties with the same quantities averaged over the ensemble of graphs.

Let us conclude this section with some remarks: In the whole manuscript we employed the exact result of our procedure, thus the average over the ensemble truly reproduces the score degree sequence observed in the real network. Nevertheless, the null model's calibration (i.e., the determination of a numerical value for the Lagrangian multipliers $\vec{x}$ and $\vec{y}$) may easily become costly, since the number of unknowns of the problem grows linearly with the number of nodes in the network N and observed scores $\beta_{\max}$ [26]. For this reason, for extremely large and sparse systems, a possible approximation is provided in the Appendix. As in the Chung-Lu model [27], we relax the constraints required by considering expected values and we approximate the Lagrangian multipliers to be proportional to the degree of the node for the given score. We will show that for high degrees, this approximation systematically overestimates the exact probability, but there is a quite good agreement for lower degrees.

It is worth noticing that no assumption has been made so far concerning the nature of the different entries of the adjacency matrix, but for the fact that they are mutually exclusive. Therefore, our method is completely general and can be employed in many different applications. For instance, this framework can be intended as describing a multiedge network in which $\beta_{\max}$ is the maximum number of edges allowed between any pair of nodes. Other possible extensions can be toward the cases of signed networks or categorical networks. In the former case, the different values of scores would be $\beta \in \{+1,-1\}$, indicating, respectively, presence of a positive and negative link. In the latter case instead, each score β is simply assigned one of the possible realisations of a categorical variable: we will present an application to this kind of data in one of the following sections.

A. Higher-order topological benchmark

Whenever dealing with numerical scores (lets us exclude the case of categorical scores for the moment; we will discuss this situation later on), we are able to distinguish "positive" from "negative" reviews. More specifically, since in all cases $\beta_{\max}$ is known, it is possible to fix a threshold score to $\beta_{\text{th}} < \beta_{\max}$ and interpret each review as "positive" whenever the reviewed product has received a score greater or equal than the threshold, meaning that it has been appreciated by the user. Clearly, the "negative" reviews are analogously defined as those that have been assigned a score smaller than the threshold, meaning that the user who assigned the review was not satisfied by the purchase. Therefore, at the end of the described randomization procedure, we are able to define a signed version of the original adjacency matrix, indicated as $\overline{\mathbf{M}}$. This new matrix has entries $\overline{m}_{i,\alpha} = +1$ or $\overline{m}_{i,\alpha} = -1$, whenever a positive (i.e., $\beta_{\text{th}} \leqslant \beta \leqslant \beta_{\max}$) or, respectively, negative (i.e., $1 \leqslant \beta < \beta_{\text{th}}$) review was registered in the observed data $\mathbf{M}^*$. In what follows, the aforementioned positive or negative entries will be indicated using the shortcuts $m^+_{i,\alpha}$ or $m^-_{i,\alpha}$, respectively, denoting the presence of a positive or negative review, i.e., $m^+_{i,\alpha} = \delta(\overline{m}_{i,\alpha},+1)$ or $m^-_{i,\alpha} = \delta(\overline{m}_{i,\alpha},-1)$, where δ is again the Kronecker δ function. Moreover, we denote the quantities $k^+_i = \sum_\alpha m^+_{i,\alpha}$ and $k^-_i = \sum_\alpha m^-_{i,\alpha}$, respectively, *positive degree* and *negative degree*, to indicate the number of edges with positive or negative sign incident to node i. The previous quantities are equivalently defined for node α. Finally, the entries of the related probability matrices

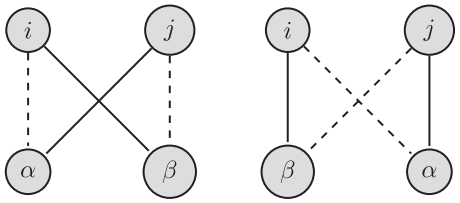

FIG. 2. Checkerboardlike motifs.

have been indicated as $\langle m^+_{i,\alpha}\rangle$ and $\langle m^-_{i,\alpha}\rangle$ with

$$\langle m^+_{i,\alpha}\rangle = p^+_{i,\alpha} = \sum_{\beta \geqslant \beta_{\rm th}} p_{i,\alpha,\beta},$$
$$\langle m^-_{i,\alpha}\rangle = p^-_{i,\alpha} = \sum_{\beta < \beta_{\rm th}} p_{i,\alpha,\beta}$$

for all i, α and β. The previous terms, respectively, represent the probability to observe a positive or negative review between a pair of nodes (i, α) and can be recovered from the quantities $p_{i,\alpha,\beta}$ defined in Eq. (6) and representing the probability that user α reviews product i with a numerical score equal to β.

On the datasets, we first analyze the correlation between neighbor nodes' degrees, introducing a signed version of the classical average nearest-neighbor degree (ANND). We separately analyze all possible combinations of positive and negative neighbors and positive and negative degrees, as follows:

$$k^{pp}_i(\overline{\mathbf{M}}) = \frac{\sum_\alpha m^+_{i,\alpha}k^+_\alpha}{k^+_i}, \quad k^{pn}_i(\overline{\mathbf{M}}) = \frac{\sum_\alpha m^+_{i,\alpha}k^-_\alpha}{k^+_i},$$
$$k^{np}_i(\overline{\mathbf{M}}) = \frac{\sum_\alpha m^-_{i,\alpha}k^+_\alpha}{k^-_i}, \quad k^{nn}_i(\overline{\mathbf{M}}) = \frac{\sum_\alpha m^-_{i,\alpha}k^-_\alpha}{k^-_i}. \tag{8}$$

In the previous equations, the first apex letter is referred to the sign of the edges incident to i, while the second one indicates the sign of i's neighbor degree. In other words, the terms k^{pp}_i or k^{pn}_i in Eq. (8), respectively, identify the average positive or negative degree of node i's positive neighbors, i.e., the positive or negative degree of the other nodes that have assigned a positive review to i. Clearly, replacing positive with negative sign, we obtain the analogous interpretation for k^{np}_i and k^{nn}_i.

Then, we compute the number of signed checkerboard-like motifs c_i [28] for each node, as follows:

$$c_i(\overline{\mathbf{M}}) = \sum_{\alpha,\beta}\sum_j m^+_{i,\beta}m^+_{j,\alpha}m^-_{i,\alpha}m^-_{j,\beta}. \tag{9}$$

The number of checkerboardlike motifs node i is involved represents the number of times we observe a pair of products (i, j) that receives conflicting reviews from a pair of users (α, β). Their graphical representation is provided in Fig. 2, where continuous edges represent positive reviews while dashed represent negative ones. In this case, user α has appreciated product j and instead disliked i, while for β we observe the exact opposite behavior. Clearly, both Eqs. (8) and (9) can be analogously defined for column nodes α.

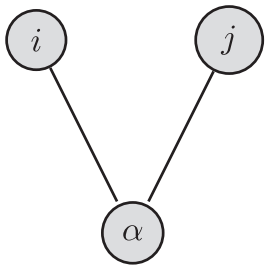

FIG. 3. A V-motif.

B. Monopartite projections

The traditional way to analyze collaboration systems [29] (e.g., actors in the movie system), is to project the information contained in a bipartite network on one of the layers by considering the statistical significance of their common connections. From various attempts on boards [30] to more recent approaches [31], several works [22,32–34] applied a similar idea, making use of the bipartite configuration model. Summarizing, once the probability for the single bipartite link is calculated, it is possible to compute the probability that a pair of nodes shares a link with an item on the opposite layer. Such a pattern can be represented by a V-motif (see Fig. 3): If the probabilities per link are independent, then the probability of observing the single V-motif of Fig. 3 is simply given by the product $P(V^{i,j}_\alpha = 1) = p_{i,\alpha}p_{j,\alpha}$. Thus, the number of common neighbors between i and j is Poisson-binomially distributed [35,36], i.e., the distribution of N_Γ independent Bernoulli events, each with different (in general) probabilities. Comparing the observation on the real network with their theoretical Poisson binomial distribution, it is possible to calculate a p value for each pair of nodes on the same layer. After a multiple hypothesis testing procedure it is possible to state which connections of the monopartite projection are statistically relevant, i.e., which are the nodes that share more connections than expected by the null model.

In the present paper, as an application of the BiSCM null-model as a benchmark, we extend such a procedure to rating networks, considering pairs of items that receive both positive reviews from the same customer. Indeed, the inclusion of extra constraints in the randomization phase, allows to construct a more restrictive null model. This fact is confirmed by the results of our validation procedure, since all validated networks are characterized by a very low connectance and the communities observed have a very precise interpretation (as shown in the following sections). If we set the threshold for positive reviews at $\beta_{\rm th}$, then the probability to simultaneously observe positive reviews for the items (i, j) from the same customer α is given by

$$P\left(V^{i,j}_\alpha = 1\right) = p^+_{i,\alpha}\, p^+_{j,\alpha}.$$

Now, the algorithm follows exactly the same steps of the original procedure: From the probability that both items receive a positive review from the same customer we calculate the distribution of common good reviews. Although several methods are available in the literature, we employ the false discovery rate procedure [37] to validate the previously calculated p values, since it permits to have a stricter control on the false

TABLE I. Data description

	N	N_L	N_Γ	E	ρ	$+$	$-$
(ML) MovieLens	2 588	1 645	943	100 000	6.36×10^{-2}	0.83	0.17
(MI) Musical Instruments	2 329	900	1 429	10 261	7.98×10^{-3}	0.95	0.05
(SM) Smartphones	16 172	2 256	13 916	15 817	5.04×10^{-4}	0.73	0.27
(DM) Digital Music	7 000	2 500	4 500	36 774	3.27×10^{-3}	0.91	0.09
(AN) Anvur dataset	1 396	15	1 381	14 903	7.19×10^{-1}		

positives. The result of the algorithm is a threshold p value, used to validate all the hypotheses at a time. For a quick recap, the method's recipe is the following:

(1) sort the vector of p values to be tested in ascending order;

(2) select the largest integer $\widehat{i}$ satisfying

$$p\,\text{value}_{\widehat{i}} \leqslant \frac{\widehat{i}t}{N_L N_\Gamma}, \tag{10}$$

where t is the chosen significance level;

(3) consider $p\text{value}_{\widehat{i}}$ as the threshold value.

Then, all hypotheses which p value is smaller than or equal to the threshold must be rejected, while we are not able to reject all hypotheses whose p value is greater than $p\text{value}_{\widehat{i}}$. In all the applications we will consider $t = 0.05$ as the significance value.

III. DATASETS

The following datasets have been employed to test the presented randomization procedure.

(1) *MovieLens 100k*: Bipartite network that collects 100 000 movies' ratings. The website's users are characterized by some individual features, such as age, job, sex, state, and zip code. For the set of movies we have information on the release year, title, and genre. Each user can review a movie with a numerical score $\beta \in \{1, 2, 3, 4, 5\}$, according to her level of appreciation. We consider as positive all reviews assigned a score greater than or equal to $\beta_{\text{th}} = 3$. The data has been downloaded from the repository [38] while any additional information is provided in Ref. [39].

(2) *Amazon*: We collected three datasets involving different categories of products. From Ref. [38] we downloaded the musical instruments and digital music datasets that, respectively, collect purchases of musical instruments and CDs or vinyls (the latter had to be further sampled due to its high dimensions). The data about smartphones and related products has instead been downloaded from Ref. [40]. For all of them, the possible numerical ratings for each purchase are $\beta \in \{1, 2, 3, 4, 5\}$. As in the previous case, we consider as positive all reviews that receive a score greater than or equal to $\beta_{\text{th}} = 3$.

(3) *Anvur* is the acronym for "Associazione Nazionale di Valutazione del sistema Universitario e della Ricerca," an Italian agency that evaluates the quality of the Universities and research systems and determines which journals are considered top journals (and so the most influent) in any scientific area. For this application we have downloaded the free dataset [41] regarding the journals' classification in the scientific area of economics and statistics (Area 13). We then constructed the bipartite network of journals and scientific areas, in which a link exists if the journal is considered a top journal related to the scientific area of interest. The two different scores are indicated in the original table as $\beta \in \{\text{green}, \text{red}\}$. The first one indicates that the journal is currently considered a "class A" journal, the second one instead indicates that the journal was considered "class A until December 2017"; all later publications will not receive the same classification.

A more detailed description of the datasets is provided in Table I, where ρ denotes the connectance of the networks, while the symbols $+$ and $-$ indicate the percentage of positive and negative edges in each dataset (when available).

IV. RESULTS

For each dataset we employ the procedure described in Sec. II to construct the benchmark model. So we obtain a set of $\beta_{\max}$ probability matrices (one for every rating level), collecting the probability to observe the different ratings for each pair of nodes in the network.

Once the Lagrangian multipliers' values $(\vec{x}^*, \vec{y}^*)$ are obtained from the maximization of the likelihood function in Eq. (7), the expected quantities $\langle k_i^{pp}\rangle$, $\langle k_i^{pn}\rangle$, $\langle k_i^{np}\rangle$, $\langle k_i^{nn}\rangle$, and $\langle c_i\rangle$ across the ensemble can be analytically computed starting from their formulation in Eqs. (8) and (9). However, for the following analysis, instead of the exact computation of these topological quantities, we have considered the approximation obtained simply replacing the terms $m_{i,\alpha}^+$ and $m_{i,\alpha}^-$ in Eqs. (8) and (9) with their expectation values $\langle m_{i,\alpha}^+\rangle$ and $\langle m_{i,\alpha}^-\rangle$. The same procedure has been followed by Refs. [20,24]. Moreover, again following the instructions in [20], we have identified a confident region of two standard deviations around the average values. The comparison of observed and expected quantities indicates whether these higher order network properties can be directly explained by lower order topological structures, i.e., the constraints imposed on nodes' degrees, or require further investigation since they represent an indication of some correlation patterns in the observed network.

Figures 4 and 5 show the results of this comparative analysis on the MovieLens dataset. All the expectations have been computed averaging the values of k_i^{pp}, k_i^{pn}, k_i^{np}, k_i^{nn}, and c_i over the number of nodes having the same degree in the network. To provide a reliable analysis of the proposed model, we compare the BiSCM performance with some alternative ones: weighted configuration model (WCM), partial bipartite score configuration model (PCM), and Erdös-Rényi random graph (RG). The main difference among their construction relies in the specification of the imposed constraints. However, we refer to the Appendix for a full description of these alternative models. Red lines show the average connectivity

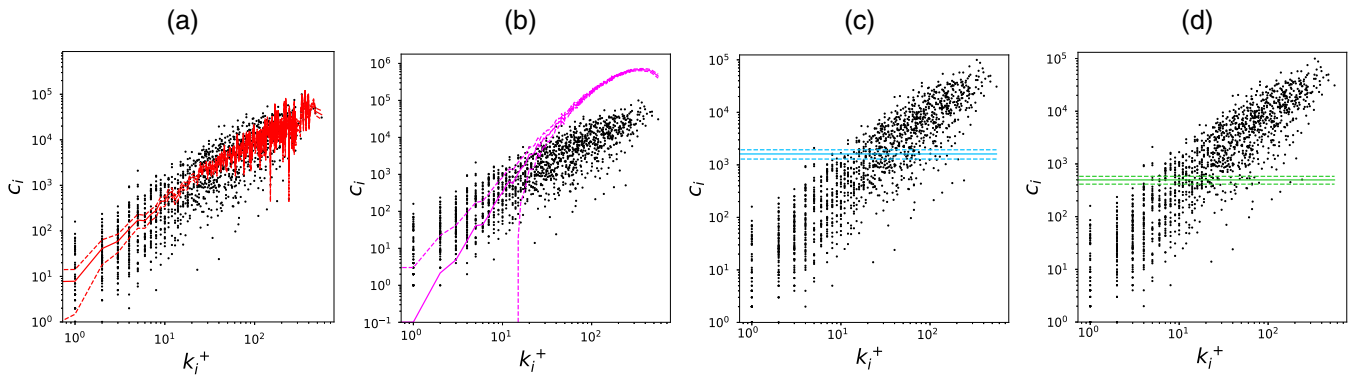

FIG. 4. Application of the method to the ML network. The panels report c_i versus k_i^+. The red line (a) shows the expectation values computed with our method. Instead magenta, blue and green lines (b, c, d) are the expectations under WCM, PCM and Erdös-Rényi RG respectively. The area of ± 2 standard deviations around the average value has been reproduced with dashed linestyle.

and number of checkerboards in the ensemble estimated by BiSCM. Magenta, blue, and green lines represent instead the same quantities estimated by WCM, PCM, and RG, respectively. In most cases, the overall data trend is well captured by our ensemble. For the case of k_i^{pp}, some observations remain outside the two standard deviations range, suggesting the possibility of extra correlations that cannot be directly traced back to the degree sequence alone, despite the full specification of scores' distribution. The analysis of the other null models would lead to completely unreliable conclusions, since in most of the cases, the induced ANND baseline is not able to capture the data overall trend. This is true especially for the cases of WCM and RG. The best performing alternative null model is the partial BiSCM, which relies on the same type of constraints upon which the BiSCM is based, but imposed on a reduced set of nodes.

Due to the evident difference on the percentage of positive and negative observed reviews, a different type of analysis has been performed on the remaining datasets, taking into consideration positive reviews only: The BiSCM outperforms even in this case, as shown in Fig. 10 of the Appendix.

V. MONOPARTITE COMMUNITIES

For three of our datasets we have reported the results of the projection analysis. The ML and DM networks have been binarized and then projected on the products layer, i.e., the movies and musical products layers, respectively. The AN dataset has instead been projected onto both layers, considering separately the two sets of edges representing "green" and "red." All projection algorithms require to connect a pair of nodes in the monopartite network, whenever they share at least a common neighbor in the bipartite graph. However, our projected edges have been further validated using the procedure presented in Ref. [22] and explained in Sec. IIB. To discuss the structure of the validated networks, we apply the Louvain modularity-based community detection algorithm [42]. However, this method is known to be order-dependent [43]. Therefore, to overcome this limitation, we consider the outcome of several runs, obtained by reshuffling the initial order of the nodes, following the recipe of Ref. [22]. Interestingly enough, this approach greatly increases the performance of the original algorithm and performs similarly to the Combo algorithm [44]. In the following we will consider the partition in communities detected by the better performing algorithm between the reshuffled Louvain and the Combo.

A. MovieLens

The partition in communities (obtained with the reshuffled Louvain algorithm, with a modularity of 0.451, $\sim$0.2% higher than the analogous Combo value) does not follow any genre-based division, as previously observed in Ref. [22] but rather identifies some characteristics shared by the movies audience. The result of the community detection procedure is shown in Fig. 6 (top panel). Our method is able to detect movies released in 1996–'97 the year before the survey (in orange), such as "Mission Impossible," "Independence Day," and "Donnie Brasco." So this group of movies is characterized by the curiosity of users toward new releases. A second group collects *family movies* (as they were called in Ref. [22]), including "Cinderella," "101 Dalmatians," "Home Alone," or "Mrs. Doubtfire" (in green). In the blue community we find more "adult" movies, such as the "Alien" saga, the episodes of "Die Hard," "Escape from New York," "Judge Dredd," "Conan the Barbarian," as well as "Terminator" episodes, and some westerns like "The Good, the Bad, and the Ugly" and "Young Guns." In this block we have also cult movies, such as "Blade Runner," "Star Wars," and "Back to the Future." In the lime community we can find horror titles, like "Tales From the Crypt" episodes, "A Nightmare on Elm Street," and "Bram Stoker's Dracula." The red community groups together European production movies ("Cinema Paradiso," "Mediterraneo," "Four Weddings and a Funeral," "Jean de Florette," "Como agua para chocolate"), but also U.S. Independent production (such as "Raising Arizona," "Clerks," or "Night on Earth"). Movies inspired by books or theatrical plays ("Emma," "Richard III," "Sense and Sensibility," and "Othello") can be found in the pink community. In the last relevant block we can find classical Hollywood movies (in yellow) such as "Casablanca," "Ben Hur," "Once Upon a Time in America," and "Taxi Driver." Interestingly, in this group we can find Hitchcock's filmography ("Vertigo," "Psycho," "Rebecca," and "Rear Window").

Our results are in substantial agreement with those of Ref. [22] but for the different connectance values of the

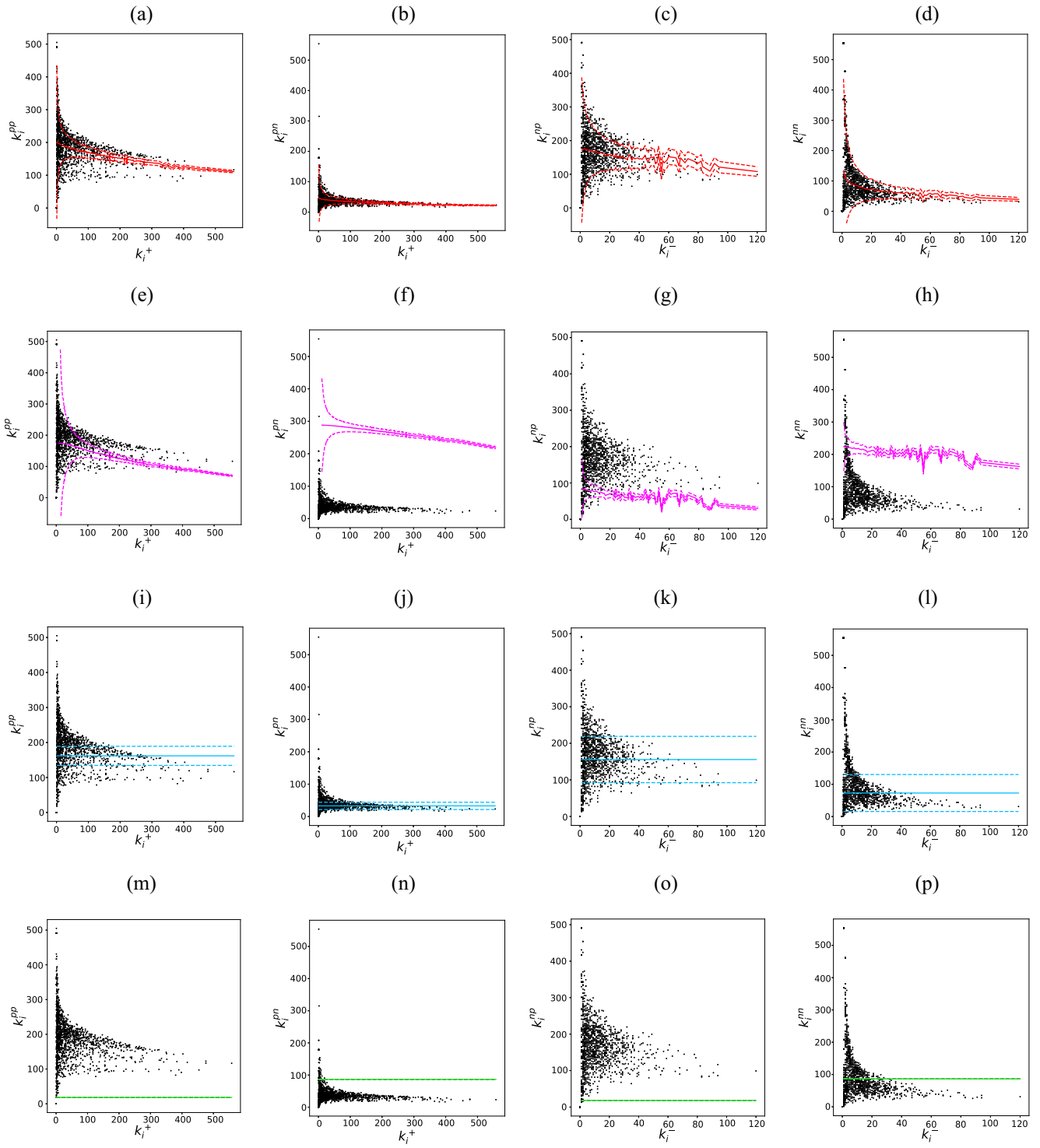

FIG. 5. Application of the method to the ML network. In all panels we show ANND versus k_i. The first column shows k_i^{pp} versus k_i^+. Proceeding on the right we find k_i^{pn} vs. k_i^+, k_i^{np} vs. k_i^-, and k_i^{nn} vs. k_i^-. Red lines (a, b, c,d) show the expectation values computed with our method. Magenta (e, f, g, h), blue (i, j, k, l), and green lines (m, n, o, p) are instead the expectations under WCM, PCM, and Erdös-Rényi RG, respectively. The area of ± 2 standard deviations around the average value is reproduced in dashed-line style.

BiCM- and the BiSCM-induced projection networks. Such behavior is not surprising, due to the different constraints imposed by the two null models. Indeed the BiSCM is more restrictive, fixing the degree sequence of each rating in the bipartite network, while the BiCM fixes the degree sequence of positive ratings only (i.e., merging the cases of $\beta = 3, 4, 5$), thus allowing for greater fluctuations. This effect can be observed in the connectance of the validated projection, which is 0.87% for BiSCM against a value of 1.17% for the BiCM.

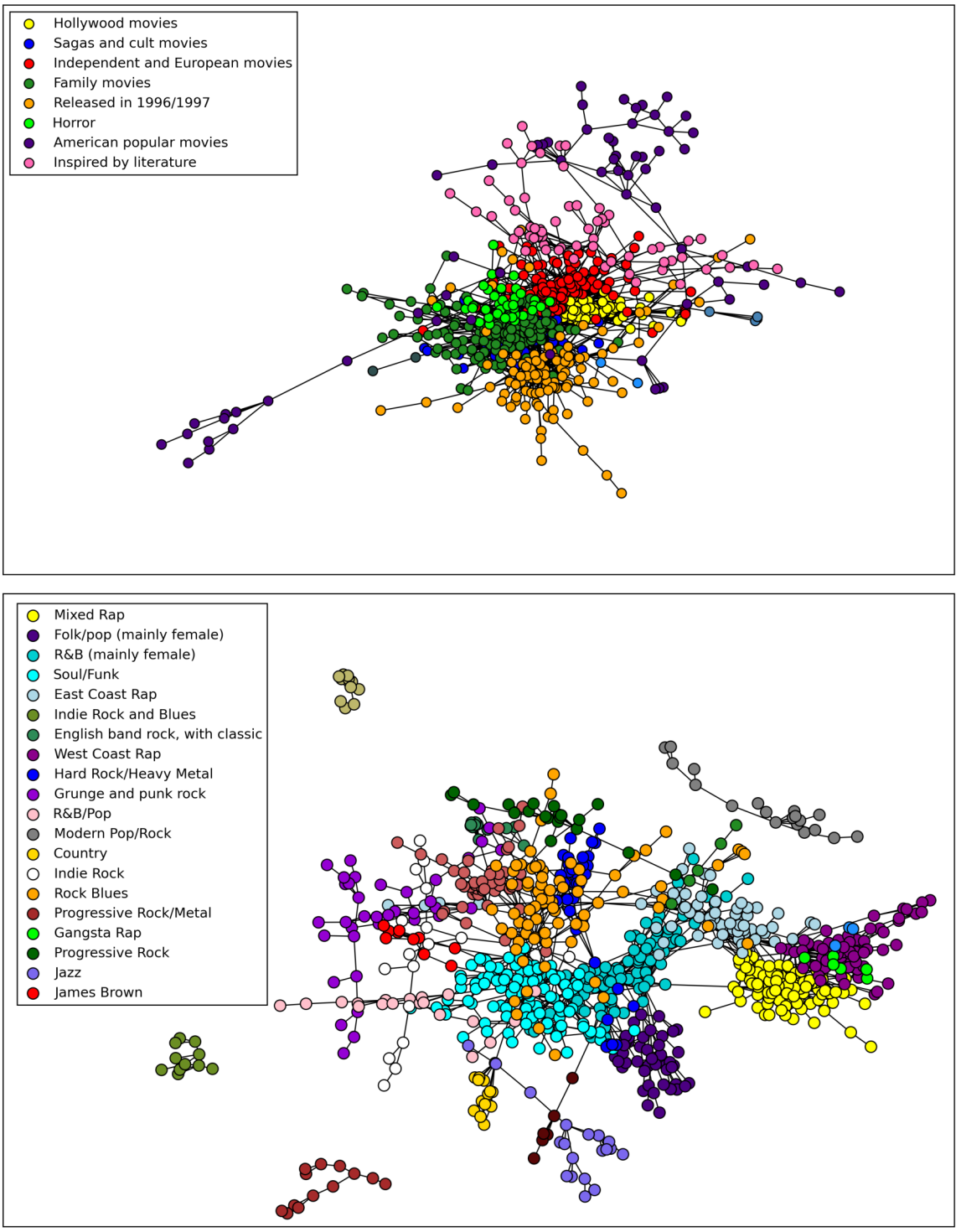

FIG. 6. Graphical representation of the most numerous communities for ML and DM networks. After the validation procedure, a standard modularity-based community detection algorithm is performed and the communities are here represented.

B. Digital music

With respect to the previous ML case, we obtain smaller and more precise groups of artists. In this case, the communities were found via the Combo community detection algorithm. The final configuration has a modularity of 0.874, $\sim$0.001% higher than the reshuffled Louvain best partition value. For the DM network, each community reveals a specific genre or combination of genres. A pictorial representation of the most numerous communities of the validated network is provided in Fig. 6 (bottom panel).

We have the small light-green community with two classic rock English bands, both of them characterized by a fusion with classical arrangements (Moody Blues and Electric Light Orchestra). Different groups collect different shades of rock: the hard rock and heavy metal community is in blue (Loverboy, Alice Cooper, Van Halen, Scorpions, Deep

Purple, Lynyrd Skynyrd) while the progressive rock is in green (Premiata Forneria Marconi, Soft Machine), and experimental rock in magenta. In dark violet we can find the grunge rock and related tendencies: Alice in Chains, Pearl Jam, Soundgarden, as well as Red Hot Chilli Peppers, Iggy and the Stooges, and the MC5. In indian red there is a community including Elton John, Billy Joel, Genesis, as well as Phil Collins and Peter Gabriel in their solo careers. The sea green and indigo groups represent, respectively, female R&B singers (Whitney Houston, Aretha Franklin, Alicia Keys, Nelly Furtado) and female folk/pop singers (Alanis Morisette, Anastacia, Vanessa Carlton, Dido). The rap genre is divided between East Coast and West Coast hip-hop, gangsta rap, and a mixed community with the most famous artists (Eminem, Jay Z, 2 Pac, 50 Cent, D12), depicted in light blue, dark magenta, lime, and yellow, respectively. The isolated community in violet collects, respectively, jazz (Thelonious Monk, Miles Davis, Cannonball Adderley, Charles Mingus, Sonny Rollins), while the red one contains almost exclusively James Brown albums. A folk/country community (almost exclusively composed by John Denver and Gordon Lightfoot albums) is represented in gold. We finally have the gray and white groups with indie rock artists (Radiohead, Bon Iver, Of Monsters and Men), the R&B singers and songwriters in pink (Marvin Gaye, Johnny Gill, Luther Vandross). In orange we find the community of folk/rock/blues, including Bob Dylan, Jimi Hendrix, Eric Clapton, the Who, Paul Simon but also the subsequent Elvis Costello, Bruce Springsteen. It is interesting to find here even Robert Johnson, the legendary bluesman, who was a source of inspiration for the artists in this community. The community of soul-funk (the Jackson 5, Barry White, Stevie Wonder, the Commodores, the Parliament, Sly and the Family Stone, Prince, the Isley Brothers) is in cyan. It is interesting to note that some Jamiroquai albums can be found in this latter community: Indeed, several experts compared the first production of this artist to Stevie Wonder [45]. Some smaller communities have not been included in the plot due to their low number of participants. However, their interpretation is still clear, since they generally collect single artists (Leonard Cohen) or identify a very specific music genre (such as the group of white rappers Insane Clown Posse and Anybody Killa of the horrorcore genre).

C. The case of categorical scores

As previously highlighted, in this section we present one of the possible extensions of the methodology. More specifically, we show an application to the case in which distinct scores represent different realizations of a nominal variable. In this case, we use the Anvur dataset presented in Sec. III, where the categorical scores represent the classification of journals in the considered scientific areas, i.e., currently "class A" or "class A only until December 2017."

First of all, we construct the benchmark model following the procedure explained in Sec. II. By doing so, we obtain two probability matrices, one for each β, representing the probability that each journal is currently considered a top journal in each scientific sector, or it is considered a top journal only until December 2017. Then, the analysis proceeds differently with respect to the previous cases. Indeed, due to the impossibility to fix a threshold beta given the nature of the analyzed scores, we validate the projected networks using the two probability matrices as separate benchmarks. Through the validation on the scientific sectors layer we are able to understand which groups of subjects share a significant number of common neighbors; i.e., they both consider the same journals as relevant for the topic of interest. The validation on the other layer instead identifies the groups of journals that are considered top journals (or not) for the same scientific area in the economics and statistics field.

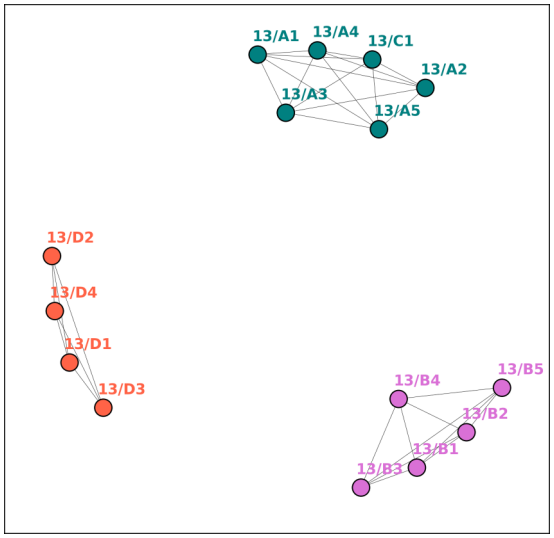

FIG. 7. Graphical representation of the communities identified in the Anvur dataset. The validation procedure has been performed on the scientific sectors layer.

Figure 7 shows the communities identified after the validation on the scientific sectors layer, using the benchmark model obtained with $\beta =$ green. However, we got the same result with the other benchmark model, with $\beta =$ red. Clearly the three disconnected communities represent the division in different research topics: the sea green community identifies the two macro areas 13/A and 13/C, respectively, corresponding to economics and economic history; the other two pink and orange groups instead represent the sectors 13/B and 13/D, respectively, corresponding to the research fields of business administration and statistics. Conversely, the validation on the other layer does not recognize any significant link between pairs of journals, with both the considered benchmarks. This result is not so surprising, since a similar behavior was previously observed in the projection of the World Trade Web dataset [22,34]: When the rectangularity of the bipartite network, i.e., the ratio between the dimension of the layers, is particularly high, the longest layer has a lower variability due to the smaller dimension of the support of the Poisson-binomial distribution. Moreover, since the effective significance threshold of FDR is even reduced by the total number of nodes' pairs on the layer, the possibility of finding a significant p value is further reduced.

VI. CONCLUSIONS

In everyday web experience, it is possible to encounter many different examples of online review platforms: from

Amazon customer ratings, to Tripadvisor and Anobii, just to mention the most famous ones. All these services provide an incredible source of information: Indeed, they are currently used to train recommendation systems, to focus possible advertisements about items that resemble the customers' tastes [23,46–50]. Nevertheless, to the best of our knowledge, a proper randomization of this kind of system is not available in the literature.

To fill this gap, we follow the research line of entropy-based null models [18–20], which provides an unbiased framework. In the case of score networks, the main difficulty resides in considering mutually exclusive outputs for each entry of the biadjacency matrix, i.e., having different possible scores with different probabilities. Our approach resembles the one presented in Ref. [51] for the reciprocal configuration model. In that case, four mutually exclusive possibilities were observable for every pair of nodes: no link, an exclusively outgoing link, an exclusively ingoing link, or a reciprocal link. Following this track, we were able to extend the configuration model framework to rating networks. We have shown that our method can be applied to many other types of networks, such as signed networks as well as categorical bipartite network.

Once the randomization procedure is complete, many types of analysis are possible. We first show that the obtained benchmark ensemble is able to capture some nontrivial network information, like the abundance of topological patterns such as the extensions of the ANND and bipartite motifs to rating networks. Then, another application was proposed: The model was employed as benchmark for the validation procedure presented in Ref. [22], to filter the information contained in one of the two layers. Otherwise stated, if co-occurrences cannot be explained by the model only, i.e., cannot be explained by the score degree sequence, they are validated. In this sense, we compare the real network with the expected value of our BiSCM randomization: The disagreements are signals of nontrivial similarities among nodes on the same layer. The result of such a procedure is a monopartite undirected network describing nodes belonging to the same layer performing similarly in the bipartite system. To have a clear understanding of the structure of the projected network, we run the Louvain community detection on it. Analyzing the Amazon Digital Music dataset [38], we were able to uncover communities of music based on customers taste. Analyzing the dataset of Ref. [26] it was instead possible to refine the community detection of Ref. [22]: Indeed, our model is more constrained than the one proposed there and filters more than the simple BiCM, after the binarization. Finally, we provide an example of possible application to the case of categorical networks. In this case, the standard division of research areas has been recovered just observing the number of top journals they share. Indeed, the knowledge of the success probability, for every score level, for every pair of nodes in the system, may provide significant insights regarding customers' purchase habits and preferences. Moreover, our methodology recovers these values unbiasedly and with no *a priori* information about the involved users, but simply observing the topology of the network itself and therefore, it may be suitable for different future applications. As an example, in the past year entropy-based configuration models [18–20] were successfully used for link prediction [52]. Indeed, a similar approach can be addressed to tackle the problem of predicting link in a review context using BiSCM, taking advantage of the general properties of configuration models. We leave this analysis for future developments of the research.

ACKNOWLEDGMENTS

This work was supported by the EU projects CoeGSS (Grant No. 676547), Openmaker (Grant No. 687941), SoBigData (Grant No. 654024), and the FET project DOLFINS 713 (Grant No. 640772). The authors are grateful to Giulio Cimini and Tiziano Squartini for useful discussions.

APPENDIX

1. Entropy maximization in BiSCM

Let us maximize the entropy Eq. (3) under the constraints on the degrees for each possible rating. The Hamiltonian reads as

$$H(\mathbf{G}) = \sum_{\beta}\left(\sum_{i} k_{i,\beta}\cdot\eta_{i,\beta} + \sum_{\alpha} k_{\alpha,\beta}\cdot\theta_{\alpha,\beta}\right) = \sum_{\beta}\sum_{i,\alpha} m_{i,\alpha,\beta}(\eta_{i,\beta}+\theta_{\alpha,\beta}),$$

meaning that the expectation value of the degree per rating is conserved. As in this class of systems, the solution is quite straightforward:

$$P(\mathbf{G}|\vec{\eta}_{\beta},\vec{\theta}_{\beta}) = \frac{e^{-H(\mathbf{G})}}{Z}, \tag{A1}$$

where Z is a normalizing factor called partition function. The computation of the partition function returns instead

$$\begin{aligned} Z &= \sum_{\mathbf{G}\in\mathcal{G}}\prod_{\beta}\prod_{i,\alpha}(x_{i,\beta}\,y_{\alpha,\beta})^{m_{i,\alpha,\beta}(\mathbf{G})} \\ &= \prod_{\beta}\prod_{i,\alpha}\sum_{\mathbf{G}\in\mathcal{G}}(x_{i,\beta}\,y_{\alpha,\beta})^{m_{i,\alpha,\beta}(\mathbf{G})} \\ &= \prod_{i,\alpha}\left(1+\sum_{\beta}x_{i,\beta}\,y_{\alpha,\beta}\right), \end{aligned} \tag{A2}$$

where $x_{i,\beta} = e^{-\eta_{i,\beta}}$, $y_{\alpha,\beta} = e^{-\theta_{\alpha,\beta}}$ and the last step is justified by the fact that all $m_{i,\alpha,\beta}$ are mutually exclusive, so the presence of an edge with rating $\widehat{\beta}$ excludes all the others (i.e., $m_{i,\alpha,\widehat{\beta}} = 1 \Rightarrow m_{i,\alpha,\beta} = 0,\ \forall\beta\neq\widehat{\beta}$). Implementing Eqs. (A2) into Eq. (A1) we get Eq. (4).

2. Entropy maximization in truncated WCM

In this case the entropy Eq. (3) is maximized under the constraints on the observed strengths. The Hamiltonian of the problem is

$$H(\mathbf{G}) = \sum_{i} s_i\eta_i + \sum_{\alpha} s_\alpha\theta_\alpha = \sum_{i,\alpha} m_{i,\alpha}(\eta_i+\theta_\alpha),$$

meaning that we preserve the expectation value of nodes' strengths. Again, the solution is straightforward and given by

$$P(\mathbf{G}|\vec{\eta},\vec{\theta}) = \frac{e^{-H(\mathbf{G})}}{Z}. \tag{A3}$$

However, taking into account that $m_{i,\alpha}$ can only vary into the range $1,\dots,\beta_{\max}$, the partition function Z returns

$$\begin{aligned} Z &= \sum_{\mathbf{G}\in\mathcal{G}}\prod_{i,\alpha}(x_i\,y_\alpha)^{m_{i,\alpha}(\mathbf{G})} = \prod_{i,\alpha}\sum_{\mathbf{G}\in\mathcal{G}}(x_i\,y_\alpha)^{m_{i,\alpha}(\mathbf{G})} \\ &= \prod_{i,\alpha}\frac{1-(x_i\,y_\alpha)^{\beta_{\max}+1}}{1-x_i\,y_\alpha}, \end{aligned} \tag{A4}$$

where $x_i = e^{-\eta_i}$ and $y_\alpha = e^{-\theta_\alpha}$. As in the previous case, the last passage is justified by the fact that edges are mutually exclusive and the presence of a score $\widehat{\beta}$ excludes all the others. With the implementation of Eqs. (A3) and (A4) it is possible to get the following probability distribution on the graphs ensemble:

$$P(\mathbf{G}|\vec{x},\vec{y}) = \prod_{i,\alpha}\frac{(1-x_i\,y_\alpha)(x_i\,y_\alpha)^{m_{i,\alpha}}}{1-(x_i\,y_\alpha)^{\beta_{\max}+1}}.$$

The probability to observe a link with rating β between nodes i and α reads as follows:

$$p_{i,\alpha,\beta} = \frac{(1-x_i\,y_\alpha)(x_i\,y_\alpha)^{\beta}}{1-(x_i\,y_\alpha)^{\beta_{\max}+1}}. \tag{A5}$$

Note that the previous Eq. (A5) identifies a truncated geometric distribution with parameter $x_i y_\alpha$ and $\beta \in \{1,\dots,\beta_{\max}\}$.

3. Entropy maximization in PCM

The entropy maximization procedure in the PCM framework follows exactly the same steps presented in the BiSCM section but with a reduced number of imposed constraints. Indeed, in this framework we just preserve the expectation value of the degrees per rating on one single predefined layer. The solution of this problem is

$$P(\mathbf{G}|\vec{\eta}) = \frac{e^{-H(\mathbf{G})}}{Z}, \tag{A6}$$

where the Hamiltonian reads as follows:

$$H(\mathbf{G}) = \sum_{i,\beta}k_{i,\beta}\,\eta_{i,\beta} = \sum_{\beta}\sum_{i,\alpha}m_{i,\alpha,\beta}\,\eta_{i,\beta}, \tag{A7}$$

while the partition function Z becomes

$$\begin{aligned} Z &= \sum_{\mathbf{G}\in\mathcal{G}}\prod_{\beta}\prod_{i,\alpha}(x_{i,\beta})^{m_{i,\alpha,\beta}(\mathbf{G})} = \prod_{\beta}\prod_{i,\alpha}\sum_{\mathbf{G}\in\mathcal{G}}(x_{i,\beta})^{m_{i,\alpha,\beta}(\mathbf{G})} \\ &= \prod_{i,\alpha}\left(1+\sum_{\beta}x_{i,\beta}\right), \end{aligned} \tag{A8}$$

where $x_{i,\beta} = e^{-\eta_{i,\beta}}$. Combining Eqs. (A7) and (A8) into Eq. (A6) we obtain the following probability distribution over the graphs ensemble:

$$P(\mathbf{G}|\vec{x}) = \prod_{i,\alpha}\frac{\prod_{\beta}(x_{i,\beta})^{m_{i,\alpha,\beta}}}{1+\sum_{\beta}x_{i,\beta}} = \prod_{i}\frac{\prod_{\beta}(x_{i,\beta})^{k_{i,\beta}}}{\left(1+\sum_{\beta}x_{i,\beta}\right)^{N_\Gamma}},$$

where the single term inside the product,

$$p_{i,\beta} = \frac{x_{i,\beta}}{1+\sum_{\beta}x_{i,\beta}} = \frac{k_{i,\beta}}{N_\Gamma}, \tag{A9}$$

simply identifies the probability to observe a link with score β incident to node i and coincides with the empirical observed frequency for score β. Notice that the previous model has been defined considering the degrees $k_{i,\beta}$ for $i \in \{1,\dots,N_L\}$ as constraints. However, the analogous counterpart can be implemented imposing the degrees of nodes on the other layer k_α, for $\alpha \in \{1,\dots,N_\Gamma\}$.

4. Entropy maximization in random graph

In the last considered null model, the entropy is maximized under the constraint on the observed number of edges per score only, denoted as E_β. The solution to the problem is given by

$$P(\mathbf{G}|\vec{\theta}) = \frac{e^{-H(\mathbf{G})}}{Z}, \tag{A10}$$

where the Hamiltonian of the problem is

$$H(\mathbf{G}) = \sum_{\beta}\theta_\beta\cdot E_\beta(\mathbf{G}) = \sum_{\beta}\theta_\beta\sum_{i,\alpha}m_{i,\alpha,\beta}, \tag{A11}$$

meaning that, for each score, we want to preserve the observed number of edges, while the partition function Z reads as follows:

$$Z = \sum_{\mathbf{G}\in\mathcal{G}}\prod_{i,\alpha,\beta}x_\beta = \prod_{i,\alpha}\left(1+\sum_{\beta}x_\beta\right) = \left(1+\sum_{\beta}x_\beta\right)^{N_L N_\Gamma}, \tag{A12}$$

where $x_\beta = e^{-\theta_\beta}$. Implementing Eqs. (A11) and (A12) into Eq. (A10) we obtain the following probability distribution:

$$P(\mathbf{G}|\vec{x}) = \frac{\prod_{\beta}x_\beta^{E_\beta}}{\left(1+\sum_{\beta}x_\beta\right)^{N_L N_\Gamma}},$$

where the term

$$p_\beta = \frac{x_\beta}{1+\sum_{\beta}x_\beta} = \frac{E_\beta}{N_L N_\Gamma} \tag{A13}$$

denotes the probability to observe a link of score β between any pair of nodes. Notice that the previous probability is invariant for all pairs of nodes and coincides with the empirical frequency of observed edges for the considered rating.

5. Chung-Lu approximation

As in the standard model presented in Ref. [27], we define the fitness variables associated to each node to be proportional to nodes' degrees as follows:

$$x^{CL}_{i,\beta} = \frac{k_{i,\beta}}{\sqrt{E_\beta}} \quad \text{and} \quad x^{CL}_{\alpha,\beta} = \frac{k_{\alpha,\beta}}{\sqrt{E_\beta}}. \tag{A14}$$

Then, to relax the constraints required, the connection probability between each pair of nodes in the network are obtained as

$$p^{CL}_{i,\alpha,\beta} = x^{CL}_{i,\beta}\,x^{CL}_{\alpha,\beta} = \frac{k_{i,\beta}\,k_{\alpha,\beta}}{E_\beta} \tag{A15}$$

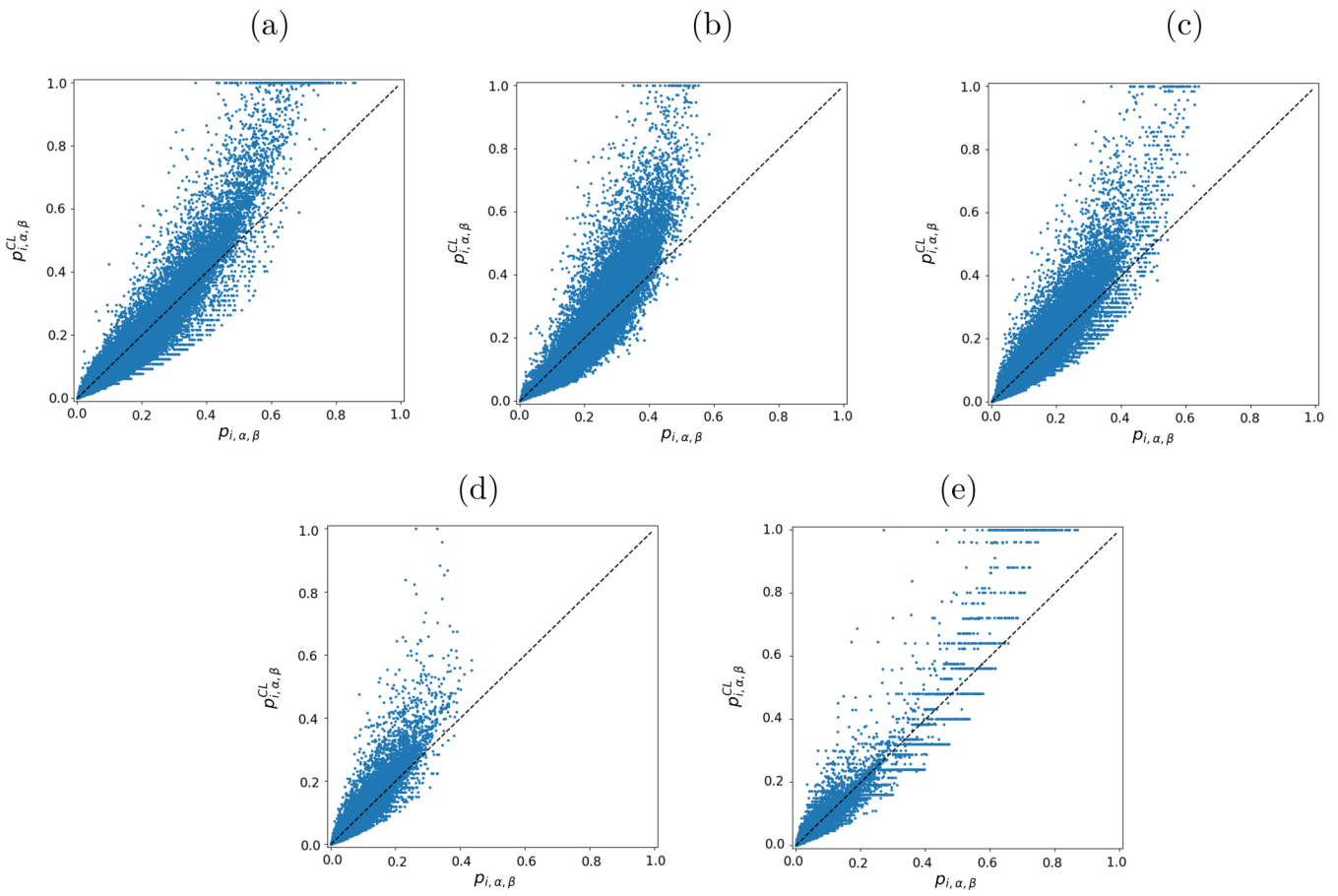

FIG. 8. Graphical comparison of the two definitions of probabilities in the ML network. All panels show $p_{i,\alpha,\beta}$ on the x axis and $p^{CL}_{i,\alpha,\beta}$ on the y axis. The plots reproduce the scores in decreasing order, starting from $\beta = 5$ in (a) to $\beta = 1$ in (e).

for all i, α, and β. The term $E_\beta = \sum_{i,\alpha} m_{i,\alpha,\beta}$ in Eqs. (A14) and (A15), identifies the total number of observed links for each score present in the data. Figure 8 provides a graphical comparison of the two definitions of probability for the ML network. It is evident that Eq. (A15) systematically overestimates the BiSCM values, especially for high probabilities.

6. Degree sequence recovery

As an additional safety check, in this section we show that our method numerically converges to the maximum likelihood solution, i.e., the vector of Lagrangian multipliers $(\vec{x}^*, \vec{y}^*)$ ensuring that the expected degree of each node over the ensemble, computed under the conditional probability distribution $P(\mathbf{M}|\vec{x}^*, \vec{y}^*)$ in Eq. (4), is equal to the observed value for each score level. We graphically represent this result in Fig. 9. Figures 9(a) and 9(b) compare the observed degree sequence obtained for each score β with the expected values of the degrees obtained with the BiSCM ensemble, for the ML and DM datasets, respectively. Instead, Figs. 9(c) and 9(d) show a comparison of the observed degree sequence for each

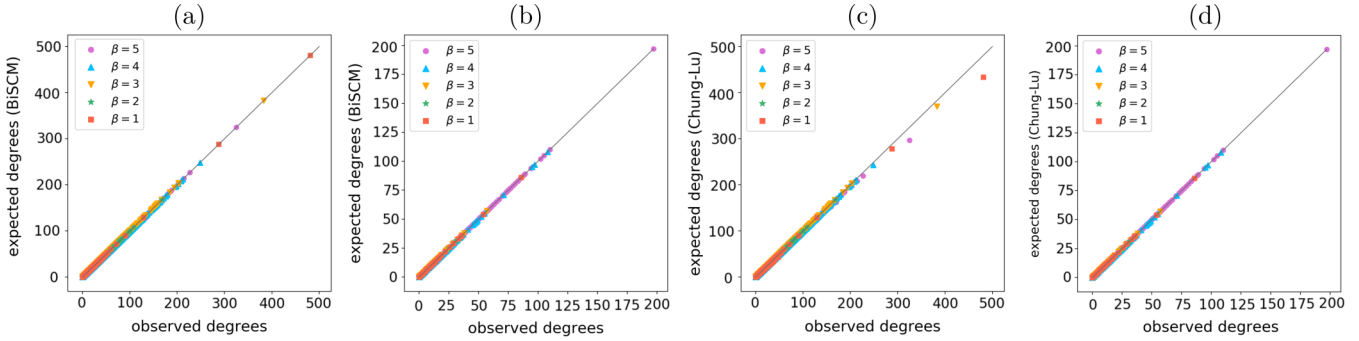

FIG. 9. Each plot provides a comparison of observed degrees on the x axis and expected values of the degrees in the ensemble on the y axis. Each value is computed for every available score. The expected degrees are recovered for the ensemble constructed with the BiSCM model (a), (b) and for its Chung-Lu approximation (c), (d). Panels (a) and (c) show the results for the ML dataset, while (b) and (d) provide the results for the DM dataset.

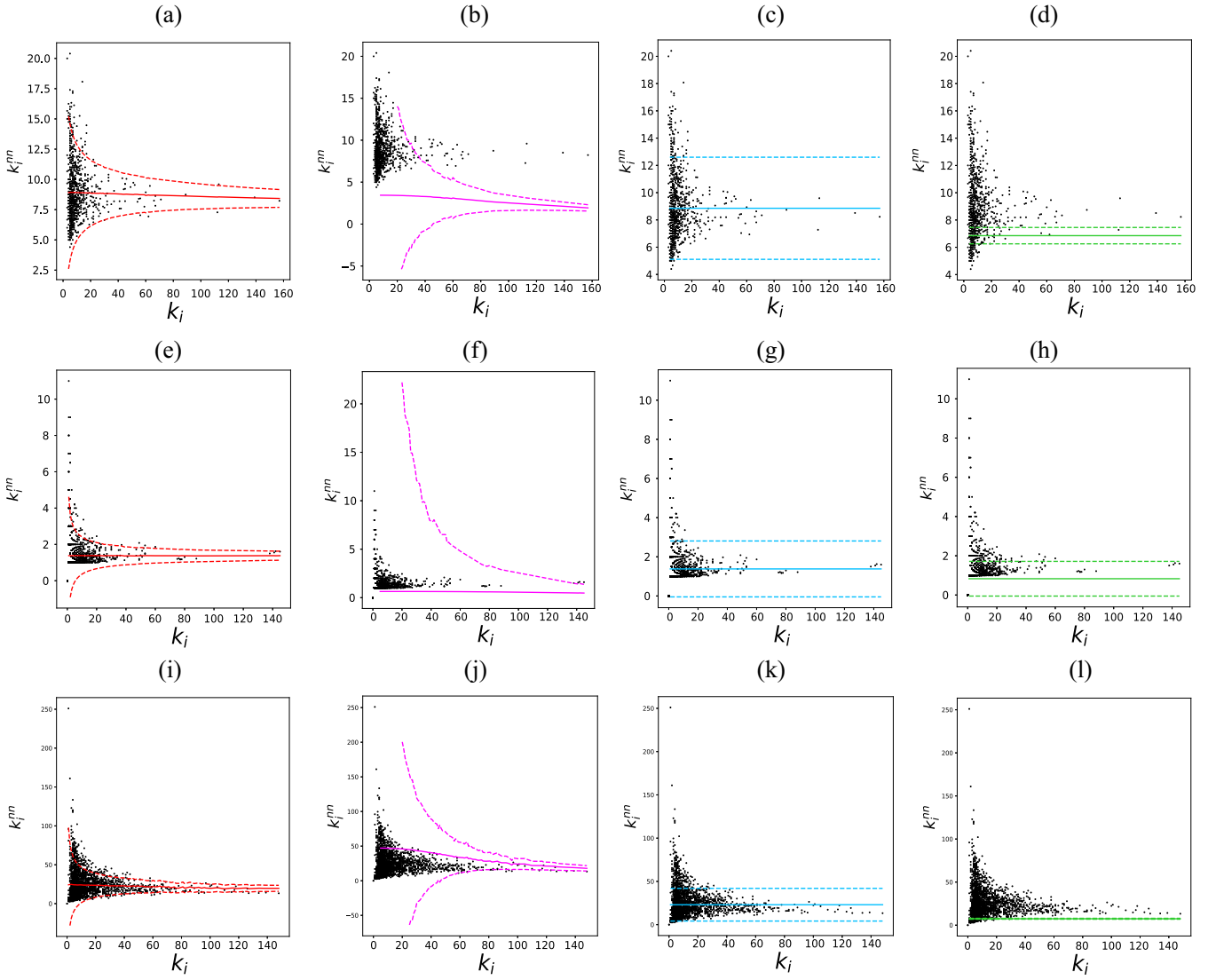

FIG. 10. Application of the method to the MI, SM, DM networks, from top to bottom line. All plots represent k_i^{nn} versus k_i. Red lines (a), (e), (i) show the expectation values computed with our method. Magenta (b), (f), (j), blue (c), (g), (k), and green lines (d), (h), (l) are instead the expectation values under WCM, PCM, and Erdös-Rényi RG. The area of ± 2 standard deviations around the average values has been represented in dashed-line style.

score and the expected values of the degrees obtained with the Chung-Lu approximation, again for the datasets ML and DM, respectively. All dots are almost aligned along the diagonal. This means that BiSCM is perfectly able to reproduce the observed degrees for each vertex and for each available score. On the contrary, the Chung-Lu model is *almost* able to reproduce the observed degrees. Indeed, in the case of connections between hubs, the probability per link saturates (as shown in Fig. 8) and therefore the degree is not perfectly replicated.

7. Further analyses of higher topological quantities: Positive reviews only

To check the capability of reproducing just positive ratings, a binarized version of the original matrix has been computed: The matrix $\mathbf{M}^{\mathrm{o}}$ has entries $m^{\mathrm{o}}_{i,\alpha} = 1$ whenever a positive review is registered and $m^{\mathrm{o}}_{i,\alpha} = 0$ otherwise. The associated entries of the probability matrix are indicated as $\langle m^{\mathrm{o}}_{i,\alpha} \rangle$ for all i, α and β. These terms represent the probability to observe a positive score (i.e., with $\beta = 3, 4, 5$) from user α to node i. In this setting, the degree of a node is defined as usual as $k_i = \sum_\alpha m^{\mathrm{o}}_{i,\alpha}$ and the correlation between neighbors' degrees can be studied with the standard definition of ANND,

$$k_i^{nn} = \frac{\sum_\alpha \sum_j m^{\mathrm{o}}_{i,\alpha} m^{\mathrm{o}}_{j,\alpha}}{\sum_\alpha m^{\mathrm{o}}_{i,\alpha}}.$$

The results are presented in Fig. 10. In all cases our method is perfectly able to identify the data's overall trend. Despite the fact that few observations still lie outside the confidence region, we can state that the correlation between neighbors' degrees may be interpreted as a consequence of the network topology, whenever the full distribution of received scores is specified in the null model construction, that is the case of the BiSCM. However, the other null models do not provide

satisfactory results: the RG model significantly underestimates the observed average traces [as shown in Figs. 10(d), 10(h) and 10(l)], while the WCM provides huge confidence intervals for the estimates [see Figs. 10(b), 10(f) and 10(j)]. Also in this case, the best performing alternative is the PCM null model [see Figs. 10(c), 10(g) and 10(k)] that uses the same constraints as the BiSCM but enforced on a reduced set of nodes.

[1] G. Caldarelli, *Oxford University Press Catalogue* (Oxford University Press, New York, 2007).

[2] M. E. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).

[3] A. L. Barabási, Science **325**, 412 (2009).

[4] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).

[5] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).

[6] R. Meusel, S. Vigna, O. Lehmberg, and C. Bizer, J. Web Sci. **1**, 33 (2015).

[7] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno, Sci. Rep. **1**, 197 (2011).

[8] D. Lazer, A. Pentland, L. A. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, Science **323**, 721 (2009).

[9] M. Buchanan, G. Caldarelli, P. De Los Rios, F. Rao, and M. Vendruscolo, *Networks in Cell Biology* (Cambridge University Press, Cambridge, 2010), pp. i–iv.

[10] A. Avena-Koenigsberger, B. Misic, and O. Sporns, Nat. Rev. Neurosci. **19**, 17 (2017).

[11] R. Mastrandrea, A. Gabrielli, F. Piras, G. Spalletta, G. Caldarelli, and T. Gili, Sci. Rep. **7**, 4888 (2017).

[12] A. Scala, P. Auconi, M. Scazzocchio, G. Caldarelli, J. A. McNamara, and L. Franchi, PLoS One **7**, e44521 (2012).

[13] J. A. Greene and J. Loscalzo, N. Engl. J. Med. **377**, 2493 (2017).

[14] G. De Masi, G. Iori, and G. Caldarelli, Phys. Rev. E **74**, 066112 (2006).

[15] I. Vodenska, H. Aoyama, A. P. Becker, Y. Fujiwara, H. Iyetomi, and E. Lungu, Systemic Risk and Vulnerabilities of Bank Networks (2017), doi: 10.2139/ssrn.3049976.

[16] D. B. Larremore, A. Clauset, and A. Z. Jacobs, Phys. Rev. E **90**, 012805 (2014).

[17] J. Bennett, C. Elkan, B. Liu, P. Smyth, and D. Tikk, ACM SIGKDD Explorations Newsletter **9**, 51 (2007).

[18] J. Park and M. E. J. Newman, Phys. Rev. E **70**, 066117 (2004).

[19] D. Garlaschelli and M. I. Loffredo, Phys. Rev. E **78**, 015101 (2008).

[20] T. Squartini and D. Garlaschelli, New J. Phys. **13**, 083001 (2011).

[21] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, Phys. Rev. E **90**, 062804 (2014).

[22] F. Saracco, M. J. Straka, R. D. Clemente, A. Gabrielli, G. Caldarelli, and T. Squartini, New J. Phys. **19**, 053022 (2017).

[23] J. D. M. Rennie and N. Srebro, in *Proceedings of the 22nd International Conference on Machine Learning (ICML'05)* (ACM, New York, 2005), pp. 713–719.

[24] F. Saracco, R. Di Clemente, A. Gabrielli, and T. Squartini, Sci. Rep. **5**, 10595 (2015).

[25] S. Boriah, V. Chandola, and V. Kumar, in *Society for Industrial and Applied Mathematics - 8th SIAM International Conference on Data Mining 2008, Proceedings in Applied Mathematics 130* (SIAM, 2008), Vol. 1, p. 243.

[26] As a rule of thumb, the time taken for obtaining the link probabilities for a network of 525×25 nodes with a connectance of order 0.5 (i.e., really high) is almost 3 h on a single process of an Intel Xeon CPU E5-2650 v2, 2.60 GHz server. Indeed, this value should be considered as a high limit for such a number of nodes, since the higher the connectance, the higher the calculation times and 0.5 is a pretty high value. For the fitnesses (i.e., the Lagrangian multipliers) calculation, the complexity of the problem scales as $O[\beta_{\max}(N_L + N_\Gamma)]$.

[27] F. Chung and L. Lu, Ann. Comb. **6**, 125 (2002).

[28] J. M. Diamond, *Assembly of Species Communities* (Harvard University Press, Boston, MA, 1975).

[29] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[30] G. Caldarelli and M. Catanzaro, Physica A: Stat. Mech. Appl. **338**, 98 (2004).

[31] M. Tumminello, S. Micciché, F. Lillo, J. Piilo, and R. N. Mantegna, PLoS One **6**, e17994 (2011).

[32] S. Gualdi, G. Cimini, K. Primicerio, R. Di Clemente, and D. Challet, Sci. Rep. **6**, 39467 (2016).

[33] N. Dianati, arXiv:1607.01735.

[34] M. J. Straka, G. Caldarelli, and F. Saracco, Phys. Rev. E **96**, 022306 (2017).

[35] A. Y. Volkova, Theory Probab. Appl. **40**, 791 (1996).

[36] Y. Hong, Comput. Stat. Data Anal. **59**, 41 (2013).

[37] Y. Benjamini and Y. Hochberg, J. Roy. Stat. Soc. B **57**, 289 (1995).

[38] http://konect.uni- koblenz.de/networks/.

[39] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, in *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (CSCW '94), Chapel Hill, NC, 22–26 October 1994* (ACM, New York, 1994), pp. 175–186.

[40] http://times.cs.uiuc.edu/∼wang296/Data/.

[41] http://www.anvur.it/wp-content/uploads/2018/04/Area_13_CLA_V_quad.pdf.

[42] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, J. Stat. Mech. (2008) P10008.

[43] S. Fortunato, Phys. Rep. **486**, 75 (2010).

[44] S. Sobolevsky, R. Campari, A. Belyi, and C. Ratti, Phys. Rev. E **90**, 012811 (2014).

[45] https://www.allmusic.com/artist/jamiroquai-mn0000176358/biography.

[46] J. L. Herlocker, J. A. Konstan, and J. Riedl, in *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00), Philadelphia, PA* (ACM, New York, 2000), pp. 241–250.

[47] G. Karypis, in *Proceedings of the Tenth International Conference on Information and Knowledge Management (CIKM '01),*

Atlanta, Georgia, 5–10 October 2001 (ACM, New York, 2001), pp. 247–254.

[48] R. Salakhutdinov and A. Mnih, in *Proceedings of the Conference on Advances in Neural Information Processing Systems 20 (NIPS'07)* (2007), p. 1257.

[49] Y. Koren, R. Bell, and C. Volinsky, Matrix factorization techniques for recommender systems (2009).

[50] J. Tang, C. Aggarwal, and H. Liu, in *Proceedings of the 25th International Conference on World Wide Web (WWW '16), Montreal, Canada, 11–15 April 2016* (ACM, 2016), pp. 31–40.

[51] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, New J. Phys. **16**, 043022 (2014).

[52] F. Parisi, G. Caldarelli, and T. Squartini, App. Netw. Sci. **3**, 17 (2018).