

**Tensorial and bipartite block models for link prediction in layered networks and temporal networks**Marc Tarrés-Deulofeu,<sup>1,\*</sup>† Antonia Godoy-Lorite,<sup>2,\*</sup>‡ Roger Guimerà,<sup>1,3,§</sup> and Marta Sales-Pardo<sup>1,||</sup><sup>1</sup>*Departament d'Enginyeria Química, Universitat Rovira i Virgili, 43006 Tarragona, Catalonia*<sup>2</sup>*Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom*<sup>3</sup>*ICREA, 08010 Barcelona, Catalonia*

(Received 18 January 2018; revised manuscript received 24 October 2018; published 27 March 2019)

Many real-world complex systems are well represented as multilayer networks; predicting interactions in those systems is one of the most pressing problems in predictive network science. To address this challenge, we introduce two stochastic block models for multilayer and temporal networks; one of them uses nodes as its fundamental unit, whereas the other focuses on links. We also develop scalable algorithms for inferring the parameters of these models. Because our models describe all layers simultaneously, our approach takes full advantage of the information contained in the whole network when making predictions about any particular layer. We illustrate the potential of our approach by analyzing two empirical data sets: a temporal network of e-mail communications, and a network of drug interactions for treating different cancer types. We find that multilayer models consistently outperform their single-layer counterparts, but that the most predictive model depends on the data set under consideration; whereas the node-based model is more appropriate for predicting drug interactions, the link-based model is more appropriate for predicting e-mail communication.

DOI: [10.1103/PhysRevE.99.032307](https://doi.org/10.1103/PhysRevE.99.032307)**I. INTRODUCTION**

Imagine a team of researchers looking for promising drug combinations to treat a specific cancer type for which current treatments are ineffective. The team has data on the effect of certain pairs of drugs on other cancer types, but the data are very sparse—only a few drug pairs have been tested on each cancer type, and each drug pair is tested in a few cancer types, at best, or has never been tested at all. The challenge is to select the most promising drug pairs for testing with the target cancer type, so as to minimize the cost associated to unsuccessful tests.

We can formalize this challenge as the following inference problem: We have a partial observation of the pairwise interactions between a set of nodes (drugs) in different network layers (cancer types), and we need to infer which are the unobserved interactions within each layer (drug interactions in each cancer type). This challenge is relevant for the many systems that can be represented as multilayer networks [1–4], and is also formally analogous to the challenge of predicting the existence of interactions between nodes in time-resolved networks [5–11]. For instance, we would face the same situation if we had data about the daily e-mail or phone communications between users, and wanted to infer the existence of interactions between pairs of users on a certain unobserved day; in this case each layer would be a different day.

Here, we introduce new generative models that are suitable to address the challenge above. We model all layers concurrently, so that our approach takes full advantage of the information contained in all layers to make predictions for any one of them. Our approach relies on the fact that having information on the interactions in different layers aids the inference process; in other words, that the interactions in layers different from the one we are interested in are informative about the interactions in the query layer. For instance, biologically similar cancer types are likely to show similar responses to the same drug pairs, and similar days of the week (for instance weekdays versus weekends) are also likely to display similar communication patterns for pairs of users.

Our approach is based on recent results on probabilistic inference on stochastic block models, which has been successful at modeling the structure of complex networks [12–14] and at predicting the behavior in biological [15] and social [16,17] systems. In particular, we focus on mixed-membership stochastic block models [18], in which nodes are allowed to belong to multiple groups simultaneously. With these models it is possible to model large complex networks with millions of links and, because they are more expressive than their fixed-membership counterparts, their predictive power is often superior [17]. We propose two different mixed-membership multilayer network models: a tensorial model that takes nodes as the basic unit to describe interactions in different layers, and a bipartite model that takes links (or pairs of nodes) as the basic unit. In our models layers, as well as nodes or links, are grouped based on the interaction patterns observed in them. This is in contrast to existing approaches, which do not take full advantage of the information that each layer carries about the structure of other layers.

We illustrate our models and inference approaches by analyzing two data sets: a network of drug interactions in different

\*These authors contributed equally to this work.

†marc.tarres@urv.cat

‡a.godoy-lorite@ucl.ac.uk

§Corresponding author: roger.guimera@urv.cat

||marta.sales@urv.cat

cancer types, and a temporal network of e-mail communications [19]. We find that modeling all layers simultaneously, and assuming that they can be grouped, results in link predictions that are more accurate (in terms of standard metrics such as precision and recall) than those of single-layer models and of simpler multilayer models. However, which of the two models (node-based or link-based) is the most predictive depends on the data set under consideration. Indeed, whereas for drug interactions drug groups are very informative and, therefore, node-based models are most predictive, temporal e-mail networks are best described in terms of links, that is, in terms of the relationships between pairs of individuals rather than the individuals themselves.

## II. TENSORIAL AND BIPARTITE MIXED-MEMBERSHIP BLOCK MODELS FOR LAYERED NETWORKS

We aim to model  $N$  nodes interacting by pairs in  $M$  different layers; these layers correspond to the different contexts in which the nodes interact (for example, different cancer types or time windows). We represent these interactions as a layered graph  $G$  whose links  $(i, j, \ell)$  represent interactions between nodes  $i$  and  $j$  in layer (or at time)  $\ell$ . Moreover, we allow for multivalued interactions so that  $(i, j, \ell)$  can be of different types  $r_{ij\ell} \in R$ , where  $R$  is a finite set. Note that we can use this formalism to model labels, attributes or ratings associated to the interactions [15,17]; graphs with binary interactions are therefore a particular case within this general framework in which  $r_{ij\ell} = 1$  if the interaction occurs and  $r_{ij\ell} = 0$  if it does not.

We consider two types of generative models: one that takes individual nodes as its basic unit, and one that models links (or node pairs). The first generative model, based on individual nodes, is as follows. There are  $K$  groups of nodes and  $L$  groups of layers. We assume that the probability that a node in group  $\alpha$  has an interaction of type  $r$  with a node in group  $\beta$  in a layer in group  $\gamma$  is  $p_{\alpha\beta\gamma}(r)$ . Furthermore, we assume that both nodes and layers can belong to more than one group. To model such mixed-group memberships [18], to each node  $i$  we assign a vector  $\theta_i \in \mathbb{R}^K$ , where  $\theta_{i\alpha} \in [0, 1]$  denotes the probability that node  $i$  belongs to group  $\alpha$ . Similarly, to each layer  $\ell$  we assign a vector  $\eta_{\ell\gamma} \in \mathbb{R}^L$ . These vectors are normalized so that  $\sum_{\alpha} \theta_{i\alpha} = \sum_{\gamma} \eta_{\ell\gamma} = 1$ . The probability that link  $(i, j, \ell)$  is of type  $r$  is then

$$\Pr[r_{ij\ell} = r] = \sum_{\alpha\beta\gamma} \theta_{i\alpha} \theta_{j\beta} \eta_{\ell\gamma} p_{\alpha\beta\gamma}(r). \quad (1)$$

Note that if link types are exclusive (i.e., each edge can be of only one type), the probability tensor must satisfy the constraint  $\sum_{r \in R} p_{\alpha\beta\gamma}(r) = 1$ . Since this model is an extension of the mixed-membership stochastic block model [17,18] where the probability matrices become tensors because of the multiple layers [1], we call it the tensorial mixed-membership stochastic block model (T-MBM).

Our second generative model for layered networks is as follows. Instead of assuming that nodes belong to groups, we assume that it is links (or pairs of nodes, rather than individual nodes) that belong to groups [8]. In this model we have  $J$  groups of links, and the probability that a link  $e_{ij} \equiv e$  in group  $\alpha$  is of type  $r$  in a layer  $\ell$  in group  $\gamma$  is  $p_{\alpha\gamma}(r)$ . We also assume

that links can belong to more than one group so that  $\zeta_{e\alpha}$  is the probability that link  $e$  belongs to group  $\alpha$  and  $\sum_{\alpha} \zeta_{e\alpha} = 1$ . As before, to each layer  $\ell$  we also assign a vector  $\eta_{\ell} \in \mathbb{R}^L$  of group memberships. Then, the probability that a given link in a particular layer is of type  $r$  is

$$\Pr[r_{ij\ell} = r] = \Pr[r_{e\ell} = r] = \sum_{\alpha\gamma} \zeta_{e\alpha} \eta_{\ell\gamma} p_{\alpha\gamma}(r), \quad (2)$$

where, as before, if link types are exclusive the probability matrices satisfy the condition  $\sum_{r \in R} p_{\alpha\gamma}(r) = 1$ . This model can be seen as a bipartite model with two types of elements, links and layers. In this representation, a link  $e_{ij}$  has a connection of type  $r$  to a layer  $\ell$  if  $r_{ij\ell} = r_{e\ell} = r$ . Therefore, we call this model the bipartite mixed-membership stochastic block model (B-MBM).

These models are novel in a number of ways. First, unlike other models of multilayer networks [8,9], they do not assume any particular order in the layers, and therefore do not impose any restrictions to how layers should be grouped. This is in contrast to approaches for temporal networks that can only group layers corresponding to consecutive times. While such restriction simplifies the task of grouping layers, it also eliminates the possibility of identifying, for example, periodicities in temporal networks. More importantly, this restriction prevents models from being applicable to nontemporal multilayer networks. Our models eliminate this restriction.

Second, unlike other models of multilayer and temporal networks, our models do not assume that: (i) the group-to-group connectivity matrices are independent from one layer to another, as do models based on the original multigraph stochastic block model [20–22]; (ii) layers belong to a single group of layers [8,9,23]. In the former case, all layers are completely different from each other, whereas in the latter layers are statistically identical to the others in the same group because they are assumed to come from the exact same generative model. Rather, in our models, layers come from mixtures of these generative models, so that they can be all distinct but still share some features with other layers. This means that our models are able to describe and predict missing links in layers that are unlike all other observed layers in the multiplex network, or predict interactions in layers that are only partially similar to other layers. The mixed membership also allows us to develop efficient expectation-maximization algorithms that can be massively parallelized [24] and, at the same time, provide better predictions than single-group models [17].

Our models are also different from matrix and tensor factorization approaches for temporal networks [25,26] in that they allow for richer and flexible representations of the data. Indeed, our tensorial model is a generalization of tensor factorization, in the same way that single-layer stochastic block models are the appropriate network generalization of matrix factorization [17]. Our models also differ from Bayesian Poisson Tucker decomposition [27] in that they can be used on networks with labeled links (for example, synergistic, additive, and antagonistic links in the cancer drug-interaction data set), which is not possible in Poisson-based models.

Finally, it is worth noting that all these related models assume the existence of node groups. The link-based B-MBM

is different from all of them in that it assumes that it is links that belong to groups, and in that sense is closer in spirit to the approach taken in Ref. [28].

### III. INFERENCE EQUATIONS AND EXPECTATION-MAXIMIZATION ALGORITHMS

Given a set  $G^O$  of observed links types, our goal is to predict the types  $r_{ij\ell}$  of links  $(i, j, \ell) \notin G^O$  whose type is unknown. Because marginalizing over the parameters in our models [Eqs. (1) and (2)] is too time consuming, here we present a maximum-likelihood approach (and the corresponding expectation-maximization algorithms) for the two models above. Note that for complex data sets our models are bound to be highly dimensional. As a result, we expect the likelihood landscape to be very rugged with many local maxima with comparable likelihood values that correspond to different model parameters. To take such quasidegeneracy into account, instead on making predictions based on a single maximal-likelihood solution, we make predictions by averaging over the (local) maximum-likelihood model solutions obtained for different random initializations of the model parameters (see Appendix C and Ref. [29]).

#### A. Tensorial model

Given the generative T-MBM model in Eq. (1), and abbreviating its parameters as  $\theta, \eta, \mathbf{p}$ , the likelihood of the model is

$$P(G^O | \theta, \eta, \mathbf{p}) = \prod_{(ij\ell) \in G^O} \sum_{\alpha\beta\gamma} \theta_{i\alpha} \theta_{j\beta} \eta_{\ell\gamma} p_{\alpha\beta\gamma}(r_{ij\ell}). \quad (3)$$

As we show below (Appendix A), the values of the parameters that maximize this likelihood satisfy the following equations:

$$\theta_{i\alpha} = \frac{\sum_{(j\ell) \in \partial i} \sum_{\beta\gamma} \omega_{ij\ell}(\alpha, \beta, \gamma)}{d_i}, \quad (4)$$

$$\eta_{\ell\gamma} = \frac{\sum_{(ij) \in \partial \ell} \sum_{\alpha\beta} \omega_{ij\ell}(\alpha, \beta, \gamma)}{d_\ell}, \quad (5)$$

$$p_{\alpha\beta\gamma}(r) = \frac{\sum_{(i,j,\ell) \in G^O | r_{ij\ell}=r} \omega_{ij\ell}(\alpha, \beta, \gamma)}{\sum_{(i,j,\ell) \in G^O} \omega_{ij\ell}(\alpha, \beta, \gamma)}. \quad (6)$$

Here,  $\partial i = \{(j, \ell) | (i, j, \ell) \in G^O\}$  are the set of observed layer-specific neighbors of node  $i$  and  $d_i = |\partial i|$  is the total degree of the node in all the layers. Similarly,  $\partial \ell = \{(i, j) | (i, j, \ell) \in G^O\}$  is the set of observed links in layer  $\ell$  and  $d_\ell = |\partial \ell|$ . Finally,  $\omega_{ij\ell}(\alpha, \beta, \gamma)$  is the estimated probability that the type of a given link  $r_{ij\ell}$  is due to  $i, j,$  and  $\ell$  belonging to groups  $\alpha, \beta,$  and  $\gamma$  respectively, and is given by

$$\omega_{ij\ell}(\alpha, \beta, \gamma) = \frac{\theta_{i\alpha} \theta_{j\beta} \eta_{\ell\gamma} p_{\alpha\beta\gamma}(r_{ij\ell})}{\sum_{\alpha'\beta'\gamma'} \theta_{i\alpha'} \theta_{j\beta'} \eta_{\ell\gamma'} p_{\alpha'\beta'\gamma'}(r_{ij\ell})}. \quad (7)$$

These equations can be solved iteratively with an expectation-maximization algorithm, starting with an initial estimate of  $\theta, \eta,$  and  $\mathbf{p}$  and, then, repeating the following steps: (i) use Eq. (7) to compute  $\omega_{ij\ell}(\alpha, \beta, \gamma)$  for  $(i, j, \ell) \in G^O$  (expectation step); (ii) use Eqs. (4)–(6) to compute  $\theta, \eta,$  and  $\mathbf{p}$  (maximization step).

#### B. Bipartite model

Similarly, the likelihood of the B-MBM is

$$P(G^O | \zeta, \eta, \mathbf{p}) = \prod_{(e,\ell) \in G^O} \sum_{\alpha\gamma} \zeta_{e\alpha} \eta_{\ell\gamma} p_{\alpha\gamma}(r_{e\ell}), \quad (8)$$

and the maximum-likelihood estimators of the parameters satisfy

$$\zeta_{e\alpha} = \frac{\sum_{\ell \in \partial e} \sum_{\beta\gamma} \phi_{e\ell}(\alpha, \gamma)}{d_e}, \quad (9)$$

$$\eta_{\ell\gamma} = \frac{\sum_{e \in \partial \ell} \sum_{\alpha} \phi_{e\ell}(\alpha, \gamma)}{d_\ell}, \quad (10)$$

$$p_{\alpha\gamma}(r) = \frac{\sum_{(e\ell) \in G^O | r_{e\ell}=r} \phi_{e\ell}(\alpha, \gamma)}{\sum_{(e\ell) \in G^O} \phi_{e\ell}(\alpha, \gamma)}. \quad (11)$$

Here  $\partial e = \{\ell | (e, \ell) \in G^O\}$  are the observations of link  $e_{ij}$  in all layers and  $d_e = |\partial e|$ . As before,  $\partial \ell = \{e | (e, \ell) \in G^O\}$  are the observed links in layer  $\ell$  and  $d_\ell = |\partial \ell|$ . Finally,  $\phi_{e\ell}(\alpha, \gamma)$  is the estimated probability that the type of a specific link  $r_{e\ell}$  is due to  $e$  and  $\ell$  belonging to groups  $\alpha$  and  $\gamma$ , respectively; we can compute  $\phi_{e\ell}(\alpha, \gamma)$  as

$$\phi_{e\ell}(\alpha, \gamma) = \frac{\zeta_{e\alpha} \eta_{\ell\gamma} p_{\alpha\gamma}(r_{e\ell})}{\sum_{\alpha'\gamma'} \zeta_{e\alpha'} \eta_{\ell\gamma'} p_{\alpha'\gamma'}(r_{e\ell})}. \quad (12)$$

Like in the tensorial model, these equations can be solved iteratively using an expectation-maximization algorithm.

### IV. VALIDATION ON SYNTHETIC DATA

We start by testing the performance of our expectation-maximization inference algorithms on synthetic networks generated with the T-MBM and the B-MBM. In each case we generate one network with 12 layers, one with 24, and one with 36, all of them with 128 nodes and two types of link, active and inactive (which can be interpreted as presence or absence of link, respectively). The number of groups of nodes, node pairs, and layers is always  $K = J = L = 4$ . The values of  $\theta_{i\alpha}, \zeta_{e\alpha},$  and  $\eta_{\ell\gamma}$  are drawn from a beta distribution  $\text{Beta}(a, b)$  with  $a = 0.2$  and  $b = 0.3$ , and then normalized. Similarly, the values of the elements of  $\mathbf{p}$  are drawn from a beta distribution with  $a = 0.2$  and  $b = 1.0$ , and then normalized.

To evaluate the ability of the expectation-maximization inference algorithms to recover the original networks, we calculate the area under the ROC curve (AUC), the precision, and the recall in fivefold cross-validation experiments (see Appendix C for details). The AUC measures how well a model separates active from inactive links. In particular, it measures the frequency with which an active unobserved link is assigned a higher probability to be active than an inactive unobserved link. Precision accounts for the fraction of links predicted to be active that are indeed active. Recall gives the fraction of active links that are predicted to be active. To calculate both precision and recall, we need to set a threshold  $T$  that allows us to map probabilities  $P[r_{ij\ell} = 1]$  into a binary variable, so that the model predicts that the link is active if  $P[r_{ij\ell} = 1] \geq T$ . In what follows, we choose  $T$  in such a way

that the density of active links in the test set is predicted to be the same as in the training set.<sup>1</sup>

We compare our inference algorithms to two different baselines. The naive baseline takes into account all the observations of a link  $(i, j)$  in the training set. Then, it makes predictions for the unobserved link types  $r_{ijl}$  (active or inactive) based on the fraction of times link  $(i, j)$  has been observed to be of type  $r$  in the training set

$$P_{\text{naive}}[r_{ijl} = r] = \frac{\sum_{s|(ijs) \in G^o} \delta r, r_{ijs}}{N_{ij}}, \quad (13)$$

where  $N_{ij}$  is the number of times the link  $(i, j)$  is observed in the training set.

The second baseline, which we call the independent-layer naive, estimates the probability of link  $(i, j, l)$  being of type  $r$  as the fraction of links of type  $r$  observed in layer  $l$ ,

$$P_{\text{naive-IL}}[r_{ijl} = r] = \frac{\sum_{(kn)|(knl) \in G^o} \delta r, r_{knl}}{N_l}, \quad (14)$$

where  $N_l$  is the number of links of any type observed in layer  $l$  of the training set. The validation experiments confirm that the inference algorithms are, in most situations, able to recover the networks with as much accuracy as one could possibly expect (Fig. 1). More precisely, when the inference is carried out assuming that the number of groups is equal or greater than the planted value  $K = J = L = 4$ , then AUC, precision, and recall converge to the values one would obtain by assigning to each link the exact probability that was used to create the network using the model (although, in the B-MBM, the number of groups used for the inference sometimes needs to be larger than the planted number of groups). The only exception to this is the B-MBM model when the number of layers is 12; in this case some layers are extremely sparse and some links are only active in one or very few layers, so recovering the true model is virtually impossible. With typical networks containing tens of layers (such as those we investigate in the next section), we expect this limitation not to be very significant.

### V. MODEL COMPARISON ON REAL DATA

We perform experiments on two different data sets: the time-resolved e-mail network of an organization spanning one year [19], and a network of drug-drug interactions in different cancer cell lines [30]. In the e-mail data set, we represent each day as a different layer of the multilayer network, and two users are considered to interact in a given day if they send at least one e-mail in either direction during that day. We consider several e-mail networks that correspond to e-mail communications within organizational units (see Table I).

In the drug-drug interactions data set, each layer corresponds to a different cancer cell line and we have information on the effects of some drug pair combinations on some cancer cell lines [30]. In contrast to the e-mail data set, in which all

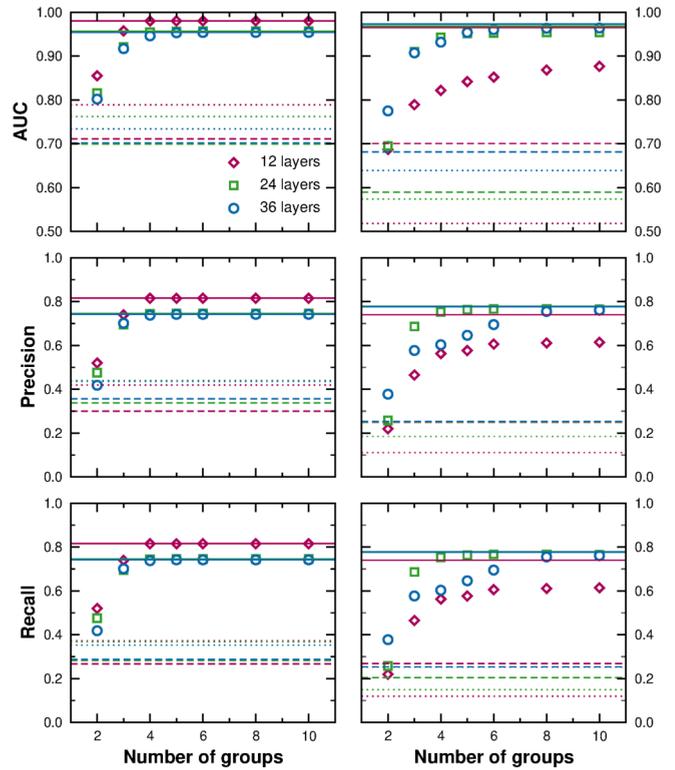


FIG. 1. Validation of the inference algorithms on synthetic networks. We generate networks using the tensorial model, T-MBM (left column), and the bipartite model, B-MBM (right column). All networks have  $K = J = L = 4$  groups, 128 nodes, and 12, 24, or 36 layers (see text for details). We plot the AUC (top), precision (middle), and recall (bottom) as a function of the number of groups assumed by the inference algorithm. Each point shows the average of a fivefold cross validation, and the error bars represent the standard error of the mean (they are typically smaller than symbol sizes). The dashed line represents the naive baseline, and the dotted line the naive independent layers baseline (see text). The solid line indicates the maximum possible performance, obtained by assigning to each link a probability that is exactly the probability that was used to create the network using the model.

the interactions (or lack of interaction) are observed, this data set is sparsely observed—we have information about 1.5% of the drug pairs. Specifically, the available experimental data is a real-valued magnitude representing the combined efficiency of two drugs on a particular cell line. These magnitudes range from large absolute values, in which case the interaction is categorized as synergistic (if it is positive) or antagonistic (if it is negative), to small absolute values, in which case the interaction is categorized as additive. In an additive interaction, the application of the two drugs together has an efficiency equal or similar to the sum of the efficiencies of each drug administered separately. By contrast, in a synergistic (antagonistic) interaction the efficiency of the two drugs administered together is significantly higher (lower) than the sum of the efficiencies of each drug administered separately.

In Table I we show the characteristics of each data set in terms of the types of links  $R$ , the total number of nodes, the total number of layers, the total number of possible links, and the number and fraction of actually observed links. In all

<sup>1</sup>This particular choice of threshold leads, when models are properly calibrated in a frequentist sense, to both precision and recall having very similar values (see Ref. [29]).

TABLE I. Data set characteristics. The e-mail networks we consider are complete networks where no-links are treated as links of type 0, so all potential links in the network are observed links. The drug-drug interaction network is a sparse data set, where we only have information about 1.4% of the links. Each observed drug-drug interaction can be of three types: antagonistic (ANT), additive (ADD), or synergistic (SYN).

| Data set               | Types of links $R$ | #Nodes | #Layers | #Observables | Fraction observed | #Observed                     |
|------------------------|--------------------|--------|---------|--------------|-------------------|-------------------------------|
| E-mail Unit 1          | {0, 1}             | 104    | 365     | 1954940      | 100%              | $ G^O _1 = 20, 807$           |
| E-mail Unit 2          | {0, 1}             | 114    | 365     | 2350965      | 100%              | $ G^O _1 = 27, 180$           |
| E-mail Unit 3          | {0, 1}             | 116    | 365     | 2434550      | 100%              | $ G^O _1 = 23, 979$           |
| E-mail Unit 4          | {0, 1}             | 118    | 365     | 2519595      | 100%              | $ G^O _1 = 17, 508$           |
| E-mail Unit 5          | {0, 1}             | 141    | 365     | 3602550      | 100%              | $ G^O _1 = 23, 923$           |
| E-mail Unit 6          | {0, 1}             | 161    | 365     | 4701200      | 100%              | $ G^O _1 = 20, 790$           |
| E-mail Unit 7          | {0, 1}             | 225    | 365     | 9198000      | 100%              | $ G^O _1 = 60, 238$           |
| Drug-drug interactions | {ANT, ADD, SYN}    | 69     | 85      | 199410       | 1.37%             | $ G^O _{\text{ANT}} = 385$    |
| Drug-drug interactions | {ANT, non-ANT}     | 69     | 85      | 199410       | 1.37%             | $ G^O _{\text{ADD}} = 1, 543$ |
| Drug-drug interactions | {SYN, non-SYN}     | 69     | 85      | 199410       | 1.37%             | $ G^O _{\text{SYN}} = 863$    |

cases, we validate our models using a fivefold cross-validation scheme as described for synthetic networks (Appendix C).

In addition to the two baselines described in the previous section, here we consider an independent-layer mixed-membership stochastic block model for each layer so that,

$$P_{\text{IL}}[r_{ijl} = r] = \sum_{\alpha\beta} \theta_{i\alpha}^l \theta_{j\beta}^l p_{\alpha\beta}^l(r), \quad (15)$$

where the superindex  $l$  denotes that each layer has its own set of parameters. As in the tensorial and bipartite layered models, parameters are subject to the constraints  $\sum_{\alpha} \theta_{i\alpha}^l = 1, \forall l$  and  $\sum_r p_{\alpha\beta}^l(r) = 1 \forall l$ . The parameters for this model are obtained using the same method as in the tensorial and bipartite mixed-membership models, but considering each layer separately (see also Ref. [17]).

**A. E-mail networks**

We first consider the ability of each model and baseline to predict unobserved links in the e-mail networks listed in Table I. As before, to assess the performance of each model for each network we measure AUC, precision, and recall (see Appendix C for details).

In Fig. 2, we show that the bipartite link-based model outperforms the tensorial and baseline models in all metrics (see Fig. S1 for all other e-mail units). In these e-mail networks, the AUC is quite high even for the naive baseline because most pairs of individuals never exchange an e-mail and therefore it is easy to predict links for which  $r_{ijl} = 0$  in all observed layers. The situation changes when we look at precision and recall, which clearly show that the bipartite model is consistently and significantly superior at predicting links that are active. Somewhat surprisingly, we also find that the tensorial node-based model often gives lower values than the naive baseline model. The explanation lies in the fact that, contrary to both the naive baseline and the bipartite models, the tensorial model focuses primarily on nodes rather than on links and is thus less likely to account for the fact that many pairs of nodes in the network never communicate. More precisely, the probabilities assigned by the tensorial model depend on the product of the membership of the involved nodes, and these memberships are rarely equal to zero. Hence, according to the tensorial model most links have a nonzero

probability of existing, including those that are inactive for all observations in the training set.

To further investigate the workings of each approach, we analyze whether they are properly calibrated in a frequentist sense, that is whether the fitted models are able to reproduce statistical features of the training data set [31]. In particular, we consider the marginal and probabilistic calibration of all models. A model is probabilistically calibrated if events to

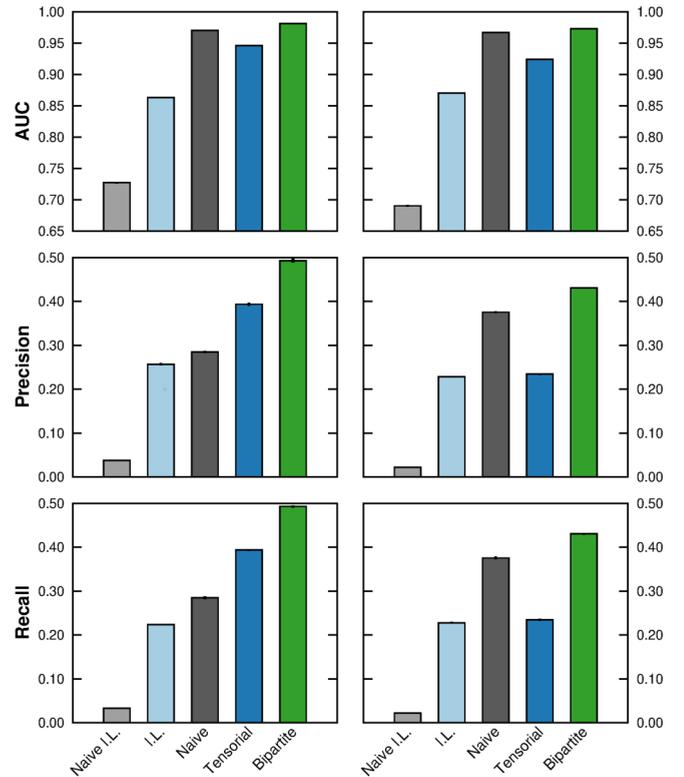


FIG. 2. Predictive performance of the models for e-mail networks. The left column shows results for Unit 1 and the right column shows results for Unit 5. Top: AUC; middle: precision; bottom: recall. Each bar represents the average of the fivefold cross validation for a given model (see Appendix C). The error bars (shown as a vertical line, which is small and not visible in some cases) represent the standard error of the mean.

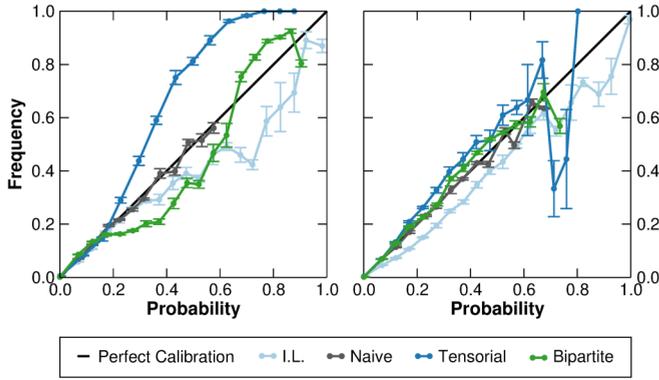


FIG. 3. Probabilistic calibration of the models for e-mail networks. We show results for Unit 1 (left) and Unit 5 (right). Each point in each line represents the average of the fivefold cross validation for a given model, with error bars representing the standard error of the mean. The Naive I.L. model is not included as it only assigned tiny probabilities that resulted in a single data point near the origin.

which the model assigns a probability  $p$  are observed with frequency  $p$  [31]. In our case, a model is calibrated if a fraction  $p$  of the links for which  $P[r_{ijl} = 1] = p$  actually exist. A model is marginally calibrated if, on average, each type of event is assigned a probability that is equal to the actual frequency of such events in the training set. In our case, a model is calibrated if the mean  $P[r_{ijl} = 1]$  assigned to links coincides with the density of the observed network [31].

In Fig. 3, we show that all models are relatively well calibrated probabilistically (higher probabilities correspond to higher frequencies), although the calibration is noticeably worse for the network obtained for Unit 1 (see Fig. S2 for the remaining units). In general, the bipartite model is better calibrated than the tensorial model, which is consistent with the higher predictive accuracy of the bipartite, link-based model. Perhaps surprisingly, the naive baseline model appears to have an even better probabilistic calibration across all units. Figure 4 also shows that all models are marginally calibrated.

In light of these observations, the difference in performance between bipartite and naive models must come from the fact that the bipartite model is able to detect temporal patterns that are relevant for the prediction of active links. Indeed, we find that for all the e-mail networks we consider, temporal layers (days) are classified either as week days or as weekend days (and holidays), so that it is more likely for any link to be active on a week day. Interestingly, this is all the temporal information required to be able to accurately predict whether a specific link is going to be active or not on a certain day.<sup>2</sup>

### B. Drug-drug interactions in cancer

Links in the drug-drug interaction network are of three different types: synergistic, antagonistic, and additive; we

<sup>2</sup>Note that our results do not depend on the number of latent dimensions allowed for the temporal layers  $L$ , since for  $L > 2$  we also find that temporal layers have  $\eta_{\ell\gamma} \neq 0$  only for two latent groups  $\gamma$ .

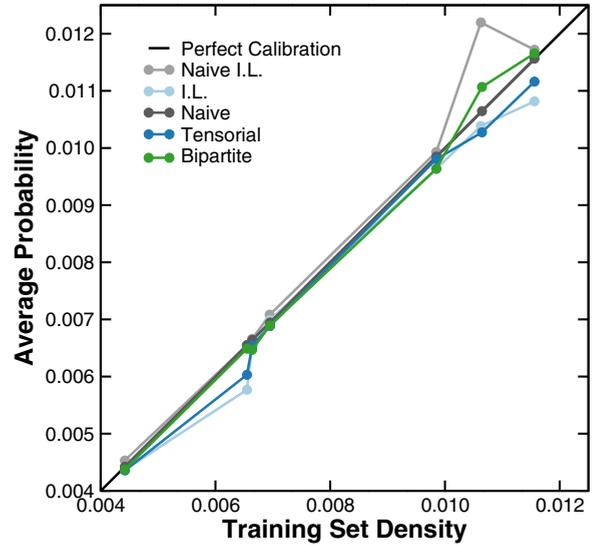


FIG. 4. Marginal calibration of the models for e-mail networks. Each line corresponds to a different model, and each point in a line corresponds to a different e-mail network (see Table I). Each point represents the average over the fivefold cross validation for a given model. Error bars are smaller than symbols.

trained the models considering the three types of interactions. However, because the interesting question is whether synergistic or antagonistic interactions can be predicted, we evaluated the performance of each model for each one of these two tasks. For instance, to evaluate the accuracy of a model at predicting synergistic interactions, we binarized model predictions into synergistic and nonsynergistic. We then computed the metrics over this binary outcome as we did for e-mail networks [Figs. 5–7; Fig. S3 shows that all of the results below are qualitatively similar when training our models on networks with only two types of interactions: synergistic vs nonsynergistic interactions or antagonistic vs nonantagonistic interactions].<sup>3</sup>

Contrary to what we observed for the e-mail networks, we find that the tensorial model performs better than the bipartite model. Our results thus suggest that for this data set, grouping nodes (drugs) into groups summarizes more parsimoniously the information relevant for prediction. This is consistent with previous findings that show that mechanisms of action and target pathways of drugs are related to the effect they display when combined with other drugs, an information that is best captured by node memberships than by link memberships [15].

<sup>3</sup>Due to the sparsity of observations in these networks many interactions were never observed in the training sets, and thus no group memberships could be assigned to the links  $(e_{ij})$  corresponding to those interactions. We solved this cold start problem by, at each iteration, assigning them the average membership of the observed interactions  $\zeta_e = \langle \zeta_f \rangle_{f \in G^O}$ . Analogously, if a node  $i$  had no observed interactions in the training set, at each iteration we set its membership vector as the average of membership vectors for nodes with observed interactions  $\theta_i = \langle \theta_k \rangle_{k \in G^O}$ .

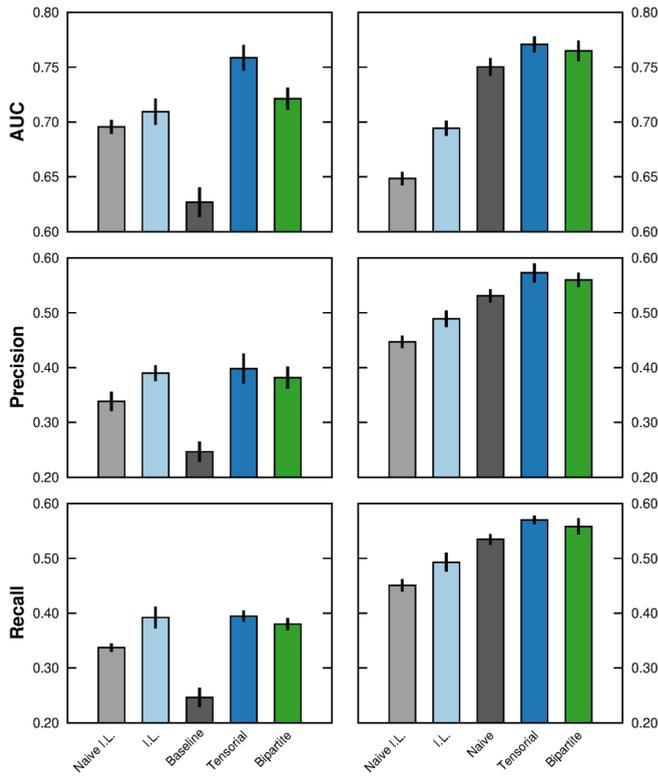


FIG. 5. Predictive performance of the models for drug-drug interaction networks. We show results for the prediction of antagonistic interactions (left column) and synergistic interactions (right column). Top: AUC statistic. Middle: precision. Bottom: recall. Each bar represents the average of the fivefold cross validation for a given model, with error bars (shown as a vertical line for clarity) representing the standard error of the mean.

Interestingly, we observe differences in performance at detecting antagonistic and synergistic interactions. For the synergistic interaction network, we find that the tensorial model consistently outperforms the bipartite and baseline models in all metrics (AUC, precision, and recall), although its marginal calibration is slightly worse than that of the other models. For antagonistic interactions, the tensorial model also performs better than the bipartite and baseline models in terms of AUC. However, the tensorial model has a precision and recall that are similar to those of the independent-layers baseline model. The generalized decrease in precision and recall with respect to synergistic network does not come as a surprise since none of the models is perfectly calibrated for probabilities lower than the density of the training set (Fig. 6). In fact, we observe that the fraction of antagonistic interactions for which  $P[r_{ijk} = 1] < T$  is larger than desired. As a result, some antagonistic interactions are counted as nonantagonistic interactions in terms of precision and recall. This effect is exacerbated by the fact that, due to the sparsity of the network, a large fraction of interactions are assigned low probability values by all of the models.

The fact that the independent-layer model has prediction and recall values similar to those of the tensorial model can be explained by the fact that antagonistic interactions are more localized to specific layers than synergistic

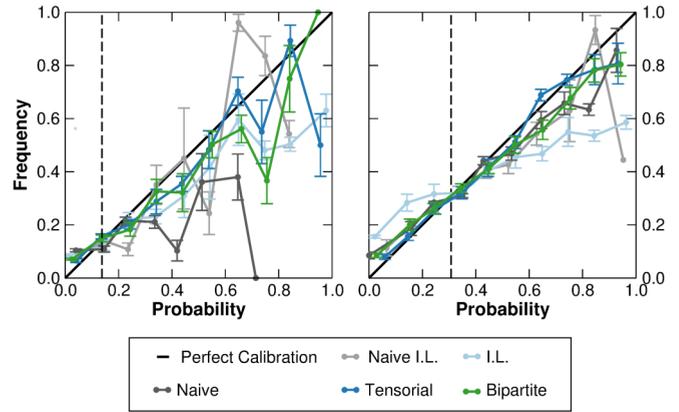


FIG. 6. Probabilistic calibration of the models for drug-drug interaction networks. We show results for the prediction of antagonistic interactions (left column) and synergistic interactions (right column). Each point in each line represents the average of the fivefold cross validation for a given model. Error bars represent the standard error of the mean. The vertical dashed lines show the density of each training set.

interactions are (see Fig. S6). This situation makes it easier for the independent-layer baseline model to make more accurate predictions for these layers. Note, however, that if more information on antagonistic interactions was available, the performance of the tensorial model would likely be comparable to that of the synergistic case.

VI. DISCUSSION

We have presented two mixed-membership multilayer network models that can be applied to any multilayer networks,

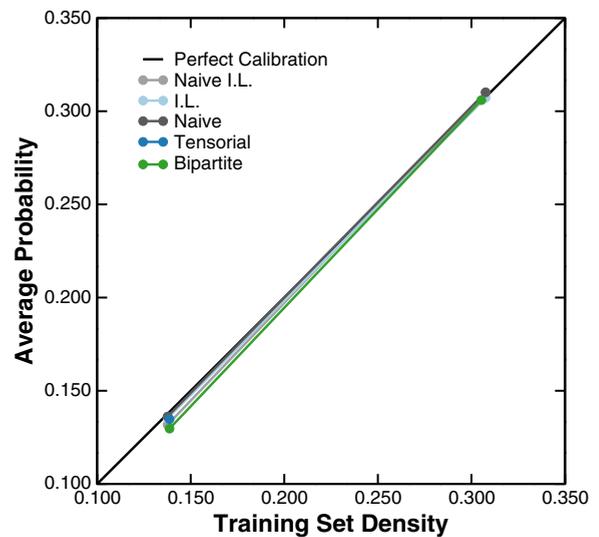


FIG. 7. Marginal calibration of the models for drug-drug interaction networks. For each of the models we consider (see legend) we plot the average probability for links being of a certain type (antagonistic or synergistic) with respect to the density of links of that type in the training set. Each point represents the average over the five training sets for a given model. Error bars are smaller than symbols.

with layers representing temporal snapshots of the interactions or different contexts for the interactions. By extending the mixed-membership paradigm to the layers themselves, and by not making any prior assumption about them, our models can detect and take advantage of interlayer correlations in the network of interactions to make better predictions. As a result, both our multilayer models outperform the baseline models in almost all the studied cases, except for the cases in which information is too sparse for the multilayer model to recover unobserved interactions with precision.

Importantly, none of the models we present—the tensorial node-based model nor the bipartite link-based model—is intrinsically better than the other; however, they can hold clues as to the mechanisms that are predictive of interaction types. Our results precisely illustrate this fact. We find that the bipartite model works better for e-mail networks in which the communication between pairs of users (links), rather than the users themselves, together with their temporal evolution are the relevant description unit for prediction. This could be due to the fact that we are analyzing communication at a rather small scale (people working within the same unit of an organization), and it is possible that a node-based model could be better for communication between users at a larger scale. Moreover, as the network grows the number of  $\theta$  parameters for the tensorial model scales linearly with the number of nodes, whereas the number of  $\eta$  parameters for the bipartite model scales quadratically. In really big networks it is then plausible that the tensorial becomes more parsimonious.

Conversely, our results show that for the drug-drug interaction network the relevant unit of description are drugs (nodes). This is consistent with the fact that the action-target mechanism that determines how a drug will interact with another one; this information is encapsulated in the node (and its observed interactions). The use of the interactions of nodes in different cancer types (layers) boosts our ability to predict the type of type-dependent interactions more precisely. In contrast, the description of these networks in terms of drug pair interactions completely misses the drug-specific information that is relevant for prediction in this context.

Our results unambiguously show that using the information of the interactions on other layers helps obtain better models. Remarkably, the flexibility of the models we propose make this approach suitable to analyze multilayer networks in any context. A natural step to further improve the model and prediction accuracy would be to include auxiliary data (i.e., metadata such as node or link attributes) into the modeling process. This problem has just started being explored in the literature [10,32,33], so there is no general framework on how to introduce auxiliary data into the inference process yet. Nonetheless, recent results show that single-layer mixed-membership models are suitable models to incorporate specific types of auxiliary data into the inference process without adding methodological complexity [34], thus opening the window to developing general inference frameworks that consider different types of metadata also in multilayer contexts.

Finally, the models we have presented offer an alternative approach toward the problem of finding the number of effective layers in a multilayer network [23,35]. While in this

paper we have focused solely on showing the suitability of our approach for multilayered data, the simultaneous modeling of nodes or links and layers we propose has the potential to result in better methods to summarize and extract information from multilayer data.

## ACKNOWLEDGMENTS

This work was supported by the Spanish Ministerio de Economía y Competitividad (MINECO) Grant No. FIS2016-78904-C3-1-P, and European Union FET Grant No. 317532 (MULTIPLEX). The authors acknowledge the computer resources at MinoTauro and the technical support provided by the Barcelona Supercomputing Center (FI-2015-3-0024, FI-2016-3-0026). The authors acknowledge AstraZeneca UK Limited, Sanger and Sage Bionetworks-DREAM for providing the cancer drug interaction data.

## APPENDIX A: DERIVATION OF THE EXPECTATION MAXIMIZATION EQUATIONS FOR THE T-MBM

In the tensorial mixed-membership stochastic block model, we assign membership vectors  $\theta_{i\alpha}$ ,  $\eta_{\ell\gamma}$  to each node  $i$  and each layer  $\ell$ , respectively. These membership vectors are properly normalized, therefore represent the probability that each node or layer belongs to a specific node or layer group:

$$\forall i : \sum_{\alpha=1}^K \theta_{i\alpha} = 1, \quad \forall \ell : \sum_{\gamma=1}^L \eta_{\ell\gamma} = 1. \quad (\text{A1})$$

Because we consider that links can not take different values  $r \in R$ , to ensure that each observed interaction has probability 1 of receiving any rating, we normalize probability matrices  $p_{\alpha\beta\gamma}(r)$

$$\forall \alpha, \beta, \gamma : \sum_{r \in R} p_{\alpha\beta\gamma}(r) = 1. \quad (\text{A2})$$

Note that if  $R = \{0, 1\}$ , then  $p_{\alpha\beta\gamma}(0) = 1 - p_{\alpha\beta\gamma}(1)$ .

We maximize the likelihood (3) as a function of  $\theta$ ,  $\eta$ ,  $\mathbf{p}$  using an expectation maximization (EM) algorithm. We start with a standard variational and use Jensen's inequality  $\log \bar{x} \geq \overline{\log x}$  in order to transform the logarithm of a sum into a sum of logarithms

$$\begin{aligned} & \log P(G^O | \theta, \eta, \mathbf{p}) \\ &= \sum_{(ij\ell) \in G^O} \log \sum_{\alpha\beta\gamma} \theta_{i\alpha} \theta_{j\beta} \eta_{\ell\gamma} p_{\alpha\beta\gamma}(r_{ij\ell}) \\ &= \sum_{(ij\ell) \in G^O} \log \sum_{\alpha\beta\gamma} \omega_{ij\ell}(\alpha\beta\gamma) \frac{\theta_{i\alpha} \theta_{j\beta} \eta_{\ell\gamma} p_{\alpha\beta\gamma}(r_{ij\ell})}{\omega_{ij\ell}(\alpha\beta\gamma)} \\ &\geq \sum_{(ij\ell) \in G^O} \sum_{\alpha\beta\gamma} \omega_{ij\ell}(\alpha\beta\gamma) \log \frac{\theta_{i\alpha} \theta_{j\beta} \eta_{\ell\gamma} p_{\alpha\beta\gamma}(r_{ij\ell})}{\omega_{ij\ell}(\alpha\beta\gamma)}. \quad (\text{A3}) \end{aligned}$$

Here we have introduced the auxiliary variable  $\omega_{ij\ell}(\alpha\beta\gamma)$ , which is the estimated probability that a given link's type  $r_{ij\ell}$  is due to  $i$ ,  $j$ , and  $\ell$  belonging to groups  $\alpha$ ,  $\beta$ , and  $\gamma$  respectively. Note in the expression above, equality holds

when

$$\omega_{ij\ell}(\alpha\beta\gamma) = \frac{\theta_{i\alpha}\theta_{j\beta}\eta_{\ell\gamma}p_{\alpha\beta\gamma}(r_{ij\ell})}{\sum_{\alpha'\beta'\gamma'}\theta_{i\alpha'}\theta_{j\beta'}\eta_{\ell\gamma'}p_{\alpha'\beta'\gamma'}(r_{ij\ell})}. \quad (\text{A4})$$

This is precisely the equation for the expectation step.

For the maximization step, we derive update equations for the parameters  $\theta$ ,  $\eta$ ,  $\mathbf{p}$  by taking derivatives of the log likelihood (A3). Including Lagrange multipliers for the normalization constraints (A1), we obtain for  $\theta_{i\alpha}$

$$\theta_{i\alpha} = \frac{\sum_{j\ell\in\partial i}\sum_{\beta\gamma}\omega_{ij\ell}(\alpha\beta\gamma)}{\sum_{j\ell\in\partial i}\sum_{\alpha\beta\gamma}\omega_{ij\ell}(\alpha\beta\gamma)} = \frac{\sum_{j\ell\in\partial i}\sum_{\beta\gamma}\omega_{ij\ell}(\alpha\beta\gamma)}{d_i}, \quad (\text{A5})$$

where  $\partial i = \{j, \ell | (ij\ell) \in G^O\}$  and  $d_i = |\partial i|$  is the degree of node  $i$  in all the layers for any type of link. Similarly, for  $\eta_{\ell\gamma}$  we obtain

$$\eta_{\ell\gamma} = \frac{\sum_{ij\in\partial\ell}\sum_{\alpha\beta}\omega_{ij\ell}(\alpha\beta\gamma)}{\sum_{ij\in\partial\ell}\sum_{\alpha\beta\gamma}\omega_{ij\ell}(\alpha\beta\gamma)} = \frac{\sum_{ij\in\partial\ell}\sum_{\alpha\beta}\omega_{ij\ell}(\alpha\beta\gamma)}{d_\ell}, \quad (\text{A6})$$

where  $\partial\ell = \{i, j | (ij\ell) \in G^O\}$  and  $d_\ell = |\partial\ell|$  is the number of observed links of any type in layer  $\ell$ .

Finally, including a Lagrange multiplier for (A2), we have for  $p_{\alpha\beta\gamma}(r)$

$$p_{\alpha\beta\gamma}(r) = \frac{\sum_{(ij\ell)\in G^O|_{r_{ij\ell}=r}}\omega_{ij\ell}(\alpha\beta\gamma)}{\sum_{(ij\ell)\in G^O}\omega_{ij\ell}(\alpha\beta\gamma)}. \quad (\text{A7})$$

## APPENDIX B: DERIVATION OF THE EXPECTATION MAXIMIZATION EQUATIONS FOR THE B-MBM

As in the tensorial model, we assign normalized membership vectors  $\zeta_{e\alpha}$ ,  $\eta_{\ell\gamma}$  to links and layers, respectively. We also consider probability matrices  $p_{\alpha\gamma}(r)$  that are as well normalized ( $\sum_{r\in R}p_{\alpha\gamma}(r) = 1$ ).

In order to maximize the likelihood, we again use Jensen's inequality to transform the logarithm of a sum into a sum of logarithms and introduce an auxiliary variable  $\phi_{e\ell}(\alpha, \gamma)$ :

$$\begin{aligned} & \log P(G^O | \zeta, \eta, \mathbf{p}) \\ &= \sum_{(ij\ell)\in G^O} \log \sum_{\alpha\gamma} \zeta_{e\alpha}\eta_{\ell\gamma}p_{\alpha\gamma}(r_{e\ell}) \\ &= \sum_{(ij\ell)\in G^O} \log \sum_{\alpha\gamma} \phi_{e\ell}(\alpha, \gamma) \frac{\zeta_{e\alpha}\eta_{\ell\gamma}p_{\alpha\gamma}(r_{e\ell})}{\phi_{e\ell}(\alpha, \gamma)} \\ &\geq \sum_{(e\ell)\in G^O} \sum_{\alpha\gamma} \phi_{e\ell}(\alpha, \gamma) \log \frac{\zeta_{e\alpha}\eta_{\ell\gamma}p_{\alpha\gamma}(r_{e\ell})}{\phi_{e\ell}(\alpha, \gamma)}. \end{aligned} \quad (\text{B1})$$

where again the equality holds when

$$\phi_{e\ell}(\alpha, \gamma) = \frac{\zeta_{e\alpha}\eta_{\ell\gamma}p_{\alpha\gamma}(r_{e\ell})}{\sum_{\alpha'\gamma'}\zeta_{e\alpha'}\eta_{\ell\gamma'}p_{\alpha'\gamma'}(r_{e\ell})}, \quad (\text{B2})$$

giving us the update equation (B2) for the expectation step.

For the maximization step, we derive update equations for the parameters  $\zeta$ ,  $\eta$ ,  $\mathbf{p}$  by taken derivatives of the log likelihood (B1). Including Lagrange multipliers for the

normalization constraints, we obtain

$$\zeta_{e\alpha} = \frac{\sum_{\ell\in\partial e}\sum_{\gamma}\phi_{e\ell}(\alpha, \gamma)}{\sum_{\ell\in\partial e}\sum_{\alpha\gamma}\phi_{e\ell}(\alpha, \gamma)} = \frac{\sum_{\ell\in\partial e}\sum_{\beta\gamma}\phi_{e\ell}(\alpha, \gamma)}{d_e}, \quad (\text{B3})$$

where  $\partial e = \{\ell | (e\ell) \in G^O\}$  are the set of layers in which we observe link  $e_{ij}$  and  $d_e = |\partial e|$  is the total number of layers in which we observe link  $e_{ij}$ . Similarly,

$$\eta_{\ell\gamma} = \frac{\sum_{e\in\partial\ell}\sum_{\alpha}\phi_{e\ell}(\alpha, \gamma)}{\sum_{e\in\partial\ell}\sum_{\alpha\gamma}\phi_{e\ell}(\alpha, \gamma)} = \frac{\sum_{e\in\partial\ell}\sum_{\alpha}\phi_{e\ell}(\alpha, \gamma)}{d_\ell}, \quad (\text{B4})$$

where  $\partial\ell = \{e | (e\ell) \in G^O\}$  and  $d_\ell = |\partial\ell|$ . Finally, including a Lagrange multiplier for the normalization of  $p_{\alpha\gamma}(r)$ , we have

$$p_{\alpha\gamma}(r) = \frac{\sum_{(e\ell)\in G^O|_{r_{e\ell}=r}}\phi_{e\ell}(\alpha, \gamma)}{\sum_{(e\ell)\in G^O}\phi_{e\ell}(\alpha, \gamma)}. \quad (\text{B5})$$

Equations (B2)–(B5) are solved iteratively with an EM algorithm following the same procedure as in the tensorial model. The bipartite model also scales linearly with the size of the data set, but in this case the number of parameters of the model is  $IK + ML + |G^O|K \cdot L$ , where the number of links  $I \leq N \cdot (N - 1)/2$ , thus, even though it increases the number of parameters (number of nodes  $N$  is typically smaller than number of links  $I$ ), there is one dimension less to run over all observed links in all layers  $|G^O|$ .

## APPENDIX C: EXPERIMENTAL DETAILS

In the drug-drug interactions data set, we divided the continuous values of efficiency into three categories (synergistic, additive, and antagonistic) by setting two thresholds as suggested in the original experimental data. These thresholds are  $-20.0$  and  $20.0$ , so that interactions with an efficiency lower than  $-20.0$  are classified as antagonistic, those with an efficiency higher than  $20.0$  are classified as synergistic, and those in between are considered additive [30].

For both data sets, we validated our models using a fivefold cross-validation scheme. We first divided the data into five equal splits. Then for each fold we considered four splits as the training set to which we fitted the model, and the remaining split was kept as the test set on which we made predictions. For each fold, we repeated the fitting processes between 100 and 500 times with different random initializations, and averaged over all of them. This procedure yields better results than using a single initialization, even if the single initialization is the one with maximum likelihood (see Ref. [29]). The results we present throughout the paper correspond to the average over the results for the five folds.

In order to select the number of latent groups  $K$ ,  $J$ , and  $L$ , we analyzed the evolution of AUC, precision, and recall as a function of these numbers, always maintaining  $K = L$  and  $J = L$  (Figs. 8 and 9). We selected the smallest values

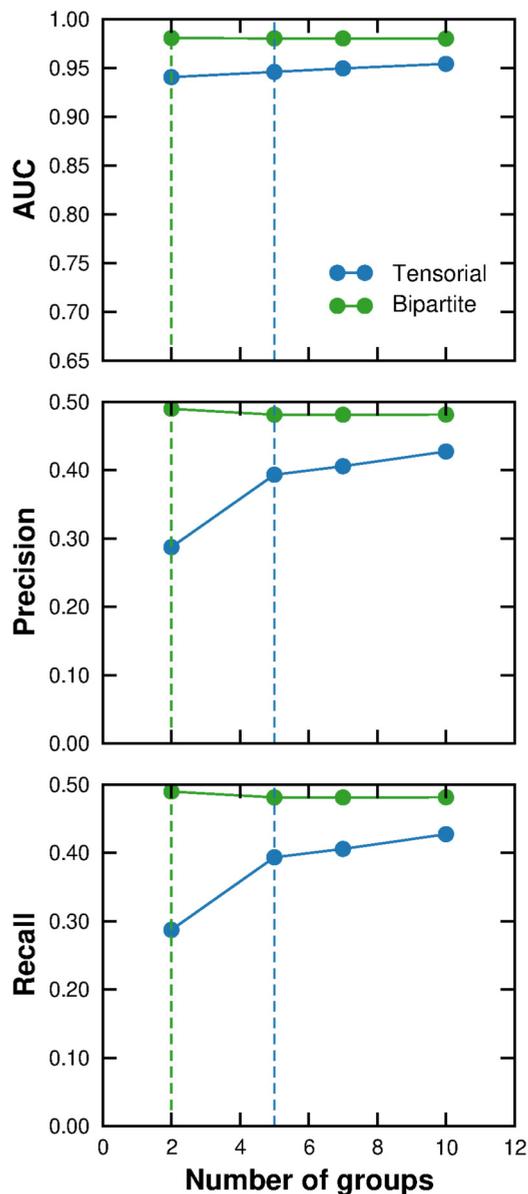


FIG. 8. Performance as a function of the number of groups for e-mail networks. Each point represents the average of the fivefold cross validation for a given model. The error bars (shown as a vertical line, which is small and not visible in some cases) represent the standard error of the mean. The dashed vertical lines indicate the selected values for  $K$  and  $L$  in the T-MBM, and  $J$  and  $L$  in the B-MBM. The results we show correspond to Unit 1.

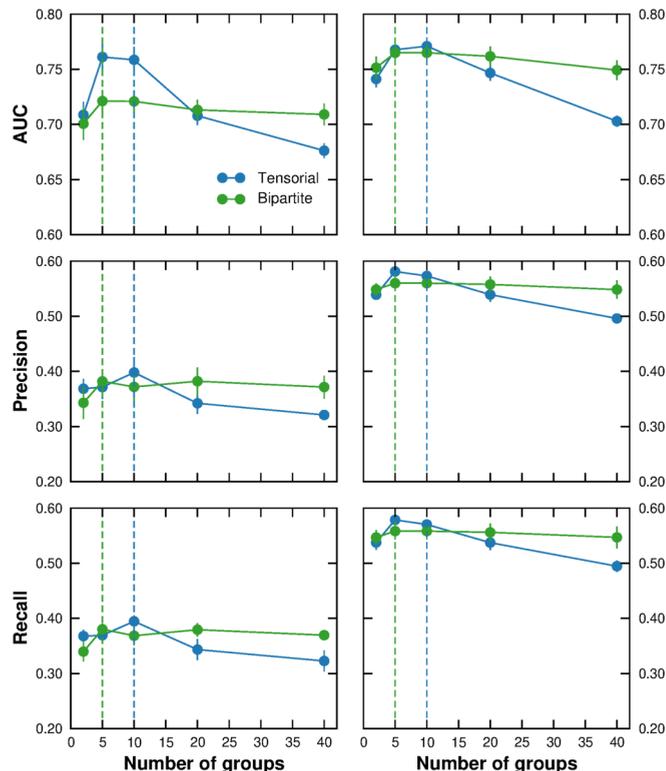


FIG. 9. Performance as a function of the number of groups for drug interaction networks. We show results for the prediction of antagonistic interactions (left column) and synergistic interactions (right column). Each point represents the average of the fivefold cross validation for a given model. The error bars (shown as a vertical line, which is small and not visible in some cases) represent the standard error of the mean. The dashed vertical lines indicate the selected values for  $K$  and  $L$  in the T-MBM, and  $J$  and  $L$  in the B-MBM.

for which the prediction accuracy was compatible with the maximal observed values. For e-mail networks, precision and recall still seem to increase slightly between  $J = L = 5$  and  $J = L = 10$ , but the computational cost becomes prohibitive for the number of repetitions that our validation requires (seven units, five folds, and 100–500 initializations per fold).

Ultimately, the selected values for the e-mail networks are  $K = L = 5$  for the T-MBM and  $J = L = 2$  for the B-MBM. The selected values for the drug interaction networks are  $K = L = 10$  for the T-MBM and  $J = L = 5$  for the B-MBM [36].

[1] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivela, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, Mathematical Formulation of Multilayer Networks, *Phys. Rev. X* **3**, 041022 (2013).

[2] D. S. Bassett and O. Sporns, Network neuroscience, *Nature Neurosci.* **20**, 353 (2017).

[3] S. Kéfi, V. Miele, E. A. Wieters, S. A. Navarrete, and E. L. Berlow, How structured is the entangled bank? The surprisingly simple organization of multiplex ecological networks leads to increased persistence and resilience, *PLoS Biology* **14**, e1002527 (2016).

[4] S. Pilosof, M. A. Porter, M. Pascual, and S. Kéfi, The multilayer nature of ecological networks, *Nature Ecology & Evolution* **1**, 0101 (2017).

[5] J. L. Iribarren and E. Moro, Impact of Human Activity Patterns on the Dynamics of Information Diffusion, *Phys. Rev. Lett.* **103**, 038702 (2009).

[6] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, Community structure in time-dependent, multi-scale, and multiplex networks, *Science* **328**, 876 (2010).

[7] L. Gauvin, A. Panisson, and C. Cattuto, Detecting the community structure and activity patterns of temporal networks:

- A non-negative tensor factorization approach, *PLoS ONE* **9**, e86028 (2014).
- [8] T. P. Peixoto, Inferring the mesoscale structure of layered, edge-valued, and time-varying networks, *Phys. Rev. E* **92**, 042807 (2015).
- [9] A. Ghasemian, P. Zhang, A. Clauset, C. Moore, and L. Peel, Detectability Thresholds and Optimal Algorithms for Community Structure in Dynamic Networks, *Phys. Rev. X* **6**, 031005 (2016).
- [10] A. Sapienza, A. Barrat, C. Cattuto, and L. Gauvin, Estimating the outcome of spreading processes on networks with incomplete information: A mesoscale approach, *Phys. Rev. E* **98**, 012317 (2018).
- [11] A. Li, S. P. Cornelius, Y. Y. Liu, L. Wang, and A. L. Barabási, The fundamental advantages of temporal networks, *Science* **358**, 1042 (2017).
- [12] R. Guimerà and M. Sales-Pardo, Missing and spurious interactions and the reconstruction of complex networks, *Proc. Natl. Acad. Sci. USA* **106**, 22073 (2009).
- [13] T. P. Peixoto, Hierarchical Block Structures and High-Resolution Model Selection in Large Networks, *Phys. Rev. X* **4**, 011047 (2014).
- [14] T. Vallès-Català, T. P. Peixoto, M. Sales-Pardo, and R. Guimerà, *Phys. Rev. E* **97**, 062316 (2018).
- [15] R. Guimerà and M. Sales-Pardo, A network inference method for large-scale unsupervised identification of novel drug-drug interactions, *PLoS Comput. Biol.* **9**, e1003374 (2013).
- [16] N. Rovira-Asenjo, T. Gumí, M. Sales-Pardo, and R. Guimerà, Predicting future conflict between team-members with parameter-free models of social networks, *Sci. Rep.* **3**, 1999 (2013).
- [17] A. Godoy-Lorite, R. Guimerà, C. Moore, and M. Sales-Pardo, Accurate and scalable social recommendation using mixed-membership stochastic block models, *Proc. Natl. Acad. Sci. USA* **113**, 14207 (2016).
- [18] E. M. Airolidi, D. M. Blei, S. E. Fienberg, and E. P. Xing, Mixed membership stochastic blockmodels, *J. Mach. Learn. Res.* **9**, 1981 (2008).
- [19] A. Godoy-Lorite, R. Guimerà, and M. Sales-Pardo, Long-term evolution of e-mail networks: Statistical regularities, predictability and stability of social behaviors, *PLoS ONE* **11**, e0146113 (2016).
- [20] P. W. Holland, K. B. Laskey, and S. Leinhardt, Stochastic blockmodels: First steps, *Soc. Networks* **5**, 109 (1983).
- [21] Q. Han, K. S. Xu, and E. M. Airolidi, Consistent estimation of dynamic and multi-layer block models, in *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, 2015, pp. 1511–1520.
- [22] C. De Bacco, E. A. Power, D. B. Larremore, and C. Moore, Community detection, link prediction, and layer interdependence in multilayer networks, *Phys. Rev. E* **95**, 042317 (2017).
- [23] N. Stanley, S. Shai, D. Taylor, and P. J. Mucha, Clustering Network Layers with the Strata Multilayer Stochastic Block Model, *IEEE Trans. Network Sci. Eng.* **3**, 95 (2016).
- [24] W. Jeffries, MMSBM for sparks, <https://github.com/billjeffries/mixMemRec>, 2017.
- [25] D. M. Dunlavy, T. G. Kolda, and E. Acar, Temporal link prediction using matrix and tensor factorizations, *ACM Trans. Knowl. Discov. Data* **5**, 10 (2011).
- [26] A. Schein, J. Paisley, D. M. Blei, and H. Wallach, Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15 (ACM, New York, 2015), pp. 1045–1054.
- [27] A. Schein, M. Zhou, D. M. Blei, and H. Wallach, Bayesian poisson tucker decomposition for learning the structure of international relations, in *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16 (JMLR.org, 2016), pp. 2810–2819.
- [28] T. P. Peixoto and M. Rosvall, Modelling sequences and temporal networks with dynamic community structures, *Nat. Commun.* **8**, 582 (2017).
- [29] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.99.032307> for additional figures and computational details not shown in the main text.
- [30] M. P. Menden, D. Wang, Y. Guan, M. Mason, B. Szalai, K. C. Bulusu, T. Yu, J. Kang, M. Jeon, R. Wolfinger, T. Nguyen, M. Zaslavskiy, AstraZeneca-Sanger Drug Combination DREAM Consortium, I. S. Jang, Z. Ghazoui, M. E. Ahsen, R. Vogel, E. C. Neto, T. Norman, E. K. Y. Tang, M. J. Garnett, G. Di Veroli, S. Fawell, G. Stolovitzky, J. Guinney, J. R. Dry, and J. Saez-Rodriguez (unpublished).
- [31] T. Gneiting, F. Balabdaoui, and A. E. Raftery, Probabilistic forecasts, calibration and sharpness, *J. R. Statist. Soc. B* **69**, 243 (2007).
- [32] M. E. J. Newman and A. Clauset, Structure and inference in annotated networks, *Nat. Commun.* **7**, 11863 (2016).
- [33] D. Hric, T. P. Peixoto, and S. Fortunato, Network Structure, Metadata, and the Prediction of Missing Nodes and Annotations, *Phys. Rev. X* **6**, 031038 (2016).
- [34] S. Cobo, A. Godoy-Lorite, M. Sales-Pardo, and R. Guimerà, Optimal prediction of decisions and model selection in social dilemmas using block models with metadata, *EPJ Data Sci.* **7**, 48 (2018).
- [35] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, Structural reducibility of multilayer networks, *Nat. Commun.* **6**, 6864 (2015).
- [36] The code is available at <https://github.com/seeslab/MMmultilayer.git>