Multilayer stochastic block models reveal the multilayer structure of complex networks

Toni Vallès-Català,[1] Francesco A. Massucci,[1] Roger Guimerà,[2,1,*] and Marta Sales-Pardo[1,†]

[1]*Departament d'Enginyeria Química, Universitat Rovira i Virgili, 43007 Tarragona, Catalonia*
[2]*Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Catalonia*

In complex systems, the network of interactions we observe between system's components is the aggregate of the interactions that occur through different mechanisms or layers. Recent studies reveal that the existence of multiple interaction layers can have a dramatic impact in the dynamical processes occurring on these systems. However, these studies assume that the interactions between systems components in each one of the layers are known, while typically for real-world systems we do not have that information. Here, we address the issue of uncovering the different interaction layers from aggregate data by introducing multilayer stochastic block models (SBMs), a generalization of single-layer SBMs that considers different mechanisms of layer aggregation. First, we find the complete probabilistic solution to the problem of finding the optimal multilayer SBM for a given aggregate observed network. Because this solution is computationally intractable, we propose an approximation that enables us to verify that multilayer SBMs are more predictive of network structure in real-world complex systems.

The development of tools for the analysis of real-world complex networks has significantly advanced our understanding of complex systems in fields as diverse as molecular and cell biology [1], neuroscience [2], biomedicine [3, 4], ecology [5, 6], economics [7], and sociology [8]. One of the main successes of the network approach has been to unravel the relationship between the modular organization of interactions within a complex system [9], and the function and temporal evolution of the system [10–13]. As a result, a large body of research has been devoted to the detection of the modular structure (or community structure) of complex networks, that is, to the division of the nodes of the network into densely connected subgroups [14].

Stochastic block models (SBMs) [15–17] are a class of probabilistic generative network models that provide a more general description of the (mesoscopic) group structure of real-world networks than modular models. In SBMs, nodes are assumed to belong to groups and connect to each other with probabilities that depend only on their group memberships. The simple mathematical form of SBMs has enabled not only the identification of generalized community structures in networks [17–26], but also to make network inference a predictive tool to detect missing and spurious links in empirical network data [27], to predict human decisions [28, 29] and the appearance of conflict in work teams [30], and for the identification of unknown interactions between drugs [31].

While these approaches have pushed forward our understanding of complex network structure, a limitation is that they rely on the premise that there is a single mechanism that describes the connectivity of the network, even though we know that real-world networks are often the result of processes occurring on different "layers" (for example, social networks comprise relationships that arise in the familiar layer, and others that arise in the professional layer) [32]. Moreover, it is increasingly clear that the multilayer structure of complex networks can have a dramatic impact on the dynamical processes

that take place on them [33–37]. Unfortunately, we often lack information about the different layers of interaction and can only observe projections of these multilayer interactions into an aggregate network in which all links are equivalent.

Here, we precisely address the problem of unraveling the underlying multilayer structure in real-world networks. To do so, we first introduce the family of multilayer SBMs that generalizes single-layer SBMs to situations where links arise in different layers and are aggregated through different mechanisms. Although there have been proposals to extend the concept of modularity to multilayer networks [38], ours represents a pioneering attempt to extend generative group-based models to multilayer systems, and to study those models rigorously using tools from statistical physics.

Second, we give the probabilistically complete solution to the problem of inferring the optimal multilayer SBM for a given aggregate network. Because this solution is computationally intractable, we propose an approximation which enables us to objectively address the question of whether an observed network is likely to be the projection of multiple layers. Our results suggest that many real-world networks are indeed projections.

I. MULTILAYER STOCHASTIC BLOCK MODELS

In our approach, nodes interact in different layers. In each one of these layers $\ell = 1, \ldots, L$ we define a SBM as follows: each node i belongs to a specific group σ_i^ℓ, and links between pairs of nodes belonging to groups α and β in layer ℓ exist with probability $q_{\alpha\beta}^\ell$. The observed adjacency matrix $A^\varnothing$ is an aggregate that results from the combination of the links in each of the layers, but where all information of the layers has been lost (Fig. 1). We call this model the multilayer SBM.

Here we consider the simplest case of two layers, $L = 2$. In such case, there are two combinations with a plausible physical interpretation: i) the *AND* combination of layers, in which $A_{ij}^\varnothing = 1$ if, and only if, (i, j) are connected in both layers (Fig. 1(a)); ii) the *OR* combination of layers, in which $A_{ij}^\varnothing = 1$ if (i, j) are connected in at least one layer (Fig. 1(b)). Indeed,

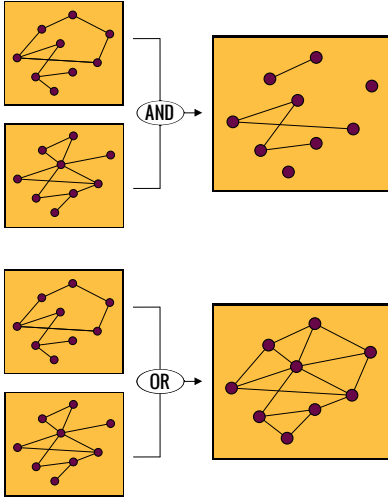arXiv:1411.1098v1 [physics.soc-ph] 4 Nov 2014

FIG. 1. In aggregated multilayer networks, several networks with the same nodes but different links are combined into an observed network $A^{\mathcal{O}}$ that has no information about the original layers. Two-layer networks can be aggregated using an AND combination of the layers or an OR combination of the layers. **(a)** The AND aggregation has link $A_{ij}^{\mathcal{O}} = 1$ if, and only if, i and j are connected in both layers. **(b)** The OR aggregation has link $A_{ij}^{\mathcal{O}} = 1$ if, and only if, i and j are connected in at least one layer.

each of these two mechanisms is plausible for specific scenarios. For example, the AND model is a plausible model for *in vivo* protein interactions, because in order for proteins to interact in the cell it is necessary for them to be capable of physically interacting (that is, to be linked in the layer of *in vitro* physical interactions) and to be expressed simultaneously in the same cellular compartment (that is, to be linked in the co-expression layer). The OR model is a plausible model for the effective on-line social network through which *memes* spread [39], because some people use Facebook to share memes, others use Twttier, and others use both.

In principle, we would like to identify which is the pair of partitions $(\mathcal{P}_1, \mathcal{P}_2)$ (in layers 1 and 2, respectively) that best describe the observed aggregate topology. The probabilistically complete way to solve this problem is to obtain the joint probability $P(\mathcal{P}_1, \mathcal{P}_2|A^{\mathcal{O}})$ that $\mathcal{P}_1$ and $\mathcal{P}_2$ are the true partitions of the nodes given the aggregate observed network. This distribution is given by

$$P(\mathcal{P}_1, \mathcal{P}_2|A^{\mathcal{O}}) \propto \qquad (1)$$
$$\int DQ_1 \int DQ_2 \, P(A^{\mathcal{O}}|Q_1, Q_2, \mathcal{P}_1, \mathcal{P}_2) P(Q_1, Q_2, \mathcal{P}_1, \mathcal{P}_2)$$

where Q_ℓ is a matrix whose elements $q_{\alpha\beta}^\ell$ represent the probability that a link exists between a pair of nodes belonging to groups α and β in layer ℓ, and $\int DQ_\ell \equiv \prod_{\alpha \leq \beta} \int_0^1 dq_{\alpha\beta}^\ell$ is the integral over all possible values of these probabilities.

This integral can be computed both for AND combinations and for OR combinations of the two layers; for simplicity, here we focus on the AND model and discuss the OR model in the Appendices. Because in a SBM each links is independent of each other and in the AND model a link has to be present in

both layers to appear in the observed aggregate network $A^{\mathcal{O}}$, the AND likelihood is

$$P_{\text{AND}}(A^{\mathcal{O}}|Q_1, Q_2, \mathcal{P}_1, \mathcal{P}_2) =$$
$$= \prod_{\left[\substack{\alpha \leq \beta \\ \gamma \leq \delta}\right]} \left(q_{\alpha\beta}^1 q_{\gamma\delta}^2\right)^{n_{\alpha\beta\gamma\delta}^1} \left(1 - q_{\alpha\beta}^1 q_{\gamma\delta}^2\right)^{n_{\alpha\beta\gamma\delta}^0}, \qquad (2)$$

where $n_{\alpha\beta\gamma\delta}^1$ is the number of links between pairs of nodes that are in groups α and β respectively in layer 1, and in groups γ and δ respectively in layer 2 ($n_{\alpha\beta\gamma\delta}^1 = \sum_{i<j} A_{ij}^{\mathcal{O}} \delta_{\sigma_i^1 \alpha} \delta_{\sigma_j^1 \beta} \delta_{\sigma_i^2 \gamma} \delta_{\sigma_j^2 \delta}$); and $n_{\alpha\beta\gamma\delta}^0$ is the number of non-links between such pairs of nodes ($n_{\alpha\beta\gamma\delta}^0 = \sum_{i<j}(1 - A_{ij}^{\mathcal{O}}) \delta_{\sigma_i^1 \alpha} \delta_{\sigma_j^1 \beta} \delta_{\sigma_i^2 \gamma} \delta_{\sigma_j^2 \delta}$).

Assuming a uniform distribution for the prior $P(Q_1, Q_2, \mathcal{P}_1, \mathcal{P}_2) = \text{const}$ [27] [40], we can plug Eq. (2) into Eq. (1) and integrate to find (Appendices)

$$P_{\text{AND}}(\mathcal{P}_1, \mathcal{P}_2|A^{\mathcal{O}}) \propto \qquad (3)$$
$$\sum_{\left[\substack{\{m_{rs}\} \\ m_{rs}=0,\ldots,n_{rs}^0}\right]} \prod_{r,s} \frac{(-)^{m_{rs}}}{(n_r^1 + m_r + 1)(n_s^1 + m_s + 1)} \binom{n_{rs}^0}{m_{rs}}$$

where, for clarity, we have used the shorthand $r \equiv \alpha\beta$ and $s \equiv \gamma\delta$, $m_r \equiv \sum_s m_{rs}$ and $m_s \equiv \sum_r m_{rs}$.

Given Eq. (3), which is the complete probabilistic description of the multilayer SBM, one could in principle find the partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ that maximize $P_{\text{AND}}(\mathcal{P}_1, \mathcal{P}_2|A^{\mathcal{O}})$. If this were possible, one would be able to perfectly disentangle the two SBMs responsible for the observed links, even though the observation did not have explicit information about the layers. It would also be possible to compare regular SBMs to multilayer SBMs to determine if a multilayer model is more or less appropriate to describe a given network. Unfortunately, the expression above becomes numerically intractable even for a small number of groups and therefore one needs to make approximations that simplify the problem.

II. LINK RELIABILITY WITH APPROXIMATE MULTILAYER STOCHASTIC BLOCK MODELS

We propose an approximation that makes it possible to work with multilayer SBMs. We start by noting that any multilayer SBM can be represented as a single-layer SBM (Fig. 2(a)) [41]. In the single-layer SBM, each group comprises the nodes that belong to the same pair of groups α in $\mathcal{P}_1$ and β in $\mathcal{P}_2$ in the multilayer SBM (and only those); we call the single-layer partition the *intersection partition*. Moreover, if group r in the intersection partition corresponds to groups α in $\mathcal{P}_1$ and β in $\mathcal{P}_2$, and group s in the intersection partition corresponds to groups γ in $\mathcal{P}_1$ and δ in $\mathcal{P}_2$, then the probability of connection in the single-layer SBM is $q_{rs}^{\text{AND}} = q_{\alpha\gamma}^1 q_{\beta\delta}^2$ (for simplicity, we again focus on the AND model and leave the OR model for the Appendices). This fully determines the single-layer SBM.

Here, we make the following approximation: we keep the information of the partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ in the intersection partition, but consider that the matrix elements q_{rs}^{AND}, while each
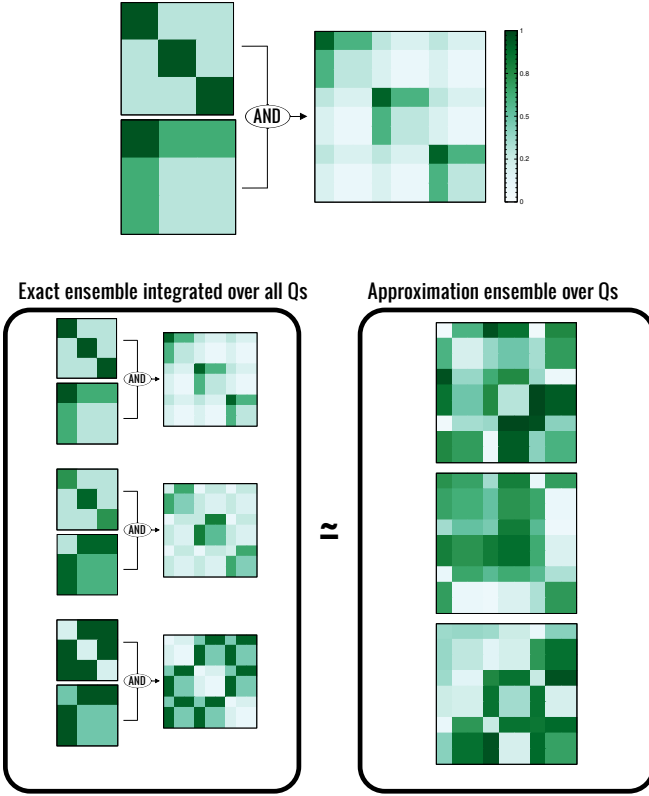
FIG. 2. Exact and approximate multilayer SBM ensembles. **(a)** Two independent single-layer SBMs aggregated using the AND mechanism. Each single-layer SBM is represented by its node-to-node connection probability matrix (represented by the shades of green; note that node ordering is different in each SMB). The aggregation of the two layers can also be represented as a single-layer SBM, in which each group comprises the nodes that belong to the same pair of groups α in layer 1 and β in layer 2; this is the *intersection partition* $\mathcal{P}_I$. Moreover, if group r in $\mathcal{P}_I$ corresponds to groups α in $\mathcal{P}_1$ and β in $\mathcal{P}_2$, and group s in $\mathcal{P}_I$ corresponds to groups γ in $\mathcal{P}_1$ and δ in $\mathcal{P}_2$, then the probability of connection in the single-layer SBM is $q_{rs}^{\text{AND}} = q_{\alpha\gamma}^1 q_{\beta\delta}^2$. **(b)** For a fixed pair of partitions $\mathcal{P}_1$ and $\mathcal{P}_2$, we integrate over the ensemble of all possible probability matrices Q_1 and Q_2 (Eq. (3)). For each pair (Q_1, Q_2), the resulting q_{rs}^{AND} are highly correlated. In our approximation, we assume that the elements q_{rs}^{AND} are randomly drawn and independent of each other.

being the result of the product of two factors, are all independent of each other (see Fig. 2(b)). Since this approximation is equivalent to integrating separately every term with a different $(\alpha, \beta, \gamma, \delta)$ combination in Eq. (2), it follows that the integrated likelihood depends exclusively on the intersection partition. In other words, within this approximation all pairs of partitions $(\mathcal{P}_1, \mathcal{P}_2)$ with the same intersection partition $\mathcal{P}_I$ are equally likely, and it is not possible anymore to uniquely determine the multilayer SBM that best describes the observed topology.

Despite this limitation, our approximation still enables us to address the fundamental question of whether real-world networks are better described by single-layer or multilayer models. Specifically, in what follows we compare the predictive power of single-layer and multilayer SBMs in the problem of detecting missing and spurious links in noisy networks [27]; we argue that, if (approximate) multilayer SBMs yield better predictions on real networks, then there is evidence to suggest that these networks are likely the outcome of multilayer processes (despite being observed as single-layer aggregates).

In the problem of assessing link reliability [27, 42], the goal is to compute the probability $P(A_{ij} = 1|A^{\varnothing})$ that a link between nodes i and j truly exists ($A_{ij} = 1$) given a noisy network observation $A^{\varnothing}$, which contains false positives (spurious interactions that are reported but do not truly exist) and false negatives (missing interactions that truly exist but are not reported). We call the probability $R_{ij} = P(A_{ij} = 1|A^{\varnothing})$ the *reliability* of the link. In general, for any set $\mathcal{M}$ of models (single-layer SBMs, AND-multilayer SBMs or OR-multilayer SBMs), the reliability is [27]

$$R_{ij}^{\mathcal{M}} = \frac{\int_{\mathcal{M}} dM P(A_{ij} = 1|M) P(A^{\varnothing}|M) P(M)}{Z}, \quad (4)$$

where Z is a normalization constant.

In the case of multilayer SBMs, the integral over the ensemble of models $\mathcal{M}$ requires: i) the integration over the connection probabilities Q_1 and Q_2 (akin to what we did to obtain Eq. (1)); ii) the sum over all pairs of partitions $\mathcal{P}_1$ and $\mathcal{P}_2$. Within our approximation, the first step can be carried out analytically but the second cannot (Appendices). However, always within our approximation, one can exploit the fact that the integral in Eq. (4) depends exclusively on the intersection partition $\mathcal{P}_I$ and map the sum over pairs of partitions onto a sum over a single partition. By doing so we obtain the following expression for the link reliability (see Appendices for the analogous expression for the OR model)

$$R_{ij}^{\text{AND}} = \quad (5)$$

$$\frac{1}{Z} \sum_{\mathcal{P}_I} \left(\frac{n_{\sigma_i \sigma_j}^1 + 1}{n_{\sigma_i \sigma_j} + 2} \cdot \frac{\sum_{k=n_{\sigma_i \sigma_j}^1 + 2}^{n_{\sigma_i \sigma_j} + 2} \frac{1}{k}}{\sum_{k=n_{\sigma_i \sigma_j}^1 + 1}^{n_{\sigma_i \sigma_j} + 1} \frac{1}{k}} \cdot D(\mathcal{P}_I) \cdot e^{-\mathcal{H}(\mathcal{P}_I)} \right)$$

where the sum is over all possible intersection partitions (that is, all single-level partitions), $n_{\alpha\beta}^1$ is the number of links between groups α and β in the intersection partition, $n_{\alpha\beta} = n_{\alpha\beta}^0 + n_{\alpha\beta}^1$ is the number of pairs of nodes in groups α and β, and $D(\mathcal{P}_I)$ is the number of pairs $(\mathcal{P}_1, \mathcal{P}_2)$ that have the same intersection partition $\mathcal{P}_I$ (see Appendices). The energy function $\mathcal{H}$ is

$$\mathcal{H}(\mathcal{P}_I) = \quad (6)$$

$$\sum_{\alpha \leq \beta} \left(\ln(n_{\alpha\beta} + 1) + \ln \binom{n_{\alpha\beta}}{n_{\alpha\beta}^1} - \ln \left(\sum_{k=n_{\alpha\beta}^1 + 1}^{n_{\alpha\beta} + 1} \frac{1}{k} \right) \right)$$

where the sum is over all distinct pairs of groups in $\mathcal{P}_I$.

As in [27], the expression for the link reliability (Eq. 5) is analogous to an ensemble average of an observable in statistical mechanics, giving $\mathcal{H}(\mathcal{P}_I)$ the meaning of an energy associated to a specific intersection partition. We can use a Markov Chain Monte Carlo algorithm to compute numerically R_{ij} (see Supplementary Material for details). As it turns out,
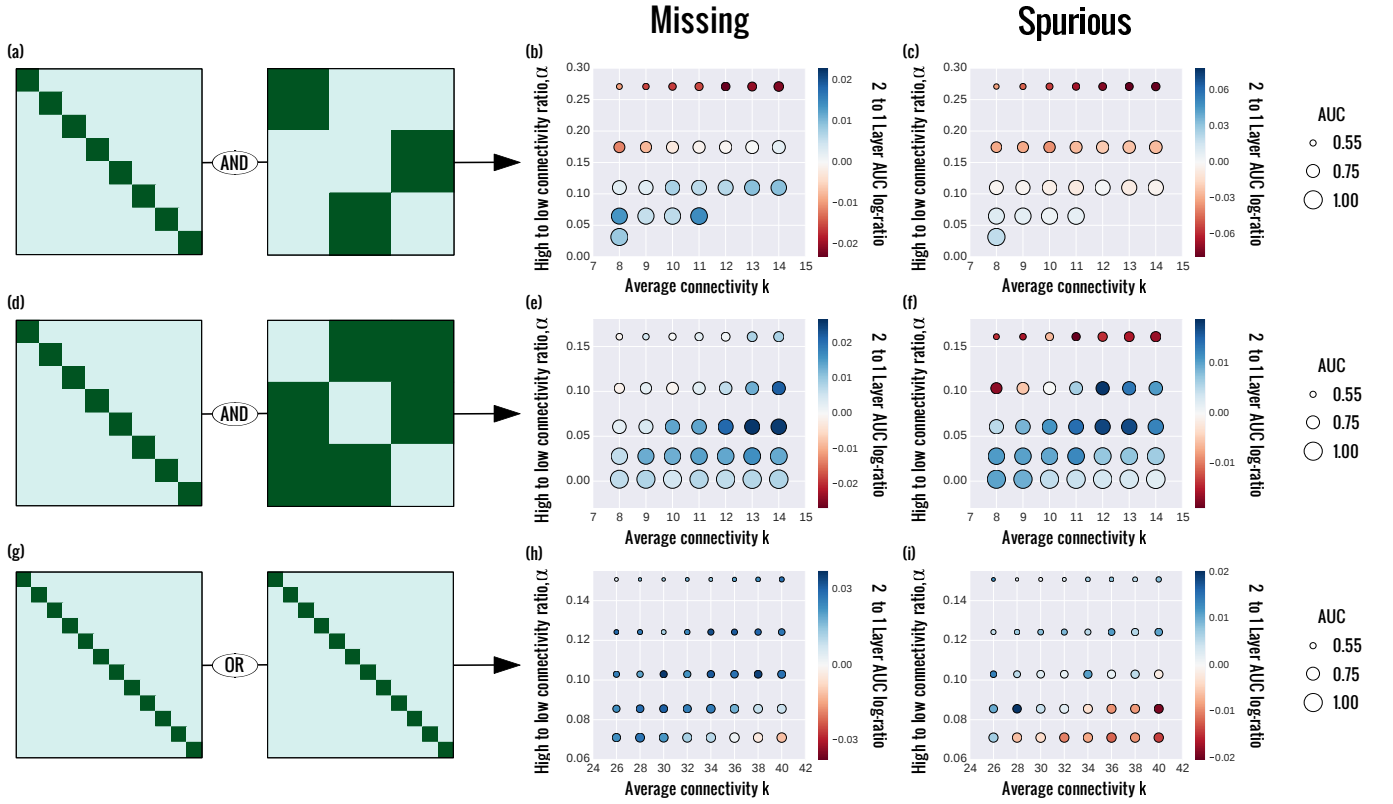
FIG. 3. Performance of missing and spurious link identification on synthetic aggregated 2-layer networks. Each row corresponds to a different collection of 2-layer SBMs, which are illustrated in **(a, d, g)**. Dark green corresponds to high connection probability hp and light green to low connection probability lp, and we generate synthetic networks varying two parameters: the high-to-low connectivity ratio $\alpha = lp/hp < 1$, and the average connectivity k (Appendices). We compare the performance (AUC) at detecting missing links **(b, e, h)** and spurious links **(c, f, i)** of the approximate multilayer SBM approach, AUC_{2L}, against that of the the single-layer SBM approach, AUC_{1L}. The size of the circles represents the AUC_{2L} of the multilayer approach. The color of the circles represents the logarithm of the ratio $\frac{AUC_{2L}}{AUC_{1L}}$, so that blue circles correspond to instances where the multilayer approach outperforms the single-layer approach, and conversely for red circles.

$\mathcal{H}(\mathcal{P}_I)$ is equal to the energy obtained assuming a single SBM (Eq. S2, [27]) plus a term that arises because of the fact that the probability matrix elements associated to the intersection SBM are the result of a product of two probabilities. In a Bayesian context, we can interpret this term and the degeneration $D(\mathcal{P}_i)$ as non-uniform priors for the intersection partitions.

III. VALIDATION OF LINK RELIABILITY ESTIMATION IN MODEL NETWORKS

Now that we are able to estimate link reliabilities using single-layer SBMs [27] and our approximation to two-layer (AND and OR) SBMs (Eq. (5)), we compare the performance of these approaches at detecting missing and spurious interactions. Our expectation is that if real-world networks are truly the result of the aggregation of multiple layers, assuming a two layer structure should result in higher accuracy.

To identify the limits of detectability of the 2-layer SBM model, we first construct a set of multilayer test networks that have a clearly differentiated block structure in each of two layers, and that are aggregated using the AND and OR models

(see Methods and Fig. 3). We consider the predictive power of each of the approaches at detecting [27, 42]: i) missing links (we remove a fraction f of the links and compute the fraction of times that a removed link has a higher reliability than a link not present in the original network, that is the AUC statistic); ii) spurious links (we add a fraction f of links and compute the fraction of times that an added link has a lower reliability than a link present in the original network, that is the AUC statistic).

For AND networks (Fig. 3(a-f)) we find that, for the detection of both missing and spurious links, the 2-layer approach outperforms the single-layer approach, especially: (i) when the number of distinct node groups in the intersection partition and the connectivity grow; (ii) for small or moderate noise levels (fraction of removed/added links. Only when the structure of the blocks becomes very blurry do we observe that the single-layer approach works better (but in this region all approaches do in fact work poorly).

For OR networks (Fig. 3(g-i)), the 2-layer approach again outperforms its single-layer counterpart in most situations. In this case, however, the largest improvements in performance happen for the hard cases with lower connectivity. This can be explained by noting that the OR model tends to generate
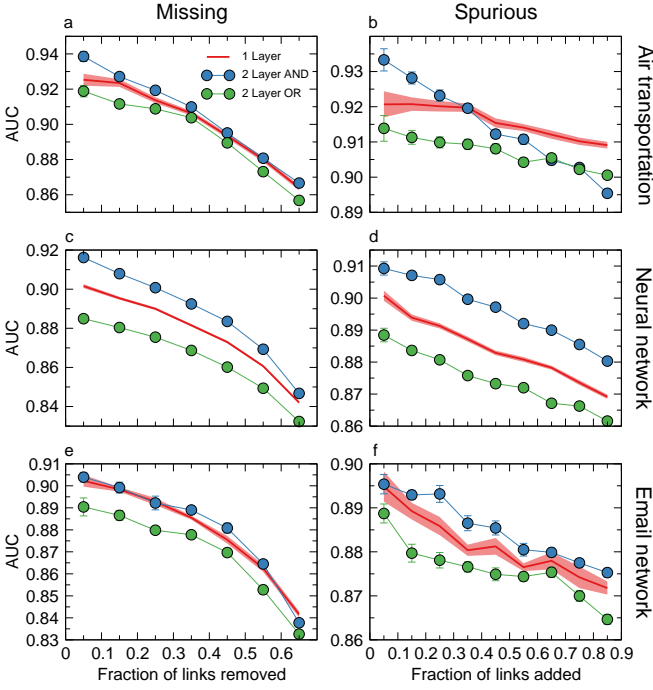
FIG. 4. Performance of missing and spurious link identification on real networks. To compare the performance of the different approaches at detecting missing links **(a, c, e)**, we randomly remove a fraction of the links (false negatives) from the real network and calculate the reliability of each unobserved link. Then we rank the links by decreasing score and calculate how often a removed link (false negative) has a higher reliability that a link that is truly non-existent in the real network (true negative). Analogously, to detect spurious links **(b, d, f)** we randomly add a fraction of links (false positives), calculate the reliability of the observed links, and calculate how often an added link (false positive) has a lower reliability that a link that is truly existent in the real network (true positive). We tested the analysis on three real networks: **(a, b)** the air transportation network in Eastern Europe; **(c, d)** the neural network of *(*C. elegans); **(e, f)** the email network within an organization.

very dense networks, whereas aggregate AND networks are sparser than the networks in each of the layers. Therefore, in general we expect the AND model to produce better results in real-world networks.

IV. MULTILAYER STOCHASTIC BLOCK MODELS ARE MORE PREDICTIVE FOR REAL NETWORKS

Finally, we compare the performance of the single-layer and multilayer approaches on three real-world networks: (i) the air transportation network in Eastern Europe [43]; (ii) the neural network of *C. elegans* [44]; and (iii) the email network within a university [45]. Our results show that the two-layer AND model provides a better description of these real-world networks since both missing and spurious interactions

are more accurately detected by the multilayer SBM approach consistently (the improvement is slight but, in most cases, significant), especially for low observational noise.

V. DISCUSSION

We have introduced the family of multilayer SBMs, which generalizes single-layer SBMs to situations where links arise in different layers and are aggregated through different mechanisms. We have also given the probabilistically complete solution to the problem of inferring the optimal multilayer SBM for a given aggregate network, and proposed a tractable approximation which enables us to objectively address the question of whether an observed network is best described as the projection of multiple layers or as a single layer. Our results suggest that many real-world networks are indeed projections.

Although, as mentioned above, there have been proposals to extend the concept of modularity to multilayer networks [38], ours represents a pioneering attempt to extend stochastic block models to multilayer systems. In this regard, it is important to stress that in this work we are concerned with the learning of multilayer models from aggregate networks where all information about the layers has been lost; in this sense, our work is different from previous attempts to do inference of stochastic block models on multigraphs where the layers themselves are observed [29].

Our work is also different from works on link prediction using latent feature models [46–48]. An important difference between latent feature approaches and ours is that the latent feature model considers that the probability of existence of a link is a function of the weighted sum of the interactions at the different layers; therefore, the latent feature model does not allow a physical interpretation of what each layer is and of how layers are combined. All in all, latent feature models are very well suited for the inference of unobserved links, but due to the intricacies of the model and the difficulty to interpret its "parameters," it is not clear whether they are appropriate to address the question of whether a real network is really the outcome of a multilayer process or not (and may it may also be prone to overfitting when observational data is noisy).

Our multilayer SBM is the simplest group-based multilayer model one can propose. We believe that its detailed analysis will open the door to better understand the structure of real complex networks.

[1] A.-L. Barabási and Z. N. Oltvai, Nat. Rev. Genet. **5**, 101 (2004).

[2] E. Bullmore and O. Sporns, Nat. Rev. Neurosci. **10**, 186 (2009).

[3] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, Nat. Rev. Genet. **12**, 58 (2011).

[4] P. Csermely, T. Korcsmáros, H. J. M. Kiss, G. London, and R. Nussinov, Pharmacol. Therapeut. **138**, 333 (2013).

[5] R. M. Thompson, U. Brose, J. A. Dunne, R. O. Hall, S. Hladyz, R. L. Kitching, N. D. Martinez, H. Rantala, T. N. Romanuk, D. B. Stouffer, and J. M. Tylianakis, Trends Ecol. Evol. **27**, 689 (2012).

[6] R. P. Rohr, S. Saavedra, and J. Bascompte, Science **345**, 416 (2014).

[7] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, Science **325**, 422 (2009).

[8] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, Science **323**, 892 (2009).

[9] M. E. J. Newman, Nat. Phys. **8**, 25 (2011).

[10] R. Guimerà and L. A. N. Amaral, Nature **433**, 895 (2005).

[11] A. Arenas, A. Díaz-Guilera, and C. J. Pérez-Vicente, Phys. Rev. Lett. **96**, art. no. 114102 (2006).

[12] R. Guimerà, M. Sales-Pardo, and L. Amaral, Nature Phys. **3**, 63 (2007).

[13] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, Nature **466**, 761 (2010).

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

[15] H. C. White, S. A. Boorman, and R. L. Breiger, Am. J. Sociol. **81**, 730 (1976).

[16] P. W. Holland, K. B. Laskey, and S. Leinhardt, Soc. Networks **5**, 109 (1983).

[17] K. Nowicki and T. A. B. Snijders, J. Am. Stat. Assoc. **96**, 1077 (2001).

[18] B. Karrer and M. Newman, Phys. Rev. E **83**, 016107 (2011).

[19] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Phys. Rev. Lett. **107**, 065701 (2011).

[20] M. N. Schmidt and M. Mørup, IEEE Signal Processing Magazine **30**, 110 (2013).

[21] T. P. Peixoto, Phys. Rev. Lett. **110**, 148701 (2013).

[22] T. P. Peixoto, Phys. Rev. X **4**, 011047 (2014).

[23] T. P. Peixoto, Phys. Rev. E **89**, 012804 (2014).

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

[25] C. Aicher, A. Z. Jacobs, and A. Clauset, J. Compl. Netw. (in press, 2014).

[26] X. Yan, C. Shalizi, J. E. Jensen, F. Krzakala, C. Moore, L. Zdeborová, P. Zhang, and Y. Zhu, J. Stat. Mech.: Theor. Exp. , P05007 (2014).

[27] R. Guimerà and M. Sales-Pardo, Proc. Natl. Acad. Sci. U. S. A. **106**, 22073 (2009).

[28] R. Guimerà and M. Sales-Pardo, PLOS ONE **6**, e27188 (2011).

[29] R. Guimerà, A. Llorente, E. Moro, and M. Sales-Pardo, PLOS ONE **7**, e44620 (2012).

[30] N. Rovira-Asenjo, T. Gumí, M. Sales-Pardo, and R. Guimerà, Sci. Rep. **3**, 1999 (2013).

[31] R. Guimerà and M. Sales-Pardo, PLOS Comput. Biol. **9**, e1003374 (2013).

[32] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, J. Complex Netw. **2**, 203 (2014).

[33] R. Morris and M. Barthelemy, Phys. Rev. Lett. **109**, 128703 (2012).

[34] F. Radicchi and A. Arenas, Nat. Phys. **9**, 717 (2013).

[35] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas, Phys. Rev. Lett. **110**, 028701 (2013).

[36] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, Phys. Rev. X **3**, 041022 (2013).

[37] M. De Domenico, G. S. Solé-Ribalta, A, , and A. Arenas, Proc. Natl. Acad. Sci. U.S.A. , 83518356 (2014).

[38] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, Science **328**, 876 (2010).

[39] L. Weng, F. Menczer, and Y.-Y. Ahn, Scientific Reports **3**, 2522 (2013).

[40] Other possibilities include choosing non-uniform priors for the connection probabilities [20–22] or different priors for the partitions [19–23, 26].

[41] The reverse is also true, so the possible network models one can generate with single-layer SBMs and multilayer SBMSs are, in fact, identical. However, it is important to note that each of them gives different weights to different models, so that a model that is relatively probable in the multilayer SBM family might be relatively rare in the single-layer SBM family, and vice versa.

[42] A. Clauset, C. Moore, and M. E. J. Newman, Nature **453**, 98 (2008).

[43] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, Proc. Natl. Acad. Sci. USA **102**, 7794 (2005).

[44] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, Philos. T. R. Soc. B. **314**, 1 (1986).

[45] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, art. no. 065103 (2003).

[46] K. Miller, M. I. Jordan, and T. L. Griffiths, in *Advances in Neural Information Processing Systems 22*, edited by Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta (Curran Associates, Inc., 2009) pp. 1276–1284.

[47] K. Palla, D. Knowles, and Z. Ghahramani, in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)* (2012) pp. 1607–1614.

[48] M. Kim and J. Leskovec, in *Advances in Neural Information Processing Systems 26*, edited by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Curran Associates, Inc., 2013) pp. 1385–1393.