



NETWORK INFERENCE BASED ON STOCHASTIC BLOCK MODELS: MODEL EXTENSIONS, INFERENCE APPROACHES AND APPLICATIONS

Toni Vallès Català

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

PhD THESIS

**Network inference based on
Stochastic Block Models:
Model extensions,
inference approaches
and applications**

Toni Vallès Català



UNIVERSITAT ROVIRA i VIRGILI

UNIVERSITAT ROVIRA I VIRGILI
NETWORK INFERENCE BASED ON STOCHASTIC BLOCK MODELS: MODEL EXTENSIONS, INFERENCE APPROACHES
AND APPLICATIONS
Toni Vallès Català

UNIVERSITAT ROVIRA I VIRGILI

PhD THESIS

**Network inference based on
Stochastic Block Models:
Model extensions,
inference approaches
and applications**

Author

Toni Vallès Català

Supervisors

Marta Sales-Pardo

Roger Guimerà

DEPARTMENT OF CHEMICAL ENGINEERING


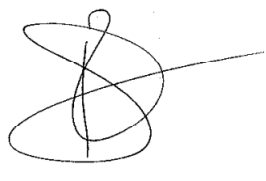
Tarragona, 2016

UNIVERSITAT ROVIRA I VIRGILI
NETWORK INFERENCE BASED ON STOCHASTIC BLOCK MODELS: MODEL EXTENSIONS, INFERENCE APPROACHES
AND APPLICATIONS
Toni Vallès Català



WE STATE that the present study, entitled “Network inference based on Stochastic Block Models: Model extensions, inference approaches and applications”, presented by Toni Vallès Català for the award of the degree of Doctor, has been carried out under our supervision at the Department of Chemical Engineering of this university.

Doctoral Thesis Supervisor/s



Dr. Roger Guimerà Manrique

Dra. Marta Sales Pardo

Tarragona, 2nd September 2016

UNIVERSITAT ROVIRA I VIRGILI
NETWORK INFERENCE BASED ON STOCHASTIC BLOCK MODELS: MODEL EXTENSIONS, INFERENCE APPROACHES
AND APPLICATIONS
Toni Vallès Català

Agraïments

El procés de quatre anys dedicats a aquesta tesi ha estat possible gràcies a les persones que m'envolten:

En primer lloc, voldria agrair als meus supervisors Roger i Marta per confiar en mi i ensenyarme multitud d'eines i d'idees: apart de conceptes tècnics de física i diverses metodologies, també m'han mostrat com treballar rigurosament en ciència i com comunicar millor els resultats obtinguts. El seu treball i esforç rigurós, humil i continuat son un exemple de com fer bé les coses, sempre dedicant gran part del seu temps a la formació de nous doctorats.

M'ha encantat formar part del grup SEES Lab, on agraeixo a tots i cadascun dels seus membres, sense ells la feina s'hauria tornat més feixuga. A Núria, Arnau, Francesco, Manu, Toni Aguilar, Toñi, Oriol, Marc, Pedro i Sergio. Les nostres discussions sobre tot tipus de temes, tant científics com intrascendentals, m'han forjat una millor cultura i m'han plantejat preguntes des d'altres punts de vista. Dins d'aquest grup és fàcil venir a treballar cada matí, on els companys de feina es transformen paulativament en amics. També agrair a membres d'altres labs, que han contribuït en aquest ambient dins de la universitat (Janire, Joan, Àngel, Patricia, Carmen, Dani, Carlos, Victor, Alberto, Yonhara, Judith, Noelia, Sandra).

Moltes gràcies a Tiago i Borja per les seves enriquidores col·laboracions, demostrant que la distància no impedeix un bon intercanvi d'idees.

També voldria agrair a totes les persones que m'he anat trobant en congressos i escoles d'estiu. Pol, Oleguer, Mario, Guille, Fede, Massimo, Emmanuelle, Manlio, Manu, Bruno Pace, Bruno Medeiros, Mika, Milo, Marc Suñé. Les intenses jornades d'interessants xerrades necessiten l'oxigen que ens proporcionem entre companys durant els 'coffee breaks'.

Per acabar, res del que mai pugui fer no seria possible sense tot el que els meus pares i família m'han donat. Ells han permès la meva formació professional i personal

per poder optar a fer un doctorat sense oblidar els nostres humils principis. En especial, mai agrairé prou tot el que l'Alba fa per mi, dia a dia, amb la enorme paciència que comporta escoltar xerrades sobre xarxes i conviure amb algú que escriu una tesis.

Summary

Introduction

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 [6], neuroscience [11], biomedicine [5, 14], ecology [100, 85], economics [92], anatomy [23] and sociology [9]. One of the main successes of the complex networks approach has been to unravel the relationship between the modular organization of interactions within a complex system [66], and the function and temporal evolution of the system [37, 4, 42, 2]. 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 (partition) of the nodes of the network into densely connected subgroups [28].

Stochastic block models (SBMs) [105, 46, 70] are a class of probabilistic generative network models that provide a more general description of the large-scale 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 [70, 50, 18, 91, 74, 76, 75, 56, 3, 108], but also to make network inference a predictive tool to detect missing and spurious links in empirical network data [44], to predict human decisions [41, 39] and the appearance of conflict in work teams [86], and for the identification of unknown interactions between drugs [45].

While these approaches have pushed forward our understanding of complex network structure, there are some limitations: (i) it is not clear which inference methodology yields better predictions, (ii) they rely on the premise that there is a single mechanism that describes the connectivity of the network. The goal of this thesis is to develop novel inference approaches that will improve our understanding of complex systems, decide which inference models to use and apply it to real world problems.

Approaches to network inference with stochastic block models

The reliability of missing links is the likelihood that a link exists on a given network, it can be compute by $p(A_{ij} = 1 | A^{\mathcal{O}}) = \frac{\int_{\mathcal{M}} dM p(A_{ij}=1|M) p(A^{\mathcal{O}}|M) p(M)}{\int_{\mathcal{M}} dM p(A^{\mathcal{O}}|M) p(M)}$ [44], given an observed network $A^{\mathcal{O}}$ with an adjacency matrix A_{ij} and integrating over all the family of models \mathcal{M} .

The most popular inference approach to estimate this equation focuses on finding a single-point estimate for the most likely set of parameters [32, 13, 67, 18, 74, 75]. However, Bayesian inference theory suggests that the correct approach goes through integrating the whole ensemble of possible sets of parameters [49]; we will refer to *single-point* for the former method and *sampling* for the latter. The single-point method is mostly used because the likelihood distribution is usually peaked around the optimal single-point estimate, then the rest of non-optimal models are considered negligible. In this chapter we investigate to what extent sampling over several models improves the predictive power than considering the optimal set of parameters alone, hence quantifying the error at estimating the equation when applying the single-point method.

To adress this question we adopt the definition of SBM described in Ref. [74] that minimizes the entropy of SBM to find the most likely set of parameters, and we compare it with an approach that samples several sets of likely parameters instead of scoping for the optimal one. Although the approach is based on Ref. [44], it was never applied before on the model defined in Ref. [74]. Then we apply both inference approaches on four real-world networks to tackle the same problem: the accuracy at predicting missing links. Such accuracy measure is consistent with the model, since we observed better accuracies for those partitions that are more likely. We discovered that the sampling approach significantly outperforms the single-point, suggesting that the error at estimating the equation when applying the single-point method is not negligible.

We also examine the predictive power of different SBM specializations found in the literature (degree-corrected [50, 73] and hierarchical nested [76]). We observed that the simple sampling approach predicts better missing links than the single-point for each of the different model specializations. Furthermore, we noticed that the degree-corrected version underperformed on some real world networks, suggesting that this specialization should not always be used.

Suture fusion in normal and pathological development is constrained by the network architecture of the human skull

To show the power of the inference methodology we apply it to a novel problem.

Sutures fuse as part of the normal developmental process of the skull when taking place at the right time. However, deviations from the normal process of suture pattern formation in the human skull usually cause birth defects, such as cleft palate and craniosynostosis. We still do not know which factors predispose some sutures but no others to fuse pathologically or to not form at all.

Here we address this question by modeling the skull as a network in which nodes and links formalize bones and their articulations. Anatomical network models have been used before, for example to model the growth of human skull bones [25]. We use the reliability formalism [44] to infer the susceptibility of craniofacial sutures to be lost in pathological conditions.

We found that sutures that normally fuse have significantly lower reliability scores than those that do not, which is in agreement with our hypothesis that during normal development there is a tendency to lose articulations that are topologically rare in the newborn skull. Interestingly, we discovered that sutures associated with pathological conditions have significantly lower reliability scores than sutures that are not, which shows that sutures associated to pathological conditions are also unexpected from a topological point of view.

Multilayer stochastic block models reveal the multilayer structure of complex networks

In complex systems we often observe a network of interactions between systems components that is the aggregate of the interactions that occur through different mechanisms or layers (for example, social networks encompass relationships that arise on the familiar layer and relationships that arise in the professional layer) [54]. Recent studies reveal that the existence of multiple interaction layers can have a dramatic impact in the dynamical processes occurring on these systems [79, 80, 34, 16, 15, 90]. 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. In this chapter, we address the issue of uncovering the different interaction layers from aggregate data by introducing multilayer 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 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 intersection of each pair of groups in layer 1 and layer 2. For each pair of groups, the resulting probabilities of connection are therefore correlated; in our approximation, we assume that the elements of the intersection are randomly drawn and independent of each other.

Our approximation multilayer SBMs yielded better predictions on real networks, suggesting that these networks are likely the outcome of multilayer processes, despite being observed as single-layer aggregates.

Conclusions

- We have found that sampling over models significantly improves the predictive power than considering an optimal set of parameters alone, suggesting that the error at estimating the reliability of missing links when applying the single-point method is not negligible. Additionally, we noticed that the degree-corrected version should not be used in all the networks, since it may retreat the predictions in some cases.
- We discovered that sutures associated with pathological conditions have significantly lower reliability scores than sutures that are not.
- Our results suggest that networks that are presented as single-layer may indeed be projections of multilayer networks.

Contents

1	Introduction	1
1.1	Networks unravel patterns in data	1
1.2	The large-scale structure of complex networks	3
1.2.1	Model validation	5
1.3	Inference methodology	6
1.3.1	Analogy between inference and statistical mechanics	7
1.3.2	Markov Chain Monte Carlo methods	9
1.3.3	Maximal a posteriori	10
1.4	Inference with Stochastic Block Models	11
1.5	Scope of the work	12
2	Approaches to network inference with Stochastic Block Models	15
2.1	Introduction	15
2.2	Description length applied on Stochastic Block Models	16
2.3	Specializations on SBM	17
2.3.1	Degree-corrected	17
2.3.2	Hierarchical priors	18
2.4	Consistency of the Area Under the Curve (AUC) measure	18
2.5	Scoping for the best model strategy	21
2.6	Similarity between partitions sampled	23

2.7	Discussion	24
3	Suture fusion in normal and pathological development is constrained by the network architecture of the human skull	29
3.1	Introduction	29
3.2	Methods	31
	3.2.1 A network model of the skull	31
	3.2.2 Statistical analysis	32
3.3	Results	33
3.4	Modularity	33
3.5	Evolution	34
3.6	Discussion	39
4	Multilayer stochastic block models reveal the multilayer structure of complex networks	41
4.1	Introduction	41
4.2	Multilayer stochastic block models	42
	4.2.1 Computation of degeneracies	44
	4.2.2 Link reliability with approximate multilayer stochastic block models	45
	4.2.3 OR combination of layers	48
4.3	Validation of link reliability estimation in model networks	49
	4.3.1 Ensemble of two-layer synthetic networks	58
4.4	Convergence in accuracy of single-layer and two-layer AND models for sparse networks	62
	4.4.1 Performance of the OR model in the ensemble of networks of our approach	64
4.5	Multilayer stochastic block models are more predictive for real networks	65
	4.5.1 Alternative approximation for multi-layer SBMs: A single layer SBM with degeneracies	68
4.6	Quantification of the preference for multilayer models	70
	4.6.1 Computation of the Bayes factor	71
	4.6.2 Bayes factor quantifies the preference for multilayer models in real-world networks	71
4.7	Discussion	73
5	Conclusions and perspectives	75
A	Appendices	79
	References	85

Introduction

1.1 Networks unravel patterns in data

Science is based on empirical observations and therefore it depends on our ability to interpret data. Precisely, from that knowledge the world has been modified to enhance human welfare. Recent technological advances have enabled the generation and access to increasing amounts of data, a phenomenon called Big Data; one of the objectives of science is to use data to better comprehend a constantly changing reality, thus maintaining and enhancing the welfare of society.

The availability of large amounts of data holds the promise of bringing the solutions to problems in many fields (such as health). In fact, probability theory proves that an increase in available information must help to diminish uncertainty [49]. However, there is an ever increasing gap between information and knowledge, and appropriate representations of data is necessary. Mathematical models are designed to encounter patterns on data can help us to narrow the gap between information and knowledge. A rather succesful approach is that of complex networks: a system of interconnected items organize data properly, facilitating its analysis. A network consists of a set of nodes (or vertices) connected through links (or edges), representing the relation between two nodes. By providing different meanings to nodes and their connections one can get different types of networks. For instance, in social networks nodes may represent people that are connected if they know each other, and in protein-protein interaction networks proteins are connected if they interact. Therefore, a network approach can be applied to all those cases were one can relate a set of items, or even when this relation is no that trivial -for instance, in metabolic networks chemical reactions connect several metabolites, and the morphology of skulls can be studied from networks where bones connect through sutures [26]-. The study of real-world networks have pushed towards to the understanding of complex systems in a wide range of fields as molec-

ular and cell biology [6], neuroscience [11], biomedicine [5, 14], ecology [100, 85], economics [92], and sociology [9].

The study of networks is able to find common behaviors in different fields [87]: for instance, in social networks people tend to connect to more popular people, and in cell biology highly connected proteins are more likely to obtain links from duplicated genes [6]; appears a common tendency of highly connected nodes to assemble even higher amounts of connections, a pattern so-called as preferential attachment. Therefore, network science is able to disentangle universal properties that emerge from the topology of the network, i.e. how links are distributed over the nodes. Thus we can apply the same methods and models to different types of networks to obtain reliable predictions, which later on can be used for specialists in each field.

One of the most successfully studied problems in networks science is that of finding the community structure, it consists in dividing the network into densely connected groups to summarize the information contained inside the network. Recently, alternative generative models that generalize the idea of communities have been proposed. These probabilistic models allow for the use of inference techniques to divide the nodes into groups, not necessarily communities. Then arises the question of finding the best model inference technique to use.

There are several types of networks [68]: directed networks with unidirectional links, multigraph networks that may contain self-loops and multiple edges between a pair of nodes, although the vast majority of networks studied are undirected and simple (without self-loops nor multiple edges). However, the assumption that systems can be described by a single network is almost always an oversimplification, instead the information is usually compressed throwing away crucial information. To overcome such fact, multilayer/multiplex networks were recently defined as a set of interconnected networks where many -or even all- nodes have a counterpart in different layers [16], different layers enables the opportunity to study more accurately real-world networks (for instance, temporal networks [47, 30], social networks [98] or transportation networks [79])

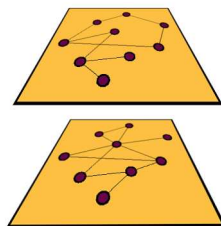


Figure 1.1: **Multilayer network.** A multilayer network with two disconnected layers, all nodes appear in both layers but they contain different edges

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 [79, 80, 34, 16, 15, 90]. Nevertheless, in such studies layers are assumed to be known, while generally real-world networks are presented to us as single-layered networks. It remains as an open question whether the topology of single-layered real world networks is actually affected by an intrinsic multilayer structure.

It is clear that networks enable a manageable representation of data in a wide variety of fields, and that statistical inference on networks is a powerful predictive tool concerning high interest recently. For such purpose, in this thesis we enhance this predictive power by introducing novel approaches on simple network inference, from a single layer point of view and especially from a multilayer perspective. To adress this topic, we will handle with a family of models that accurately capture the information contained in complex systems, enabling a high flexibility for its wide appliance. Particularly, we will answer those questions by focusing on exploring a set of models that, combined, yield better predictions than a single model alone. Our novel techniques will enable further knowledge in a high variety of fields: molecular and cell biology, disentangling the metabolic pathways to comprehend the whole genome; anatomy, model the growth of human skull bones; neuroscience, quantifying the brain's structural and functional systems; ecology, managing the effects of biodiversity on ecosystem function and services; economics, reducing conflicts between individual interests and the risk of global failure; and sociology, understanding social behavior.

1.2 The large-scale structure of complex networks

The high amounts of data that networks contain can be summarized by partition the nodes into representative groups, methods are developed in the same way that clustering methods do but exploiting network topology instead.

Communities

A huge proportion of networks are composed of diverse communities of nodes: a network can be partitioned in groups densely connected inside (communities) and loosely connected inbetween them. This community structure (or modular structure) brings out much information about the network; for instance, communities in metabolic networks represent biological functions of the cell [6], and the community structure in neuronal networks explain the interconnection between brain regions [11]. This fact is well explained by the nodes tendency to a preferential attachment on some specific nodes (hubs), creating an assortative structure that is well modeled with community structure based models. Therefore, a large body of research has been devoted to develop the detection of which communities comprise a given network [28, 66]. Typically, communities are obtained through the modularity function introduced by [67], defined as

$$Q_{mod} = \sum_{\alpha} (e_{\alpha\alpha} - (\sum_{\beta} e_{\alpha\beta})^2) \quad (1.1)$$

where $e_{\alpha\beta}$ are the fraction of edges that link vertices from communities α and β . It quantifies the strenght of community structure.

Hierarchical structure

; Besides a community structure, networks may have a hierarchical structure with several nested groupings of the nodes, were small clusters are included inside larger clusters that in turn are included in larger ones. Examples where this structure is recognized in real-world networks are the airport transportation networks [89], or the protein interaction networks [59]. The classic algorithm to extract the hierarchical organization of networks is the hierarchical clustering. First define a similarity measure between nodes, then there are two choices: (i) agglomerative algorithms in which clusters are iteratively merged until a cluster containing all the nodes is reached [78]; (ii) divisive algorithms in which clusters are iteratively divided until all nodes are isolated [32]; the merge/division is dependent on the previously defined measure which yields a dendrogram of the process; afterwards, a cut throughout the dendrogram -a partition inbetween the process of merging/division- is usually proposed as the more representative community partition.

Stochastic block models

At this point only models to detect community detection were proposed -the presented hierarchical structures conceals nested communities-, but actually other sctructures appear in nature. Communities are not always enclosed hierarchaly but they usually overlap in social networks. Moreover, disassortative networks are poorly described by communities, where a convenient representation abides a partition with sparse groups that are densely connected between them. For instance, in the bipartite network of prey-predators is more reasonable to group together predators that ate the same preys although there are no connection between predators [88]. Furthermore, core-periphery structures are commonly found in economic and social networks, they consist of a dense core in contrast with a sparse periphery.

The major drawbacks concerning community detection models and hierarchical clustering based models are its stubborn resilience to exclusively detect the structure they were defined for. In [81] is highlighted the need of a more precise definition of community to sharpen the use of detection algorithms, yet they only discern between weak or strong communities while in reality networks can have a continuum varieties of structural organizations. Additionally, the number of clusters is commonly unknown a priori and it is often added as an extra parameter of the model [69, 50, 17]. Regarding hierarchical clustering methods, they are only accurate at a local level, the output is always a hierarchical tree regardless of whether the system is indeed hierarchically or not, and there is no statistically general criterion to determine the relevant levels on the

hierarchy [89]. Hence a need for a flexible model that is capable to adapt to any kind of structure (assortative, disassortative, core-periphery and mixed) surges.

Generative Bayesian models are a handfully family of models that consider the whole set of data without structure constrictions, an inference approach that communities were not able to embrace. Stochastic Block Models (SBM) are the most used generative models, enabling to capture all those possible variabilities of structural organizations, not only discerning between weak and strong communities but capturing along the whole possible probability spectrum [44, 56, 76].

SBM were first formalized in the social sciences [105, 46, 70, 102], the goal of this approach is to order the diversity of actor behaviors, each actor being a node in a social network. It identifies actors that are *structurally equivalent*, i.e. actors that have the same relationships to all other actors, and group them together into the same block actors. The notion of structurally equivalent actors is slightly relaxed into *stochastically equivalent* in the sense that all members from a group do not need to have exactly the same relationships but instead the same probability of connection; two stochastically equivalent actors are exchangeable without modifying the topology of the network significantly. This *actor behavior diversity* can be translated into *pattern connection diversity*, thus the same notion of stochastic block models can be generalized outside the social science field and applied to any type of network to order the patterns of connection heterogeneity into blocks holding homogeneous patterns.

A SBM is totally determined by: (i) a partition P which distributes N nodes into blocks; (ii) the matrix Q , whose elements $q_{\alpha\beta}$ indicate the probability of connection between blocks α and β .

1.2.1 Model validation

Different models will inevitably lead to different predictions, a validation of the models is required in order to distinguish models and trust in their predictions.

The first step towards a reliable validation usually initiates by building a synthetic network imposing a known structure with random generation of links, where the randomness factor enables some uncertainty degree over the choice of a proposed partition. As the structure was imposed on the beginning -but let unknown for the model- the validation is straightforward when comparing the proposed partition with the imposed one.

This situation becomes intricate when validating real world networks, in the sense that the structure is usually not known at before hand. Many studies overcome that situation by applying the model onto real networks with a known partition considered as a ground truth. Karate club network [110] is the paradigm of real networks to validate community structure, this social network consist of members of a karate classroom that ended up splitting into two different academies, thus two communities are known in advance. Similar methodology can be applied on other real networks: a college football teams network [32] were teams of american football of Division I connect depending on the 2000 regular-season games, were teams are subjectively assigned to communi-

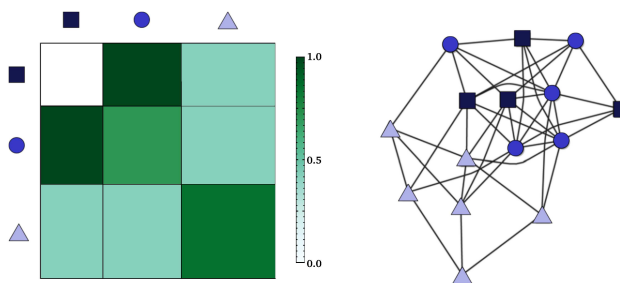


Figure 1.2: **Stochastic block models.** A stochastic block model is fully specified by a partition of nodes into groups and a matrix Q in which each element $Q_{\alpha\beta}$ represents the probability that a node in group α connects to a node in group β . (A) A simple matrix of probabilities Q . Nodes are divided in three groups (which contain 4, 5, and 6 nodes, respectively) and are represented as squares, circles, and triangles depending on their group. The value of each element $Q_{\alpha\beta}$ is indicated by the shade of green; for example, squares do not connect to other squares, and connect to triangles with small probability, but squares connect to circles with high probability. (B) A realization of the model in (A). In this realization, the number of links between the square and the triangle group is $n^1 = 4$, whereas the maximum possible number of links between these groups is $n = 24$

ties lean on geographical location; or the political blogs network [1] enclosing blogs about U.S. politics and the web links between them, the blogs have known political leanings that represent the established communities.

Bayesian approaches are a more reliable alternative to test the model on real world networks, its performance at inferring some particular property of the network is not based in any assumed ground truth but probability theory, for instance the prediction of missing/spurious links of the network that will be explained further in the introduction.

1.3 Inference methodology

Inference exploits logic reasoning to draw conclusions from existing knowledge. In particular, statistical inference takes advantage of mathematical formulation (logic reasoning) to analyse an observed data (the existing knowledge) in order to obtain the posterior probability (the conclusions drawn). Uncertainty emerges from observed data due to our usual ignorance of the whole truth, therefore methods are required to handle with such uncertainty and provide reliable posteriors. Identifying the underlying distribution of the observable data one can easily predict any particular posterior feature related with the observable, hence is crucial to define a statistical model enclosing parameters that shape the underlying distribution, where the parameters estimation plays a key role on inference. Not all models have a fixed number of parameters, non-parametric models are not a model without parameters but a model that enables the amount of parameters to vary depending on the observable data.

Traditional frequentists consider model parameters as unknown but fixed quantities, then the likelihood function of a set of parameters θ of a specified model $M(\theta)$ -that take values over the parameter space Θ_M - is equal to the probability that the parameters yield the \mathcal{O} observables: $p(\mathcal{O}|M(\theta))$, i.e., the likelihood measures how likely are the observables to be generated from the model parameters. Some criticism arise from the fact that traditional frequentism assume to know the frequencies which parameters are fixed [49]. Bayesian modeling circumvents this issue by introducing a degree of belief about parameters, although parameters are still thought of as unknown fixed quantities they are modeled as random variables. The uncertainty of the parameters knowledge prior to the observed data is measured by a prior probability distribution $p(M(\theta))$.

Once the prior and the likelihood have been decided, the model is completely specified and the posterior distribution can be computed:

$$p(M(\theta)|\mathcal{O}) = \frac{p(\mathcal{O}|M(\theta))p(M(\theta))}{\int_{\Theta_M} d\theta p(\mathcal{O}|M(\theta))p(M(\theta))} \quad (1.2)$$

Such posterior probability quantifies the degree of belief about the model on fitting the data. Finally, the general problem to solve is given a dataset \mathcal{O} , find a model $M(\theta)$ that predicts unobservables x . Then, the probability of the prediction is computed by marginalizing over the parameter space as follows,

$$p(x|\mathcal{O}, M(\theta)) = \int_{\Theta_M} d\theta p(x|M(\theta), \mathcal{O})p(M(\theta)|\mathcal{O}) \quad (1.3)$$

Substituting Eq. 1.2 into Eq. 1.3 yields a similar expression obtained in [44]

$$p(x|\mathcal{O}, M(\theta)) = \frac{\int_{\Theta_M} d\theta p(x|M(\theta), \mathcal{O})p(\mathcal{O}|M(\theta))p(M(\theta))}{\int_{\Theta_M} d\theta p(\mathcal{O}|M(\theta))p(M(\theta))} \quad (1.4)$$

Such approach is followed by generative models, that benefit from the joint probability $p(\mathcal{O}, M(\theta))$ and the Bayes rules to estimate the posterior probability $p(M(\theta)|\mathcal{O})$, in contrast with discriminative models that directly model the conditional posterior probability.

Often, large parameters space makes unfeasible to analytically compute Eq. 1.4, therefore several methodologies have been developed to approximate this integral. Some of them are based on statistical mechanics that is shortly explained in the following section.

1.3.1 Analogy between inference and statistical mechanics

Statistical mechanics is a branch of physics that is capable to deduce the behavior of large systems by using probability theory. Usually items from a system can be in different states: for instance electrons from a magnetic field may have positive charge (spin-up) or negative charge (spin-down); or regarding an infectious disease where people may be infected, susceptible or recovered. Then, a given system and its items

possible states are constraining a finite probability phase space that include all the possible combinations of the items different states. The same word *state* can be referred to for an item or for the whole system, thus a state of the system will be a particular configuration of the items. Surely some states will be more probable than others and the entropy function is able to measure such probability -a distribution of electrons with mixed spins (high entropy) is more expected than an all spins-up distribution (low entropy)- To predict the behavior of large systems it would be useful to know the more probable states, since the dynamical system will converge into one of those.

This measure of the probability of the states is anything but trivial. Statistical mechanics provides a function that associates each different state to a different value (so-called Hamiltonian) that measures the energy of the system, and its existence inherently determine a free-energy landscape of values representing all the possible configurations of the phase space. The second law of thermodynamics entangles energy with entropy of a system at equilibrium, therefore the study of this free-energy landscape is of high concern, since exploring the free-energy landscape in order to minimize the Hamiltonian function allows to spot the ground state with maximum entropy, the most probable configuration of the system.

From probability theory, the most probable distribution without any prior knowledge is provided by the exponential function [49]. This theory agrees with the Boltzmann distribution, the probability distribution of particles in a system over various possible states. It gives the probability that a system will be in a certain state as a function of that state energy.

$$p_i = \frac{1}{Z} \cdot e^{-\frac{\varepsilon_i}{k \cdot T}} \quad (1.5)$$

where ε_i is the energy of state i , k is the Boltzmann constant, T is the temperature, and $Z = \sum_{i=1}^M e^{-\frac{\varepsilon_i}{k \cdot T}}$ is the partition function that normalizes the probability over M accessible states.

Several models were created in order to describe the free-energy of different systems: Ising model for systems with binary states items (to model an electron magnetic field), or its generalized Potts model for systems with more than two possible states (to model an infectious disease) [109].

Particularly, the aforementioned partitions of a network can be seen from a dynamical systems point of view, to partition the nodes into disjoint groups is in fact a model of the system, whose items states are defined for the different labels representing the several groups a node can belongs to, therefore the state of the whole system coincides with a particular partition. A change of the network's state can be carried out by changing nodes labellings. Following the statistical mechanics theory, by finding a Hamiltonian that depends on the partition of the network one can scrutinize the free-energy landscape to find an optimal state, that intuitively relates with the partition that better fits the network [82].

Since the number of possible configurations of the model increases exponentially with the number of nodes of a network, it is usually infeasible to compute all the

possible values of the Hamiltonian even for small networks, hence a common practice is to take use of statistical/stochastic methods to approximate the landscape of the free energy associated to the Hamiltonian.

1.3.2 Markov Chain Monte Carlo methods

For those solutions of Eq. 1.4 that are computationally untractable, Markov Chain Monte Carlo methods (MCMC) generate draws from a desired probability distribution over the parameters θ in order to estimate it. Starting with an initial state $\theta^{(0)}$, MCMC computationally simulates S transitions of a Markov chain in θ to explore the whole parameter space Θ , finally obtaining an estimation for Eq. 1.4

$$I \approx \hat{I} = \frac{1}{S} \sum_{i=1}^S f(\theta^{(i)}) \quad (1.6)$$

MCMC takes use of randomness to explore and approximate the free-energy landscape of a system's Hamiltonian. To ensure a fair MCMC convergent to a steady state, the sampling must be ergodic -all accessible states of the parameter space are equiprobable over a long period of time-, and detailed balance must be guaranteed -equilibrium is invariant under time reversal, i.e., probability transition from state θ_i to state θ_j must be equivalent to the inverse transition-. Without those properties, the sampling could get stuck in some subgroup of states which could bias the sample.

In what follows, two specific MCMC methods to compute Eq. 1.6 are explained.

Metropolis-Hastings

The Metropolis-Hastings algorithm is probably the most generic scheme from the MCMCs [60], each Markov chain transition is split into three steps: (i) from a given state θ generate a proposal $\hat{\theta}$; (ii) Compute the acceptance distribution $a(\theta, \hat{\theta}) = \min\left\{1, \frac{p(\hat{\theta}|\mathcal{O})}{p(\theta|\mathcal{O})} \cdot \frac{q(\theta|\hat{\theta})}{q(\hat{\theta}|\theta)}\right\}$, with the proposal distribution $q(\hat{\theta}|\theta)$ which is the conditional probability of proposing a state $\hat{\theta}$ given θ ; (iii) With probability a replace θ with the proposal $\hat{\theta}$. Otherwise, leave θ unchanged. Detailed balance causes a symmetry that makes $\frac{q(\theta|\hat{\theta})}{q(\hat{\theta}|\theta)}$ to cancel out.

Gibbs sampling

The Gibbs sampling [31] is an special case of Metropolis-Hastings which advocates for estimate a probability distribution by sampling from a conditional distribution, instead of marginalizing by integrating over a joint distribution. The Gibbs sampler becomes specially handy as in many statistical models the conditional posterior distributions takes the form of some familiar distributions. The Gibbs sampling begins with some initial parameter vector $\theta^{(0)} = \theta_1^{(0)}, \dots, \theta_n^{(0)}$, then proceeds by iteratively generating

from the conditional posterior distribution $\theta_j^{(t+1)} \sim p(\theta_j | \theta_1^{(t+1)}, \dots, \theta_{j-1}^{(t+1)}, \theta_{j+1}^{(t)}, \dots, \theta_n^{(t)}, \mathcal{O})$, for $j = 1 \dots n$, and in repeating this process the samples will eventually converge to an approximation of the posterior distribution $p(\theta | \mathcal{O})$ [91].

1.3.3 Maximal a posteriori

Maximal a posteriori (MAP) is a method to approximate Eq. 1.4 which, instead of intensively exploring a likely subset of the parameter space, it leads directly towards the specific set of parameters that maximize the posterior prior likelihood. While this method is computationally faster than the MCMC, it remains unclear whether a single point estimate accurately predicts as well as sampling from a conditional distribution.

Several greedy algorithms follow the MAP approach.

Expectation-maximization

The expectation–maximization (EM) [20] algorithm is an iterative method used to locally find the maximum likelihood parameters of a statistical model in cases where the equations cannot be solved directly. The EM iteration alternates between performing an expectation (E) step -which estimates a posterior function evaluated with the current parameters estimation- and a maximization (M) step -which search those parameters that maximize the expected log-likelihood function from the E step-. Usually EM starts with arbitrary estimates of parameters and the algorithm iterates until a fixed point is reached.

Variational Bayesian method

Variational Bayes can be seen as an extension of the EM algorithm which computes an approximation to the entire posterior distribution of all the variables, that enclose not only the parameters (as EM) but also the unobserved variables. Specifically, while EM computes optimum values of the parameters the variational approach first fit a prior distribution to these parameters, iteratively computing optimum values for the hyperparameters of the prior distribution.

Simulated annealing

The method of simulated annealing [53] is an adaptation of the Metropolis-Hastings algorithm for computing maximum likelihood parameters. It is particularly useful to find the global optimum of a given function in a large search space, since it is more resilient to get stuck on local optimum than other greedy algorithms. Specifically, simulated annealing can estimate the ground state of a system, if the given function is the probability of a state in a system (Eq. 1.5) and a state determines a specific configuration of parameters. Temperature in Eq. 1.5 plays a crucial role in here, since higher temperatures enable the algorithm to avoid local optimum energy and lower

temperatures permit to locally find the optimum state, thus the algorithm must balance between high and low temperature to find the global optimum.

1.4 Inference with Stochastic Block Models

The previously defined Stochastic Block Models (SBM) are a family of non-parametric generative models that enables us to explore the whole configuration of possible partitions from a given network, because its flexibility at capturing different partitions settings does not constrain any specific configuration. For a given network, convenient SBMs group nodes that have similar patterns of connection. Mathematically, we formalize this intuition in a non-parametric Bayesian framework as follows. A SBM is totally determined by: (i) a partition P which distributes N nodes into blocks; (ii) the matrix Q , whose elements $q_{\alpha\beta}$ indicate the probability of connection between blocks α and β . It is non-parametric because we let the number of blocks to vary.

Once the model is set, Eq. 1.4 can be applied to predict a network property x where the model M ranges over all possible SBMs \mathcal{M} and the observable \mathcal{O} corresponds with the adjacency matrix $A^{\mathcal{O}}$ of a given network -a simple network adjacency matrix A consists in a binary matrix with 1 in the position A_{ij} representing an edge between nodes i and j and 0 otherwise-; since $M(\theta) = (P(\theta_P), Q(\theta_Q))$, the parameter space $\Theta = \mathcal{P} \times \mathcal{Q}$ consists in the cartesian product of all possible partitions \mathcal{P} times all possible probability matrices $\mathcal{Q} = q_{\alpha\beta} \in [0, 1], \forall \alpha, \beta \in \mathcal{P} \times \mathcal{P}$. In particular, if the property x is set to the probability that a link exists (reliability of a link) between two particular nodes i, j within an adjacency matrix A , the equation becomes

$$p(A_{ij} = 1 | A^{\mathcal{O}}) = \frac{\int_{\Theta} d\theta p(A_{ij} = 1 | M(\theta)) p(A^{\mathcal{O}} | M(\theta)) p(M(\theta))}{\int_{\Theta} d\theta p(A^{\mathcal{O}} | M(\theta)) p(M(\theta))}. \quad (1.7)$$

Here, $p(A^{\mathcal{O}} | M(\theta))$ is the probability of the observed interactions given model $M(\theta)$ and $p(M(\theta))$ is the *a priori* probability of a model, which we assume to be model-independent $p(M(\theta)) = \text{const.}$

Thus, if i belongs to group σ_i and j to group σ_j then [39]

$$p(A_{ij} = 1 | M(\theta)) = q_{\sigma_i \sigma_j}; \quad (1.8)$$

and

$$p(A^{\mathcal{O}} | M(\theta)) = \prod_{\alpha < \beta} q_{\alpha\beta}^{n_{\alpha\beta}^1} (1 - q_{\alpha\beta}^{n_{\alpha\beta}^0}), \quad (1.9)$$

where $n_{\alpha\beta}^1$ is the number of links between nodes in groups α and β and $n_{\alpha\beta}^0$ is the number of disconnections (no-links) between nodes in groups α and β .

The integral over all the parameters set Θ can be split into a discrete sum over all possible partitions of nodes into groups, and a continuous integral over all possible

values of each $q_{\alpha\beta}$. Using this together with Eqs. (1.7), (1.8) and (1.9), and under the assumption of no prior knowledge about the models ($p(M(\theta)) = \text{const.}$), we have

$$p(A_{ij} = 1|A^\circ) = \frac{1}{Z} \sum_{P \in \mathcal{P}} \int_0^1 dQ q_{\sigma_i \sigma_j} \prod_{\alpha \leq \beta} q_{\alpha\beta}^{n^1(\alpha,\beta)} (1 - q_{\alpha\beta}^{n^0}), \quad (1.10)$$

where the integral is over all $q_{\alpha\beta}$ and Z is the normalizing constant (or partition function). Using these expressions in Eq. (1.11), one obtains the reliability of a link

$$R_{ij} = p(A_{ij} = 1|A^\circ) = \frac{1}{Z} \sum_P \left(\frac{n_{\sigma_i \sigma_j}^1 + 1}{n_{\sigma_i \sigma_j} + 2} \right) \exp(-H(P)), \quad (1.11)$$

where the sum is over all partitions of nodes into groups, $n_{\sigma_i \sigma_j} = n_{\sigma_i \sigma_j}^1 + n_{\sigma_i \sigma_j}^0$ is the total number of possible links between groups σ_i and σ_j , and $H(P)$ is a function that depends uniquely on the partition

$$H(P) = \sum_{\alpha \leq \beta} \left[\ln(n_{\alpha\beta} + 1) + \ln \left(\frac{n_{\alpha\beta}}{n_{\alpha\beta}^1} \right) \right], \quad (1.12)$$

To avoid overfitting with an unnecessary amount of parameters we include a penalty term into the Harmonic function, thus instead of $H(P)$ we use $H'(P) = H(P) - \log[(N - k)]$ with N the number of nodes and k the number of non-empty groups in partition P , a similar Bayesian Information Criterion was used in [39].

In comparison with previous models, SBMs provides an accurate prediction on estimating missing links in a network, Fig. 1.3 compares the SBM approach with previously defined models on predicting missing links: the Clauset et al. [13] and a local algorithm based on the number of common neighbors between pairs of nodes [57].

1.5 Scope of the work

This thesis deepens in the statistical inference on complex networks, a better understanding of the mechanisms that govern systems interactions will enable us to design proper models that accurately describe a given network. The main objective of this thesis is to improve the predictive power of complex networks via the family of Stochastic Block Models (SBMs).

In Chapter 2 we revise previously defined approaches based on SBMs in order to comprehend which ones perform better at predicting different real world networks. Particularly, we adress the open question of how to validate different models, and we prove that the sampling methodology defined in Section 1.3.2 perform better predictions than a greedy algorithm (Section 1.3.3).

Once we know which model is capable to describe better a given network, in Chapter 3 we apply such method in a particular real world network case: a network based

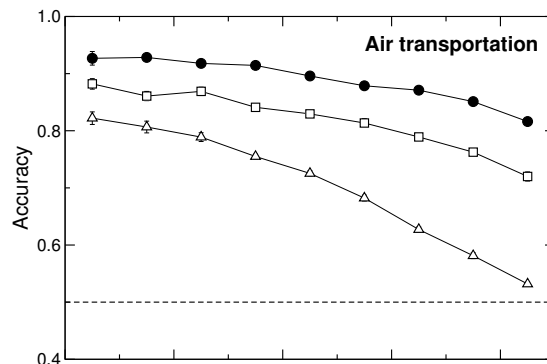


Figure 1.3: **Identification of missing links in the European air transportation network.** We compare the SBM approach presented in [44] (black circles), to the approach of Clauset et al. [13] (white squares) and to a local algorithm based on the number of common neighbors between pairs of nodes [57] (white triangles). For each true network A we remove a fraction f of its links to generate an observed network A^O , calculate the link reliability R_{ij} for each pair of nodes, and rank pairs of nodes in order of decreasing reliability. The dashed line indicates the baseline accuracy when false negatives and true negatives are randomly ranked

on the interactions/sutures between bones in newborn skulls. Statistical inference was never applied before in such networks, under our knowledge. Notably, we discovered that sutures fused due to a pathological disease in human newborn were less likely, from a morphological point of view, that those sutures that fused under a normal development.

Finally, from our observation on all the work done, we suspect that a single SBM is not sufficient to fully describe all the mechanisms governing single-layered real networks. Additionally, recent research on multilayer networks has concluded that the behavior of single-layered networks are different than those of multilayer ones; notwithstanding, real world networks are presented to us as single-layered networks. Intuitively, it could be that such networks conceal a multilayered structure inside; and, since multilayer networks behave differently than single-layered ones, this could explain why a single SBM is not sufficient to describe real-world networks. To prove so, we design a novel approach where two separate SBMs simultaneously describe a given single-layered network, and we importantly find that it predicts better missing/spurious links that the single SBM approach.

UNIVERSITAT ROVIRA I VIRGILI
NETWORK INFERENCE BASED ON STOCHASTIC BLOCK MODELS: MODEL EXTENSIONS, INFERENCE APPROACHES
AND APPLICATIONS
Toni Vallès Català

2

Approaches to network inference with Stochastic Block Models

2.1 Introduction

One of the most interesting problems on data analysis consists on detecting missing data (false positives), on complex networks this problem is translated into detect missing links. To tackle this problem Ref. [44] suggests to compute the reliability of a link, that is the likelihood that a link exists on a given network; it can be compute by eq. 1.7 in the introduction chapter 1. The most popular inference approach to estimate eq. 1.7 focuses on finding a single-point estimate for the most likely set of parameters [32, 13, 67, 18, 74, 75], following a Maximal A Posteriori (MAP) methodology. However, Bayesian inference theory suggests that the correct approach goes through integrating the whole ensemble of possible sets of parameters [49]; few approaches have followed such methodology by sampling over all the ensemble of the parameter space [44, 91, 95, 10]. We will refer to *single-point* for the former method and *sampling* for the latter. The single-point method is mostly used because the likelihood distribution is usually peaked around the optimal single-point estimate, then the rest of non-optimal models are considered negligible. In this chapter we investigate to what extent sampling over several models improves the predictive power than considering the optimal set of parameters alone, hence quantifying the error at estimating eq. 1.7 when applying the single-point method.

To adress this question we adopt the definition of SBM described in Ref. [74] that minimizes the entropy of SBM to find the most likely set of parameters, and we compare it with an approach that samples several sets of likely parameters instead of scoping for the optimal one. Although the approach is based on Ref. [44], it was never applied before on the model defined in Ref. [74].

2.2 Description length applied on Stochastic Block Models

We can parametrize a SBM as follows: N vertices are distributed into B blocks with n_α nodes in each block $\alpha \in [0, B - 1]$, the matrix $e_{\alpha\beta}$ indicates the number of edges between blocks α and β (twice this number when $\alpha = \beta$, for convenience). However, the model does not specify which node in α is connected to a node in β , thus there are different edge choices of the $e_{\alpha\beta}$. Therefore, a SBM with specified parameters generates an ensemble of different graphs realizations, call Ω the cardinal of such ensemble

$$\Omega = \prod_{\alpha \geq \beta} \Omega_{\alpha\beta}, \quad \Omega_{\alpha\beta} = \binom{n_\alpha \cdot n_\beta}{e_{\alpha\beta}}, \quad \Omega_{\alpha\alpha} = \binom{\binom{n_\alpha}{2}}{\frac{e_{\alpha\alpha}}{2}}. \quad (2.1)$$

The microcanonical entropy is defined as the logarithm of all possible graphs generated by a SBM $S = \ln \Omega$. Additionally, the log-likelihood function infers the most probable SBM grouping, $\mathcal{L} = \ln \mathcal{P}$, where \mathcal{P} is the probability of observing a particular network occurrence. With no prior information a uniform distribution is commonly adopted, hence $\mathcal{P} = \frac{1}{\Omega}$ and in such particular case $\mathcal{L} = -S$. Therefore, the most likely partitions (with maximum likelihood) are equivalent to those partitions with minimum entropy. However, S becomes an strictly decreasing function of B and minimizing the entropy unavoidably heads to the trivial partition $B = N$. Such overfitting can be evaded by including a penalty term in the minimizing function. In Ref.[74] they redefine the entropy function into Eq. 2.2 introducing a penalty term, the description length \mathcal{DL} .

$$\Sigma = S + \mathcal{DL}. \quad (2.2)$$

The description length \mathcal{DL} is the amount of information required to describe the model. SBM is established by two parameters: the connectivity matrix $e_{\alpha\beta}$ and the block partition b_i . The $e_{\alpha\beta}$ matrix can be viewed as the adjacency matrix of a multi-graph with B nodes and E edges, where the blocks are the nodes and self-loops are

allowed. The total number of $e_{\alpha\beta}$ matrices is then simply $\binom{\binom{\binom{B}{2}}{E}}{E}$, and the total number of block partitions is B^N . We obtain \mathcal{DL} by multiplying these numbers and taking the logarithm. Then the optimal partition can be found with the Occam's razor principle, in which the best hypothesis for a given set of data is the one that leads to the best compression of the data, in our cases the minimum of Σ yields the most probable partition (minimum entropy) with minimum description length.

In the Introduction chapter 1 a slightly different modelization of SBMs was employed: instead of the parameters set n_α and $e_{\alpha\beta}$ we used a partition P and the probability matrix Q consisting in the probabilities of connection between the groups $q_{\alpha\beta}$. Both are equally valid generative models since $\langle e_{\alpha\beta} \rangle = n_\alpha n_\beta q_{\alpha\beta}$ are constrained on average, as long as the edge counts are sufficiently large they are fully equivalent (Appendix A in Ref. [76])

2.3 Specializations on SBM

2.3.1 Degree-corrected

The degree k_i of a node i is the number of neighbours that this node is connected to. A drawback of SBM suggested in the literature [50], indicates that in networks with substantial degree heterogeneity SBM rather to split networks into groups of high and low degree. Logically, two nodes with sufficiently distinct degrees have actually different patterns of connection, so they are reasonably grouped in different blocks for the SBMs. Such partition does not coincide with the ground truth of some real networks as they did in Fig. 2.1, yet from mosts networks a ground truth is not available.

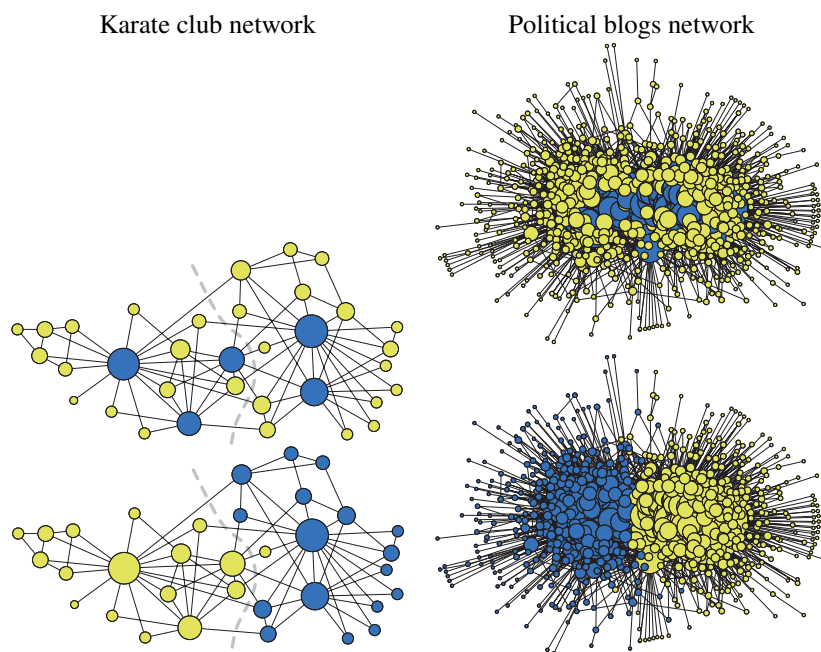


Figure 2.1: **Degree-corrected SBM divide networks in concordance with ground truth.** Divisions of the karate club network [110] (left) and the political blogs network [1] found using the uncorrected (top) and corrected (bottom) SBM in [50]. The size of each vertex is proportional to its degree and the shading reflects inferred group membership. The dashed line indicates the split observed in real life: the two academies that the class of karate split into and the division between liberal and conservative blogs.

To overcome such deviation, Ref.[50, 73] propose a degree-corrected SBM version of the model where the degree sequence k_i is provided as an additional set of parameters. The actual degree sequence can be easily computed, but fixing it as an extra parameter increases the complexity at counting the number of possible graphs realiza-

tions Ω , in fact, even the simpler asymptotic counting of graphs with a uniform degree sequence ($k_i = k, \forall i$) is an open problem in combinatorics [107].

Therefore, we will use the refinements on [50, 73] where the imposed degree on each node is only an average over the ensemble Ω , and their values over sampled realizations are allowed to fluctuate. In the corrected version, each block is split in different subgroups with the possible degrees averages, enabling the same block to confine degree heterogeneity. Instead of k_i for each node i , the needed set of parameters is the fraction of nodes with degree k , q_k . However, since the degrees are correlated with the block partitions, it is rather convenient to use the parameter set q_k^α -the degree distribution of nodes belonging to block α - as suggested in Ref. [76].

Note that Eq. 2.2 must be adjusted to introduce the late parameter set q_k in the degree-corrected description length and the entropy corrected version in Ref. [73]

The degree-corrected version provides a partition that matches with the real-world networks ground truth, on the other hand remains to be studied whether this corrected version enhance or hinders the predictive power with respect to the uncorrected version.

2.3.2 Hierarchical priors

To this achieve the detection of much smaller blocks in very large networks, Ref. [76] increased the resolution limit by designing a nested hierarchy of SBMs, where an upper level of the hierarchy serves as prior information to a lower level, rather than assuming that all possibilities should occur with the same probability.

The methodology to build the nested version resembles the idea of Ref. [8] by iteratively transforming a simple network to a multigraph, where each block of the network is converted to a node of the multigraph and the $e_{\alpha\beta}$ into multiple edges or self-loops, the main difference in Ref. [76] lies on that each iteration is saved into a different layer conforming a nested hierarchy. Since each multigraph may also be described via a SBM of its own, from the block network on the lower level we obtain another smaller block multigraph at a higher level, and so on recursively, until we finally reach a model with only one block. This forms a nested stochastic block model hierarchy, which describes a given network at detached resolution levels (Fig. 2.2).

In the degree-corrected version of the nested model, it is only required to specify the degrees at the lowest level of the hierarchy.

2.4 Consistency of the Area Under the Curve (AUC) measure

In order to contrast between various models a reliable measure is required. Despite some approaches take advantage of ground truth, this knowledge is not always available and the reliability of the ground truth is not always guaranteed, maybe causing misleading conclusions particularly in those cases without such valuable information. Rather measures involving predictive inference are preferred, without the need of assuming any ground truth. We will contrast the models on predicting missing links, an approach regularly employed in the literature. To predict missing links you can randomly remove some existing links in the network, then compute the reliability scores

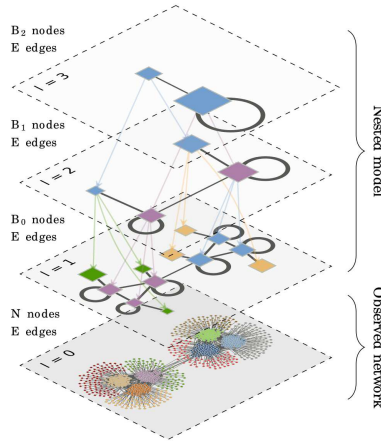


Figure 2.2: **Nested stochastic block model.** Example of a nested stochastic block model with three levels, and its generated network at the bottom. Figure extracted from [76]

(Eq. 1.11) for all pairs of nodes not connected in the observed network, including those you had removed. Therefore, we assume an error-free “true” network A , it is divided into $A = A^{\mathcal{O}} \cup \delta A$, where δA is a fraction of the edges in A . We observe $A^{\mathcal{O}}$ and we want to predict δA .

We consider two different measures at predicting missing links defined as follows:

The AUC

Sort the scores and quantify how much likely are removed links with respect to actual no-links, i.e. count how many no-links scores lay beneath a removed link, for each removed, obtaining an Area Under the Curve (AUC) measuring the accuracy at detecting missing links.

If we treat this as a classification scenario, the AUC is

$$\text{AUC} = \frac{1}{|TP||TN|} \sum_{(i,j) \in \delta A} \sum_{(u,v) \in \bar{A}} \Theta(q_{ij} - q_{uv}), \quad (2.3)$$

where $\Theta(x)$ is the Heaviside step function counting how many removed edges $(i, j) \in A$ have higher probability q_{ij} than actual no-links $(u, v) \in \bar{A}$, normalized by $|TP| = E - E^{\mathcal{O}}$ and $|TN| = \binom{N}{2} - E$, where E and $E^{\mathcal{O}}$ are the number of edges in A and $A^{\mathcal{O}}$, respectively.

Bayesian predictive held-out likelihood

Given observed network $A^{\mathcal{O}}$, the reliability of a link is the difference in Σ of from Eq. 2.2 before and after adding such link on $A^{\mathcal{O}}$.

The nonparametric model used to compute Σ is a joint distribution on the model and its parameters, which allows us to compute a likelihood conditioned on the network partition $\{b_i\}$

$$p(A^\mathcal{O}|\{b_i\}) = \sum_{\theta} p(A^\mathcal{O}|\{b_i\}, \theta)p(\theta), \quad (2.4)$$

where θ is the set of remaining model parameters. For instance, for the degree-corrected model it corresponds to the edge counts $\{e_{\alpha\beta}\}$ and the degree sequence $\{k_i\}$. Based on this, we can write the likelihood of non-observed edges δA , given the observed graph and its partition, which reads

$$p(\delta A|A^\mathcal{O}, \{b_i\}) = \frac{\sum_{\theta} p(A^\mathcal{O} \cup \delta A|\{b_i\}, \theta)p(\theta)}{\sum_{\delta A} \sum_{\theta} p(A^\mathcal{O} \cup \delta A|\{b_i\}, \theta)p(\theta)}, \quad (2.5)$$

where the denominator is a sum over all possible missing edges δA . Note that for the microcanonical models, the sum over θ is trivial, since given any particular partition of a graph, there is only one choice of θ that generates it with nonzero probability. Therefore, we have simply

$$p(\delta A|A^\mathcal{O}, \{b_i\}) = \frac{p(A^\mathcal{O} \cup \delta A|\{b_i\}, \hat{\theta})p(\hat{\theta})}{\sum_{\delta A} p(A^\mathcal{O} \cup \delta A|\{b_i\}, \hat{\theta})p(\hat{\theta})}. \quad (2.6)$$

Although the denominator can be computed numerically for the removal of a single edge, it is not actually needed, since it is just a multiplicative constant that does not affect the AUC scores. We need only to compute the numerator, which is just the joint likelihood of the data and parameters, conditioned on the partition.

Then the held-out likelihood is computed via the difference of Σ before and after adding δA ,

$$e^{-\Delta\Sigma} = p(\delta A|A^\mathcal{O}, \{b_i\}) \quad (2.7)$$

Note that differently from the approximation of the previous section, the scores computed like the above will never be strictly zero, even if the edge counts or degrees in the observed network become zero. This is an important property of such Bayesian methods, and this could in general give a better predictions than the previous scheme.

We should expect a consistency between model selection done by minimizing Σ from Eq. 2.2 and the predictive performance as measured by AUC at least in some limit of sufficient data. For that purpose, in Fig. 2.3 we can observe a consistency between the AUC measure and the Σ , observe that in general, the larger the AUC the smaller the Σ , thus such model specializations that get the minimum entropy (with description length) are those that achieve higher accuracies at predicting missing links. Note that the nested version usually outperforms the non-nested version, both in AUC and Σ . Regarding the degree-corrected version, the improvement is larger than the nested specification, however it is not applicable to all networks studied: observe that in the neural network of the *C. elegans* it is preferable to use the non corrected version.

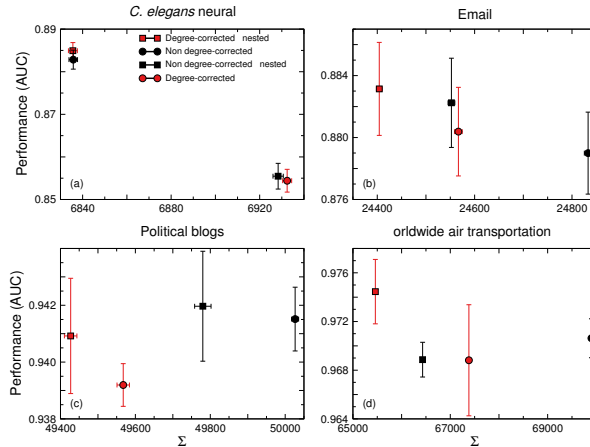


Figure 2.3: **Consistency of the AUC.** We show the AUC at predicting missing links computed, versus its Σ from Eq. 2.2 applied for each model specialization (nested: squares, non-nested: triangles, degree-corrected: solid, non-degree-corrected: empty), on four different real networks: the neural network of *C. elegans* [106], the email network within an organization [38], the web links between U.S. political blogs [1], the worldwide air transportation network [43]. Observe a tendency of larger AUC with smaller Σ , hence the AUC measure is consistent with minimal Σ .

2.5 Scoping for the best model strategy

Once the model is properly defined, we want to check whether there is an improvement at predicting the same posterior probability when considering a bunch of likely parameters sets (sampling method) in comparison with a single but optimal set of parameters (single-point).

To test the resemblance applying both methodologies we need a measure and some datasets. In particular, we will test both methodologies at tackling the same problem: predicting missing links. For the dataset, we consider four real world networks of different types, sizes and degree-distributions: the neural network of *Caenorhabditis elegans* [106], the Email network within a university [38], the web links between U.S. political blogs [1], and the worldwide air transportation network [43].

We apply both methods based on the same SBM, and we observe on Fig. 2.4 that the sampling methodology (dashed bars) outperforms the single-point (solid bars) when predicting missing links on all datasets contemplated.

Furthermore, in Fig. 2.5 we can observe that even the simple sampling approach without any specification performs better at predicting missing links than all single-point approaches with sophistications. Indeed, we were able to compute a sampling version of the degree-corrected SBM, note that such approach outperforms the simple sampling, at least for those networks that degree-corrected SBM worked properly for single-point approaches. However, to build a sampling nested version gets more intricate because we have to decide how to do the sampling in all the levels of the hierarchy.

Approaches to network inference with Stochastic Block Models

Network	Number of nodes	Number of edges	Average degree	Degrees standard deviation
<i>Caenorhabditis elegans neural</i>	297	2148	13.99	0.7425
<i>Email</i>	1133	5451	9.61	0.2775
<i>Political blogs</i>	1222	16714	27.36	1.0990
<i>Worldwide air transportation</i>	3227	19675	11.98	0.4493

Table 2.1: **Real world networks of different types.** Number of nodes and edges and degree distributions conforming each network.

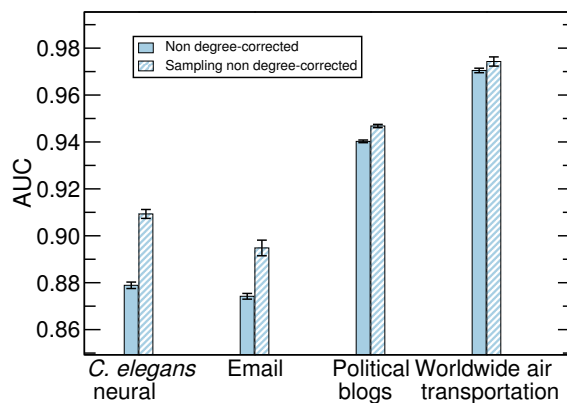


Figure 2.4: **Performance of sampling and single-point methodologies.** We show the performance at predicting missing links for the sampling (dashed) and the single-point (solid) inference applied onto different real networks: the neural network of *C. elegans* [106], the email network within an organization [38], the web links between U.S. political blogs [1], the worldwide air transportation network [43]. Note that sampling methodology provides better accuracies at predicting missing links than single-point methods.

Although statistical mechanics suggests that the optimal state is by far the most likely, we found that a bunch of likely states result to be more predictable than the optimal alone. A plausible explanation can be suggested if we analyze the sum in Eq. 1.11, statistical mechanics demonstrates that the optimal state has the highest order in such sum, say $\mathcal{O}(u)$, in contrast with the rest of the states that contribute on average with an order of $\mathcal{O}(v) \ll \mathcal{O}(u)$, such large difference may indicate that the sum is uniquely determined by the optimal state, with negligible contributions of all the rest lower order states to the sum. However, we have found that exists a subset S of intermediate states that contribute enough to be considered, each state $s \in S$ contributes to the sum with a higher order than the average $\mathcal{O}(v) \ll \mathcal{O}(w_s)$ but still much lower than the optimal state $\mathcal{O}(u) > \mathcal{O}(w_s)$; nevertheless, the large amount of possible intermediate states makes the cardinal $|S|$ close to $\mathcal{O}(u - v)$, hence Eq. 1.11 is no longer uniquely

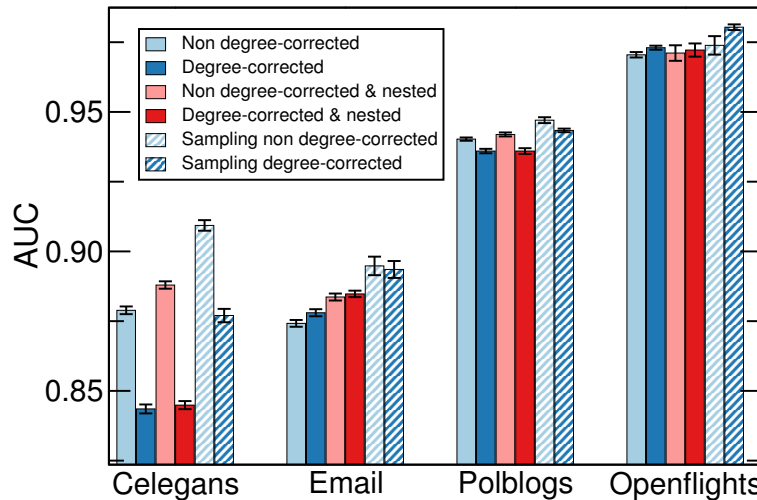


Figure 2.5: **Performance of sampling and single-point methodologies.** We show the performance at predicting missing links for the sampling (dashed) and the single-point (solid) inference, for several SBMs specializations (naive, light blue; degree-corrected, dark blue; nested, light red; degree-corrected and nested, dark red) applied onto different real networks (the neural network of *C. elegans* [106], the email network within an organization [38], the web links between U.S. political blogs [1], the worldwide air transportation network [43]). Note that sampling methodology provides better accuracies at predicting missing links than single-point methods.

determined by the optimal state, but additionally with a small contribution of the whole subset of intermediate states S .

2.6 Similarity between partitions sampled

To understand in more depth how the partitions varied throughout the sampling process, we saved 100 partitions equidistant in time during the sampling. We additionally compute the optimal partition to test its similarity between all partitions sampled. Then we compute the Normalized Mutual Information (NMI) similarity measure for all pairs of saved partitions, obtaining a NMI matrix for each sampling process. The NMI is defined as $2I(b_i, c_i)/[H(b_i) + H(c_i)]$, where the mutual information of two partitions b_i and c_i , with joint probability distribution p_{bc} and marginal probabilities p_b and p_c , is $I(b_i, c_i) = \sum_{\alpha \in b_i} \sum_{\beta \in c_i} p_{bc}(\alpha, \beta) \ln(p_{bc}(\alpha, \beta)/p_b(\alpha)p_c(\beta))$, and where the entropy is $H(b_i) = \sum_{\alpha} p_b(\alpha) \ln p_b(\alpha)$.

Fig. 2.6 and Fig. 2.7 show the sampling process applied to the *C. elegans* network [106] and the email network [38], respectively. We order the NMI matrices to cluster the partitions by similarity (on the right side in Figs. 2.6, 2.7), but keeping the chronological order to understand the actual transition of the process (on the left side

in Figs. 2.6,2.7) We repeat 20 independent realizations of the process for the *C. elegans* network and 10 for the email network. Generally, we observe two differentiate behaviors:

- the vast majority of partitions are quite similar, contrary with what one would expect, the partitions explored are different from the partition with optimal likelihood (top in Figs. 2.6,2.7).
- the sampling transitate through several significatively different partitions (bottom in Figs. 2.6,2.7).

Therefore, the sampling explores around similar partitions occasionally transiting to another different set of similar partitions, all of them different from the optimal one.

Furthermore, to understand the dissimilarity between the partitions sampled and the optimal partition, we computed the number of groups for each of them. We obtain a higher amount of groups for the partitions sampled than the optimal one (Table. ??).

Network	Sampled partitions	Optimal partition	Number of repetitions r
<i>Caenorhabditis elegans neural</i>	(22.1, 23.7)	(12.0, 13.4)	20
<i>Email</i>	(28.0, 29.4)	(19.6, 24.4)	10
<i>Political blogs</i>	(30.2, 32.2)	(17.8, 23.0)	5
<i>Worldwide air transportation</i>	(55.9, 61.3)	(33.8, 47.0)	5

Table 2.2: **Confidence Interval (CI) of the mean number of groups in the partitions.** For each sampling we compute the mean number of groups over the partitions sampled, then we compute the CI over the r independent samplings executed. For the optimal partition we compute the CI for the number of groups over the r repetitions at finding an optimal partitions. We obtain a higher amount of groups for the partitions sampled than the optimal one

Additionally, note that partitions in the email network ($0.84 > NMI(b_i, c_i) > 0.72$ for all pairs of partitions b_i, c_i) show lower similarity than the ones in the *C. elegans* network ($0.95 > NMI(b_i, c_i) > 0.82$ for all pairs of partitions b_i, c_i). Larger networks have higher amounts of possible partitions, increasing the range of partitions to be explored. The sampling approach is able to adapt to such large networks systematically and tend to explore more diverse partitions.

2.7 Discussion

We set out a novel question to compare the predictive power of two differentiate inference techniques:

- the vastly employed single-point estimation considers the most likely state (call it *single-point*)
- the Bayesian approach rather to contemplate a sample of likely states (call it *sampling*).

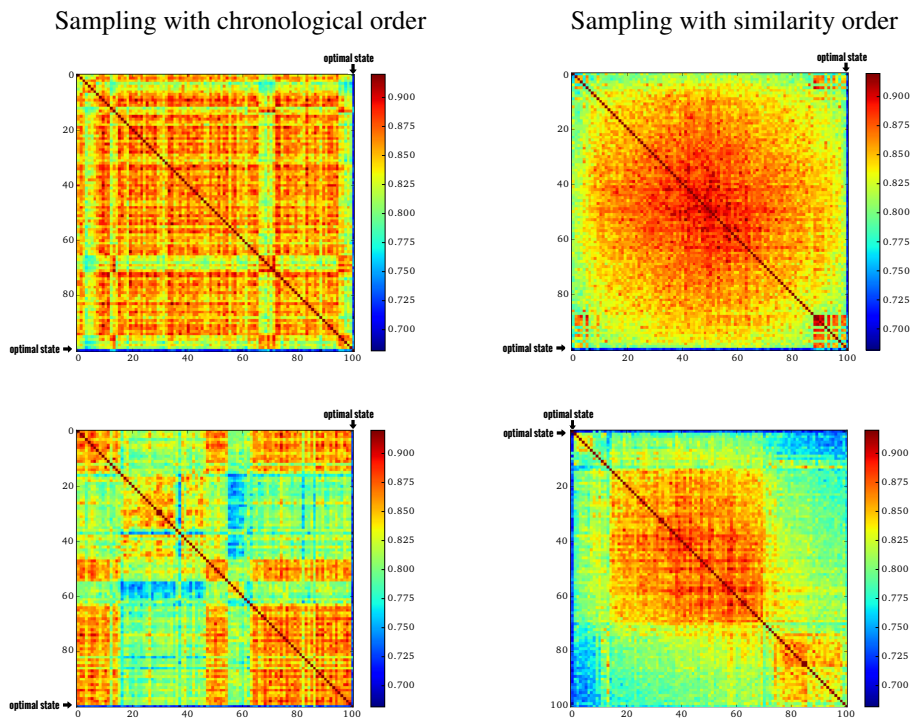


Figure 2.6: **Normalized Mutual Information (NMI) for each pair of partitions sampled.** We apply a sampling approach to compute the reliability links for the neural network of *C. elegans* [106]. From the total 10000 partitions sampled we save 100, one partition saved for every 100 sampled. We show the Normalized Mutual Information (NMI) similarity measure for all the pairs of partitions saved, in two differentiate orderings: chronological order as the partitions appear during the sampling (left), and reordering them depending on their similarity (right). We repeat the process 20 times and we observe two clear behaviors. On the first behavior (top), the vast majority of partitions are quite similar ($0.95 > NMI(b_i, c_i) > 0.82$ for all pairs of partitions b_i, c_i), but all of them are different from the optimal partition. On the second behavior (bottom), observe that the sampling transitate through several significantly different partitions.

Approaches to network inference with Stochastic Block Models

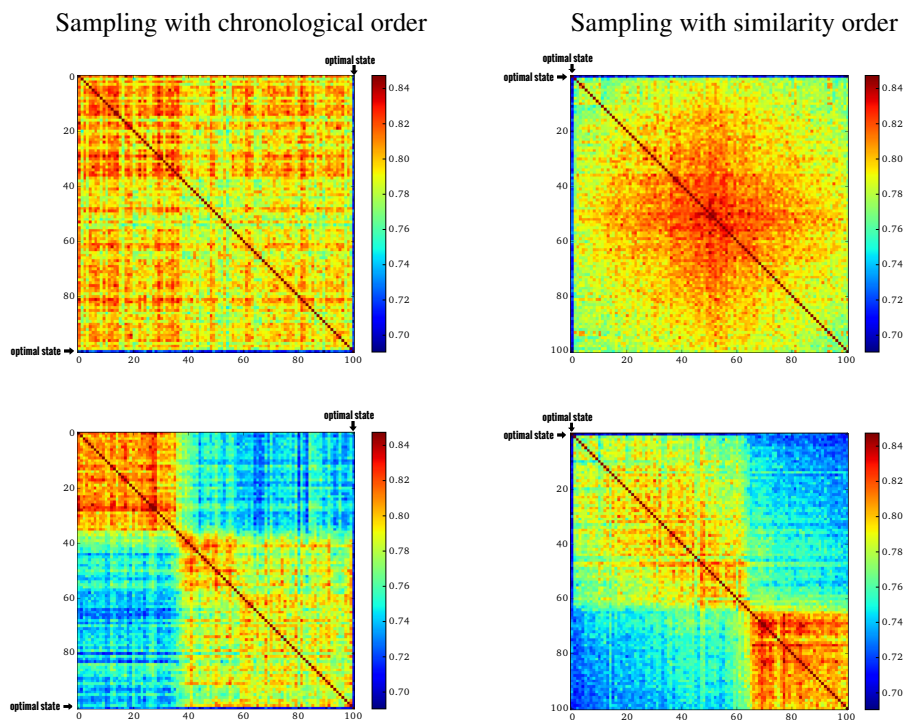


Figure 2.7: **Normalized Mutual Information (NMI) for each pair of partitions sampled.** We apply a sampling approach to compute the reliability links for the email network within an organization [38]. From the total 10000 partitions sampled we save 100, one partition saved for every 100 sampled. We show the Normalized Mutual Information (NMI) similarity measure for all the pairs of partitions saved, in two differentiate orderings: chronological order as the partitions appear during the sampling (left), and reordering them depending on their similarity (right). We repeat the process 10 times and we observe two clear behaviors. On the first behavior (top), the vast majority of partitions are quite similar ($0.84 > NMI(b_i, c_i) > 0.72$ for all pairs of partitions b_i, c_i), but all of them are different from the optimal partition. On the second behavior (bottom), observe that the sampling transitate through several significantly different partitions.

To do so, we apply both techniques on the same family of models (SBMs) and addressing the same predictive problem (detecting missing links in the same real networks). We discovered that sampling over models significantly improves the predictive power than considering an optimal set of parameters alone, suggesting that the error at estimating eq. 1.7 when applying the single-point method is not negligible.

Moreover, we compare the predictive power of several specializations of the SBM provided in the literature: degree-corrected [50, 73] and/or hierarchical nested [76]. We observe that even the simplest sampling approach (without specializations) outperforms the single-point with specializations.

Regarding the measure used (accuracy at predicting missing links), we verify its consistency with Σ from Eq. 2.2. The optimal point-estimate has minimum Σ , so we expect better accuracies for those partitions that are more likely. We observe that the larger the accuracy the lower the Σ , for those real world networks considered.

Interestingly, we observed that the degree-corrected version defined in previous studies underperform the uncorrected version in two out of four real world networks considered, suggesting a further revision on which cases does this specification actually fits better the data. Additionally, we shape the sampling inference with the degree-corrected specialization, a similar improvement/retreat observed in the single-point is confirmed with the sampling degree-corrected version.

Finally, we investigate further in the difference between the sampling and the single-point. We examined in more detail the sampling process by observing the likely partitions explored. The sampling explores around similar partitions occasionally transiting to another different set of similar partitions, all of them different from the optimal one.

Our study provides a reliable framework to compare different methodologies.

UNIVERSITAT ROVIRA I VIRGILI
NETWORK INFERENCE BASED ON STOCHASTIC BLOCK MODELS: MODEL EXTENSIONS, INFERENCE APPROACHES
AND APPLICATIONS
Toni Vallès Català

Suture fusion in normal and pathological development is constrained by the network architecture of the human skull

3.1 Introduction

Craniofacial sutures are primary sites of bone growth and remodeling; adequate formation and maintenance of sutures is therefore important for a healthy development of the head and brain. Sutures fuse as part of the normal developmental process of the skull when taking place at the right time. However, deviations from the normal process of suture pattern formation in the human skull usually cause birth defects, such as cleft palate and craniosynostosis. The former is a condition in which the bones of the hard palate (maxilla and palatine) fail to form the midpalatal suture, with a prevalence of about 15 in 10,000 live births [103]; this condition is often associated with a cleft lip, causing health and social problems for the child [103]. The latter is a condition in which one or more sutures between cranial bones (frontal, parietal, temporal, and occipital) fuse prematurely, with a prevalence of about 5 in 10,000 live births [21]; when these premature fusions are not treated surgically, they can cause head malformations due to compensatory growth of other sutures [19], sometimes provoking severe brain damage due to an increase of intracranial pressure [48]. Both conditions can occur in isolation [29, 103], or as part of a variety of congenital disorders, such as Van der Woude and X-linked intellectual disability syndromes [103], or Apert and Crouzon syndromes [83].

Genetic and epigenetic factors participate in the formation and maintenance of craniofacial sutures. The number of genes identified carrying mutations associated with these two pathologies has grown in the last two decades [103, 101]. For example, more than 60 genes are now known to carry mutations associated with craniosynostosis [101]: some of them show specificity for a suture in the context of a syndrome (e.g.,

Suture fusion in normal and pathological development is constrained by the network architecture
of the human skull

ASXL1 and metopic suture in the Bohring-Opitz syndrome), others predispose to more than one type of craniosynostosis (e.g., *FGFR2* in coronal, sagittal, and multisuture synostoses), while most of them are not specifically associated with suture development, but to osteogenesis in general (e.g., *ALX4*, *EFNA4*, and *TGFBR2*). Epigenetic factors include, among many others, bio-mechanical stress, hypoxia, and use of drugs during pregnancy [71, 77, 103]. Epigenetic factors are even less specific than genetic ones; for example, maternal smoking has been associated to a predisposition for both cleft palate and some craniosynostoses [12, 93].

Only a small fraction of the more than 80 articulations that make up the human skull before birth are associated with these birth defects. However, we still do not know which factors predispose some sutures but no others to fuse pathologically or to not form at all. Here we address this question by modeling the skull as a network in which nodes and links formalize bones and their articulations (Fig. 3.1). We use the reliability formalism developed for network models [44] to infer the susceptibility of craniofacial sutures to be lost in pathological conditions.

Anatomical network models have been used before, for example, to identify developmental constraints in skull evolution [23, 24], analyze the evolution of tetrapod disparity in morphospace across phylogeny [26], and model the growth of human skull bones [25]. A recent comparison of network models of craniosynostosis conditions showed that, despite the associated abnormal shape variation, skulls with different types of craniosynostosis share a same general pattern of network modules [27].

Here, we want to investigate whether the topological arrangement of bones and articulations predicts which sutures are more susceptible to be lost; in other words, we want to assess if the architecture of the skull acts as an agent that constraints the formation and fusion of sutures. A common feature of the topology of complex networks such as the skull is that one can identify groups of nodes (bones) that have well-defined patterns of connections (i.e., articulations) with other groups of nodes [44]. Such realization allows one to identify links that are topologically unexpected. If the architecture of the skull is driving the fusion of articulations, we surmise that there is a relationship between the susceptibility of an articulation to fuse and the topological 'unexpectedness' of such articulation. To quantify such susceptibility, we use the *link reliability* score, that is the probability that a connection exists in the network given the observed (newborn) topology of the skull [44]. A low score means that the presence of this articulation is rare, that is, not commonly expected in the given arrangement of bones (see *Methods* for details on how this is estimated). Importantly, the link reliability formalism has been used in other complex systems to accurately predicting missing and spurious interactions in social, neural, and molecular networks [44], to predict harmful interactions between pairs of drugs [45] and to predict the apparition of conflicts in teams [86].

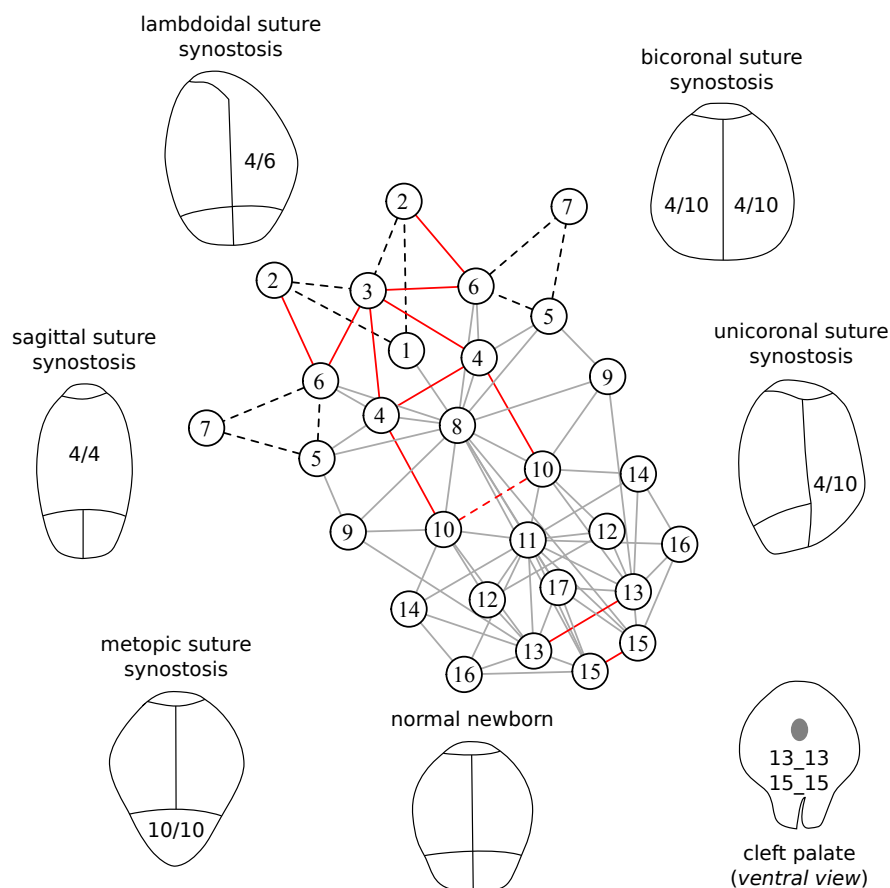


Figure 3.1: The arrangement of bones in the human skull at birth modeled as a network; nodes and links represent bones and articulations (sutures and synchondroses). Red links indicate sutures associated with cleft palate and craniosynostosis conditions; dashed links are articulations lost during the normal development due to fusion. Note that the metopic suture between the left and right frontal bones fuses in pathological and normal development. Drawings illustrate the shape of the head in some of the conditions studied. *Labels: 1 basilar, 2 lateral occipital, 3 occipital plane, 4 parietal, 5 squama, 6 petromastoid, 7 tympanic ring, 8 sphenoidal, 9 zygomatic, 10 frontal, 11 ethmoidal, 12 nasal, 13 maxilla, 14 lacrimal, 15 palatine, 16 nasal concha, 17 vomer.*

3.2 Methods

3.2.1 A network model of the skull

We built a network model of the human skull at birth based on anatomical descriptions [35] and information of ossification timing and fusion events [96]. The nodes and

Suture fusion in normal and pathological development is constrained by the network architecture
 of the human skull

links of the network model formalize the bones and articulations (sutures and synchondroses) of the skull, respectively (Fig. 3.1).

3.2.2 Statistical analysis

We divided the links of the network model of the skull into three groups: (group 1) associated with cleft palate and craniosynostosis, (group 2) fused during a normal development, and (group 3) not fused during normal development. Groups follow the literature reviews listed in the references and common knowledge. TABLE 3.1 shows the sutures most commonly associated with cleft palate and craniosynostosis, as well as those sutures that fuse in a normal development of the human skull.

We performed an independent one-sided Mann-Whitney U test for the following pairs of groups: group 2 vs. group 3, group 1 vs. group 3, and group 1 vs. group 2. We tested the null hypothesis of equal distribution between groups against the alternative hypothesis that:

1. sutures in group 2 have lower reliability scores than sutures in groups 1+3;
2. sutures in group 1 have lower reliability than those in group 3;
3. sutures in group 2 have lower reliability scores than those in group 1.

We estimated the median of the difference between groups and a non-parametric 95% confidence interval. The statistical analysis was performed with function `wilcox.test` in R [99].

Table 3.1: **Articulations modified during cleft palate, craniosynostosis, and normal development.**

Craniofacial joint	Between	Condition
intermaxillary	left and right maxilla	cleft palate
interpalatal	left and right palatine	cleft palate
sagittal	left and right parietal	craniosynostosis
coronal	frontal and parietal	craniosynostosis
lambdoid	parietal and occipital plane	craniosynostosis
occipitomastoid	petromastoid and occipital plate	craniosynostosis
	petromastoid and lateral occipital	
metopic	left and right frontal	craniosynostosis
		normal development of the frontal bone
petrosquamosal	petromastoid and squamosal	normal development of the temporal bone
petrotympanic	petromastoid and tympanic ring	normal development of the temporal bone
squamotympanic	squamosal and tympanic ring	normal development of the temporal bone
basilateral	basilar and lateral occipital	normal development of the occipital bone
occipitolateral	lateral occipital and occipital plane	normal development of the occipital bone

3.3 Results

First we investigated the relationship between the link reliability score and the susceptibility of an articulation to fuse during normal development. To that end, we compared the reliability score of those articulations that fuse during the normal development of the skull to those that do not. We find that sutures that normally fuse have significantly lower reliability scores than those that do not (Mann-Whitney-Wilcoxon: one sided $W=206.0$, $p\text{-value} = 0.0055$; $\text{Mean}(\text{fused}) = 0.3485$; $95\% \text{ CI}(\text{non-fused}) = (0.4124, \text{inf})$)(Fig. 3.2); which is in agreement with our hypothesis that during normal development there is a tendency to lose articulations that are topologically rare in the newborn skull.

Next, we investigated whether sutures that fuse in pathological conditions are topologically different from sutures that do not form (including no-links, i.e., articulations that never appear among bones but that can be analyzed thanks to the network model of the skull). To that end, we compared the reliability score of sutures that occur in cleft palate and craniosynostosis conditions to that of those sutures that do not form or do not fuse during the normal development of the skull (Fig. 3.2). We found that sutures associated with pathological conditions have significantly lower reliability scores than sutures that are not (one-sided, $W = 116$, $p\text{-value} = 1.022E^{-4}$; $\text{Mean}(\text{pathological}) = 0.3244$; $95\% \text{ CI}(\text{non-pathological and non-fused}) = (0.4417, \text{inf})$); which shows that sutures associated to pathological conditions are also unexpected from a topological point of view.

Interestingly, we find no statistical difference between the reliability scores of sutures that are associated to pathological conditions and those that fuse during normal development (two-sided $W = 44.5$, $p\text{-value} = 0.196$; $95\% \text{ CI}(\text{fused}) = (0.3389, 0.3668)$). This finding suggests that while skull architecture is an important factor in the loss of sutures during both pathological and normal development, there are non-topological factors that discriminate between normal and pathological loss of sutures.

3.4 Modularity

Modularity of the human skull is studied by establishing a priori developmental or functional hypothesis, that is then tested by means of patterns of covariation and correlation using different morphometric tools. Conversely, [22] suggested to use such morphological information to build networks and recognize modules in the human skulls, without a priori assumptions. We are interested whether if we can recover this modularity with the SBM approach, a fact that will support the idea that SBM strongly capture the structure of the human skull network.

When applying the approach suggested by [44] we group bones into different blocks depending on their patterns of connection, during the whole process we obtain a sample of several SBM. To find a fair partition of the bones into groups we count the number of times two bones are grouped in the same block, the counting is done for each possible pair of bones (shown in the co-classification matrix in Fig. 3.3). Then a higher proportion of times that two bones are grouped together is represented in a

Suture fusion in normal and pathological development is constrained by the network architecture
 of the human skull

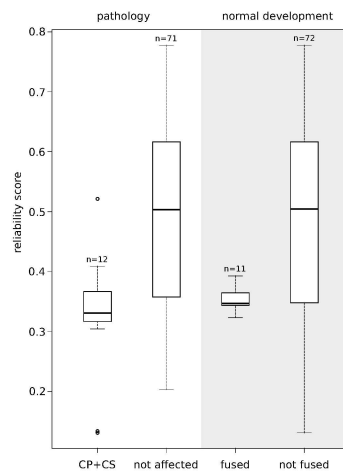


Figure 3.2: **Box plot comparing link reliability scores.** Sutures associated with cleft palate and craniosynostosis (CP+CS) have lower reliability than those that are not associated (*left*, white panel). Sutures that fuse during normal development also have lower reliability than those that do not do so (*right*, gray panel).

more red color in Fig. 3.3, meaning that is fair to put those two bones into the same group; on the other hand, a blue tendency represents that two bones are rarely grouped together. Note that we obtain a partition of the bones into a modularity structure in two groups (relative to *cranial* and *facial*), one of them integrating a hierarchical structure with three subgroups, this finding is consistent with previous research on modularity in the human skull [22].

Each pair of nodes represent a possible linkage, thus we can highlight those pair of bones whose suture is associated with cleft palate and craniosynostosis (white squares) and those fused during a normal development (black squares), we observe that bones fused in normal development are grouped together a higher amount of times than sutures associated to the mentioned pathologies.

We have found a similar modularity than that previously yielded by [22] with hierarchical clustering. Hence our approach is capable to describe closely the structure of skulls networks, without any prior information except the given sutures itself.

3.5 Evolution

Given the influence of the skull topology on the susceptibility of sutures to fuse, we speculate on the topological relevance on animals skulls. “Willinston’s Law” is an evolutionary theory which states that an organism tend to reduce into fewer parts more specialised in function. Supporting this theory, the tetrapod skull has undergone a simplification by loss and fusion of bones in all major lineages since the origin of

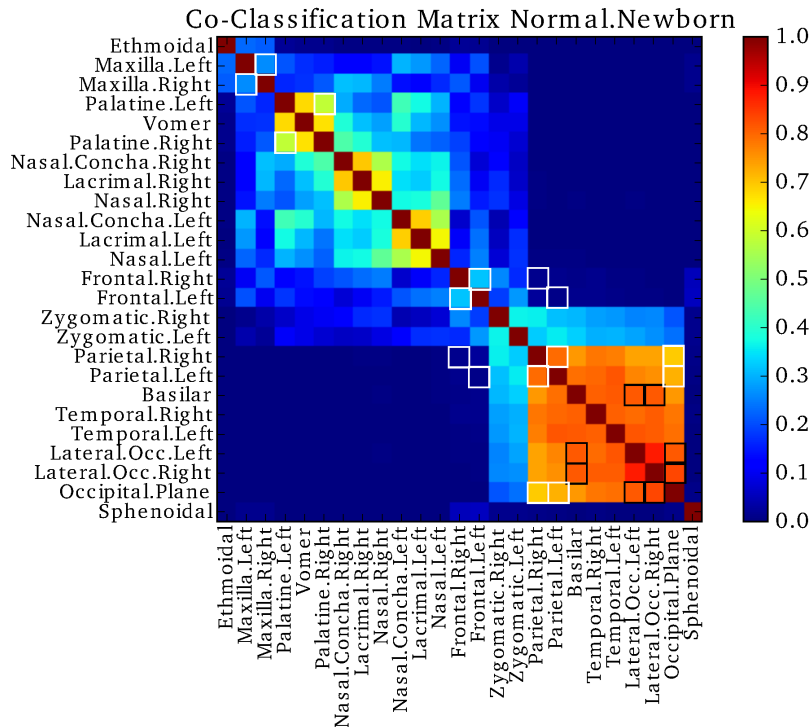


Figure 3.3: **Co-Classification matrix of the bones in the normal newborn skull.** Blue to red scale represents the amount of times two bones are grouped together during the whole sampling, normalized by the total number of steps. White(black) squares enclose those sutures associated with cleft palate and craniosynostosis (fused in normal development).

vertebrates, this reduction in the skull’s number of bones generated an evolutionary trend toward an increase in morphological complexity [24]. We will apply our new approach to study the topological effect on those tetrapod skull networks.

First we split all our available networks into extant (3.2) and extinct (3.3). Given a network we can compute the reliability of both the links and the no-links, in Fig. 3.5 we provide the boxplots of the links scores for each network. Additionally, by joining together all the scores of the links of extant species we consider the whole distribution of *extant species sutures* as a whole, conversely we can analogously join all the *extinct species sutures* (Fig. 3.4). We found that sutures in extant species have significantly higher reliability scores than sutures from extinct species (two-sided, $W = 1.432E^6$, $p\text{-value} = 3.887E^{-75}$; 95% CI(extant) (0.4045, 0.4172); 95% CI(extinct) (0.3199,

Suture fusion in normal and pathological development is constrained by the network architecture
of the human skull

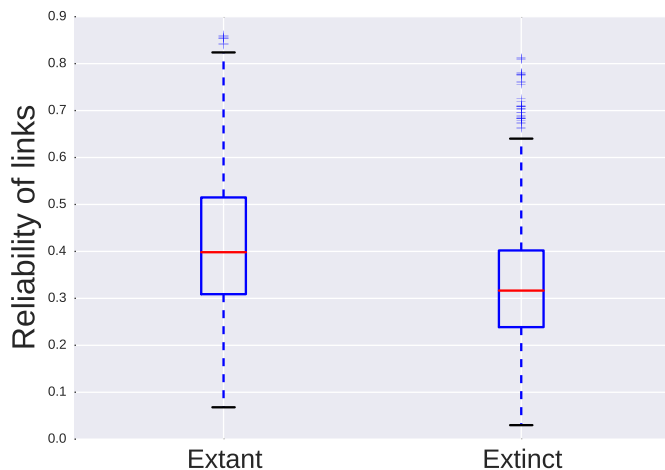


Figure 3.4: **Box plot comparing link reliability scores of all sutures from extant species together versus all sutures from extinct species together.** Sutures from extant species have a higher reliability mean than those that are extinct.

0.3315)); which shows that sutures from extinct animals are more unexpected from a topological point of view than those from extant animals.

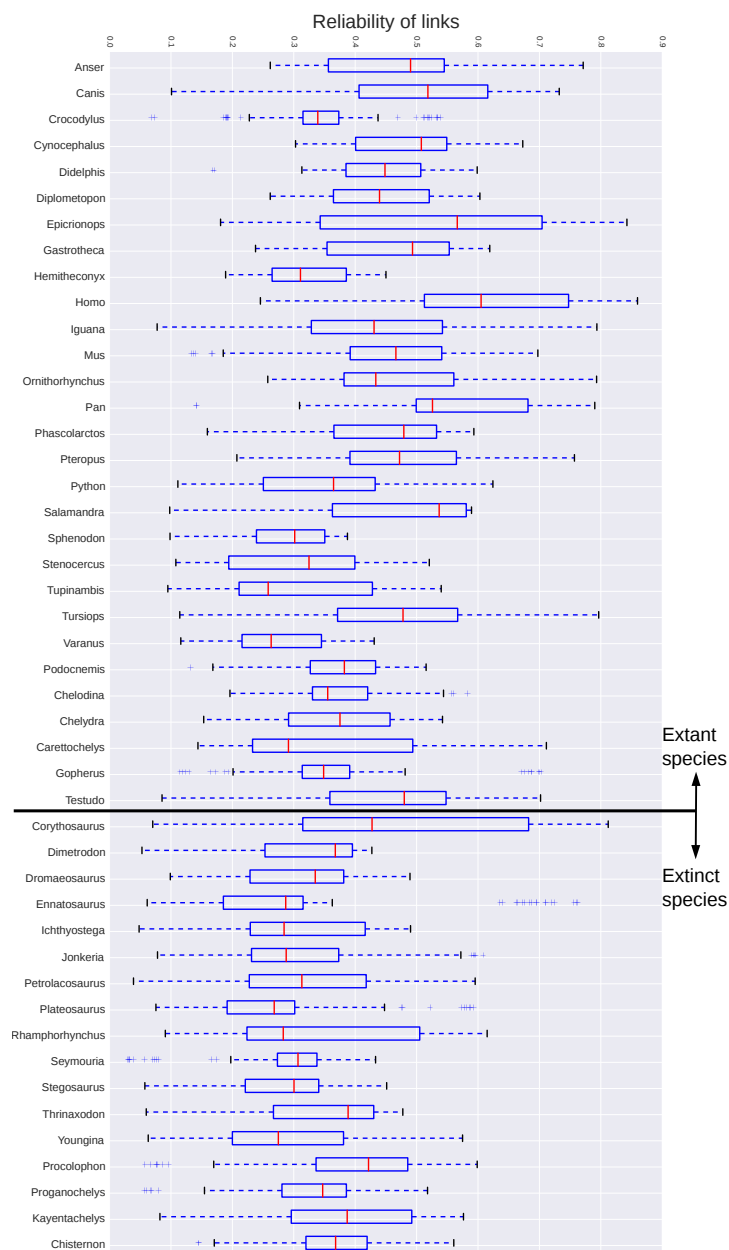


Figure 3.5: **Box plot comparing link reliability scores of all the extant/extinct species.** Su-
 tures from extant species have a tendency to higher reliability than those that are extinct.

Suture fusion in normal and pathological development is constrained by the network architecture
 of the human skull

Tetrapod specie	Number of nodes	Number of edges
<i>Anser anser</i> (Greylag goose)	18	27
<i>Canis lupus</i> (Gray wolf)	29	90
<i>Carettochelys insculpta</i> (Pig-nosed turtle)	36	92
<i>Chelodina longicollis</i>	33	80
<i>Chelydra serpentina</i> (Common snapping turtle)	36	91
<i>Crocodylus moreletii</i> (Morelet's crocodile)	39	97
<i>Cynocephalus</i> (Philippine flying lemur)	24	68
<i>Didelphis virginiana</i> (Virginia opossum)	26	66
<i>Diplometopon zarudnyi</i>	26	57
<i>Epicrionops petersi</i> (Peters' caecilian)	23	51
<i>Gastrotheca walkeri</i> (Walker's marsupial frog)	22	43
<i>Gopherus polyphemus</i> (Gopher tortoise)	36	90
<i>Hemitheconyx caudicinctus</i> (African fat-tailed gecko)	34	72
<i>Homo sapiens</i>	21	64
<i>Iguana iguana</i> (Green iguana)	42	12
<i>Mus musculus</i> (House mouse)	28	78
<i>Ornithorhynchus anatinus</i> (Platypus)	26	65
<i>Pan troglodytes</i> (Common chimpanzee)	23	60
<i>Phascolarctos cinereus</i> (Koala)	31	87
<i>Podocnemis unifilis</i> (Yellow-spotted river turtle)	34	90
<i>Pteropus lylei</i> (Lyle's flying fox)	21	47
<i>Python regius</i> (Ball python)	35	68
<i>Salamandra salamandra</i> (Fire salamander)	25	52
<i>Sphenodon punctatus</i> (Tuatara)	38	78
<i>Stenocercus guentheri</i>	44	97
<i>Testudo graeca</i> (Spur-thighed tortoise)	34	94
<i>Tupinambis teguixin</i> (Gold tegu)	42	94
<i>Tursiops truncatus</i> (Common bottlenose dolphin)	32	99
<i>Varanus salvator</i> (Asian water monitor)	42	85

Table 3.2: **Extant species from which skull networks are reconstructed.** Number of nodes and edges conforming each network.

Tetrapod specie	Geologic period	Number of nodes	Number of edges
<i>Chisternon sp</i>	Eocene	36	98
<i>Corythosaurus casuarius</i>	Upper Cretaceous Period	33	77
<i>Dimetrodon gigas</i>	Early Permian period,	45	11
<i>Dromaeosaurus albertensis</i>	Late Cretaceous	41	99
<i>Ennatosaurus tecton</i>	Wordian	52	12
<i>Ichthyostega sp</i>	Upper Devonian	56	14
<i>Kayentachelys aprix</i>	Early Jurassic	38	10
<i>Jonkeria ingens</i>	Last Permian	51	13
<i>Petrolacosaurus kansensis</i>	Late Carboniferous	55	13
<i>Plateosaurus engelhardti</i>	Late Triassic period	49	11
<i>Procolophon pricei</i>	Early Triassic	45	12
<i>Proganochelys quenstedti</i>	Late Triassic	43	11
<i>Rhamphorhynchus sp</i>	Late Jurassic	41	95
<i>Seymouria baylorensis</i>	Early Permian	56	14
<i>Stegosaurus armatus</i>	Late Jurassic	47	11
<i>Thrinaxodon liorhinus</i>	Permian and mid-Triassic	44	10
<i>Youngina capensis</i>	Late Permian	53	12

Table 3.3: **Extinct species from which skull networks are reconstructed and the geological period it belongs.** Number of nodes and edges conforming each network.

3.6 Discussion

Our results suggest that the whole architecture of craniofacial articulations of the skull might act itself as an epigenetic factor, making some sutures to be more susceptible to be lost than others. That some regions of the skull act epigenetically (e.g., via bio-mechanical signaling) to predispose sutures to a premature fusion was already proposed by Moss in the context of the functional matrix hypothesis [64]. Here we show that the most susceptible sutures to be prematurely fused (i.e., those with low reliability scores) are precisely the ones associated with cleft palate and craniosynostosis. Thus, we propose that the very arrangement of bones in the skull predisposes epigenetically some sutures as targets of pathological conditions. We are not yet in a position to offer a mechanistic explanation for the relationship reported here, which we believe may be related to the same developmental mechanism that regulate compensatory growth of bones after premature synostoses [19, 63, 58]. However, our results also suggest that such mechanisms should not be different between normal and pathological conditions, since sutures and synchondroses that are lost during normal development of the skull also show low reliability scores.

If, as our results suggest, the system of articulations of skull bones is able to self-regulate epigenetically the formation and maintenance of sutures and synchondroses, this might have consequences also at an evolutionary scale. In craniosynostosis con-

Suture fusion in normal and pathological development is constrained by the network architecture
of the human skull

ditions, the number of bones is reduced due to the early suture fusions, much in the same way as the net reduction in the number of bones during vertebrate evolution [36, 94, 23]; as a consequence, it has been postulated that craniosynostosis could be used as an informative model for skull evolution [84]. Our results suggest that this is not a mere analogy, but that similar epigenetic processes might act in regulating the configuration of bone arrangements in the skull both in development and in evolution.

Pathological conditions of the human skull, including cleft palate and craniosynostosis, are a medical and social problem that needs special attention from the research community. In addition, they represent medical examples of more general developmental and evolutionary processes found in all tetrapods. Both aspects, the medical and the biological, need and can be integrated in order to reach a better understanding that could lead to improve treatments as well as to further our knowledge about fundamental evolutionary questions.

4

Multilayer stochastic block models reveal the multilayer structure of complex networks

4.1 Introduction

In the introduction chapter 1 and previous chapters 2, 3, we highlighted the convenience of using Stochastic Block Models (SBM) based approaches at modeling real-world networks.

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 encompass relationships that arise on the familiar layer and relationships that arise in the professional layer) [54]. 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 [79, 80, 34, 16, 15, 90]. 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.

In this chapter we precisely address the problem of unraveling the underlying multilayer structure in real-world networks. First, we introduce the family of multilayer SBMs that generalizes single-layer SBMs to situations where links arise in different layers and are aggregated. Although there have been proposals to extend the concept of modularity to multilayer networks [65], 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. Our approach is also different from so-called latent feature models [61, 72, 52] in that SBMs allow to answer the fundamental

question of whether an observed network is the outcome of a multilayer process, while in latent feature models it is impossible to disentangle the contributions of each layer.

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. The analysis of complex networks from different contexts suggests that many real-world networks are indeed projections.

4.2 Multilayer stochastic block models

In our approach, nodes interact in different layers. In each one of these layers $\ell = 1, \dots, L$ 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 , respectively, in layer ℓ exist with probability $q_{\alpha\beta}^\ell$. The observed adjacency matrix $A^\mathcal{O}$ is an aggregate that results from the combination of the links in each of the layers, and where all information of the layers has been lost (Fig. 4.1). We call this model the multilayer SBM.

Here we consider the simplest multilayer case and set $L = 2$. In such case, there are two combinations with a plausible physical interpretation: i) the *AND* combination of layers, in which $A_{ij}^\mathcal{O} = 1$ if, and only if, nodes i and j are connected in both layers (Fig. 4.1(a)); ii) the *OR* combination of layers, in which $A_{ij}^\mathcal{O} = 1$ if i and j are connected in at least one layer (Fig. 4.1(b)). 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 [104], because some people use Facebook to share memes, others use Twitter, 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, which has no information about the layers. 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

$$p(\mathcal{P}_1, \mathcal{P}_2 | A^\mathcal{O}) \propto \int DQ_1 \int DQ_2 p(A^\mathcal{O} | Q_1, Q_2, \mathcal{P}_1, \mathcal{P}_2) p(Q_1, Q_2, \mathcal{P}_1, \mathcal{P}_2) \quad (4.1)$$

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 $\int DQ_\ell \equiv \prod_{\alpha < \beta} \int_0^1 dq_{\alpha\beta}^\ell$ is the integral over all possible values of these probabilities.

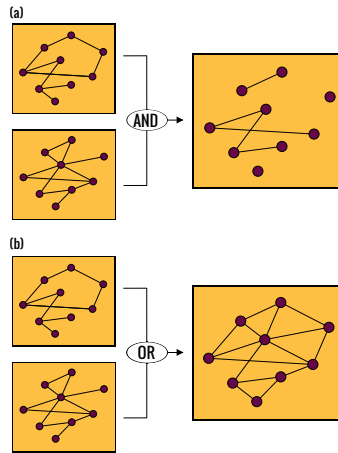


Figure 4.1: **Network aggregation mechanisms.** In aggregated multi-layer networks, different networks containing the same nodes but with different adjacency matrices are combined into an observed network with adjacency matrix $A^\mathcal{O}$ where all information about the original layers has been lost. We consider two aggregation mechanisms of two-layer networks with adjacency matrices A^1 and A^2 : **(a)** AND aggregation, in which $A_{ij}^\mathcal{O} = A_{ij}^1 A_{ij}^2$, so that $A^\mathcal{O} = 1$ if, and only if, i and j are connected in both layers. **(b)** OR aggregation in which $A_{ij}^\mathcal{O} = 1 - (1 - A_{ij}^1)(1 - A_{ij}^2)$, so that $A^\mathcal{O} = 1$ if i and j are connected in at least one layer.

This integral can be computed both for AND combinations and for OR combinations of the two layers; for clarity, we first show the calculation for the AND model and discuss the OR model in 4.2.3. Because in a SBM each link 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 likelihood for an AND model is

$$\begin{aligned}
 p_{\text{AND}}(A^\mathcal{O} | Q_1, Q_2, \mathcal{P}_1, \mathcal{P}_2) &= \\
 &= \prod_{\substack{[\alpha \leq \beta] \\ [\gamma \leq \delta]}} (q_{\alpha\beta}^1 q_{\gamma\delta}^2)^{n_{\alpha\beta\gamma\delta}^1} (1 - q_{\alpha\beta}^1 q_{\gamma\delta}^2)^{n_{\alpha\beta\gamma\delta}^0}, \quad (4.2)
 \end{aligned}$$

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 no-links between such pairs of nodes ($n_{\alpha\beta\gamma\delta}^0 = \sum_{i < j} (1 - A_{ij}^\mathcal{O}) \delta_{\sigma_i^1 \alpha} \delta_{\sigma_j^1 \beta} \delta_{\sigma_i^2 \gamma} \delta_{\sigma_j^2 \delta}$).

We can plug Eq. (4.2) into Eq. (4.1) and integrate for $P_{\text{AND}}(\mathcal{P}_1 \mathcal{P}_2 | A^\mathcal{O})$ over $q_{\alpha\beta}^1$ and $q_{\gamma\delta}^2$, assuming uniform distribution for the prior $p(Q_1, Q_2, \mathcal{P}_1, \mathcal{P}_2) = \text{const}$ [44]. Other possibilities include choosing non-uniform priors for the connection probabilities [74, 76, 91] or different priors for the partitions [18, 91, 74, 76, 75, 108]. To simplify the notation, we introduce two indices r and s , so that $r \equiv \alpha\beta$ and $s \equiv \gamma\delta$ and we drop the reference to layer 1 and 2 so that $q_r \equiv q_{\alpha\beta}^1$, and $q_s \equiv q_{\gamma\delta}^2$. In order to perform

Multilayer stochastic block models reveal the multilayer structure of complex networks

the integration over q_r , for example, we note that all the terms that contain q_r have the following form:

$$q_r^{\sum_s n_{rs}^1} \prod_s (1 - q_r q_s)^{n_{rs}^0} = q_r^{n_r^1} \prod_s \sum_{m_{rs}=0, \dots, n_{rs}^0} \binom{n_{rs}^0}{m_{rs}} (-)^{m_{rs}} (q_r q_s)^{m_{rs}} \quad (4.3)$$

where $n_r^1 = \sum_s n_{rs}^1$. Then for fixed values of $\{m_{rs}\}$ we have that the contribution to the likelihood factorizes for every q_r and q_s as follows

$$\int DQ_r \int DQ_s \prod_{r,s} \binom{n_{rs}^0}{m_{rs}} (-)^{m_{rs}} \prod_r q_r^{n_r^1 + m_r} \prod_s q_s^{n_s^1 + m_s}, \quad (4.4)$$

where $\int DQ_r = \prod_r \left(\int_0^1 dq_r \right)$ and $m_r \equiv \sum_s m_{rs}$.

Integrating out the q_r s and q_s s we obtain for the likelihood the expression in Eq. (4.5).

$$p_{\text{AND}}(\mathcal{P}_1, \mathcal{P}_2 | A^{\mathcal{O}}) \propto \sum_{\substack{\{m_{rs}\} \\ [m_{rs}=0, \dots, n_{rs}^0]}} \frac{\prod_{r,s} (-1)^{m_{rs}} \binom{n_{rs}^0}{m_{rs}}}{\prod_r (n_r^1 + m_r + 1) \prod_s (n_s^1 + m_s + 1)} \quad (4.5)$$

where the summation is over all possible values of each m_{rs} . Note that it is straightforward to apply the same formalism to directed networks by considering a bipartite graph of nodes with incoming and outgoing connections. In this case, for each layer we would have an SBM with two sets of block partitions, one for nodes with outgoing connections and one for nodes with incoming connections, and a non-symmetric connection probability matrix (see [86]).

Given Eq. (4.5), 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.

4.2.1 Computation of degeneracies

Our goal is to compute the number $D(\mathcal{P}_I)$ of pairs $(\mathcal{P}_1, \mathcal{P}_2)$ that have the same intersection partition \mathcal{P}_I , that is, the cardinality of the set $\{(\mathcal{P}_i, \mathcal{P}_j) | \mathcal{P}_i \cap \mathcal{P}_j = \mathcal{P}_I\}$. We

start by noting that a specific \mathcal{P}_I consists of n groups of nodes that we call “elements;” we make explicit the number of such elements in an intersection partition and write $\mathcal{P}_I^n = [E_1][E_2][E_3] \dots [E_n]$. By the definition of intersection partition we have that: (i) all the nodes within an element must belong to the same group in both partitions \mathcal{P}_1 and \mathcal{P}_2 (otherwise, they would not belong to the same element); (ii) two elements cannot belong to the same group in both \mathcal{P}_1 and \mathcal{P}_2 (otherwise they would be a single element). We compute the degeneracy in two steps (see Supplementary Material for details): 1) We compute all the possible unique partitions \mathcal{P}_1 combining the elements in \mathcal{P}_I^n , group them in *classes* according to the numbers of elements combined, and compute the multiplicity associated to each class; 2) For each class, we compute all the possible partitions \mathcal{P}_2 that result in a specific intersection \mathcal{P}_I^n .

4.2.2 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. 4.2(a)). The reverse is also true, so the possible network models one can generate with single-layer SBMs and multilayer SBMs are, in fact, identical. For instance, a specific one-layer SBM is equivalent to a two-layer AND model in which one of the layers has a single group and connection probability matrix with all entries equal to 1 and another layer equal to the single-layer SBM. 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.

In the single-layer SBM, each group comprises the nodes that belong to the same pair of groups α, γ in \mathcal{P}_1 and β, δ 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 Sec. 4.2.3). 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 being the result of the product of two factors, are all independent of each other (see Fig. 4.2(b)). Since this approximation is equivalent to integrating separately every term with a different $(\alpha, \beta, \gamma, \delta)$ combination in Eq. (4.2), 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

Multilayer stochastic block models reveal the multilayer structure of complex networks

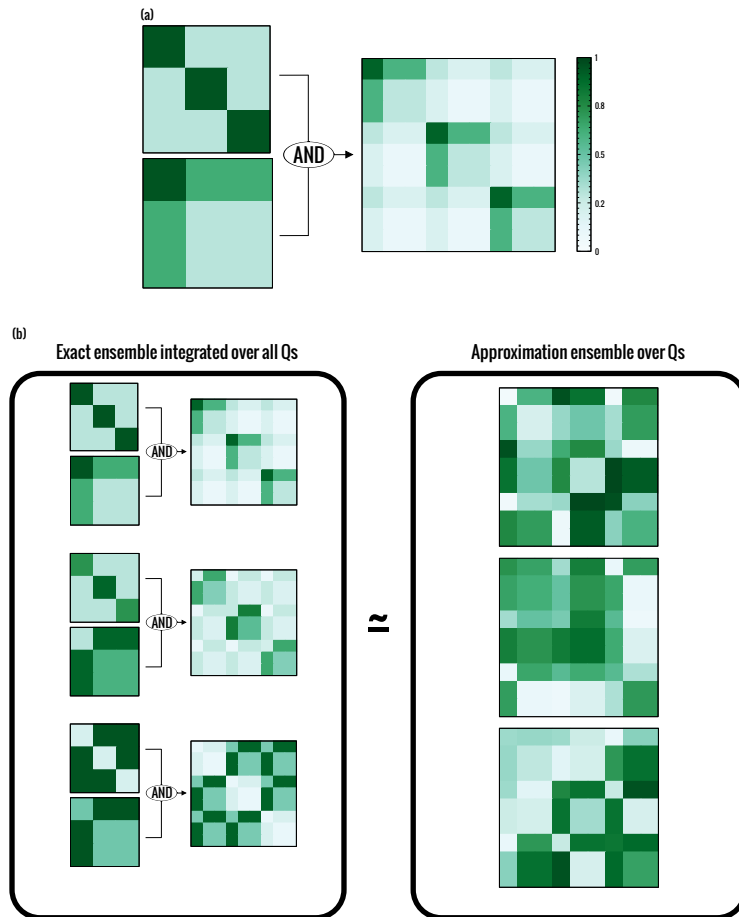


Figure 4.2: **Exact and approximate multilayer SBM ensembles.** (a) Two independent single-layer SBMs aggregated using the AND mechanism. We represent each single-layer SBM by its node-to-node connection probability matrix (indicated in the shades of green shown in the colorbar; note that node ordering is different in each SBM). 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 γ 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 γ in \mathcal{P}_2 , and group s in \mathcal{P}_I corresponds to groups β in \mathcal{P}_1 and δ in \mathcal{P}_2 , then the probability of connection in the single-layer SBM is $q_{r,s}^{\text{AND}} = q_{\alpha,\beta}^1 q_{\gamma,\delta}^2$. (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. (4.5)). For each pair (Q_1, Q_2) , the resulting $q_{r,s}^{\text{AND}}$ are therefore correlated. In our approximation, we assume that the elements of the intersection $q_{r,s}^{\text{AND}}$ are randomly drawn and independent of each other.

of single-layer and multilayer SBMs in the problem of detecting missing and spurious links in noisy networks [44]. In fact, we argue that, if (approximate) multilayer SBMs yield better predictions on real networks, then there is evidence (supported by our results) 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 [13, 44], the goal is to compute the probability $p(A_{ij} = 1|A^\circ)$ that a link between nodes i and j truly exists ($A_{ij} = 1$) given a noisy network observation A° , 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 $R_{ij} = p(A_{ij} = 1|A^\circ)$ 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 [44]

$$R_{ij}^{\mathcal{M}} = \frac{\int_{\mathcal{M}} dM p(A_{ij} = 1|M) p(A^\circ|M) p(M)}{Z}, \quad (4.6)$$

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. (4.1)); 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. However, always within our approximation, one can exploit the fact that the integral in Eq. (4.6) 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)

$$R_{ij}^{\text{AND}} = \frac{1}{Z} \sum_{\mathcal{P}_I} \left(\frac{n_{\sigma_i \sigma_j}^1 + 1}{n_{\sigma_i \sigma_j} + 2} \cdot \frac{\sum_{k=n_{\sigma_i \sigma_j}^1 + 2}^{n_{\sigma_i \sigma_j} + 2} \frac{1}{k}}{\sum_{k=n_{\sigma_i \sigma_j}^1 + 1}^{n_{\sigma_i \sigma_j} + 1} \frac{1}{k}} \cdot D(\mathcal{P}_I) \cdot e^{-\mathcal{H}(\mathcal{P}_I)} \right) \quad (4.7)$$

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 (possible links between) pairs of nodes in groups α and β , and $D(\mathcal{P}_I)$ the number of pairs $(\mathcal{P}_1, \mathcal{P}_2)$ that have the same intersection partition \mathcal{P}_I (the degeneracy of partition \mathcal{P}_I ; see Appendices). The *energy* \mathcal{H} is

$$\mathcal{H}(\mathcal{P}_I) = \sum_{\alpha < \beta} \left(\ln(n_{\alpha\beta} + 1) + \ln \binom{n_{\alpha\beta}}{n_{\alpha\beta}^1} - \ln \left(\sum_{k=n_{\alpha\beta}^1 + 1}^{n_{\alpha\beta} + 1} \frac{1}{k} \right) \right) \quad (4.8)$$

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

As in [44], the expression for the link reliability (Eq. (4.19)) 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 Appendices for details). Note that, although Eqs. (4.19)-(4.8) are the exact solution to the link inference problem with approximate multilayer stochastic block models, there is no mathematical guarantee that, in a finite amount of time, the Markov chain will sample the space of node partitions with the desired probabilities. In particular, the energy landscape may be rugged and the chain may get trapped in some region. However, we have performed equilibration tests that suggest that the chain is, indeed, sampling the space correctly. As it turns out, $\mathcal{H}(\mathcal{P}_I)$ is equal to the energy obtained assuming a single SBM (Eq. (S2), [44]), plus a term that accounts for the product of two probabilities that generate each element of the intersection probability matrix. In a Bayesian context, we can interpret this term and the degeneracy $D(\mathcal{P}_I)$ as non-uniform priors for the intersection partitions.

4.2.3 OR combination of layers

For the OR model, one can obtain an expression for the likelihood by noticing that the OR model is an AND model for the no-links, that is non-existing edges between pairs. The likelihood of the observed topology A^O given the model M_{OR} assuming two layers is then:

$$P(A^O|M_{\text{OR}}) = \prod_{\substack{[\alpha \leq \beta] \\ [\gamma \leq \delta]}} ((1 - q_{\alpha\beta}^a)(1 - q_{\gamma\delta}^b))^{n_{\alpha\beta\gamma\delta}^0} \times (1 - (1 - q_{\alpha\beta}^a)(1 - q_{\gamma\delta}^b))^{n_{\alpha\beta\gamma\delta}^1}, \quad (4.9)$$

where all quantities have the same definition as in Eq. (4.2).

Following the same steps as in the AND model, we obtain the following expression for $P_{\text{OR}}(\mathcal{P}_1\mathcal{P}_2|A^O)$:

$$P_{\text{OR}}(\mathcal{P}_1, \mathcal{P}_2|A^O) \propto \sum_{\substack{\{m_{rs}\} \\ [m_{rs}=0, \dots, n_{rs}^1]}} \frac{\prod_{r,s} (-1)^{m_{rs}} \binom{n_{rs}^1}{m_{rs}}}{\prod_r (n_r^0 + m_r + 1) \prod_s (n_s^0 + m_s + 1)} \quad (4.10)$$

where we have used the notation $r \equiv \alpha\beta$ and $s \equiv \gamma\delta$ and all the quantities have already been defined in the AND model.

Finally, one can compute the reliability for an OR combination of two layers as

$$R_{ij}^{\text{OR}} = 1 - \quad (4.11)$$

$$\frac{1}{Z} \sum_{\mathcal{P}_I} \left(\frac{n_{\sigma_i \sigma_j}^0 + 1}{n_{\sigma_i \sigma_j} + 2} \cdot \frac{\sum_{k=n_{\sigma_i \sigma_j}^0 + 2}^{n_{\sigma_i \sigma_j} + 2} \frac{1}{k}}{\sum_{k=n_{\sigma_i \sigma_j}^0 + 1}^{n_{\sigma_i \sigma_j} + 1} \frac{1}{k}} \cdot D(\mathcal{P}_I) \cdot e^{-\mathcal{H}(\mathcal{P}_I)} \right)$$

$$\mathcal{H}(\mathcal{P}_I) = \quad (4.12)$$

$$\sum_{\alpha \leq \beta \in \mathcal{P}_I} \left(\ln(n_{\alpha\beta} + 1) + \ln \left(\frac{n_{\alpha\beta}}{n_{\alpha\beta}^0} \right) - \ln \left(\sum_{k=n_{\alpha\beta}^0 + 1}^{n_{\alpha\beta} + 1} \frac{1}{k} \right) \right)$$

where, as before, the sum is over all possible (intersection) partitions, Z is a normalization constant and $D(\mathcal{P}_I)$ is the number of pairs of partitions that have the same intersection. In Eq. (4.12), the sum is over all distinct pairs of blocks within a fixed partition, $n_{\alpha\beta}^1 = \sum_{i \leq j} A_{ij} \delta_{\sigma_i \alpha} \delta_{\sigma_j \beta}$, $n_{\alpha\beta} = \sum_{i \leq j} \delta_{\sigma_i \alpha} \delta_{\sigma_j \beta}$, $n_{\alpha\beta}^0 = n_{\alpha\beta} - n_{\alpha\beta}^1$, and σ_i stands for the block to which node i belongs.

4.3 Validation of link reliability estimation in model networks

Now that we are able to estimate link reliabilities using our approximation to two-layer (AND and OR) SBMs (Eq. (4.19)), as well as single-layer SBMs [44], 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, then assuming a two layer structure should result in a higher accuracy.

Note that, because single-layer and two-layer models are identical models with a different prior, one may expect that they perform equally well in large enough networks. This is because when one has infinite available information about the system, the prior has no effect on the inference and therefore single-layer and two-layer models should be equally accurate. While this is indeed the case for simple modular SBMs whose group sizes increase with network size (see Fig. 4.11), this is not necessarily the case for real-world networks. Indeed, real-world networks have very heterogeneous connectivity patterns and groups can be arbitrarily small regardless of network size, which makes it impossible to gather infinite information about those groups. In that case, the choice of prior does affect the inference protocol so that we expect a difference in accuracy between single and two-layer SBMs. As we show in what follows, our results for all the real-world networks we study confirm that there are differences between predictions based on single-layer and two-layer models.

To identify the limits of detectability in terms of the choice of two-layer SBM model, we first construct a set of multilayer test networks that have a well-defined block structure in each of the two layers, and that are aggregated using the AND or OR models (see Fig. 4.3). We parametrize this ensemble of networks using two variables: i) the low to high connectivity ratio λ ; and ii) the average connectivity of nodes k . For

Multilayer stochastic block models reveal the multilayer structure of complex networks

a fixed value of k , we expect to obtain larger accuracies for the easy cases, that is for networks with a more marked block structure (i.e. low values of λ).

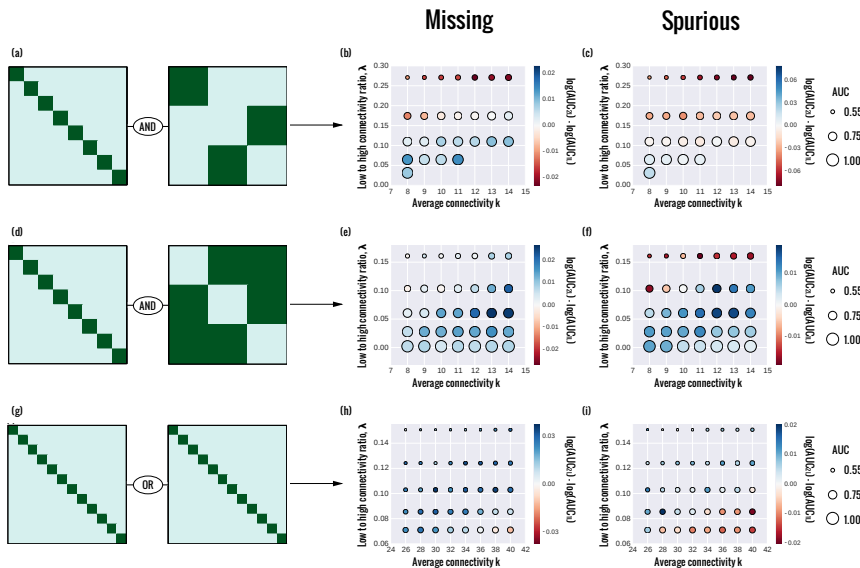


Figure 4.3: Performance of missing and spurious link identification on synthetic aggregated two-layer networks. Each row shows results for the different sets of two-layer SBMs illustrated in (a, d, g). We consider networks of $N = 168$ (a, d) and $N = 240$ (g) nodes divided into uniform groups in each layer. In the connection probability matrices, dark green represents a high connection probability p_h and light green a low connection probability p_l . We generate synthetic networks varying two parameters: the low-to-high connectivity ratio $\lambda = p_l/p_h < 1$, and the average connectivity k (see section). To compare the performance of the different approaches at detecting missing links (b, e, h), we randomly remove a fraction $f = 0.25$ of the links (false negatives) from the real network and calculate the reliability of each unobserved link. Then we calculate the AUC statistic, that is, we rank the links by decreasing reliability and calculate how often a removed link (false negative) has a higher reliability than a link that is truly non-existent in the real network (true negative). Analogously, to detect spurious links (c, f, i) we randomly add a fraction $f = 0.25$ of links (false positives), calculate the reliability of the observed links, and calculate how often an added link (false positive) has a lower reliability than a link that is truly existent in the real network (true positive). For each pair of parameter values, we generate 30 different synthetic networks. We compare the average 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 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. (See Figs. 4.4, 4.5, 4.6, 4.7, 4.8, 4.9 for results for other values of f (fraction of false negatives/false positives) and for synthetic networks generated for different numbers of nodes and/or groups).

We consider the predictive power of each of the approaches at detecting [13, 44]: 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. 4.3(a-f) and Figs. 4.4, 4.5, 4.6, 4.7) we find that, for the detection of both missing and spurious links, the two-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 both approaches do in fact work poorly).

For OR networks (Fig. 4.3(g-i) and Figs. 4.8, 4.9), the two-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 (low accuracy values) with lower connectivity.

Note that in the OR model the aggregated network is denser than each of the layers, whereas in the AND model the aggregate is sparser than each of the original layers. For this reason, we expect the AND model to produce better results in real-world networks, which are sparse. In fact, we should expect the OR model to produce better results only for networks obtained from our ensemble of OR two-layer stochastic block models, that is, networks obtained from an OR aggregation of SBMs with independent and uniformly distributed probabilities of connection between pairs of groups (according to our prior). Our results in Fig. 4.12 for such an ensemble of networks confirm that this is the case.

Multilayer stochastic block models reveal the multilayer structure of complex networks

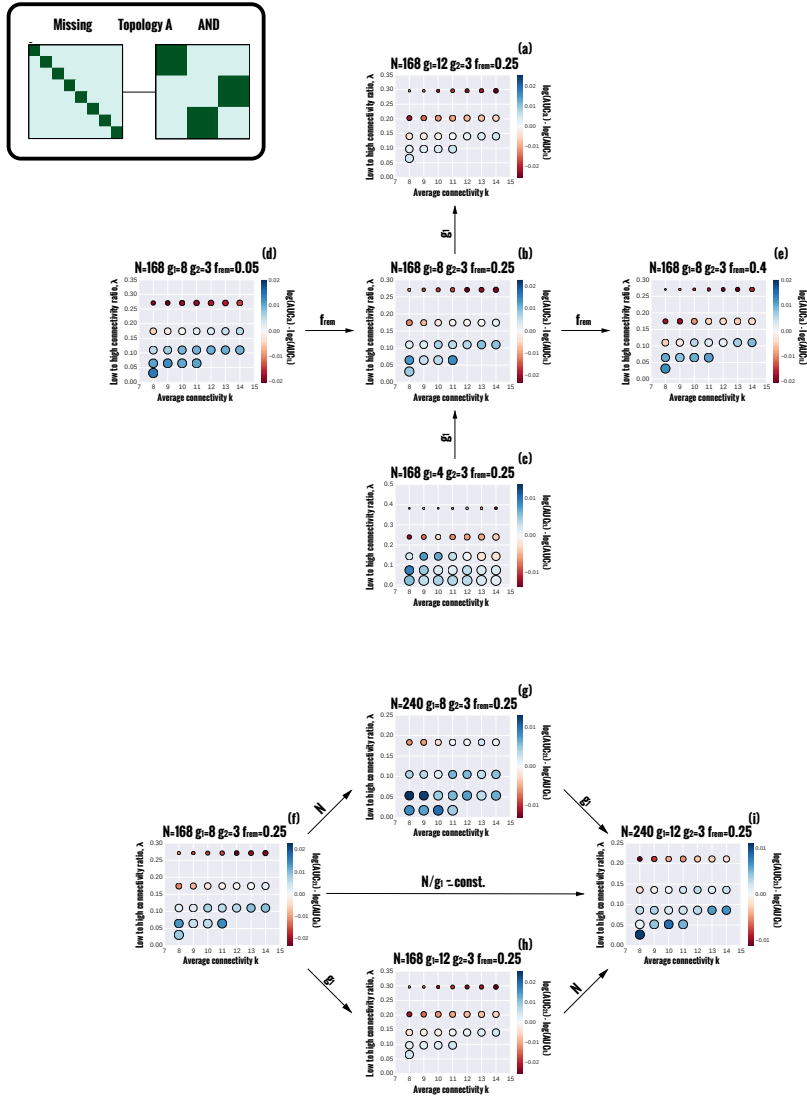


Figure 4.4: Performance of missing link identification on an ensemble of synthetic aggregated two-layer networks: AND Model; Topology A. Each panel corresponds to a different collection of two-layer SBMs, generated with fixed parameters denoted on each panel's title as done in Fig. 4.3. N : number of nodes; g_i : number of groups in layer i ; f_{rem} : probability of removed/added links. Arrows going between panels indicate which parameter is varied: f_{rem} (d,b,e), g_1 (a,b,c), N (g,f). The top left legend denote the topology of the generated networks, dark green corresponds to high connection probability p_h and light green to low connection probability p_l .

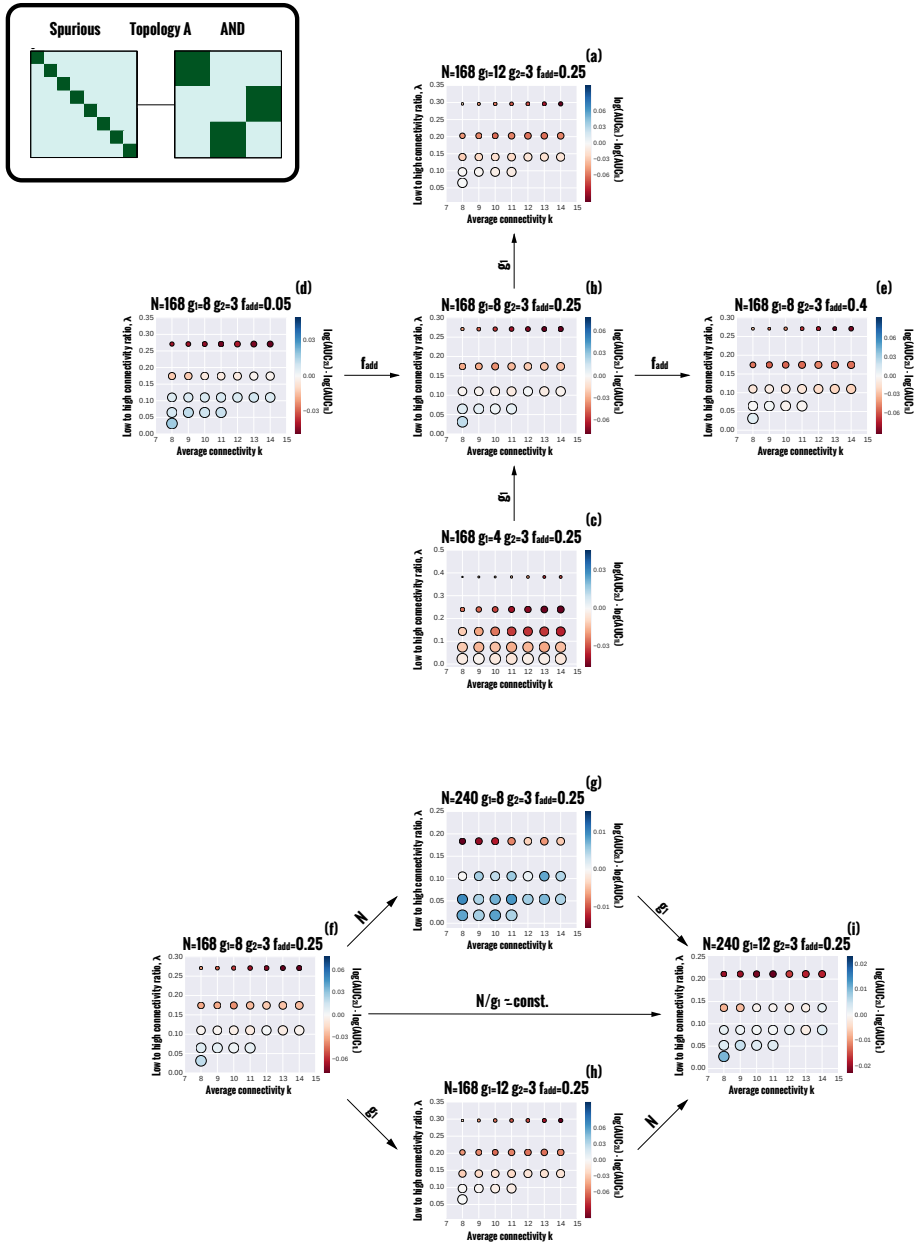


Figure 4.5: Performance of spurious link identification on an ensemble of synthetic aggregated two-layer networks: AND Model; Topology A.

Multilayer stochastic block models reveal the multilayer structure of complex networks

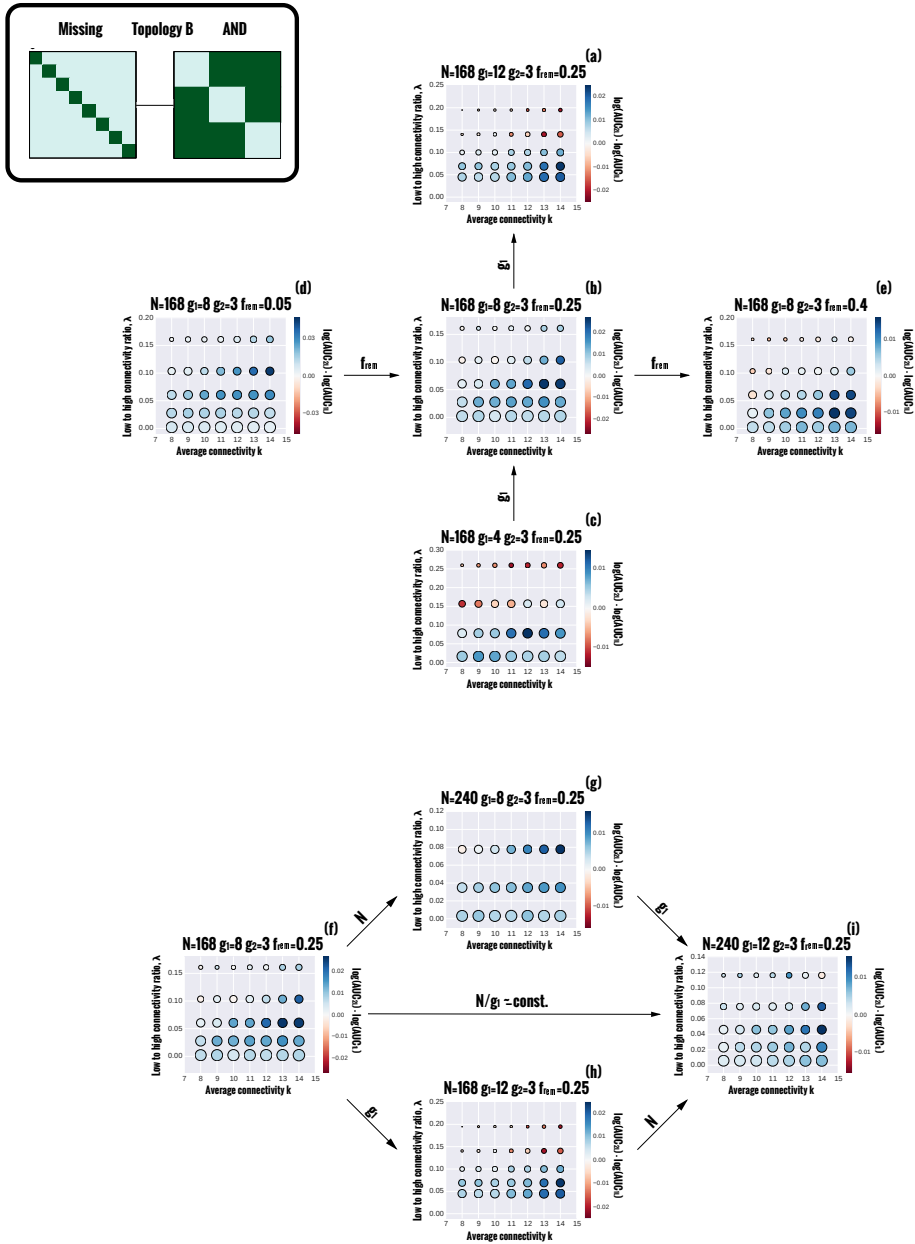


Figure 4.6: Performance of missing link identification on an ensemble of synthetic aggregated two-layer networks: AND Model; Topology B.

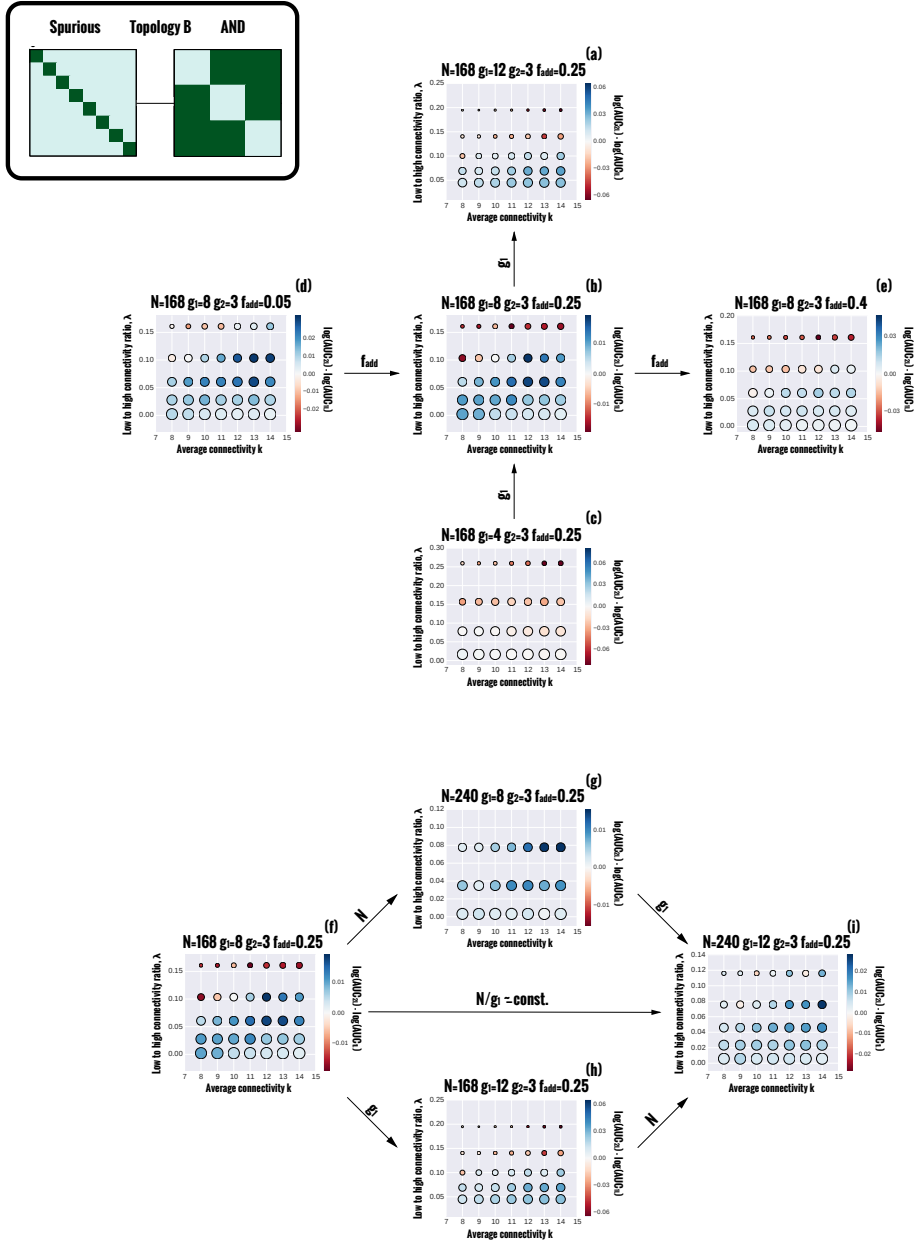


Figure 4.7: Performance of spurious link identification on an ensemble of synthetic aggregated two-layer networks: AND Model; Topology B.

Multilayer stochastic block models reveal the multilayer structure of complex networks

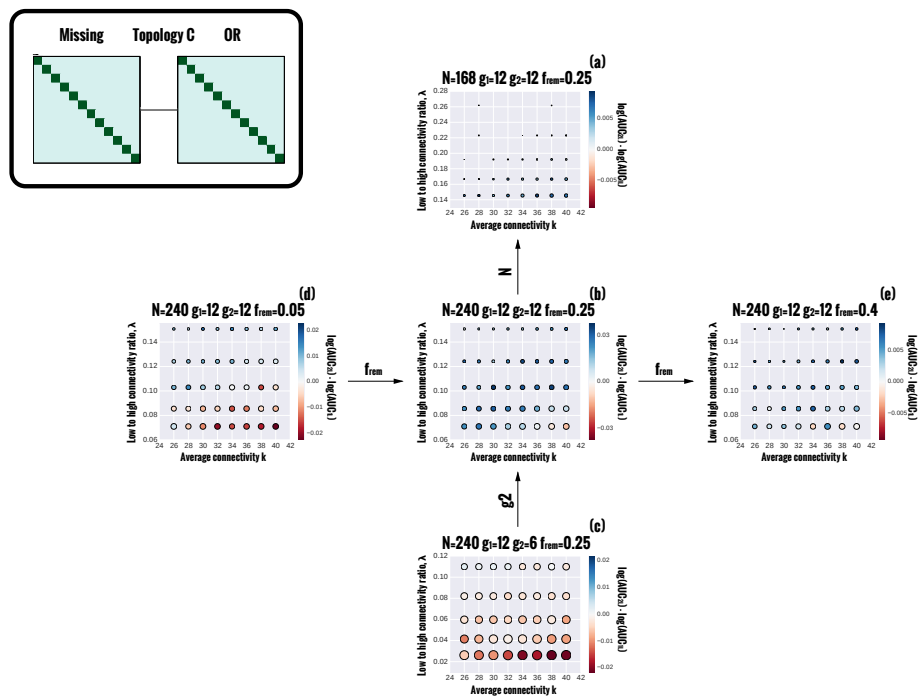


Figure 4.8: Performance of missing link identification on an ensemble of synthetic aggregated two-layer networks: OR Model; Topology C.

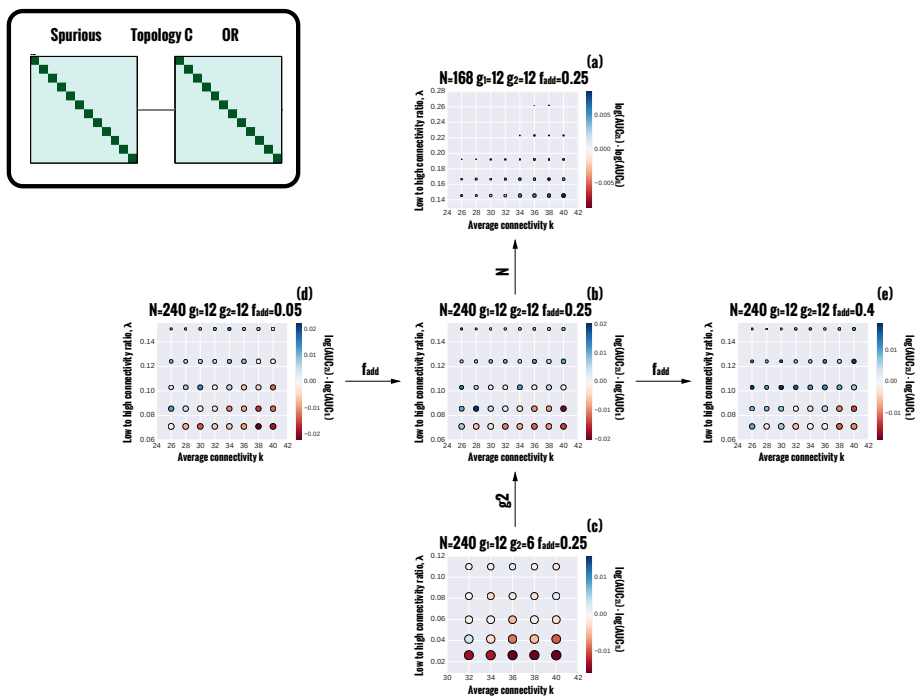


Figure 4.9: Performance of spurious link identification on an ensemble of synthetic aggregated two-layer networks: OR Model; Topology C.

4.3.1 Ensemble of two-layer synthetic networks

In order to validate our approximation, we define an ensemble of random networks that result from a planted multi stochastic block model structure with two layers that we aggregate using either the AND or the OR mechanism.

To generate the ensemble, we assume that we have two independent block models that describe the connections between a set of N nodes. In layer $l = 1, 2$, nodes are distributed into g_l groups, so that $\sigma_i^l, i = 1, \dots, N$ indicates the group to which node i belongs in layer l , and pairs of nodes (i, j) belonging to groups $(\sigma_i^l = \alpha, \sigma_j^l = \beta)$ are connected with probability $q_{\alpha\beta}^l$.

We expect that the existence of two layers will have the largest impact in the prediction of missing/spurious links when the block models in each layer are *as different as possible* in terms of the nodes they contain. Because there are a number of factors that influence the topology of the network, we make the following choices to define our ensemble:

1. For $l = 1$, we select g_1 groups comprising the same number of nodes (N/g_1); for $l = 2$, we select $g_2 \leq g_1$ (Fig. 4.10).
2. We select only two different values for $q_{\alpha\beta}^l \in \{p_h, p_l\}$, such that $p_h > p_l$. The ratio $\lambda = \frac{p_l}{p_h}$ characterizes the difference in density of connections between highly and lowly interconnected pairs of groups and has a significant impact on the topology of the single layer networks and on that of the aggregate network, respectively.
3. We select an assortative (community-like) SBM for layer $l = 1$, so that $q_{\alpha\beta}^1 = p_h$ for $\alpha = \beta$ and $q_{\alpha\beta}^1 = p_l$ otherwise. For layer $l = 2$, we select an SBM with a different profile for each group (see Fig. 4.10).

Note that once N, g_1, g_2, p_h, p_l , the $q_{\alpha\beta}^l$ s and the mechanism of aggregation (AND or OR) have been selected, the average connectivity in the aggregate network k is automatically fixed.

To assess the range of parameters in which our approach to multi-layer SBMs performs better than a single SBM at detecting missing and spurious links in synthetic networks, we sample over the following parameters: i) λ , ii) k and iii) g_1 . The reason we sample over these parameters is that we expect that the larger g_1 , the larger the number of groups in the intersection; the smaller the value of λ , the better defined the pattern of connections for each block; and, very small k (or k approaching the size of the network) results in harder detection of missing/spurious links.

Generation of ensemble networks with fixed k

Suppose we have a set of N nodes. In order to generate a network with a two-layer structure, we first generate two different networks for each layer and then we aggregate the two layers into a single one using the AND or the OR mechanisms. In the AND case, for a link to appear in the aggregated network it must exist in both layers; whilst

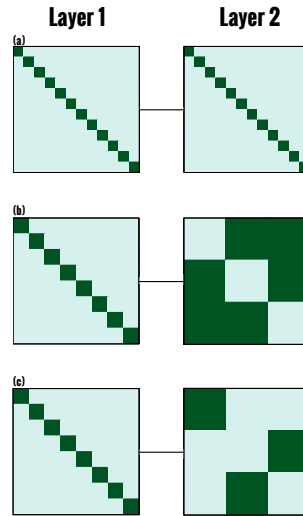


Figure 4.10: **Different collection of two-layer SBMs.** Dark green corresponds to high connection probability p_h and light green to low connection probability p_l . **(a)** For $l = 1$, we select $g_1 = g_2 = 12$ both assortative **(b)** For $l = 1$, we select $g_1 = 8$ assortative and $g_2 = 3$ disassortative **(c)** For $l = 1$, we select $g_1 = 8$ assortative and $g_2 = 3$ a mix between assortative and disassortative

in the OR case, for a link to exist in the aggregated network, it must appear in at least one of the two layers.

First we select the aggregation mechanism. Then, we randomly assign the nodes to a set of equally sized g_1 blocks and a set of equally sized g_2 blocks. We then select the $\{q_{\alpha\beta}^l\}$ s, that is the pattern of connections for each layer. Finally we select a value for the average connectivity of the aggregate network k and a value for λ and obtain the values of p_h and p_l that define the SBM.

To perform the last step, we need to obtain an expression for the average connectivity k as a function of p_h and p_l . For instance, let us consider an AND aggregated network from two assortative layers with g_1 and g_2 groups, so that $q_{\gamma\delta}^l = p_h \forall \alpha = \beta; l = 1, 2$ and $q_{\gamma\delta}^l = p_l$, otherwise. To obtain an expression for k , we consider the intersection SBM (which for the AND case is the SBM resulting from the product of the SBMs for $l = 1$ and 2). The intersection SBM will have $g_1 \cdot g_2$ blocks with $n_I = N/(g_1 \cdot g_2)$ each one. The matrix elements $q_{\gamma\beta\delta\epsilon}$ of the intersection will be as follows:

1. $q_{\gamma\beta\delta\epsilon} = p_h \cdot p_h$ if $\alpha = \beta$ and $\delta = \epsilon$. There will be $g_1 \cdot g_2$ matrix elements like this one that will be placed along the diagonal of the intersection SBM.
2. $q_{\gamma\beta\delta\epsilon} = p_h \cdot p_l$ if $\alpha = \beta$ and $\delta \neq \epsilon$. There will be $g_1 \cdot g_2 (g_2 - 1)$ matrix elements like this one.

Multilayer stochastic block models reveal the multilayer structure of complex networks

3. $q_{\gamma\beta\delta\epsilon} = p_l \cdot p_h$ if $\alpha \neq \beta$ and $\delta = \epsilon$. There will be $g_2 \cdot g_1(g_1 - 1)$ matrix elements like this one.
4. $q_{\gamma\beta\delta\epsilon} = p_l \cdot p_l$ if $\alpha \neq \beta$ and $\delta \neq \epsilon$. There will be $g_1(g_1 - 1) \cdot g_2(g_2 - 1)$ matrix elements like this one.

The average number of links expected between each pair of intersection blocks will be $l_{\gamma\beta\delta\epsilon} = q_{\gamma\beta\delta\epsilon} \cdot n_p(\gamma\beta\delta\epsilon)$, where $n_p(\gamma\beta\delta\epsilon) = n_I(n_I - 1)/2$ for $(\gamma = \beta, \delta = \epsilon)$ and $n_p(\gamma\beta\delta\epsilon) = n_I^2$ otherwise. Therefore the total number of expected links is

$$\begin{aligned}
 M &= \sum_{\alpha \leq \beta, \delta \leq \epsilon} q_{\gamma\beta\delta\epsilon} \cdot n_p(\gamma\beta\delta\epsilon) = \frac{1}{2} p_h^2 n_I(n_I - 1) g_1 g_2 \\
 &+ \frac{1}{2} p_h p_l n_I^2 (g_1(g_1 - 1)g_2 + g_1 g_2(g_2 - 1)) \\
 &+ \frac{1}{4} p_h^2 n_I^2 g_1(g_1 - 1)g_2(g_2 - 1). \tag{4.13}
 \end{aligned}$$

From here, we can express the average degree in terms of g_1, g_2, p_h and $\lambda = p_l/p_h$ as

$$\begin{aligned}
 k &= p_h^2 \cdot \left(\left(\frac{N}{g_1 \cdot g_2} - 1 \right) + \frac{N}{g_1 \cdot g_2} \cdot (g_1 + g_2 - 2) \cdot \lambda + \right. \\
 &\left. + \frac{N}{g_1 \cdot g_2} \cdot (g_1 - 1) \cdot (g_2 - 1) \cdot \lambda^2 \right). \tag{4.14}
 \end{aligned}$$

Note that we consider $0 \leq \lambda \leq 1$, therefore not all possible combinations k, p_h and λ are possible, which explains the empty regions in Fig. 3 in the main text.

Systematic validation over synthetic networks with varying parameters

We are interested in understanding which parameter choice renders the multilayer model more accurate at predicting missing and spurious links than the single-layer SBM model. Therefore we sample over synthetically generated networks to observe the influence of each parameter.

Our generation of a two-layer synthetic network depends on the following parameters:

1. Number of nodes N
2. Number of groups of each layer g_1, g_2
3. Topology of the second layer (A mixed assortative/disassortative, B disassortative, C assortative). The first layer topology is fixed as assortative. (Fig. 4.10).

4. Fraction of links removed f_{rem} (when predicting missing links), fraction of links added f_{add} (when predicting spurious links).
5. Average connectivity k .
6. Low-to-high group connection probability ratio $\lambda = \frac{p_l}{p_h}$.

For the AND model we use topologies that are not fully assortative in the second layer, since both assortative layers would yield a clearly assortative intersection difficult to distinguish from a pure single-layer SBM. For the OR model, the aggregated network is too dense unless we use assortative topologies on both layers.

The low-to-high probability ratio clearly influences the accuracy of the prediction. We show in Figs. 4.4-4.9 that the accuracy (size of the circles) increases as λ goes to 0, because as the difference between p_l and p_h increases the structure is less noisy and more easily detectable. Furthermore, for low λ in both layers the two-layer structure becomes clearer, so, in turn, $\lambda \rightarrow 0$ favors the multilayer model against the single-layer model. Note that in such networks where the accuracy is higher, the AND multilayer model outperforms the single-layer approach when predicting both missing and spurious links. On the other hand, the OR multilayer model only outperforms the single-layer model in networks where the overall accuracy is low and where p_l and p_h are closer. This could be caused by the fact that the OR model tends to generate very dense networks, whereas AND networks are sparser than the networks in each of the layers.

In general, we also observe that for moderate (but not too large) values of the average connectivity k , the multilayer model performs better than the single-layer SBM. Since we constrain the connection pattern by fixing the topology of the second layer as A, B, or C (point 3. above), adding more links make the structure of the two-layer topology clearer. In any case, excessively high k may yield the appearance of cliques, making the intersection of the two-layers similar to a one-layer network, favoring in turn the single-layer model.

To evaluate the effect of the three remaining parameters (N , g_i , and $f_{\text{rem/add}}$) we keep two of them fixed and vary the other one to observe its influence on the performance of both models in predicting missing and spurious links. First, as we show in Figs. 4.4-4.9(**d**, **b**, **e**), an increase of the fraction of removed/added links $f_{\text{rem/add}}$ generally decreases the accuracy of the models (circles become smaller as $f_{\text{rem/add}}$ increases). Moreover, a $f_{\text{rem/add}}$ increase also influences the two-to-one-layer accuracy ratio. Indeed, the AND model is comparatively better when $f_{\text{rem/add}}$ decreases (Figs. 4.4-4.7). Conversely the OR model is comparatively better when $f_{\text{rem/add}}$ increases (Figs. 4.8-4.9), that is, the OR model results seem more robust when removing/adding links than the single-layer model.

Second, if we fix all parameters but increase g_1 (Figs. 4.4-4.7(**a**, **b**, **c**)), we have a proliferation of smaller and smaller groups generally leading to smaller accuracy. Larger groups also imply better relative performance of the two-layer SBM, because the relevance of layer 1 vanishes when the groups in this layer become very small. However, in the OR model, when g_2 is increased from $g_2 = 4$ to $g_2 = 12$, the relative

accuracy of the multilayer model increases (Figs. 4.8-4.9**(b, c)**) because the single-layer SBM performance gets worse. Therefore, the OR model results are robust also when varying the number of groups.

Finally, by increasing N (Figs. 4.4-4.7**(f, g)** and Figs. 4.8-4.9**(a, b)**), we observe an overall improvement of missing/spurious link prediction, probably because there is more data to exploit in fitting the models. As before, the AND model tend to become relatively better with respect to the single-layer model when the overall performance increases. On the contrary, the performance of the OR model becomes relatively better when overall performance decreases.

4.4 Convergence in accuracy of single-layer and two-layer AND models for sparse networks

In most of the situations we explore for real-world networks, the AND model performs better than the OR model. We argue that this is because real-world networks are sparse, which favours the AND model. Nonetheless, because single-layer and two-layer models can be seen as identical models with a different prior, we expect these models to yield the same accuracies for cases in which we have gathered enough information about the system. This is because, in a Bayesian framework, the probabilities that you infer when you have infinite amounts of data about the system become independent of the choice of prior for the models. In complex networks, gathering infinite information about the system (or the underlying SBM for the network), implies having a large network with group sizes that scale with the network size. In this case, we would expect one and two-layer models to yield exactly the same accuracies. Note that this is not the typical case for real-world networks since due to the irregularities in the patterns of connections, group sizes can be arbitrarily small despite of network size. In that case, the choice of prior has an effect in the accuracy of the inference as our results for real-world complex networks show.

To assess whether our algorithm yields results that are consistent with our expectations, we have generated networks of sizes $N = 200, 500, 1000, 5000$ sampled from a sparse SBM with 4 equally sized groups and in/out connection probability ratios $p_l/p_h = 0.05, 0.15, 0.25$ with average connectivities $k = 4, 7, 10$. We generate these networks and perform missing-link prediction experiments removing a fraction $f = 0.05$ of the links. Note that we do not show results for $k = 3$ because most of the networks we were obtaining would break into small components even when we removed a small fraction of the links. For $k = 4$, the giant component remaining in those experiments comprises at least 95% of the original nodes.

Figure 4.11 shows the AUCs obtained from those experiments for three different models: i) two-layer AND SBM, ii) single-layer SBM and iii) a single-layer SBM with a degeneracy factor; iv) the planted partition of the generative SBM. In a Bayesian framework, these three models can be interpreted as the same model (a single-layer SBM) with different priors. Our results show that:

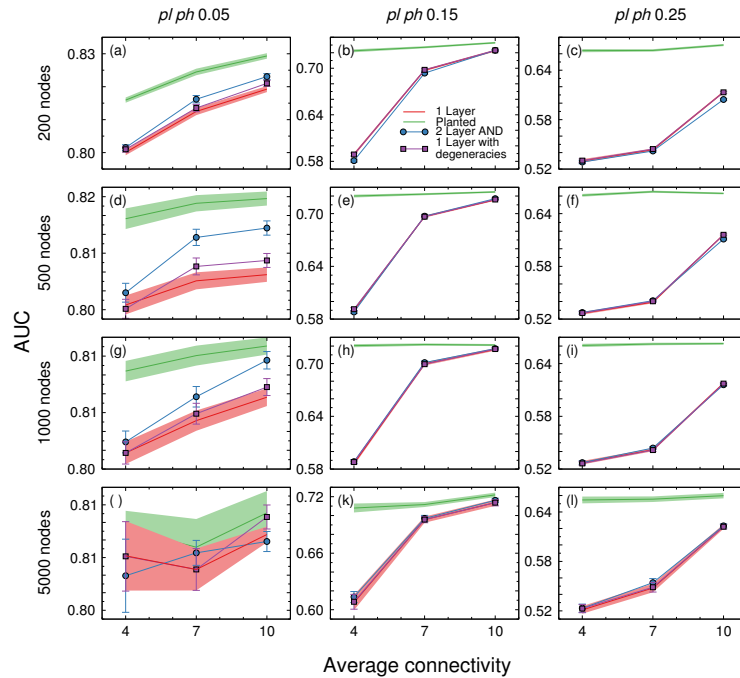


Figure 4.11: **Missing link prediction on sparse synthetic networks.** We generated 1000 different networks of size $N = 200$ (a,b,c), 600 networks with $N = 500$ (d,e,f), 200 networks with $N = 1000$ (g,h,i) and 5 networks with $N = 5000$ (j,k,l). Nodes are distributed according to an assortative stochastic block model with four equally sized groups, fixed average connectivity k , and a high to low connectivity ratio p_l/p_h . Note that as p_l/p_h increases the structure becomes less apparent. We restrict our analysis to those networks whose giant component comprises at least 95% of the nodes. For each of the networks we removed a fraction $f = 0.05$ of the total number of links to compare the average performance (AUC) of predicting missing links for three different models: single-layer SBM (red line), two-layer AND (blue circles), single-layer SBM with a degeneracy factor (purple squares), and the planted model of the synthetic network (green line).

1. As expected, as the network size increases, the prior becomes irrelevant and the the three different models yield the same accuracy. The larger the probability ratio, the lower the size of the networks needed for the accuracies to converge. For $p_l/p_h = 0.15, 0.25$, AUCs converge at a size $N = 500$, while for $p_l/p_h = 0.05$, we need to go to up to $N = 5000$ for the accuracies to converge. Note that the convergence of accuracies at these particular network sizes happens for this particular choice of SBM (that is uniform groups that scale with system size) but is bound to be much large for networks generated from more inhomogeneous SBMs.

Multilayer stochastic block models reveal the multilayer structure of complex networks

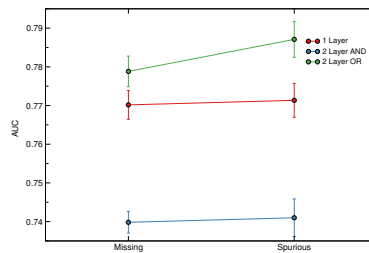


Figure 4.12: **Missing and spurious link prediction on two-layer OR generated synthetic networks.** We generate 10 different networks with $N = 500$ nodes, randomly distributed into 12 uniform groups in each of the two layers. Since all pairs of nodes have equal probability to be connected there is no average connectivity parameter here; however, each layer will have about $k = \frac{N}{2}$, with an OR resulting network with about $k = \frac{3 \cdot N}{4}$.

2. The two-layer AND SBM has a better performance than a single-layer SBM with a degeneracy factor and than a single-layer SBM for sparse networks. This is true for all the model parameter values we consider for which the AUCs obtained for the three different models have not converged. This result is consistent with our previous expectation based on the fact that the AND model effectively has a sparser prior than the single-layer SBM (with or without degeneracy factor). Also note, that the single-layer SBM with a degeneracy factor, also produces a better accuracy than the single-layer SBM. This is because while the planted partition has only four groups, the sparsity and fluctuations within the network are better described by SBMs with a larger number of groups.
3. The planted partition always works better.

4.4.1 Performance of the OR model in the ensemble of networks of our approach

In order for the OR model to work better than the AND model, we should consider networks that are generated from the true OR model in our approach, which is in fact not the set of models we investigate in Figs. 4.8 and 4.9. Specifically, because the OR model is equivalent to an AND model for the non-existing links, the ensemble of networks we assume in our approach consists of networks generated from a single layer SBM with uncorrelated connection probabilities for the non-existing links. Note that we also assume that while the probabilities of links not existing between pairs of groups are uncorrelated, they are still the result of a product of two probabilities. Figure 4.12 shows that indeed the OR model yields better accuracies in missing and spurious links experiment than the AND model for this ensemble of networks. The reason why we do not use this model to investigate the detection limits in Figs. 4.8 and 4.9 is that we think that this model is further from the structure of real-world networks than the models in Figs. 4.8 and 4.9 are.

4.5 Multilayer stochastic block models are more predictive for real networks

After showing that our approach is indeed more appropriate for model multilayer networks, we consider a real multilayer protein-protein interaction network of yeast *S. cerevisiae*. In particular, we consider two types of interactions reported in the BioGRID database [97]: those detected using “Two-hybrid” experiments, and those obtained using “Affinity-Capture Western” experiments. We aggregate the two layers using the AND mechanism; that is, we build an aggregate network comprising the interactions that are detected by both types of experiments, and only those. As we show in Fig. 4.13, the multilayer model is again more accurate than the single-layer model at detecting missing and, especially, spurious interactions.

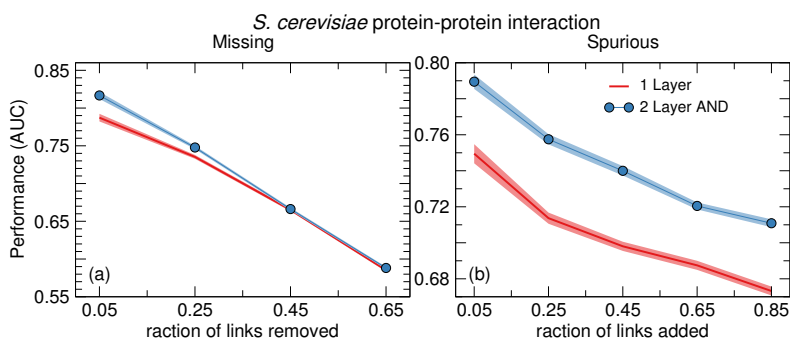


Figure 4.13: **Performance of missing and spurious link identification on a real multilayer network.** We show the accuracy at detecting (a) missing and (b) spurious interactions in the protein-protein interaction network of *S. cerevisiae*, obtained from BioGRID database [97] by aggregating two layers as described in the text. To compare the performance of the different approaches at detecting missing links (a), we randomly remove a fraction of the links (false negatives) from the real network and calculate the reliability of each unobserved link. Then we calculate the AUC statistic, that is, we rank the links by decreasing reliability 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) 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).

Finally, we turn to the question of whether real networks that are observed as single-layer networks are, in fact, better described as aggregates of multiple layers. Specifically, we compare the performance of the single-layer and multilayer approaches on eight real-world networks (Figs. 4.14): (i) the air transportation network in Eastern Europe [40]; (ii) the neural network of *C. elegans* [106]; (iii) the email network within a university [38]; (iv) the network of frequent co-purchasing of books about US politics sold by the online bookseller Amazon.com during the 2004 presidential elections [55]; (v) the transcriptional regulation network of yeast *S. cerevisiae* [62]; (vi) the air trans-

Multilayer stochastic block models reveal the multilayer structure of complex networks

portation network in the USA [7]; (vii) the collaboration network of jazz musicians, where two musicians are connected if they have played in the same band [33]; and (viii) the network of American football games between colleges during regular season Fall 2000 [32].

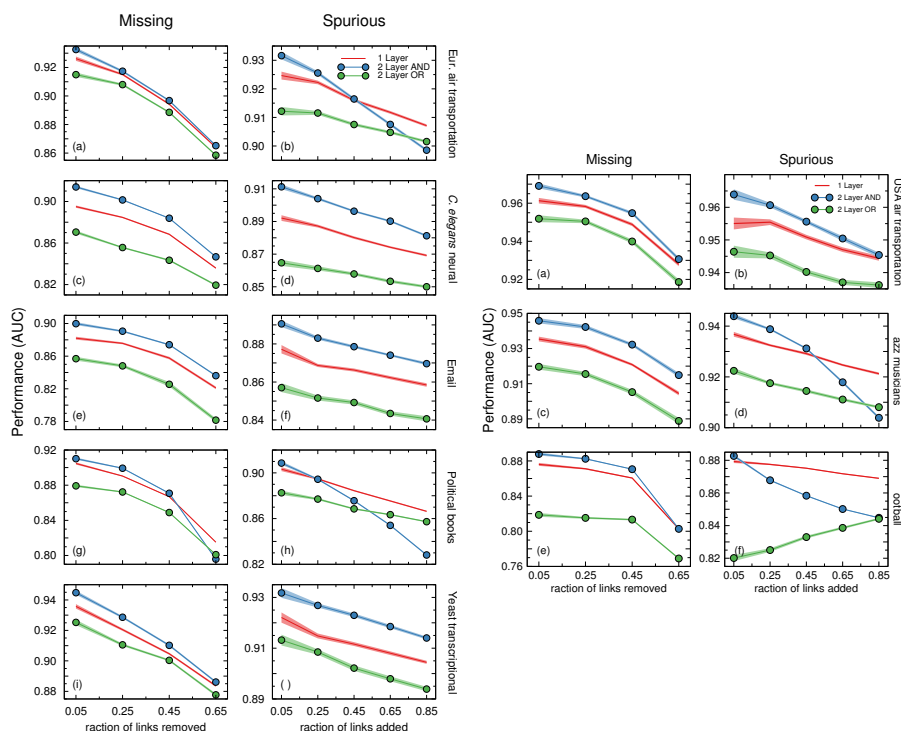


Figure 4.14: Performance of missing and spurious link identification on real aggregated networks. We proceed as in Fig. 4.13 to compare the performance of the different approaches at detecting missing links (a, c, e, g, i, k, m, o) and spurious links (b, d, f, h, j, l, n, p). We show results for eight real-world networks: (a, b) the air transportation network in Eastern Europe [40]; (c, d) the neural network of *C. elegans* [106]; (e, f) the email network within an organization [38]; (g, h) the network of books about US politics in 2004 elections [55]; (i, j) the transcriptional regulation network of yeast *S. cerevisiae* [62]; (k, l) the air transportation network in USA [7]; (m, n) the collaboration network of jazz musicians [33]; and (o, p) the network of American football games [32]. Red lines represent the AUC_{1L} obtained with single-layer SBMs, blue circles correspond to AND two-layer SBMs, and green circles to OR two-layer SBMs.

We use no prior information on how many groups we have on real networks, instead we sample over different partitions implicitly varying the number of groups. In Table 4.1 we show the mean of the number of groups over all the sampling, for each real network studied in the main text plus the recently added. In Table 4.2 we show the computation time at predicting missing/spurious links when removing/adding 5%

a) Sampling with a single SBM approach

Network	Fraction of links removed				
	0%	5%	25%	45%	65%
Air transportation network in Eastern Europe	9.78	9.79	8.44	7.66	6.82
Neural network of <i>C. elegans</i>	17.71	18.24	13.81	11.96	9.59
Email network within an organization	15.79	14.00	14.22	8.92	5.89
Network of books about USA politics	8.94	8.94	7.60	6.72	6.01
Transcriptional regulation network of yeast <i>S. cerevisiae</i>	11.99	11.36	16.60	12.57	9.99
Air transportation network in USA	16.81	15.47	11.06	9.89	9.50
Collaboration network of jazz musicians	25.05	22.70	18.68	15.15	10.94
American football games network between colleges	10.92	9.98	8.77	7.93	6.00

b) Sampling with Multilayer approach

Network	Fraction of links removed				
	0%	5%	25%	45%	65%
Air transportation network in Eastern Europe	14.48	14.60	14.10	14.86	12.04
Neural network of <i>C. elegans</i>	33.04	32.52	28.28	21.05	17.45
Email network within an organization	28.57	29.97	20.78	19.41	15.29
Network of books about USA politics	16.37	15.14	15.19	13.17	12.78
Transcriptional regulation network of yeast <i>S. cerevisiae</i>	19.05	17.67	17.24	15.82	12.79
Air transportation network in USA	24.62	23.94	21.64	20.30	16.11
Collaboration network of jazz musicians	34.22	34.14	32.96	28.31	23.50
American football games network between colleges	16.74	16.80	14.98	14.12	11.89

Table 4.1: Table of the mean number of groups over all the sampling done under the single SBM approach (a) and the Multilayer approach (b)

of the total number of links, for each real network studied in the main text plus the recently added.

Our results in Figs. 4.14 show that the two-layer AND model provides a better description of these real-world networks since both missing and spurious interactions are consistently more accurately detected by the multilayer SBM approach, especially for low observational noise.

As mentioned earlier, comparison of the two-layer approximation in Eq. (4.7) and the single-layer model in Ref. [44] shows that the two-layer model differs from the one-layer model in two ways. First, the AND model generates sparser networks than the single-layer model. Second, the two-layer model includes a degeneracy factor $D(\mathcal{P}_I)$ that favors partitions with a *larger* number of groups than the single-layer model (Table 4.2). Our results in Figs. 4.15 show that neither of the two factors alone is responsible for the improvement in accuracy we observe. In particular, we show that if we add the degeneracy factor to the single-layer model, we already improve the accuracy at

 Multilayer stochastic block models reveal the multilayer structure of complex networks

a) Computation time at predicting missing links

Network	Computation time (min.)		
	1 Layer	AND	OR
Air transportation network in Eastern Europe	0.45	1.75	0.5
Neural network of <i>C. elegans</i>	7.15	26.98	4.92
Email network within an organization	66.4	116.83	62.25
Network of books about USA politics	0.37	1.95	0.27
Transcriptional regulation network of yeast <i>S. cerevisiae</i>	10.48	23.13	11.13
Air transportation network in USA	8.42	18.77	9.43
Collaboration network of jazz musicians	8.73	21.87	6.42
American football games network between colleges	0.58	2.18	0.23

b) Computation time at predicting spurious links

Network	Computation time (min.)		
	1 Layer	AND	OR
Air transportation network in Eastern Europe	0.5	2.02	0.6
Neural network of <i>C. elegans</i>	7.62	27.0	5.97
Email network within an organization	71.0	131.33	67.73
Network of books about USA politics	0.42	2.48	0.37
Transcriptional regulation network of yeast <i>S. cerevisiae</i>	11.27	24.53	11.08
Air transportation network in USA	8.45	19.68	9.37
Collaboration network of jazz musicians	8.85	23.67	6.87
American football games network between colleges	0.73	2.63	0.33

Table 4.2: Table of the computation time in minutes for 5% links removal of links at predicting missing links **(a)** and spurious links **(b)**

detecting missing and spurious links in most cases. From our results it follows that sampling from partitions with a larger number of groups provides better models for real-world networks. This may seem counterintuitive, since one may expect a better model to have a lower number of parameters (groups in our case). However, because we expect the intersection block model resulting from a layer aggregation process to have a larger number of groups than the block models for each of the layers, this observation further reinforces our hypothesis that most real-world networks are in fact the result of an aggregation process.

4.5.1 Alternative approximation for multi-layer SBMs: A single layer SBM with degeneracies

We propose an approximation to two-layer stochastic blockmodels that assumes that the SBM resulting from the intersection of the two SBMs has independent matrix ele-

ments $q_{\alpha\beta\gamma\delta}^I$ that are the result of the product $q_{\alpha\beta\gamma\delta}^I = q_{\alpha\beta}^1 q_{\gamma\delta}^2$. As a result, if \wp is the ensemble of all possible partitions of nodes into blocks, instead of sampling over $\wp \times \wp$ pairs of partitions to compute the reliability of a link R_{ij} , we need to sample only over the space of possible intersection partitions \wp and take into account the number of pairs of partitions that result in the same intersection, or what we call degeneracy $D(\mathcal{P})$ (see Sec. A.2).

For the AND model the reliability of link l_{ij} is then given by Eq. 4.7,

$$R_{ij}^{\text{AND}} = \frac{1}{Z} \sum_{\mathcal{P}_I \in \wp} \left(\frac{n_{\sigma_i \sigma_j}^1 + 1}{n_{\sigma_i \sigma_j} + 2} \cdot \frac{\sum_{k=n_{\sigma_i \sigma_j}^1 + 2}^{\sum_{\sigma_i \sigma_j} + 2} \frac{1}{k}}{\sum_{k=n_{\sigma_i \sigma_j}^1 + 1}^{\sum_{\sigma_i \sigma_j} + 1} \frac{1}{k}} \cdot D(\mathcal{P}_I) \cdot e^{-\mathcal{H}(\mathcal{P}_I)} \right) \quad (4.15)$$

where $\mathcal{H}(\mathcal{P}_I) = \sum_{\alpha \leq \beta \in \mathcal{P}_I} \left(\ln(n_{\alpha\beta} + 1) + \ln \binom{n_{\alpha\beta}}{n_{\alpha\beta}^0} - \ln \left(\sum_{k=n_{\alpha\beta}^1 + 1}^{n_{\alpha\beta} + 1} \frac{1}{k} \right) \right)$ (4.16)

where the sum is over all possible single-layer partitions, $n_{\alpha\beta}^1$ is the number of links between groups α and β in the partition, $n_{\alpha\beta} = n_{\alpha\beta}^0 + n_{\alpha\beta}^1$ is the number of (possible links between) pairs of nodes in groups α and β .

Note that if we consider a single SBM, we compute the reliability as [44]:

$$R_{ij}^{\text{single}} = \frac{1}{Z} \sum_{\mathcal{P} \in \wp} \left(\frac{n_{\sigma_i \sigma_j}^1 + 1}{n_{\sigma_i \sigma_j} + 2} \cdot e^{-\mathcal{H}(\mathcal{P})} \right), \quad (4.17)$$

where $\mathcal{H}(\mathcal{P}) = \sum_{\alpha \leq \beta \in \mathcal{P}_I} \left(\ln(n_{\alpha\beta} + 1) + \ln \binom{n_{\alpha\beta}}{n_{\alpha\beta}^0} \right)$. (4.18)

As shown in Fig. 4.14 of the main manuscript, our approximation assuming an AND model provides a better description of real-world networks because it performs better at detecting missing/spurious interactions. In order to assess the influence of $D(\mathcal{P}_I)$ in the prediction of missing and spurious links, we consider the case in which the intersection block model is an SBM in which matrix elements are random variables uniformly distributed in the interval $[0, 1]$, rather than the product of two probabilities. In such case, we still take into account that there are many pairs of partitions with the same intersection partition, but the \mathcal{H} associated to each partition is that of a single SBM, so that we compute the reliability as

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

with \mathcal{H} given by Eq. (4.18). Note that in this case, there is no distinction between OR and AND models.

We observe in Fig. 4.15, that the model with only taking into account the degeneracies has a performance that is slightly better than that of the single-layer SBM, for

Multilayer stochastic block models reveal the multilayer structure of complex networks

real-world networks. This means that the overall performance of our approximate two-layer AND SBM is a combination of the effect of the sparsity induced by assuming that each q is in fact the product of two numbers and, to a lesser extent, the a priori preference for partitions with more groups suggested by the multi-layer model.

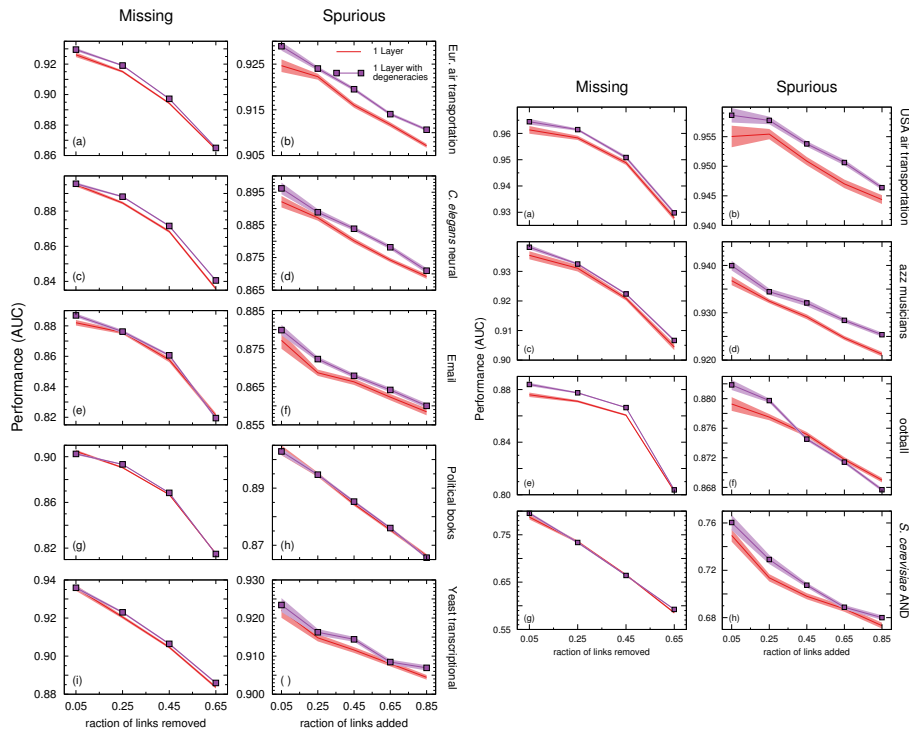


Figure 4.15: **Performance of missing and spurious link identification on real aggregated networks.** Same as Fig. 4.14, but only comparing the performance of the single-layer SBM approach with an approximation to the two-layer SBM that only takes into account the degeneracies of the intersection partition (Eqs. (4.19)).

4.6 Quantification of the preference for multilayer models

Our results demonstrate that the two-layer stochastic block model (with AND aggregation) is more predictive for real-world complex networks, thus suggesting that real-world complex networks may be the result of the projection of several layers onto a single aggregate observation. To further quantify to what extent a two-layer model provides a better description of real-world networks than a single-layer model, we use Markov chain Monte Carlo sampling to compute the Bayes factor K of the models [49]

4.6.1 Computation of the Bayes factor

To quantify up to what extent the network data we consider support more strongly the two-layer SBM than the single-layer SBM, we follow a typical approach in Bayesian model selection and compute the Bayes factor K [49]

$$K = \frac{p(A^{\mathcal{O}}|M_2)}{p(A^{\mathcal{O}}|M_1)}, \quad (4.19)$$

where we have used the shorthand 2 :=two-layer and 1 :=single-layer.

To compute K we need to perform a sampling of the model space $\mathcal{M} = \{M_1\} \cup \{M_2\}$ that consists in the union of model spaces for M_1 and M_2 . In this case, for each possible partition \mathcal{P} of the nodes into groups, we have two possible models that we sample with the following probabilities

$$p(M_1, \mathcal{P}) = \frac{1}{Z} e^{-(\mathcal{H}_I(\mathcal{P}) - \ln D(\mathcal{P}))} \equiv \frac{1}{Z} e^{-\mathcal{H}_2(\mathcal{P})} \quad (4.20)$$

$$p(M_2, \mathcal{P}) = \frac{1}{Z} e^{-\mathcal{H}_1(\mathcal{P})} \quad (4.21)$$

$$Z = \sum_{\mathcal{P}} e^{-\mathcal{H}_2(\mathcal{P})} + e^{-\mathcal{H}_1(\mathcal{P})}, \quad (4.22)$$

where $\mathcal{H}_I(\mathcal{P})$ is given by Eq. S7, $D(\mathcal{P})$ is the degeneracy factor computed in Sec. S1, and $\mathcal{H}_1(\mathcal{P})$ is given by Eq. 4.18.

Note that while for some networks independent samplings of two-layer and single-layer models might sample similar values of \mathcal{H}_1 and \mathcal{H}_2 , these correspond in general to different partitions, that is, for a specific partition there is a preferred model with a much lower \mathcal{H} . In the MC simulation, to avoid getting trapped in the subspace corresponding to one of the models, we perform two simultaneous independent samplings of M_1 and M_2 . At each step, we attempt N individual node movements (see [44] for details) in each model, so that we have one partition $(\mathcal{P}_1, \mathcal{P}_2)$ and one energy for each of the models, $\mathcal{H}_1(\mathcal{P}_1)$ and $\mathcal{H}_2(\mathcal{P}_2)$, respectively. We use a variable that stores the model that we are currently sampling (for instance `model=2`). Then, we propose a change of model with acceptance probability $p = \min\{1, \exp(-[H_1(\mathcal{P}_1) - H_2(\mathcal{P}_2)])\}$. If we accept the change, we set `model=1`.

With this setup, the Bayes factor is the ratio of the number of times we sample the two-layer model with respect to the number of times we sample the single-layer model. We follow the same procedure for the case in which we compare the single-layer model and the single-layer plus degeneracy factor model introduced in Sec. S5.

4.6.2 Bayes factor quantifies the preference for multilayer models in real-world networks

Figure 4.16 shows that for all the real-world networks we consider the Bayes factor is larger than one. Using the qualitative scale proposed by Kass and Raftery to map K values to human perception of evidence strength [51], we conclude that there is

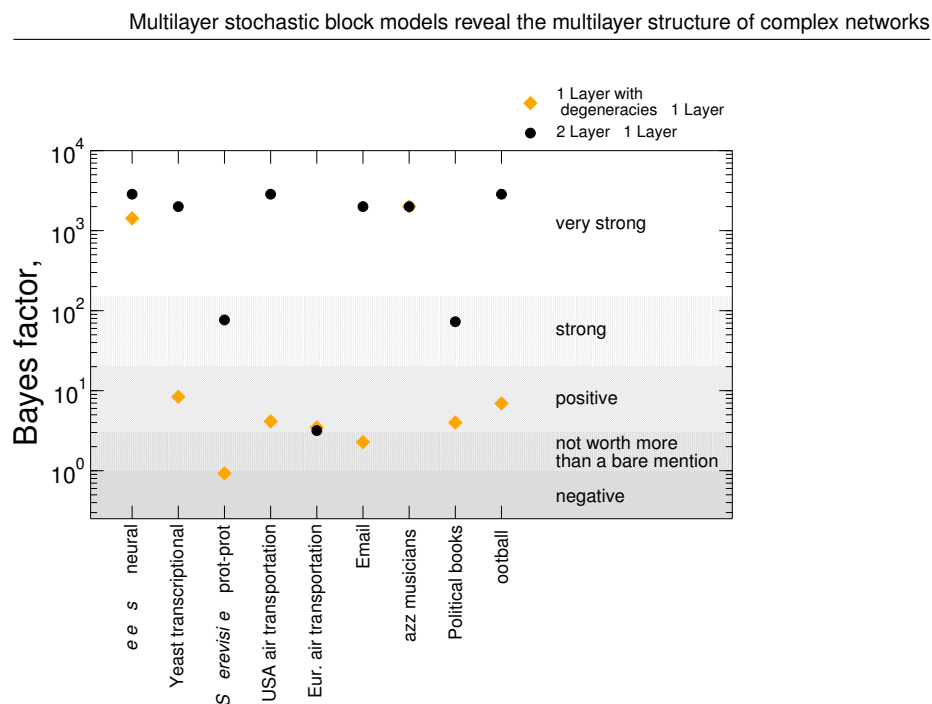


Figure 4.16: **Model comparison using Bayes factors.** We show the Bayes factor K (Eq. 4.19) for all the real-world network examples we analyzed (see text). Black dots show results for the Bayes factor considering the two-layer AND SBM and the single-layer SBM. Orange diamonds show the Bayes factor considering the single-layer SBM with a degeneracy factor and the single-layer SBM. The regions in different shades of gray indicate the qualitative scale introduced by Kass and Raftery to map K values to human perceptions of strength of evidence [51] (“negative”: $K < 1$; “not worth more than a bare mention”: $1 < K < 3$; “positive”: $3 < K < 20$; “strong”: $20 < K < 150$; “very strong”: $K > 150$).

“very strong evidence” supporting the two-layer model for most of the networks; for the *S. cerevisiae* protein-protein interaction network and for the networks of political books the evidence is “strong.” Importantly, Fig. 4.16 also shows that the preference for the two-layer model cannot be solely attributed to the sparsity induced by the AND aggregation. Indeed, simply adding to the single-layer model the degeneracy factor (which, as discussed above, introduces a preference for larger numbers of groups) also results in a model that is better supported by the data than the single-layer model, in all but one of the networks we consider (the protein-protein interaction network of *S. cerevisiae*).

Interestingly, we find that there are some discrepancies between the evidence we find for the two air transportation networks we consider. While for the USA air transportation we find very strong evidence for the two-layer model, we find barely a positive evidence for the European air transportation. From previous analyses, we know that air transportation networks have a very strong modular component driven by geo-political

factors [40, 89]; nonetheless, there is arguably a second layer that may arise from the distinction between international hubs that connect to one another and local airports that connect to hubs. Our results suggest that while the US air transportation shows strong evidence for those two layers, the European air transportation network (which is smaller and has a much lower density, and where geography may play a stronger role because of the presence of political borders) the evidence of the two layers is less conclusive.

4.7 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 [65], our approach 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 [39].

Our work is also different from works on link prediction using latent feature models [61, 72, 52]. 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 processes occurring in different layers or not (and may also be prone to overfitting when observational data is noisy).

Our multilayer SBM is the simplest group-based multilayer model one can propose. Although our approach is not exempt of limitations (for example, is computationally expensive and is therefore not suitable to handle extremely large networks) we believe that its detailed analysis will open the door to better understand the structure of real complex networks.

UNIVERSITAT ROVIRA I VIRGILI
NETWORK INFERENCE BASED ON STOCHASTIC BLOCK MODELS: MODEL EXTENSIONS, INFERENCE APPROACHES
AND APPLICATIONS
Toni Vallès Català

Conclusions and perspectives

The effort devoted on this thesis deepens on the complex networks inference via Stochastic Block Models (SBMs). Our analysis of the SBMs inference framework at the time enables us to choose the best methodology to use. Such methodology can be applied to a high variety of fields, we prove so by applying it to a newborn skull network, a field poorly studied from a complex systems point of view. Finally, the knowledge gained during the work endows us to compose a novel multilayer SBM that sheds light on the inner structure of real-world networks. The following conclusions can be drawn from the work:

- We set the framework that enables us to compare, for the first time, the predictive power of two well-known inference methodologies: the single-point estimation of the set of parameters, and the Bayesian approach of pondering over several sets of likely parameters. Such framework applies the same model (SBM) at performing the same task (prediction of missing links) on the same four real-world networks. We have found that sampling over models significantly improves the predictive power than considering an optimal set of parameters alone, suggesting that the error at estimating the reliability of missing links when applying the single-point method is not negligible.

Additionally, we considered two specifications of the SBM (degree-corrected version and the use of hierarchical priors) on the single-point estimation methodology. We discovered that the Bayesian approach inference is preferably than all the combinations of the specifications in the single-point estimate method.

During the study, we noticed that the degree-corrected version should not be used in all the networks, since it may retreat the predictions in some cases. On further work, it may be interesting to investigate on which cases it is worth to apply such correction.

Moreover, since the specifications are not mutually exclusive from the Bayesian methodology, we build a novel approach that applies the Bayesian method to the degree-corrected version of the SBM. This approach predicts better missing links than the simple Bayesian approach, at least in those real-world networks where it is worth to use the degree-corrected version.

- Our results suggest that sutures that normally fuse have significantly lower reliability scores than those that do not, which is in agreement with our hypothesis that during normal development there is a tendency to lose articulations that are topologically rare in the newborn skull. (Mann-Whitney-Wilcoxon: one sided $W=206.0$, p -value = 0.0055; Mean(fused) = 0.3485; 95% CI (non-fused) = (0.4124, inf))

Interestingly, we discovered that sutures associated with pathological conditions have significantly lower reliability scores than sutures that are not, which shows that sutures associated to pathological conditions are also unexpected from a topological point of view. (one-sided, $W = 116$, p -value = $1.022E^{-4}$; Mean(pathological) = 0.3244; 95% CI (non-pathological and non-fused) = (0.4417, inf))

Finally, we find no statistical difference between the reliability scores of sutures that are associated to pathological conditions and those that fuse during normal development (two-sided $W = 44.5$, p -value = 0.196; 95% CI (fused) = (0.3389, 0.3668))

- We have developed a new model for single aggregated networks, the multilayer SBM. The model assumes that the single-layered network under study is the outcome of two separate layers. The aggregate network can be the intersection of the layers (AND model) or the union of them (OR model). We validate AND/OR model with synthetic networks: the model detects better missing and spurious links than the single-layer SBM. Additionally, the AND model is validated with a known aggregated real-world network: the protein-protein interaction network of yeast *S. cerevisiae*, that is the AND aggregation of two techniques (“Two-hybrid” and “Affinity-Capture Western”). We apply the models to eight single-layer real-world networks from different fields and we get better predictions in both missing and spurious links with the multilayer SBM AND model than single SBM, then is more likely that the observed networks come from a multilayer network than a single-layer.

Our results suggest that networks that are presented as single-layer may indeed be projections of multilayer networks.

Our Multilayer model is defined with two layers, combined in an AND or an OR way. However, the connections in real-world networks may arise from more intricate mechanisms, maybe with more layers with AND and OR combinations. Although a small increase in number of layers imply a large increase in complexity, it may be fascinating to explore to what extent a generalised Multilayer model with no fixed layers improve the predictive power on real-world networks.

Since our model uses the intersection partition, we are not able to discern from which layers a network is aggregated into. Another interesting further work will be to disentangle which SBM are more likely from the possible ones that conform the intersection partition, then, given a network it would be possible to unravel which networks were aggregated and obtain a multiplex network.

UNIVERSITAT ROVIRA I VIRGILI
NETWORK INFERENCE BASED ON STOCHASTIC BLOCK MODELS: MODEL EXTENSIONS, INFERENCE APPROACHES
AND APPLICATIONS
Toni Vallès Català

Conclusions and perspectives



Appendices

A.1 Appendix A: Implementation details

The sum in Eq. (1.11) cannot be computed exactly because the number of possible partitions is combinatorially large, but can be estimated using the Metropolis algorithm [60, 44]. This amounts to generating a sequence of partitions in the following way. Start from an initial partition P^0 , a common practice is to start with each node in its own group [78, 8]. From such current partition, select a random node and move it to a random new group giving a new partition P^1 . If $H(P^1) < H(P^0)$, always accept the move; otherwise, accept the move only with probability $P = e^{H(P^0) - H(P^1)}$.

By doing this, one gets a sequence of partitions $\{P^i\}$ such that one can approximate the integral in 1.11 as [60]

$$p(A_{ij} = 1 | A^{\mathcal{O}}) \approx \frac{1}{S} \sum_{P \in \{P^i\}} \frac{n_{\sigma_i \sigma_j}^1 + 1}{n_{\sigma_i \sigma_j} + 2}, \quad (\text{A.1})$$

where S is the number of sampled partitions in $\{P^i\}$.

In practice, it is useful to “thin” the sample $\{P^i\}$, that is, to consider only a small fraction of evenly spaced partitions so as to avoid the computational cost of sampling very similar partitions which provide very little additional information. Moreover, one needs to make sure that sampling starts only when the sampler is “thermalized”, that is, when sampled partitions are drawn from the desired probability distribution (which in our case is given by Eq. (1.5, with T fixed to 1). The implementation automatically determines a reasonable thinning of the sample, and only starts sampling when certain thermalization conditions are met. Therefore, the whole process is completely unsupervised. The source code of this implementation of the algorithm is publicly available from <http://seeslab.info/downloads/network-c-libraries-rgraph/> and <http://github.com/seeslab/rgraph>.

A.2 Appendix B: Calculation of intersection partition degeneracies

To compute the number of pairs $(\mathcal{P}_1, \mathcal{P}_2)$ that have the same intersection partition $\mathcal{P}_I = \mathcal{P}_1 \cap \mathcal{P}_2$, we note that a specific \mathcal{P}_I consists of n groups of nodes that we call “elements;” we make explicit the number of such elements in an intersection partition and write $\mathcal{P}_I^n = [E_1][E_2][E_3] \dots [E_n]$. From the definition of intersection partition we have that: (i) all the nodes within an element must belong to the same group in both partitions \mathcal{P}_1 and \mathcal{P}_2 (otherwise, they would not belong to the same element); (ii) two elements cannot belong to the same group in both \mathcal{P}_1 and \mathcal{P}_2 (otherwise they would be a single element).

Within this setup, our goal is to compute $D(\mathcal{P}_I^n)$, the cardinality of the set $\{(\mathcal{P}_i, \mathcal{P}_j) | \mathcal{P}_i \cap \mathcal{P}_j = \mathcal{P}_I^n\}$. We compute the degeneracy in two steps: 1) We compute all the possible unique partitions \mathcal{P}_1 combining the elements in \mathcal{P}_I^n , group them in *classes* according to the number of elements combined, and compute the multiplicity associated to each class; 2) For each class, we compute all the possible partitions \mathcal{P}_2 that result in a specific intersection \mathcal{P}_I^n .

Next, we give details of the exact calculation, and we illustrate how to compute these degeneracies in an iterative manner.

A.2.1 Multiplicity of each partition class

We first make an exhaustive list of all possible partitions \mathcal{P}_1 for a given number of elements n . For instance, for $n = 4$ we have four elements $\mathcal{P}_I^4 = [A][B][C][D]$. Then, we have the following classes of partitions for \mathcal{P}_1 :

1. $[4] \equiv [ABCD]$: all elements belong to the same group;
2. $[3, 1] \equiv [ABC][D] \equiv [ABD][C] \equiv [ACD][B] \equiv [BCD][A]$: there are two groups, one comprising three elements and another one comprising 1 element.
3. $[2, 2] \equiv [AB][CD] \equiv [AC][BD] \equiv [AD][BC]$: two groups of two elements each.
4. $[1, 1, 1, 1] \equiv [A][B][C][D]$: all elements in different groups.

Note that we have introduced the notation $[X_1, \dots, X_k]$ for classes, that indicates which elements of \mathcal{P}_I^n are in each of the k groups of partition \mathcal{P}_1 . Therefore, for $n = 4$ the list of all the different partition classes is $\mathcal{C}_4 = \{[X_i, \dots, X_k] | \sum_{i=0}^k X_i = n\} = \{[4], [3, 1], [2, 2], [2, 1, 1], [1, 1, 1, 1]\}$.

Note that for any possible class $[X_1, \dots, X_k]$ there can be multiple possibilities of grouping elements from \mathcal{P}_I , therefore, in order to compute the final degeneracy, we need to take into account the class multiplicity as

$$\text{mult}([X_1, \dots, X_k]) = \frac{\binom{n}{X_1} \prod_{l=2}^k \binom{n - \sum_{i=1}^{l-1} X_i}{X_l}}{\prod_{i=1}^L |\mathfrak{N}_i|!}, \quad (\text{A.2})$$

where if $\mathcal{L} = \{x_1, \dots, x_L\}$ is the set of unique values of X_j in $[X_1, \dots, X_k]$ and L is the cardinality of this set, $\aleph_i = \{X_l \in [X_1, X_2, \dots, X_k] | X_l = x_i\}$, and $|\aleph_i|$ is its cardinality.

To illustrate how we compute a multiplicity of a class, we show the multiplicity for three different classes:

- $[3, 1]$: In this case, $n = 3 + 1 = 4$, $k = 2$ and $\mathcal{L} = \{x_1 = 3, x_2 = 1\}$ so that $|\aleph_1| = |\aleph_2| = 1$ and hence: $\text{mult}([3, 1]) = \frac{\binom{4}{1}\binom{1}{1}}{1!1!}$.
- $[5, 3, 2]$: In this case, $n = 5 + 3 + 2 = 10$, $k = 3$ and $\mathcal{L} = \{x_1 = 5, x_2 = 3, x_3 = 2\}$ so that $|\aleph_1| = |\aleph_2| = |\aleph_3| = 1$, and hence: $\text{mult}([5, 3, 2]) = \frac{\binom{10}{5}\binom{5}{3}\binom{2}{2}}{1!1!1!}$.
- $[4, 4, 1]$: In this case, $n = 4 + 4 + 1 = 9$, $k = 3$ and $\mathcal{L} = \{x_1 = 4, x_2 = 1\}$ so that $|\aleph_1| = 2$ and $|\aleph_2| = 1$, and hence: $\text{mult}([4, 4, 1]) = \frac{\binom{9}{4}\binom{5}{4}\binom{1}{1}}{2!1!}$.

A.2.2 Number of partitions \mathcal{P}_2 whose intersection with a fixed \mathcal{P}_1 is equal to \mathcal{P}_I

Once we know how to generate all possible partition classes and their multiplicity for a specific \mathcal{P}_I^n , we fix an instance of a partition \mathcal{P}_1 for each class $[X_1, \dots, X_k]$ and obtain $\text{deg}([X_1, \dots, X_k])$, the number of partitions \mathcal{P}_2 such that $\mathcal{P}_1 \cap \mathcal{P}_2 = \mathcal{P}_I^n$. Note that $\text{deg}([X_1, \dots, X_k])$ is the degeneracy of a class (of the fixed partition), so that

$$D(\mathcal{P}_I^n) = \sum_{c \in \mathcal{C}_n} \text{deg}(c) \cdot \text{mult}(c), \quad (\text{A.3})$$

where the sum is over all possible partition classes for a given n , \mathcal{C}_n .

For clarity, let us start with the case $n = 1$. In such case $\mathcal{P}_1^1 = [A]$, so for \mathcal{P}_1 there is only one possible class $\mathcal{C}_1 = \{[A]\}$ and \mathcal{P}_2 must be equal to $[A]$ in order to fulfill $\mathcal{P}_1 \cap \mathcal{P}_2 = [A]$. Therefore, $D(\mathcal{P}_1^1) = 1$.

For $n = 2$, $\mathcal{P}_I^2 = [A][B]$. There exist two possible classes for \mathcal{P}_1 : $\mathcal{C}_2 = \{[2], [1, 1]\}$. For class $[2]$ we fix $\mathcal{P}_1 = [AB]$. We then make a list of all possible \mathcal{P}_2 s: $\mathcal{P}_2 = [AB]$ or $\mathcal{P}_2 = [A][B]$. Out of these two only $[A][B] \cap \mathcal{P}_1 = [A][B]$, therefore $\text{deg}([2]) = 1$. For class $[1, 1]$, we fix $\mathcal{P}_1 : [A][B]$. Note that both \mathcal{P}_2 's have the same intersection with \mathcal{P}_1 : $\mathcal{P}_1 \cap [AB] = \mathcal{P}_1 \cap [A][B] = \mathcal{P}_I^2$, therefore $\text{deg}([1, 1]) = 2$. Putting it all together into Eq. (A.3), we obtain that $D(\mathcal{P}_I^2) = \text{deg}([2]) \cdot \text{mult}([2]) + \text{deg}([1, 1]) \cdot \text{mult}([1, 1]) = 3$, since $\text{mult}([2]) = \text{mult}([1, 1]) = 1$.

In order to compute the degeneracies for increasing n our strategy is to compute the degeneracies of a class iteratively from the degeneracies of previously computed classes. For instance, for $n = 3$ we have that $\mathcal{C}_3 = \{[3], [2, 1], [1, 1, 1]\}$. For class $[3]$, we fix $\mathcal{P}_1 = [ABC]$. Note that whenever all the n elements are grouped together only $\mathcal{P}_2 = [A][B][C]$ ($n = 3$ in this case) results in an intersection equal to \mathcal{P}_I^n , therefore $\text{deg}([n]) = 1 \forall n$.

For class $[2, 1]$ we fix $\mathcal{P}_1 = [AB][C]$. Note that we have previously computed $\text{deg}([AB]) = \text{deg}([2]) = 1$, therefore all the cases in which element $[C]$ is in its own

group can be accounted for with the degeneracies we have already computed. In the remaining cases, C is together with either A or B since in order for \mathcal{P}_2 to have the expected intersection $[A][B][C]$ with \mathcal{P}_1 , A and B must be in different groups in \mathcal{P}_2 . Thus there are only two possibilities for \mathcal{P}_2 , $\mathcal{P}_2 = [CB][A]$ or $\mathcal{P}_2 = [CA][B]$. Since C is never in its own group in this case, the degeneracy of $\mathcal{P}_2 = [CB][A]$ is that of class $[1, 1]$. Therefore $\deg([2, 1]) = \deg([2]) + \deg([1, 1]) = 3$.

For class $[1, 1, 1]$, we fix $\mathcal{P}_1 = [A], [B], [C]$ and observe that every possible partition \mathcal{P}_2 will result in the desired intersection. The number of different groupings of n elements (the number of groups going from 1 to n) is the Bell number $B(n)$, so that $\deg([1, 1, \dots, 1]) = B(n)$ and for $n = 3$, $\deg([1, 1, 1]) = 5$.

All in all, for $n = 3$ we have that $D(\mathcal{P}_1^3) = \deg([3]) \cdot \text{mult}([3]) + \deg([2, 1]) \cdot \text{mult}([2, 1]) + \deg([1, 1, 1]) \cdot \text{mult}([1, 1, 1]) = 1 \cdot 1 + 3 \cdot 3 + 5 \cdot 1 = 15$.

A.2.3 Pseudo code to compute higher order degeneracies

For illustration, we focus on the class $[3, 2, 1]$ ($n = 6$), in which only one element is grouped by itself and we assume that we have previous knowledge of all the degeneracies corresponding to all the classes with $n \leq 5$ elements.

First we fix $\mathcal{P}_1 = [ABC][DE][F]$. In the \mathcal{P}_2 's that have the target intersection with \mathcal{P}_1 , element F can be either by itself, grouped together with one element of $[ABC]$, grouped together with one element of $[DE]$, or grouped together with one element of $[ABC]$ and one of $[DE]$. A good approach is to build a binary table T_{ij} with the groups of \mathcal{P}_1 as columns and all the possible \mathcal{P}_2 configurations as rows. We let empty the column of the element grouped by itself (in our case F) and fill the rest of the table with all the possible combinations of ones and zeroes. If $T_{ij} = 0$, in partition i , element F is not grouped with any element of another group. If $T_{ij} = 1$, then in partition i , element F is grouped together with one, and only one, of the elements of group j in \mathcal{P}_1 .

For the case we consider Table A.1 shows the binary table and the corresponding degeneracy for each choice of \mathcal{P}_2 .

\mathcal{P}_2	$[ABC]$	$[DE]$	$[F]$	
1	0	0	F is grouped by itself.	2
2	0	1	F is grouped only with one element in $[DE]$.	3
3	1	0	F is grouped only with one element in $[ABC]$.	6
4	1	1	F is grouped with one element in $[ABC]$ and one element in $[DE]$.	6

Table A.1: Degeneracies for each possible arrangement of element F in \mathcal{P}_2 considering $\mathcal{P}_1 = [ABC][DE][F]$.

In what follows we use the row numbers in Table A.1 as labels for \mathcal{P}_2 . For $\mathcal{P}_2 = 1$, the degeneracy is equal to $\deg([3, 2])$ because $[F]$ is by itself, i.e. we fix $\mathcal{P}_1 = [ABC][DE]$ and we look for all the \mathcal{P}_2 's that yield $[A][B][C][D][E]$ as an intersection. Because we assume that we know all degeneracies for $n \leq 5$, we can immediately compute this degeneracy as $\deg([3, 2])$.

In $\mathcal{P}_2 = 2$, F is either with E or D . For \mathcal{P}_2 to have the target intersection, $[D]$ and $[E]$ must be in separate groups. If we fix, for instance, $[DF]$ in \mathcal{P}_2 , we just need to compute the degeneracy of $[ABC][E]$, which is equal to $\deg([3, 1])$, and also assumed to be known. The same happens if we fix $[EF]$, so that we have that the total degeneracy is $2 \cdot \deg([3, 1])$.

In an analogous way, we reach the conclusion that the degeneracy of partitions in which F is grouped with one element of $[ABC]$ ($\mathcal{P}_2 = 3$ in Table A.1) is equal to $3 \cdot \deg([2, 2])$.

For $\mathcal{P}_2 = 4$, where F is with one element of $[ABC]$ and one in $[DE]$, we can fix $[ADF]$. In this case, the associated degeneracy is $\deg([BC][E]) = \deg([2, 1])$. Because there are six different ways of grouping F with an element of $[ABC]$ and an element of $[DE]$, the total degeneracy in this case is $6 \cdot \deg([2, 2])$.

Therefore, $\deg([3, 2, 1]) = \deg([3, 2]) + 2 \cdot \deg([3, 1]) + 3 \cdot \deg([2, 2]) + \deg([2, 1])$.

This algorithm computes exact degeneracy values given the number of elements n that appear in the intersection partition \mathcal{P}_I . Since the computational time to calculate the degeneracies increase exponentially with n , we only computed the first 69 values (see Table A.2). We extrapolated values for $n > 69$, by fitting a function using the values for $n \leq 69$, so that we use $D(\mathcal{P}_I^n) = 1.4681 \cdot n \cdot (\log(n) - 1)$ (see Fig. A.1)

n	$\log(D(\mathcal{P}_I^n))$	n	$\log(D(\mathcal{P}_I^n))$	n	$\log(D(\mathcal{P}_I^n))$
1	0.0	26	87.8750	51	219.3107
2	1.0986	27	92.6064	52	225.0314
3	2.7081	28	97.3919	53	230.7813
4	4.7274	29	102.2298	54	236.5599
5	7.0501	30	107.1182	55	242.3666
6	9.6241	31	112.0557	56	248.2009
7	12.4124	32	117.0408	57	254.0625
8	15.3873	33	122.0720	58	259.9508
9	18.5278	34	127.1480	59	265.8655
10	21.8172	35	132.2676	60	271.8060
11	25.2419	36	137.4297	61	277.7720
12	28.7908	37	142.6330	62	283.7630
13	32.4543	38	147.8767	63	289.7788
14	36.2243	39	153.1596	64	295.8189
15	40.0939	40	158.4808	65	301.8829
16	44.0569	41	163.8394	66	307.9706
17	48.1079	42	169.2345	67	314.0815
18	52.2421	43	174.6654	68	320.2153
19	56.4554	44	180.1312	69	326.3718

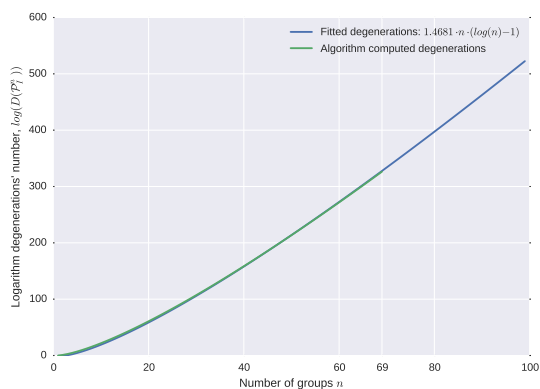


Figure A.1: **Extrapolation of degeneracy values.** We use our algorithm to compute the actual degeneracy values $D(\mathcal{P}_I^n)$ for $n = 1, \dots, 69$ (green line). Since it is computationally too expensive to use the same algorithm for $n > 69$, we extrapolate $D(\mathcal{P}_I^n)$ for $n > 69$ using the following mathematical formula $\log(D(\mathcal{P}_I^n)) = 1.4681 \cdot n \cdot (\log(n) - 1)$ (blue line).

Bibliography

- [1] Lada A Adamic. The Political Blogosphere and the 2004 U . S . Election : Divided They Blog. pages 36–43, 2005.
- [2] Yong-Yeol Ahn, James P Bagrow, and Sune Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–4, August 2010.
- [3] Christopher Aicher, Abigail Z Jacobs, and Aaron Clauset. Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3(2):221–248, 2015.
- [4] Alex Arenas, Albert Díaz-Guilera, and Conrad J. Pérez-Vicente. Synchronization reveals topological scales in complex networks. *Phys. Rev. Lett.*, 96:art. no. 114102, 2006.
- [5] Albert-László Barabási, Natali Gulbahce, and Josep Loscalzo. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.*, 12:58–68, 2011.
- [6] Albert-László Barabási and Zoltán N Oltvai. Network biology: understanding the cell’s functional organization. *Nat. Rev. Genet.*, 5(2):101–113, Feb 2004.
- [7] V. Batagelj and A. Mrvar. Pajek datasets, <http://vlado.fmf.uni-lj.si/pub/networks/data>, 2006.
- [8] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008+, October 2008.
- [9] Stephen P Borgatti, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. Network analysis in the social sciences. *Science*, 323(5916):892–895, Feb 2009.
- [10] R. J. Boys, D. J. Wilkinson, and T. B. L. Kirkwood. Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing*, 18(2):125–135, November 2007.

- [11] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.*, 10(3):186–198, March 2009.
- [12] Suzan L Carmichael, Chen Ma, Sonja A Rasmussen, Margaret A Honein, Edward J Lammer, and Gary M Shaw. Craniosynostosis and maternal smoking. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 82(2):78–85, 2008.
- [13] Aaron Clauset, Cristopher Moore, and M. E J Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, May 2008.
- [14] Peter Csermely, Tamás Korcsmáros, Huba J M Kiss, Gábor London, and Ruth Nussinov. Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. *Pharmacol. Therapeut.*, 138(3):333–408, February 2013.
- [15] M De Domenico, A Solé-Ribalta, S Gómez, and A Arenas. Navigability of interconnected networks under random failures. *Proc. Natl. Acad. Sci. U.S.A.*, page 8351–8356, 2014.
- [16] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A. Porter, Sergio Gómez, and Alex Arenas. Mathematical Formulation of Multilayer Networks. *Phys. Rev. X*, 3:041022, 2013.
- [17] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84:066106, 2011.
- [18] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Phys. Rev. Lett.*, 107(6):065701, Aug 2011.
- [19] Johnny B Delashaw, John A Persing, William C Broaddus, and John A Jane. Cranial vault growth in craniosynostosis. *Journal of Neurosurgery*, 70(2):159–165, 1989.
- [20] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B Met.*, 39(1):1–38, 1977.
- [21] Federico Di Rocco, Eric Arnaud, and Dominique Renier. Evolution in the frequency of nonsyndromic craniosynostosis. *Journal of Neurosurgery: Pediatrics*, 4(1):21–25, 2009.
- [22] Borja Esteve-Altava, Jesús Marugán-Lobón, Héctor Botella, Markus Bastir, and Diego Rasskin-Gutman. Grist for riedl’s mill: a network model perspective on the integration and modularity of the human skull. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 320(8):489–500, 2013.

-
- [23] Borja Esteve-Altava, Jesús Marugán-Lobón, Héctor Botella, and Diego Rasskin-Gutman. Structural constraints in the evolution of the tetrapod skull complexity: Williston's law revisited using network models. *Evolutionary Biology*, 40(2):209–219, 2013.
- [24] Borja Esteve-Altava, Jesús Marugán-Lobón, Héctor Botella, and Diego Rasskin-Gutman. Random loss and selective fusion of bones originate morphological complexity trends in tetrapod skull networks. *Evolutionary Biology*, 41(1):52–61, 2014.
- [25] Borja Esteve-Altava and Diego Rasskin-Gutman. Beyond the functional matrix hypothesis: a network null model of human skull growth for the formation of bone articulations. *Journal of Anatomy*, 225(3):306–316, 2014.
- [26] Borja Esteve-Altava and Diego Rasskin-Gutman. Theoretical morphology of tetrapod skull networks. *Comptes Rendus Palevol*, 13(1):41–50, 2014.
- [27] Borja Esteve-Altava and Diego Rasskin-Gutman. Evo-devo insights from pathological networks: exploring craniosynostosis as a developmental mechanism for modularity and complexity in the human skull. *Journal of Anthropological Sciences*, 93:1–15, 2015.
- [28] S. Fortunato. Community detection in graphs. *Phys. Rep.*, 486:75–174, 2010.
- [29] Rebecca M Garza and Rohit K Khosla. Nonsyndromic craniosynostosis. In *Seminars in Plastic Surgery*, volume 26, page 53, 2012.
- [30] Cattuto C Gauvin L, Panisson A. Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach. *PLoS ONE*, 2014.
- [31] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov 1984.
- [32] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99:7821–7826, 2002.
- [33] P. M. Gleiser and L. Danon. Community structure in jazz. *Adv. Complex Syst.*, 6:565–573, 2003.
- [34] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas. Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.*, 110:028701, Jan 2013.
- [35] Henry Gray. *Anatomy of the human body*. Lea & Febiger, 1918.

- [36] William K Gregory. ‘williston’s law’ relating to the evolution of skull bones in the vertebrates. *American Journal of Physical Anthropology*, 20(2):123–152, 1935.
- [37] R. Guimerà and L. A. N. Amaral. Functional cartography of complex metabolic networks. *Nature*, 433:895–900, 2005.
- [38] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in a network of human interactions. *Phys. Rev. E*, 68:art. no. 065103, 2003.
- [39] R. Guimerà, A. Llorente, E. Moro, and M. Sales-Pardo. Predicting human preferences using the block structure of complex social networks. *PLOS ONE*, 7(9):e44620, 2012.
- [40] R. Guimerà, S. Mossa, A. Turttschi, and L. A N Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles. *Proc. Natl. Acad. Sci. USA*, 102(22):7794–7799, May 2005.
- [41] R Guimerà and M Sales-Pardo. Justice blocks and predictability of U.S. Supreme Court votes. *PLOS ONE*, 6(11):e27188, 2011.
- [42] R. Guimerà, M. Sales-Pardo, and L.A.N. Amaral. Classes of complex networks defined by role-to-role connectivity profiles. *Nature Phys.*, 3:63–69, 2007.
- [43] Roger Guimerà and Luis Antonio Nunes Amaral. Modeling the world-wide airport network. *Eur. Phys. J. B*, 38:381–385, 2004.
- [44] Roger Guimerà and Marta Sales-Pardo. Missing and spurious interactions and the reconstruction of complex networks. *Proc. Natl. Acad. Sci. U. S. A.*, 106(52):22073–22078, 2009.
- [45] Roger Guimerà and Marta Sales-Pardo. A network inference method for large-scale unsupervised identification of novel drug-drug interactions. *PLOS Comput. Biol.*, 9(12):e1003374, January 2013.
- [46] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Soc. Networks*, 5:109–137, 1983.
- [47] Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, October 2012.
- [48] Takayuki Inagaki, Shigeo Kyutoku, Takatoshi Seno, Takuya Kawaguchi, Takashi Yamahara, Hideyuki Oshige, Yasuo Yamanouchi, and Keiji Kawamoto. The intracranial pressure of the patients with mild form of craniosynostosis. *Child’s Nervous System*, 23(12):1455–1459, 2007.
- [49] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

-
- [50] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83:016107, 2011.
- [51] Robert E. Kass and Adrian E. Raftery. Bayes factors. *J. Am. Stat. Assoc.*, 90(430):791, 1995.
- [52] Myunghwan Kim and J Leskovec. Nonparametric multi-group membership model for dynamic networks. In C J C Burges, L Bottou, M Welling, Z Ghahramani, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1385–1393. Curran Associates, Inc., 2013.
- [53] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.
- [54] Mikko Kivela, Alex Arenas, M Barthelemy, James P Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer Networks. *J. Complex Netw.*, 2(3):203–271, 2014.
- [55] V. Krebs. <http://www.orgnet.com>.
- [56] Daniel B. Larremore, Aaron Clauset, and Abigail Z. Jacobs. Efficiently inferring community structure in bipartite networks. *Phys. Rev. E*, 90:012805, 2014.
- [57] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Tec.*, 58:1019–1031, 2007.
- [58] Daniel Lieberman. *The evolution of the human head*. Harvard University Press, 2011.
- [59] Changning Liu, Jing Li, and Yi Zhao. Exploring hierarchical and overlapping modular structure in the yeast protein interaction network. *BMC Genomics*, 11 Suppl 4:S17, 2010.
- [60] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.
- [61] Kurt Miller, Michael I Jordan, and Thomas L Griffiths. Nonparametric Latent Feature Models for Link Prediction. In Y Bengio, D Schuurmans, J D Lafferty, C K I Williams, and A Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 1276–1284. Curran Associates, Inc., 2009.
- [62] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
- [63] Gillian M Morriss-Kay and Andrew OM Wilkie. Growth of the normal skull vault and its alteration in craniosynostosis: insights from human genetics and experimental studies. *Journal of Anatomy*, 207(5):637–653, 2005.

- [64] ML Moss. Functional anatomy of cranial synostosis. *Pediatric Neurosurgery*, 1(1):22–33, 1975.
- [65] Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, May 2010.
- [66] M. E. J. Newman. Communities, modules and large-scale structure in networks. *Nat. Phys.*, 8:25–31, 2011.
- [67] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):art. no. 026113, 2004.
- [68] Mark E. J. Newman. *Networks: An Introduction*. Oxford University Press, USA, 1 edition, 2010.
- [69] Takashi Nishikawa and Adilson E Motter. Discovering network structure beyond communities. *Scientific reports*, 1:151, January 2011.
- [70] Krzysztof Nowicki and Tom A. B. Snijders. Estimation and prediction for stochastic blockstructures. *J. Am. Stat. Assoc.*, 96:1077–1087, 2001.
- [71] Adam J Oppenheimer, Samuel T Rhee, Steven A Goldstein, and Steven R Buchman. Force-induced craniosynostosis in the murine sagittal suture. *Plastic and Reconstructive Surgery*, 124(6):1840, 2009.
- [72] Konstantina Palla, D Knowles, and Zoubin Ghahramani. An infinite latent attribute model for network data. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 1607–1614, 2012.
- [73] Tiago P. Peixoto. Entropy of stochastic blockmodel ensembles. *Physical Review E*, 85(5):056122, May 2012.
- [74] Tiago P. Peixoto. Parsimonious module inference in large networks. *Phys. Rev. Lett.*, 110:148701, Apr 2013.
- [75] Tiago P. Peixoto. Efficient monte carlo and greedy heuristic for the inference of stochastic block models. *Phys. Rev. E*, 89:012804, Jan 2014.
- [76] Tiago P. Peixoto. Hierarchical block structures and high-resolution model selection in large networks. *Phys. Rev. X*, 4:011047, Mar 2014.
- [77] C Percival and JT Richtsmeier. The epigenetics of dysmorphology. In Benedikt Hallgrímsson and Brian K Hall, editors, *Epigenetics linking genotype and phenotype in development and evolution*, pages 377–397. University Press, 2011.
- [78] P. Pons and M. Latapy. Computing communities in large networks using random walks (long version). *ArXiv Physics*, page arXiv:physics/0512106, Dec 2005.

-
- [79] R. G. Morris and M Barthelemy. Transport on coupled spatial networks. *Phys. Rev. Lett.*, 109:128703, 2012.
- [80] Filippo Radicchi and Alex Arenas. Abrupt transition in the structural formation of interconnected networks. *Nat. Phys.*, 9(11):717–720, September 2013.
- [81] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. *Proc. Natl. Acad. Sci. USA*, 101(9):2658–2663, 2004.
- [82] Joerg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Phys. Rev. E*, 74(1 Pt 2):016110, Jul 2006.
- [83] D Rice. Clinical features of syndromic craniosynostosis. In D Rice, editor, *Craniofacial sutures development, disease, and treatment*, pages 91—106. Karger, 2008.
- [84] Joan T Richtsmeier, Kristina Aldridge, Valerie B DeLeon, Jayesh Panchal, Alex A Kane, Jeffrey L Marsh, Peng Yan, and Theodore M Cole. Phenotypic integration of neurocranium and brain. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 306(4):360–378, 2006.
- [85] Rudolf P. Rohr, Serguei Saavedra, and Jordi Bascompte. On the structural stability of mutualistic systems. *Science*, 345(6195):416–425, 2014.
- [86] Núria Rovira-Asenjo, Tània Gumí, Marta Sales-Pardo, and Roger Guimerà. Predicting future conflict between team-members with parameter-free models of social networks. *Sci. Rep.*, 3:1999, June 2013.
- [87] D. J. Watts S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.
- [88] Serguei Saavedra, Felix Reed-Tsochas, and Brian Uzzi. A simple model of bipartite cooperation for ecological and organizational networks. *Nature*, 457(7228):463–6, January 2009.
- [89] M. Sales-Pardo, R. Guimerà, A. A. Moreira, and L. A. N. Amaral. Extracting the hierarchical organization of complex systems. *Proc. Natl. Acad. Sci. USA*, 104:15224–15229, 2007.
- [90] Joaquin Sanz, Cheng-Yi Xia, Sandro Meloni, and Yamir Moreno. Dynamics of interacting diseases. *Phys. Rev. X*, 4:041005, Oct 2014.
- [91] M. N. Schmidt and M. Morup. Nonparametric bayesian modeling of complex networks: an introduction. *IEEE Signal Processing Magazine*, 30(3):110–128, May 2013.

- [92] Frank Schweitzer, Giorgio Fagiolo, Didier Sornette, Fernando Vega-Redondo, Alessandro Vespignani, and Douglas R White. Economic networks: the new challenges. *Science*, 325(5939):422–425, Jul 2009.
- [93] Min Shi, George L Wehby, and Jeffrey C Murray. Review on genetic variants and maternal smoking in the etiology of oral clefts and other birth defects. *Birth Defects Research Part C: Embryo Today: Reviews*, 84(1):16–29, 2008.
- [94] Christian A Sidor. Simplification as a trend in synapsid cranial evolution. *Evolution*, 55(7):1419–1442, 2001.
- [95] John Skilling. Nested sampling. 735:395–405, July 2004.
- [96] Geoffrey H Sperber, Steven M Sperber, and Geoffrey D Guttman. *Craniofacial embryogenetics and development*. PMPH-USA, 2010.
- [97] Chris Stark, Bobby-Joe Breitzkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitzkreutz, and Mike Tyers. BioGRID: a general repository for interaction datasets. *Nucleic acids research*, 34(Database issue):D535–9, January 2006.
- [98] Michael Szell, Renaud Lambiotte, and Stefan Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences of the United States of America*, 107(31):13636–41, August 2010.
- [99] R Core Team. R: a language and environment for statistical computing. r foundation for statistical computing, vienna, austria, 2014.
- [100] Ross M. Thompson, Ulrich Brose, Jennifer A. Dunne, Robert O. Hall, Sally Hladysz, Roger L. Kitching, Neo D. Martinez, Heidi Rantala, Tamara N. Romanuk, Daniel B. Stouffer, and Jason M. Tylianakis. Food webs: reconciling the structure and function of biodiversity. *Trends Ecol. Evol.*, 27(12):689–697, 2012.
- [101] Stephen RF Twigg and Andrew OM Wilkie. A genetic-pathophysiological framework for craniosynostosis. *The American Journal of Human Genetics*, 97(3):359–377, 2015.
- [102] S. Wasserman and K. Faust. *Social Network Analysis*. Cambridge University Press, Cambridge, UK, 1994.
- [103] Stephanie E Watkins, Robert E Meyer, Ronald P Strauss, and Arthur S Aylsworth. Classification, epidemiology, and genetics of orofacial clefts. *Clinics in Plastic Surgery*, 41(2):149–163, 2014.
- [104] Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. Virality prediction and community structure in social networks. *Sci. Rep.*, 3:2522, 2013.

- [105] Harrison C. White, Scott A. Boorman, and Ronald L. Breiger. Social structure from multiple networks. i. blockmodels of roles and positions. *Am. J. Sociol.*, 81(4):730–780, 1976.
- [106] J G White, E Southgate, J N Thomson, and S Brenner. The structure of the nervous system of the nematode *c. elegans*. *Philos. T. R. Soc. B.*, 314:1–340, 1986.
- [107] Nicholas C Wormald. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, pages 239–298, 1999.
- [108] Xiaoran Yan, Cosma Shalizi, Jacob E Jensen, Florent Krzakala, Cristopher Moore, Lenka Zdeborová, Pan Zhang, and Yaojia Zhu. Model selection for degree-corrected block models. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(5):P05007, 2014.
- [109] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Understanding Belief Propagation and its Generalizations. 2002.
- [110] W. Zachary. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.*, 33(4):452–473, 1977.