Systematic assessment of the quality of fit of the stochastic block model for empirical networks

Felipe Vaca-Ramírez and Tiago P. Peixoto
Department of Network and Data Science, Central European University, 1100 Vienna, Austria

We perform a systematic analysis of the quality of fit of the stochastic block model (SBM) for 275 empirical networks spanning a wide range of domains and orders of size magnitude. We employ posterior predictive model checking as a criterion to assess the quality of fit, which involves comparing networks generated by the inferred model with the empirical network, according to a set of network descriptors. We observe that the SBM is capable of providing an accurate description for the majority of networks considered, but falls short of saturating all modeling requirements. In particular, networks possessing a large diameter and slow-mixing random walks tend to be badly described by the SBM. However, contrary to what is often assumed, networks with a high abundance of triangles can be well described by the SBM in many cases. We demonstrate that simple network descriptors can be used to evaluate whether or not the SBM can provide a sufficiently accurate representation, potentially pointing to possible model extensions that can systematically improve the expressiveness of this class of models.

I. INTRODUCTION

The stochastic block model (SBM) \cite{1, 2} is an important family of generative network models used primarily for community detection \cite{3} and link prediction \cite{4}. In its simplest formulation, it describes a network formation mechanism where the nodes are divided into discrete groups, and the probability of an edge existing between two nodes is given as a function of their group memberships. Many variations of this idea exist, including mixed-membership SBMs \cite{5}, where nodes are allowed to belong to multiple groups, the degree-corrected SBM (DCSBM) \cite{2}, where nodes are allowed to possess arbitrary degrees, as well as several extensions to other domains, such as dynamical networks \cite{6–8} and multilayer networks \cite{7, 9}, to name a few.

SBMs also serve as generalizations of more fundamental random network models. The basic SBM has the Erdős-Rényi model \cite{10} as a special case when there is a single group, and likewise the DCSBM recovers the configuration model \cite{11} in the same situation. However, differently from these more fundamental models, the SBM possesses a set of parameters — the partition of the nodes and the affinities between groups — that is not trivially recoverable from observed networks. These parameters are latent information that need to be obtained via inference algorithms, which form the basis of the community detection methods that use this approach \cite{3}. Furthermore, the SBM has a controllable level of complexity: by increasing the number of groups, we have the ability to express increasingly elaborate types of network structures, via arbitrary mixing patterns between the latent groups. In fact, despite its stylized nature, it can be shown that the SBM can approximate a broad class of generative models that are different from it \cite{12}, and its inference functions similarly to fitting a histogram to numeric data in order to estimate the underlying probability density — with the node groups playing a similar role to the histogram bins. However, the expressiveness of the SBM is not absolute, specially when the networks are sparse, i.e. when their average degree is much smaller than the total number of nodes. In such a situation, there is no guarantee that the SBM is capable of arbitrarily approximating the true underlying model, regardless of how we infer it: By increasing the model complexity we move from a situation where we are underfitting, i.e. extracting patterns that do not sufficiently capture all the features of the true model, to a situation where we are overfitting, i.e. incorporating randomness into the model description, which is also a deviation from the true model. When we find the most adequate inference that balances statistical evidence against model complexity to prevent overfitting, we might still be missing important features of the true model, simply because it cannot be sufficiently well captured under the SBM parametrization.

Here we are not interested in evaluating the SBM as a plausible generative process of networks across all domains, since it does not represent an ultimately credible mechanism for any of them. Instead, our objective is to assess how capable it is of providing a general effective description of empirical networks, and in which aspects and to what extent (and not whether) it tends to be misspecified. Understanding the limits of the SBM representation in empirical settings is therefore a nuanced undertaking that is likely to be affected by a variety of possible sources of deviations. Since the SBM tends to yield very good comparative performance in link prediction tasks \cite{13, 14}, it is therefore known that it tends to outperform alternative models in capturing the structure of networks, but we still lack a more accurate assessment of its qualities and shortcomings in absolute terms.

In this work, we evaluate the quality of fit of the SBM in empirical contexts by performing model checking on Bayesian inferences. Based on a diverse collection of 275
networks spanning various domains and several orders of size magnitude, we compare the values of many network descriptors computed on the observed network with what would be typically obtained with networks sampled from the inferred SBM. In this way, any significant discrepancy can be interpreted as a form of "residual" that points to a shortcoming of the SBM in capturing that particular network property.

Overall we find that the SBM is capable of encapsulating the network structure to a significant degree for a large fraction of the networks studied, but falls short of completely exhausting the modelling requirements in many cases. We find that for networks with very large diameter or a very slow mixing random walk the SBM tends to provide a poor description. This includes, for example, many transportation networks — which are typically embedded in a low dimensional space — as well as some economic networks. However, for other kinds of networks the quality of fit tends to be good overall.

We proceed with describing in detail the model and inference procedure (Sec. [I]), our criteria to evaluate the quality of fit (Sec. [II]), the network corpus used (Sec. [III]), and the results of our analysis for it (Sec. [IV]). We finalize in Sec. [VII] with a conclusion.

II. MODEL AND INFERENCE

For our analysis we will use the microcanonical degree-corrected SBM (DCSBM) [2, 15], which combines arbitrary mixing patterns between groups together with arbitrary degree sequence. It has as parameters the partition of the nodes into groups, \( b = \{ b_i \}, \) with \( b_i \in [1, B] \) being the group membership of node \( i, \) the degree sequence \( k = \{ k_i \}, \) where \( k_i \) is the degree of node \( i, \) and the edge counts between groups \( e = \{ e_{rs} \} \) (or twice that number for \( r = s \)), given by \( e_{rs} = \sum_{ij} A_{ij}\delta_{b_i,r}\delta_{b_j,s}. \) Given these constraints, the network is generated with probability [15]

\[
P(A|k, e, b) = \prod_{r<s} e_{rs}! \prod_r e_{rr}! \prod_i k_i! \prod_{i<j} A_{ij}! \prod_r e_{rr}!,
\]

where \( A = \{A_{ij}\} \) is the adjacency matrix of an undirected multigraph with potential self-loops, and \( e_r = \sum_s e_{rs}. \)

All the networks we will be studying are undirected simple graphs, for which the above model can give only an approximation. As demonstrated in Ref. [16], the use of multigraph models based on the Poisson distribution (or equivalently, microcanonical models based on the pairing of half-edges, as above) cannot ascribe probabilities to simple edges (i.e. \( A_{ij} = 1 \)) that are larger than \( 1/e \approx 0.37. \) This limits the applicability of such models on networks with heterogeneous density, either due to broad degree distributions or sufficiently dense communities, which are ubiquitous properties of empirical networks. To address this limitation, we use the latent multigraph model of Ref. [16], where we assume that an underlying unobserved multigraph \( A \) is in fact responsible for the observed simple graph \( G \) simply via the removal of the edge multiplicities and self-loops, i.e.

\[
P(G|A) = \prod_{i<j} \left(1 - \delta_{A_{ij},0}\right)^{G_{ij}} \delta_{A_{ij},0}^{1-G_{ij}}.
\]

Note that \( P(G|A) \) can only take a value of 0 or 1, depending on whether \( G \) and \( A \) are compatible. Via this mathematical construction, the final model

\[
P(G|k, e, b) = \sum_A P(G|A)P(A|k, e, b)
\]

can express both arbitrary mixing patterns between groups as well as degree correction, without the limitations of the multigraph model for networks with large local densities [16]. The inference of this model is performed by sampling from the posterior distribution

\[
P(A, k, e, b|G) = \frac{P(G|A)P(A|k, e, b)P(k, e, b)}{P(G)}
\]

which remains tractable. Here we use the merge-split Markov chain Monte Carlo (MCMC) algorithm described in Ref. [17] to efficiently sample from this distribution.

Note that for \( P(k, e, b) \) we use the nonparametric microcanonical hierarchical priors and hyperpriors described in Refs. [15, 18]. Importantly, this kind of approach determines the appropriate model complexity (via the number of groups) according to the statistical evidence available in the data. As has been shown in these previous works, this choice guarantees that only compressive inferences are made in a manner that prevents overfitting (finding a number of groups that is too large), but also with a substantial protection against underfitting (finding a number that is too small), which tends to happen when noninformative priors are used instead.

In addition to the DCSBM we will also use the configuration model as a comparison, obtained by reshuffling the edges of the obtained network while preserving its degree sequence (here we use the edge-switching MCMC algorithm [11]). We note that the configuration model is an approximate special case of the DCSBM considered above when there is only a single group. Therefore, whenever the Bayesian approach above identifies more than one group with a large probability, this automatically implies a selection of the DCSBM in favor of the configuration model. This happens for every network that we consider in this work, meaning that the DCSBM is the favored model for all of them. Nevertheless, the configuration model serves as a good baseline to determine to what extent the quality of fit obtained with the DCSBM can be ascribed to the degree sequence alone or to the group-based mixing patterns uncovered.

\[1\] This is only approximately true since the configuration model and the latent Poisson models are not identical, but sufficiently similar for the purposes of this work [16].
III. ASSESSING QUALITY OF FIT

The approach we use to assess the quality of fit of the DCSBM is based on obtaining the posterior predictive distribution of certain network descriptors. More precisely, for a scalar network descriptor $f(G)$, its posterior predictive distribution is given by

$$P(y|G) = \sum_{G',A,A'} \delta(y - f(G')) P(G'|A') \times P(A'|e,k,b) P(A,k,e,b(G)).$$

In other words, for each inferred parameter set $(k,e,b)$, weighted according to its posterior probability, we sample a new network $G'$ from the model defined above (which can be done in time $O(E + N)$ where $E$ and $N$ are the total number of edges and nodes, respectively, as we show in Appendix A), and obtain the descriptor value $y = f(G')$. \footnote{The posterior predictive distribution for the configuration model is analogous, i.e. $P(y|G) = \sum_{k,e} \delta(y - f(G')) P(G'|k)$, where $k$ are the observed degrees, and $P(G|k)$ is the likelihood of the configuration model.}

We can say that a model captures well the value of a descriptor if its predictive posterior distribution ascribes high probability to values that are close to what was observed in the original network. We can obtain a compact summary of the level of agreement in two different ways. The first measures the statistical significance of the deviation, e.g. via the $z$-score

$$z = \frac{f(G) - \langle y \rangle}{\sigma_y},$$

where $\langle y \rangle$ and $\sigma_y$ are the mean and standard deviation of $P(y|G)$. The second criterion is the relative deviation, which here we compute in two different ways,

$$\Delta_1 = \frac{f(G) - \langle y \rangle}{f(G)}, \quad \Delta_2 = \frac{f(G) - \langle y \rangle}{f_{\text{max}} - f_{\text{min}}},$$

depending on whether the descriptor values are bounded in a well defined interval $[f_{\text{min}}, f_{\text{max}}]$ ($\Delta_2$) or not ($\Delta_1$).

The $z$-score and relative deviation measure complementary aspects of the agreement between data and model, and represent different criteria which should be used together. While a high value of the $z$-score can be used to reject the inferred model as a plausible explanation for the data, by itself it tells us nothing about how good an approximation it is. Conversely, the relative deviation tells us how well the descriptor is being reproduced by the model, but nothing about the statistical significance of the comparison.

In Fig. 1 we show examples that illustrate how the different criteria operate. In Fig. 1(a) and (b) we see examples that show good and bad agreements between model and data, respectively, according to both criteria simultaneously. In these cases, the conclusion is unambiguous: we either see no reason whatsoever to condemn the model, or we see a definitive reason to do so. However, in Fig. 1(c) and (d) we reach mixed conclusions. Fig. 1(c) shows that the model typically yields different values than observed in the data, but it still ascribes a large probability to it. We cannot condemn the model as an implausible explanation for the data, but it is conceivable that the true generative model would be more concentrated on the observed value. Conversely, in Fig. 1(d) we see a situation where the model ascribes close to zero probability to the actual descriptor value seen in the data, but, in absolute terms, the discrepancy is quite small. Although we find evidence to condemn the plausibility of the model, we could still claim that it is a good approximation.

Overall, since we know that a model like the DCSBM cannot possibly correspond to the true generative model of empirical networks, we should expect that in situations where the network is sufficiently large, and hence there is more abundant data, the values of the $z$-score will tend to be high. Here we argue that since the objective of a model like the DCSBM is to obtain a good approxima-
Before continuing, some important considerations regarding model checking should be made. While an excellent model should fulfill both of the above criteria simultaneously, we need to observe that a model that maximally overfits, i.e. ascribes to the observed network a probability of one, and to any other a probability of zero, will achieve the best possible performance according to both relative deviation and statistical significance. This occurs because we are using the same data to perform both the model inference and evaluate its quality, which is an invalid approach for model selection. Therefore, it is important to recognize the crucial difference between model checking and model selection: the latter attempts to find the model alternative that is better justified according to statistical evidence, while the former simply finds systematic discrepancies between the inferred model and data. In our analysis, protection against overfitting is obtained via Bayesian inference, and we use model checking only to evaluate the discrepancies (indeed, the fact we find discrepancies to begin with shows that we cannot be massively overfitting). Another observation is that when performing multiple comparison over many networks and descriptors, some amount of “statistically significant” deviations are always expected, even if the models inferred correspond to the true ones, unless we incorporate the fact that we are doing multiple comparisons in our criterion of statistical significance, which would be the methodologically correct approach. We will not perform such a correction in our analysis, because we do not seek to demonstrate the absolute quality of DCSBM as a ultimately plausible hypothesis for network formation. As we will see from our results, such a correction would gain us very little.

Finally, in Table I, we list the network descriptors that are used in this work. Our approach requires scalar values, so we constrained ourselves to this category, and furthermore we chose quantities that can be computed quickly, so that robust statistics from the predictive posterior distributions can be obtained. Given these restrictions, we then chose descriptors that measure different aspects of the network structure, both at a local and global levels. Further details on the network descriptors are given in Appendix C.

### IV. NETWORK CORPUS

We base our analysis on a corpus containing 275 networks spanning various domains and several orders of size magnitude, as shown in Fig. 2. We have not collected every network at our disposal, but instead chosen networks that are as diverse as possible, both in size and domain, and avoided many networks that are closely related by belonging to the same subset. In Appendix C, we give more details about the datasets used.
Figure 3. Distribution of relative deviation (top), z-score (middle), and fraction of networks reproduced (bottom) for (a) the configuration model and (b) the DCSBM, according to their respective predictive posterior distributions for each descriptor. We also show the median and mean of the absolute values for all descriptors for each network. The solid blue lines mark the negative and positive median values, and the dashed red line marks the values of $|\Delta| = 0.05$ and $|z| = 3$. The fraction of networks reproduced correspond to those that have the absolute value of either $\Delta$ or $z$ below these thresholds. The points in green color correspond to the networks that are not reproduced according to this combined criterion.

V. RESULTS

In Fig. 3 we show the summaries of the posterior predictive checks for each descriptor and network, for both models considered. We observe a wide variety of deviation magnitudes, both for the same descriptors across networks, and across descriptors. As expected, the DCSBM results show systematically better agreement with the data when compared with the configuration model. Overall, the descriptors that show the worst agreement are the characteristic time of a random walk ($\tau$) and the diameter ($\varnothing$), both of which are particularly high for networks that are embedded in two dimensions, and for which the DCSBM is an inaccurate approximation (more on this below). Nevertheless, there is no single descriptor that the DCSBM does not capture for fewer than 50% of the networks. For descriptors like $S$, $R_t$, $R_i$, and $\langle c \rangle$, the difference between the DCSBM and the configuration model are relatively minor, indicating that those can be captured to a substantial degree by the degree sequence alone.

When considering all descriptors simultaneously for each network, either by the median or mean of the absolute values of the z-score and relative deviation, we observe that a substantial majority of the networks considered show good agreement with the DCSBM, as opposed to the small minority that agree with the configuration model. The difference between the median and the mean indicates that there is a sizeable fraction of the networks where the agreement is spoiled by a few outlier descriptors — typically $\tau$ and $\varnothing$.

The results obtained by the clustering coefficients are particularly interesting, since it is often the case that they
Figure 4. Relative deviation and $z$-score values for the global and mean local clustering coefficients, $C_g$ and $C_l$, as well as diameter and characteristic time of a random walk, $\varnothing$ and $\tau$, as a function of their empirical values, for every network in the corpus, when using the DCSBM. The dashed red line marks the values of $|\Delta| = 0.05$ and $|z| = 3$. The size of the symbol corresponds to the logarithm of the number of edges in the network, and the darkness to the mean degree.

are well reproduced by the DCSBM. This contrasts with what is commonly assumed, namely that the DCSBM should not be able to capture the abundance of triangles often seen in empirical networks, because in the limit where the number of groups is much smaller than the total number of nodes, the DCSBM becomes locally tree-like [19], with a vanishing probability of forming triangles. Therefore, we may imagine that the situations where there is an agreement with the DCSBM are those where the clustering values are low. However, as we see in Fig. 4(a) to (d), this is not quite true, and we observe good agreements even when the clustering values are high. This illustrates a point made in Ref. [20], that it is possible to obtain an abundance of triangles with the SBM simply by increasing the number of groups, in which case it can be explained as a byproduct of homophily. Indeed this is a situation we see in Fig. 4(a) to (d), where both the relative deviation and $z$-score values can be quite small even for extremal values of clustering. However, we do notice a substantial variability between agreements, and a fair amount of instances where the DCSBM cannot capture the observed clustering values, even when they are moderate or even small. This seems to indicate that there are a variety of processes capable of resulting in high clustering values, with homophily being only one of them [20]. Overall, the mean local clustering values tend to be harder to reproduce than the global clustering values. In both cases, the $z$-scores are systematically high, indicating that the clustering values are in general a good criterion to reject the DCSBM as a statistically plausible model, although the relative deviation values tend to be lower than what one would naively expect, meaning that the model can still serve as a reasonably accurate approximation for clustered networks in many cases.

The behavior seen for the clustering coefficient is different for the diameter and characteristic time of a random walk, which are the least well reproduced descriptors, as shown Fig. 4(e) to (h). For both these descriptors — which are closely related, since a network with a large diameter will also tend to result in a slow mixing random walk — it is rare to find a network with very high empirical values which the DCSBM is able to accurately describe. Therefore it seems indeed that the DCSBM offers an inadequate ansatz to describe the structure of these networks, even by optimally adjusting its complexity.

In Fig. 5 we show how the model assessment depends on the size of the network. As one could expect, the $z$-score values tend to increase for larger networks, as more evidence becomes available against the plausibility of the DCSBM as the true generative model. However, the values of the relative deviation do not change appreciably for larger networks, indicating that it remains a good approximation regardless of the size of the system.\(^3\)

\(^3\) Sampling issues with MCMC could also contribute to the elevated $z$-scores for larger networks, as we discuss in Appendix.
In Fig. 5 we show a summary of the fraction of all networks for which we obtain good agreement with either model, according to the network domains. Overall, we see that most domains show similar levels of agreements, except transportation and economic networks. Transportation networks are often embedded in two-dimensional spaces, resulting in large diameters and slow-mixing random walks. The economic networks considered also tend to show large values of these quantities, so the explanation for their discrepancy is the same.

Figure 5. Absolute value of the relative deviation (top), z-score (middle) and fraction of reproduced descriptors (bottom), as a function of the number of edges, for every network in the corpus. The dashed red line marks the values of $|\Delta| = 0.05$ and $|z| = 3$. The fraction of descriptors reproduced correspond to those that have the value of either $\Delta$ or $z$ below these thresholds. The points in green color correspond to the descriptors that are not reproduced according to this combined criterion.

Figure 6. Fraction of reproduced networks according to their domain, considering the (a) median and (b) mean values of either the z-score, the relative deviations, or their combined values, for both models (as shown in the legend). When the combined values are used, this means that a model is deemed compatible with a network when we obtain either $|\Delta| < 0.05$ or $|z| < 3$.

A. Predicting quality of fit

Now we address the question of whether it is possible to predict the quality of fit of both models considered based solely on the empirical values of the networks descriptors. If we can isolate the descriptors which are most predictive, this would give us a general direction in which more accurate models could be constructed.

In order to evaluate the predictability, we frame it as a binary classification problem, where to each network $i$ is ascribed a binary value $y_i = 0$ if we have simultaneously $|z_i| > 3$ and $|\Delta_i| > 0.05$, or otherwise $y_i = 1$. The feature vector for each network is composed of the empirical values of the descriptors, $\mathbf{x}_i = (r, c, C_1, C_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, R_r, R_t, S, E)$, with the addition of the number of edges $E$. For each network $i$, we train a random forest classifier on the entire corpus with that network removed, and evaluate the prediction score on the held-out network. We then repeat this procedure for all networks in the corpus, and evaluate how well the classifier is able to predict the binary label. We present the results of this experiment in Fig. 7 (top) which shows the receiver operating characteristic (ROC) curve, where the true positive rate and the false positive rate are plotted for all threshold values used to reach a classification. The area under the ROC curve (AUC), shown in the legend, can be equivalently interpreted as the probabil-
Figure 7. Predictiveness of the quality of fit of the generative models considered, according to the empirical descriptor values, framed as a binary classification problem, as described in the text. (a) ROC curve for a leave-one-out random-forest classifier, (b) Gini feature importance for the configuration model, (c) same as (b) but for the DCSBM. Panels (d) and (e) show the best ROC AUC obtained for a set of descriptors of a given size, for the configuration model and DSCBM, respectively. Panels (f) and (g) show the same as (d) and (e), respectively, but with the number of edges excluded from the analysis.

4 We also computed different a measure, called permutation importance, which leads to very similar results (not shown).

In order to understand what is the minimal amount of information required to predict the suitability of both models, and in this way remove the redundancy provided by the different descriptors, we computed the best ROC AUC obtained by a combination of descriptors of a given size, as shown in Fig. 7(d) and (e). In both cases we see that the predictability is saturated by only few descriptors. In the case of the configuration model most of the predictability is already achieved by a combination of \( C_l, \, \tau, \, E \). For the DCSBM we get instead \( (r, \, \varnothing, \, E) \).

If we remove the number of edges from the set of features (since it is not informative on the actual network structure), we obtain instead \( (C_g, \, \lambda_1^A, \, \tau) \) and \( (C_l, \, \lambda_1^H, \, \varnothing) \), for the configuration model and DCSBM, respectively. It should be emphasized that if a descriptor does not appear in the minimal set this does not mean it is not predictive of the quality of fit, only that it offers largely redundant information in that regard. Thus, for both models if we replace \( \varnothing \) with \( \tau \) or \( \lambda_1^H \) with \( \lambda_1^A \), etc, we get similar results. This suggests that, besides spatial embeddedness (which influence \( \varnothing \) and \( \tau \) the most), the addition of explicit mechanisms for triangle formation (which affects \( C_g, C_l, \lambda_1^H, \lambda_1^A \) directly) might improve the overall predictability, which is compatible with what we had already seen in Fig. 5(b) and (c). In both cases, we see that the number of edges is the most predictive descriptor, which is compatible with what we had already seen in Fig. 5(b) and (c).

Since we optimized exhaustively for all descriptor combinations of a given size, care should be taken to avoid overfitting, despite the leave-one-out cross-validation, because the optimization was performed the same set of networks. Because of this, we consider always the smallest set of descriptors that reaches a ROC AUC close to the optimum, not the actual optimum which is likely to be overfitting.
VI. CONCLUSION

We performed a systematic analysis of posterior predictive checks of the SBM on a diverse corpus of empirical networks, spanning a broad range of sizes and domains. Using a variety of network descriptors, we observed that the SBM is able to accurately capture the structure of the majority of networks in the corpus. The types of networks that show the worst agreement with DCSBM tend to possess a large diameter and a slow mixing of random walks — features that are commonly associated with a low-dimensional spatial embedding, and a violation of the “small-world” property. For the other kinds of networks the agreement tends to be fairly good, even for many networks with an abundance of triangles, in contradiction to what is commonly assumed to be possible with this class of models.

We have also identified the minimal set of network descriptors capable of predicting the quality of fit of the SBM, which is composed of the network diameter and characteristic time of a random walk as the most important, followed by clustering as a secondary feature. This points to the most productive directions in which this class of models could be improved.

One of the limitations of our analysis is that it is conditioned on the set of descriptors used, and thus shortcomings or successes of the model with respect to other properties not analysed are not uncovered. A natural extension of our work would be to consider an even broader set of descriptors that could reveal more relevant dimensions for the comparison. This kind of analysis is open ended, as there is no short supply of possible network descriptors. We hope our work will motivate further study in this direction, and with a larger variety of generative models within or beyond the SBM family.

AKNOWLEDGEMENTS

The computational results presented have been achieved using the Vienna Scientific Cluster (VSC).

expressiveness of the DCSBM — which in fact has been observed in a more limited dataset [20].

[1] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt, “Stochastic blockmodels: First steps,” Social Networks 5, 109–137 (1983).
[2] Brian Karrer and M. E. J. Newman, “Stochastic blockmodels and community structure in networks,” Physical Review E 83, 016107 (2011).
[3] Tiago P. Peixoto, “Bayesian stochastic blockmodeling,” arXiv:1705.10225 [cond-mat, physics:physics, stat] (2017) arXiv: 1705.10225.
[4] Roger Guimera and Marta Sales-Pardo, “Missing and spurious interactions and the reconstruction of complex networks,” Proceedings of the National Academy of Sciences 106, 22073–22078 (2009).
[5] Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing, “Mixed Membership Stochastic Blockmodels,” J. Mach. Learn. Res. 9, 1981–2014 (2008).
[6] K.S. Xu and A.O. Hero, “Dynamic Stochastic Blockmodels for Time-Evolving Social Networks,” IEEE Journal of Selected Topics in Signal Processing 8, 552–562 (2014).
[7] Tiago P. Peixoto, “Inferring the mesoscale structure of layered, edge-valued, and time-varying networks,” Physical Review E 92, 042807 (2015).
[8] Amir Ghasemian, Pan Zhang, Aaron Clauset, Christopher Moore, and Leto Peel, “Detectability Thresholds and Optimal Algorithms for Community Structure in Dynamic Networks,” Physical Review X 6, 031005 (2016).
[9] N. Stanley, S. Shai, D. Taylor, and P. J. Mucha, “Clustering Network Layers with the Strata Multilayer Stochastic Block Model,” IEEE Transactions on Network Science and Engineering 3, 95–105 (2016).
[10] Paul Erdős and Alfréd Rényi, “On random graphs, I,” Publicationes Mathematicae (Debrecen) 6, 290–297 (1959).
[11] B. Fosdick, D. Larremore, J. Nishimura, and J. Ugander, “Configuring Random Graph Models with Fixed Degree Sequences,” SIAM Review 60, 315–355 (2018).
[12] Sofia C. Olhede and Patrick J. Wolfe, “Network histograms and universality of blockmodel approximation,” Proceedings of the National Academy of Sciences 111, 14722–14727 (2014).
[13] Amir Ghasemian, Homa Hosseinmardi, and Aaron Clauset, “Evaluating Overfit and Underfit in Models of Network Community Structure,” IEEE Transactions on Knowledge and Data Engineering, 1–1 (2019).
[14] Amir Ghasemian, Homa Hosseinmardi, Aram Galstyan, Edoardo M. Airoldi, and Aaron Clauset, “Stacking models for nearly optimal link prediction in complex networks,” Proceedings of the National Academy of Sciences 117, 23393–23400 (2020).
[15] Tiago P. Peixoto, “Nonparametric Bayesian inference of the microcanonical stochastic block model,” Physical Review E 95, 012317 (2017).
[16] Tiago P. Peixoto, “Latent Poisson models for networks with heterogeneous density,” Physical Review E 102, 012309 (2020).
[17] Tiago P. Peixoto, “Merge-split Markov chain Monte Carlo for community detection,” Physical Review E 102, 012305 (2020).
[18] Tiago P. Peixoto, “Hierarchical Block Structures and High-Resolution Model Selection in Large Networks,” Physical Review X 4, 011047 (2014).
[19] Aurelien Decelle, Florent Krzakala, Christopher Moore, and Lenka Zdeborová, “Asymptotic analysis of the stochastic block model for modular networks and its al-
algorithmic applications," Physical Review E 84, 066106 (2011).

[20] Tiago P. Peixoto, “Disentangling homophily, community structure and triadic closure in networks,” arXiv:2101.02510 [physics, stat] (2021) arXiv: 2101.02510.

[21] Tiago P. Peixoto, “Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models,” Physical Review E 89, 012804 (2014).

[22] Juuso Parkkinen, Janne Sirkkonen, Adam Gyenge, and Samuel Kaski, “A block model suitable for sparse graphs,” in Proceedings of the 7th International Workshop on Mining and Learning with Graphs (MLG 2009).

[23] Karl Rohe, Jun Tao, Xin-tian Han, and Norbert Binkiewicz, “A note on quickly sampling a sparse matrix with low rank expectation,” The Journal of Machine Learning Research 19, 3040–3052 (2018), publisher: JMLR.org.

[24] Tiago P. Peixoto, “the graph-tool python library,” figshare (2014), 10.6084/m9.figshare.1164194 available at https://graph-tool.skewed.de

[25] M. E. J. Newman, “Mixing patterns in networks,” Phys. Rev. E 67, 026126 (2003).

[26] Vladimir Batagelj and Matjaž Zaveršnik, “Fast algorithms for determining (generalized) core groups in social networks,” Advances in Data Analysis and Classification 5, 129–145 (2011).

[27] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” Nature 393, 409–408 (1998).

[28] Ki-ichiro Hashimoto, “Zeta Functions of Finite Graphs and Representations of p-Adic Groups,” in Automorphic Forms and Geometry of Arithmetic Varieties, Advanced Studies in Pure Mathematics, Vol. 15, edited by K. Hashimoto and Y. Namikawa (Academic Press, 1989) pp. 211–280.

[29] T. P. Peixoto, “The Netzschleuder network cata
gologue and repository.” (2020), accessible at https://networks.skewed.de

[30] Aaron Clauset, Ellen Tucker, and Matthias Sainz, “The Colorado Index of Complex Networks,” (2016), accessible at http://sourceforge.net/projects/colorado_e.

[31] Rion Brattig Correia, Luciana P. de Araújo Kohler, K. Hashimoto, and Y. Namikawa (Academic Press, 1989) Vol. 15, edited by K. Hashimoto and Y. Namikawa (Academic Press, 1989) pp. 211–280.

[32] Steven J. Cook, Travis A. Jarrell, Christopher A. Brit
tin, Yi Wang, Adam E. Bloniarz, Maksim A. Yakovlev, Ken C. Q. Nguyen, Leo T.-H. Tang, Emily A. Bayer, Janet S. Duerr, Hannah E. Bilow, Oliver Hobert, David H. Hall, and Scott W. Emmons, “Whole-animal connectomes of both caenorhabditis elegans sexes,” Nature 571, 63–71 (2019).

[33] Nicolas Simonis, Jean-François Rual, Anne-Ruxandra Carvunis, Murat Tasan, Irma Lemmens, Tomoko Hirozane-Kishihara, Tong Hao, Julie M Sahalie, Kavitha Venkatesan, Fana Gebrab, Sebiba Cevik, Niels Klitgord, Changyu Fan, Pascal Braun, Ning Li, Nonoo Ayivi-Guedehoussou, Elizabeth Dann, Nicolas Bertin, David Szeto, Amelie Dricot, Muhammed A Yildirim, Chenwei Lin, Anne-Sophie de Smet, Huey-Ling Kao, Christophe Simon, Alex Smolyar, Jin Sook Ahn, Muneesh Tewari, Mike Boxem, Stuart Mil
stein, Haiyuan Yu, Matjia Dreeze, Jean Vandenhaute, Kristin C Gunsalus, Michael E Cusick, David E Hill, Jan Tavernier, Frederick P Roth, and Marc Vidal, “Empirically controlled mapping of the caenorhabditis elegans protein-protein interactome network.” Nat Methods 6, 47–54 (2008).

[34] Sean R. Collins, Patrick Kemmeren, Xue-Chu Zhao, Jack F. Greenblatt, Forrest Spencer, Frank C.P. Hol
stege, Jonathan S. Weissman, and Nevan J. Krogan, “Toward a comprehensive atlas of the physical interac
tome of saccharomyces cerevisiae,” Mol. Cell. Prote
tomics 6, 439–450 (2007).

[35] Xintian Han, Yi Wang, Adam E. Bloniarz, Maksim A. Yakovlev, Ken C. Q. Nguyen, Leo T.-H. Tang, Emily A. Bayer, Janet S. Duerr, Hannah E. Bilow, Oliver Hobert, David H. Hall, and Scott W. Emmons, “Whole-animal connectomes of both caenorhabditis elegans sexes,” Nature 571, 63–71 (2019).

[36] R. M. Thompson and C. R. Townsend, “Impacts on stream food webs of native and exotic forest: An inter
continental comparison,” Ecology 84, 145–161 (2003).

[37] M. De Donatucci, A. Sole-Ribalta, S. Gomez, and A. Arenas, “Navigability of interconnected networks under random failures,” Proc. Natl. Acad. Sci. 111, 8351–8356 (2014).

[38] Robert E Ulanowicz and Donald L DeAngelis, “Network analysis of trophic dynamics in south florida ecosys
tems,” US Geological Survey Program on the South Florida Ecosystem , 114 (1999).

[39] A. Arenas, M. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, “Synchronization in complex networks,” Phys. Rev. E 77, 056114 (2008).

[40] B. S. Narasimhan and T. E. Baker, “Identification and analysis of anaphase-promoting complex substrates in mammalian cells,” J. Cell Sci. 120, 4771–4780 (2007).

[41] William Gray Roncal, Zachary H. Koterba, Disa Mhembre, Dean M. Kleissas, Joshua T. Vogelstein, Randal Burns, Anita R. Bowles, Dimitrios K. Donavos, Sephira Ryman, Rex E. Jung, Lei Wu, Vincent Calhoun, and R. Jacob Vogelstein, “MIGRAINE: Mri graph reliability analysis and inference for connectomics,” in 2013 IEEE Global Conference on Signal and Information Processing (IEEE, 2013).

[42] Shai S. Chen and D. D. Clark, “A new algorithmic application of the graph-tool python library,” figshare (2014), 10.6084/m9.figshare.1164194 available at https://graph-tool.skewed.de

[43] Shai S. Chen and D. D. Clark, “A new algorithmic application of the graph-tool python library,” figshare (2014), 10.6084/m9.figshare.1164194 available at https://graph-tool.skewed.de

[44] Shai S. Chen and D. D. Clark, “A new algorithmic application of the graph-tool python library,” figshare (2014), 10.6084/m9.figshare.1164194 available at https://graph-tool.skewed.de

[45] Shai S. Chen and D. D. Clark, “A new algorithmic application of the graph-tool python library,” figshare (2014), 10.6084/m9.figshare.1164194 available at https://graph-tool.skewed.de

[46] Shai S. Chen and D. D. Clark, “A new algorithmic application of the graph-tool python library,” figshare (2014), 10.6084/m9.figshare.1164194 available at https://graph-tool.skewed.de
[45] “The organization of neural systems in the primate cerebral cortex,” Proc. R. Soc. Lond. B 252, 13–18 (1993).

[46] Daniel B. Larremore, Aaron Clauset, and Caroine O. Buckee, “A network approach to analyzing highly recombining malaria parasite genes,” PLoS Comput Biol 9, e1003268 (2013).

[47] Jennifer A. Dunne, Conrad C. Labandeira, and Richard J. Williams, “Highly resolved early Eocene food webs show development of modern trophic structure after the end-cretaceous extinction,” Proc. R. Soc. B. 20133280 (2014).

[48] Tad A. Dallas, A. Alonso Aguirre, Sarah Budischak, Colin Carlson, Vanessa Ezenwa, Barbara Han, Shan Huang, and Patrick R. Stephens, “Gauging support for macroecological patterns in helminth parasites,” Global Ecol Biogeogr 27, 1437–1447 (2018).

[49] Makoto Kato, Takeshi Kawatani, Damiji Inoue, and Takao Itino, “Insect-flower relationship in the primary beech forest of ashu, kyoto: An overview of the flowering phenomenology and the seasonal pattern of insect visits,” Nihon Gomu Kyoukoushi 61, 281–283 (1988).

[50] Charles Robertson, Flowers and insects; lists of visitors of four hundred and fifty-three flowers, by Charles Robertson. (n.p., 1928).

[51] G. Joshi-Tope, “Reactome: A knowledgebase of biological pathways,” Nucleic Acids Res. 33, D428–D432 (2004).

[52] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, “Network motifs: Simple building blocks of complex networks,” Science 298, 824–827 (2002).

[53] “The CAIDA AS relationships dataset,” (2009).

[54] Matei Ripeanu and Ian Foster, “Mapping the gnutella network: Macroscopic properties of large-scale peer-to-peer systems,” in Peer-to-Peer Systems (Springer Berlin Heidelberg, 2002) pp. 85–93.

[55] Brian Karrer, M. E. J. Newman, and Lenka Zdeborová, “Percolation on sparse networks,” Phys. Rev. Lett. 113 (2014), 10.1103/physrevlett.113.208702. Phys. Rev. Lett. 113, 208702 (2014), arXiv:1405.0483v2 [cond-mat.stat-mech].

[56] Simon Knight, Hung X. Nguyen, Nickolas Falkner, Rhys Bowden, and Matthew Roughan, “The internet topology zoo,” IEEE J. Select. Areas Commun. 29, 1765–1775 (2011).

[57] Jérôme Kunegis, “Konect,” in Proceedings of the 22nd International Conference on World Wide Web - WWW ‘13 Companion (ACM Press, 2013).

[58] Lovro Šubelj and Marko Bajec, “Software systems through complex networks science,” in Proceedings of the First International Workshop on Software Mining - SoftwareMining ’12 (ACM Press, 2012).

[59] Duncan J. Watts and Steven H. Strogatz, “Collective dynamics of ‘small-world’ networks,” Nature 393, 440–442 (1998).

[60] “University of Oregon route views project,” (2001).

[61] Lovro Šubelj and Marko Bajec, “Clustering assortativity, communities and functional modules in real-world networks,” (2012) arXiv:1202.3188v1 [physics.soc-ph].

[62] Beichuan Zhang, Raymond Liu, Daniel Massey, and Lixia Zhang, “Collecting the internet AS-level topology,” SIGCOMM Comput. Commun. Rev. 35, 53–61 (2005).

[63] M. E. J. Newman, “Finding community structure in networks using the eigenvectors of matrices,” Phys. Rev. E 74 (2006), 10.1103/physreve.74.036104.

[64] David Newman, “Bag of words data set,” (2019).

[65] D. Dua and C. Graff, “UCI machine learning repository, irvine, CA: University of california, school of information and computer science.” (2019).

[66] Xing Niu, Xinrui Sun, Haofen Wang, Shu Rong, Gulin Qi, and Yong Yu, “Zhishi.me - weaving Chinese linking open data,” in The Semantic Web – ISWC 2011 (Springer Berlin Heidelberg, 2011) pp. 205–220.

[67] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney, “Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters,” Internet Math. 6, 29–123 (2009), arXiv:0810.1355v1 [cs.DS].

[68] C. Harrison C. Römhild, “Bible cross-references,” (2019).

[69] Kurt D. Bollacker, Steve Lawrence, and C. Lee Giles, “CiteSeer,” in Proceedings of the second international conference on Autonomous agents - AGENTS ’98 (ACM Press, 1998).

[70] Andrew Kachites McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymour, Inform. Retrieval 3, 127–163 (2000).

[71] Michael Ley, “The DBLP computer science bibliography: Evolution, research issues, perspectives,” in String Processing and Information Retrieval (Springer Berlin Heidelberg, 2002) pp. 1–10.

[72] Wikipedia Foundation, “Wikipedia downloads,” (2019).

[73] GroupLens Research, “MovieLens data sets,” (2019).

[74] Lada A. Adamic and Natalie Glance, “The political blogosphere and the 2004 U.S. election,” in Proceedings of the 3rd international workshop on Link discovery - LinkKDD ’05 (ACM Press, 2005).

[75] Boris Pasternak and Ivor Ivask, “Four unpublished letters,” Books Abroad 44, 196 (1970).

[76] James H. Fowler and Sangick Jeon, “The authority of supreme court precedent,” Soc. Networks 30, 16–30 (2008).

[77] James H. Fowler, Timothy R. Johnson, James F.3 Swiggs, Sangick Jeon, and Paul J. Wahlbeck, “Network analysis and the law: Measuring the legal importance of precedents at the U.S. supreme court,” Politi. anal. 15, 324–346 (2007).

[78] Peter Bailey, Nick Craswell, and David Hawking, “Engineering a multi-purpose test collection for web retrieval experiments,” Inform. Process. Manag. 39, 853–871 (2003).

[79] Bronwyn Hall, Adam Jaffe, and Manuel Trajtenberg, The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools Tech. Rep. (National Bureau of Economic Research, 2001).

[80] Seán Slattery and Mark Craven, “Combining statistical and relational methods for learning in hypertext retrieval experiments,” in Proceedings of the International Conference on Autonomous agents - AGENTS ’98 (ACM Press, 1998).

[81] Alberto Calderone, “A Wikipedia Based Map of Science,” (2020), 10.6084/m9.figshare.11638932.v5.

[82] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenstiat, Michal Sfeiler, and Uri Alon, “Superfamilies of evolved and designed networks,” Science 303, 1538–1542 (2004).
[122] Martina Morris and Richard Rothenberg, “HIV transmission network metastudy project: An archive of data from eight network studies, 1988–2001.” (2011).
[123] R. Zafarani and H. Liu, “Social computing data repository at Arizona State University.” (2009)
[124] Pablo M. Gleiser and Leon Danon, “Community structure in jazz,” Advs. Complex Syst. 06, 565–573 (2003).
[125] Wayne W. Zachary, “An information flow model for conflict and fission in small groups,” J. Anthropol. Res. 33, 452–473 (1977).
[126] Luke M. Gerdes, Kristine Ringler, and Barbara Autin, “Assessing the abuse sayyaf’s group’s strategic and learning capacities,” Stud. Confl. Terror. 37, 267–293 (2014).
[127] Oscar Cela, “The long tail in recommender systems,” in Proceedings of the 21st ACM conference on Hypertext and hypermedia - HT ’10 (ACM, 2010).
[128] Sandra González-Bailón, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno, “The dynamics of protest recruitment through an online network,” Sci Rep 1 (2011), 10.1038/srep00197.
[129] Jérôme Kunegis, Gerd Gröner, and Thomas Göttron, “Online dating recommender systems,” in Proceedings of the 4th ACM RecSys workshop on Recommender systems and the social web - RSWeb ’12 (ACM Press, 2012).
[130] James Coleman, Elihu Katz, and Herbert Menzel, “The diffusion of an innovation among physicians,” Socometry 20, 253 (1957).
[131] Manlio De Domenico, Andrea Lancichinetti, Alex Arenas, and Martin Rosvall, “Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems,” Phys. Rev. X 5 (2015), 10.1103/physrevx.5.011027 [physics.soc-ph].
[132] Nathaniel Eagle and Alex (Sandy) Pentland, “Reality mining: Sensing complex social systems,” Pers Ubiquit Comput 10, 255–268 (2005).
[133] Linton C. Freeman, Cynthia M. Webber, and Deirdre M. Kirke, “Exploring social structure using dynamic three-dimensional color images,” Soc. Networks 20, 109–118 (1998).
[134] Rossana Mastrandrea, Julie Fournet, and Alain Barrat, “Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys,” PLoS ONE 10, e0136497 (2015).
[135] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cat-tuto, Jean-François Pinton, and Wouter Van den Broeck, “What’s in a crowd? analysis of face-to-face behavioral networks,” J. Theor. Biol. 271, 166–180 (2011).
[136] Mathieu Génois, Christian L. Vestergaard, Julie Fournet, André Panisson, Isabelle Bonmarin, and Alain Bar-rat, “Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers,” Net Sci 3, 326–347 (2015).
[137] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cat-tuto, Lorenzo Isella, Jean-François Pinton, Marco Quag-giotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems, “High-resolution measurements of face-to-face contact patterns in a primary school,” PLoS ONE 6, e23176 (2011).
[138] Michael Fire, Gilad Katz, Yuval Elovici, Bracha Shapira, and Lior Rokach, “Predicting student exam’s scores by analyzing social network data,” in Active Me-dia Technology (Springer Berlin Heidelberg, 2012) pp. 584–595.
[139] Anne-Marie Niekamp, Liesbeth A.G. Merckens, Chris-tian J.P.A. Hoeye, and Nicole H.T.M. Dukers-Muijres, “A sexual affiliation network of swingers, heterosexuals practicing risk behaviours that potentiate the spread of sexually transmitted infections: A two-mode approach,” Soc. Networks 35, 223–236 (2013).
[140] Munmun De Choudhury, “Discovery of information disseminators and receptors on online social media,” in Proceedings of the 21st ACM conference on Hypertext and hypermedia - HT ’10 (ACM, 2010).
[141] Anna Evtushenko and Michael T. Gastner, “Beyond for-mations: Inferring relationships from co-authorship, co-attendance and other co-behaviors,” in Proceedings of the 28th international conference on Human factors in computing systems - CHI ’10 (ACM Press, 2010).
[142] Munmun De Choudhury, “Discovery of information dis-seminators and receptors on online social media,” in Proceedings of the 21st ACM conference on Hypertext and hypermedia - HT ’10 (ACM, 2010).
[143] Michael Fire and Yuval Elovici, “Data mining of online genealogy datasets for revealing lifespan patterns in human population,” ACM Trans. Intell. Syst. Technol. 6, 1–22 (2015) [arXiv:1311.4276v2 [cs.SI]].
[144] L Freeman, S Freeman, and A Michaelson, “On human social intelligence,” Journal of Social and Biological Systems 11, 415–425 (1988).
[145] Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman, “The dynamics of viral marketing,” ACM Trans. Web 1, 5 (2007).
[146] Es-Feng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, and Hady Wirawan Laow, “Detecting product review spammers using rating behaviors,” in Proceedings of the 9th ACM international conference on Information and knowledge management - CIKM ’10 (ACM Press, 2010).
[147] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Kon stan, and Georg Lausen, “Improving recommendation lists through topic diversification,” in Proceedings of the 14th international conference on World Wide Web - WWW ’05 (ACM Press, 2005).
[148] Jan Y. Chang, “Beyond fortune 500: Women in a global network of directors,” [arXiv:1311.4276v2 [cs.SI]].
Appendix A: Posterior predictive sampling

As described in the main text, we obtain samples from the posterior predictive distribution of Eq. 3 by first sampling from the posterior distribution of Eq. 3 using MCMC and then generating new networks from the inferred models. More specifically, we sample \((A, k, e, b)\) from

\[
P(A, k, e, b | G) = \frac{P(G | A) P(A | k, e, b) P(k, e, b)}{P(G)}, \tag{A1}
\]

using the merge-split MCMC of Ref. [17], together with the agglomerative initialization heuristic of Refs. [18, 21], and the multigraph edge moves of Ref. [19]. For networks of size up to \(E = 10^6\) edges we observe good equilibration of the MCMC runs, but for large networks it becomes too slow. For these large networks we settle for a point estimate of the partition \(b\) obtained by several runs of the initialization algorithm and keeping the best result, and then we equilibrate the chain according to \(A\) alone (which affects \(k\) and \(e\)), which tends to happen quickly. We have verified that performing this calculation several times yields very similar results. The only noticeable outcome of this shortcut for larger networks is that it tends to reduce the variance of the posterior predictive distributions, which can potentially contribute to the elevated \(z\)-scores we obtained in our analysis. However, since the relative deviation values we obtained did not seem to depend on the size of the network, this gives us confidence that this approach does not introduce significant biases.

Given a sample \((A, k, e, b)\), we are interested only in \((k, e, b)\) (and hence samples from their marginal distribution), so we discard \(A\) and sample a new multigraph \(A'\) from the model of Eq. 1. This can be done exactly with an efficient algorithm that works similarly to what was proposed in Refs. [22, 23], but is valid for the microcanonical model: Given the parameters \((k, e, b)\) we proceed by creating for each group \(r\) a multiset of candidate nodes \(v_r\), containing \(k_i\) copies of each node \(i\) with \(b_i = r\). Then, for each group pair \((r, s)\) with \(r \neq s\) and \(e_{rs} > 0\), we repeat the following three steps for an \(e_{rs}\) number of times (or \(e_{rs}/2\) if \(r = s\)):

1. We sample a node \(i\) from the multiset \(v_r\) uniformly at random, and we remove it from the multiset.
2. We sample a node \(j\) from the multiset \(v_s\) uniformly at random, and we remove it from the multiset.
3. We add an edge \((i, j)\) to \(A\) (i.e. increment \(A_{ij}\) by one, or two if \(i = j\)).

The resulting multigraph \(A\) is sampled exactly with a probability given by Eq. 1. Since the number of nonzero entries of \(e\) cannot be larger than the total number of edges \(E\), the whole algorithm finishes in time \(O(N + E)\), where \(N\) is the number of nodes.

Given a sample \(A\), we obtain a simple graph \(G\) simply by removing all self-loops and truncating the edge multiplicities, i.e.

\[
G_{ij} = \begin{cases} 
1, & \text{if } A_{ij} > 0 \text{ and } i \neq j, \\
0, & \text{otherwise}.
\end{cases} \tag{A2}
\]

Finally, given \(G\) we compute the network descriptor \(f(G)\) of interest.

A C++ implementation of every algorithm used in this analysis is freely available as part of the \texttt{graph-tool} library [24].

Appendix B: Network descriptors

Below are the definitions of the descriptors used in our analyses.

Degree assortativity, \(r\): Defined as \[25\]

\[
r = \frac{\sum_{kk',k,k'}(m_{kk'} - m_k m_{k'})}{\sigma_k \sigma_{k'}},
\]

where \(m_{kk'}\) is the fraction of edges with endpoints of degree \(k\) and \(k'\), \(m_k = \sum_{k'} m_{kk'}\), and \(\sigma_k\) is the standard deviation of \(m_k\).

Mean \(k\)-core, \(<c>\): The \(k\)-core is a maximal set of vertices such that its induced subgraph only contains vertices with degree larger than or equal to \(k\). The \(k\)-core value \(c_i\) of node \(i\) is the largest value of \(k\) for which \(i\) belongs to the \(k\)-core. The mean value is then

\[
<c> = \frac{1}{N} \sum_i c_i.
\]

This can be computed in time \(O(N + E)\) according to the algorithm of Ref. [20].
Figure 8. Absolute value of the $z$-score versus absolute value of relative deviation, for every descriptor value and network in the corpus, according to (a) the configuration model and (b) the DCSBM. The dashed lines mark the values $|z| = 3$ and $|\Delta| = 0.05$, and the histograms the marginal distributions. The solid blue lines mark the median values.

Figure 9. (a) Kendall’s correlation coefficient $\tau$ between pairs of descriptor values across all networks in the corpus. Panels (b) and (c) show the same but for $z$-score and relative deviation values, respectively, according to the DCSBM. The insets show the correlation between coefficients from each respective panel and panel (a).

Mean local clustering coefficient, $C_l$: The local clustering coefficient \cite{27} of node $i$ is given by

$$C_i = \frac{\sum_{jk} G_{ij} G_{ki} G_{jk}}{k_i (k_i - 1)}.$$

It measures the fraction of pairs of neighbors that are also connected. The mean value is then just

$$C_l = \frac{1}{N} \sum_i C_i.$$

Global clustering coefficient, $C_g$: The global clustering coefficient of is given by

$$C_g = \frac{\sum_{ijk} G_{ij} G_{ki} G_{jk}}{\sum_i k_i (k_i - 1)}.$$

It measures the fraction of connected triads that close to form a triangle.

Leading eigenvalue of adjacency matrix, $\lambda_1$: The leading eigenvalue of the adjacency matrix is the largest value of $\lambda$ which solves

$$G x = \lambda x,$$

where $x$ is the associated eigenvector.
Characteristic time of a random walk, $\tau$:
The characteristic time of a random walk is obtained via the second largest eigenvalue $\lambda_2^T \in [0, 1]$ of the transition matrix $T$, with entries

$$T_{ij} = \frac{G_{ij}}{k_j},$$

where $k_i = \sum_j G_{ji}$. It is defined as

$$\tau = -\ln \lambda_2^T.$$

If the network is disconnected, we compute $\tau$ only on the largest component.

Pseudo-diameter, $\varnothing$: The pseudo-diameter is an approximate graph diameter. It is obtained by starting from an arbitrary source node, and finding a target node that is farthest away from the source. This process is repeated by treating the target as the new starting node, and ends when the graph distance no longer increases. This graph distance is taken to be the pseudo-diameter. The algorithm runs in time $O(N + E)$.

If the network is disconnected, $\varnothing$ is taken as the maximum of pseudo-diameters of the connected components.

Node percolation profile (random removal), $R_r$: We chose a random node order, and remove nodes sequentially from the graph according to it. If $S_i$ is the fraction of nodes in the largest component after the $i$-th removal, then the profile value is

$$R_r = \frac{1}{N} \sum_i S_i.$$

The value is averaged over several node orderings.

Fraction of nodes in the largest component, $S$: A component is a maximal set of nodes that are connected by a path. The largest component is the component with the largest number of nodes, and $S$ is the fraction of all nodes that belong to it.

In Fig. 8 we show how the z-scores and relative deviation values are related for every network descriptor, according to both models used. In Fig. 9 we show Kendall’s $\tau$ correlation coefficient among the descriptor values themselves, as well as their z-scores and relative deviations, according to the DCSBM. The insets show how the correlations among the deviations are themselves also correlated with the descriptor correlations.

Appendix C: Dataset descriptions

Below are descriptions of the network datasets used in this work. The codenames in the first row correspond to the respective entries in the Netzschleuder repository [29] where the networks can be downloaded. Some of the descriptions were obtained from the Colorado Index of Complex Networks [30].

For all networks, the versions considered in this work were transformed into simple graphs, i.e. symmetrized versions of directed networks and/or with parallel edges and self-loops removed.

| Name                        | Description                                                                 | $N$   | $E$   | Domain       |
|-----------------------------|------------------------------------------------------------------------------|-------|-------|--------------|
| blumenau_drug               | A network of drug-drug interactions, extracted from 18 months of electronic  | 75    | 181   | Biological   |
|                            | health records (EHRs) from the city of Blumenau in Southern Brazil [31].     |       |       |              |
| budapest_connectome [1]     | Brain graphs derived from connectomes of 477 people, computed from the      | 1015  | 53586 | Biological   |
|                            | Human Connectome Project [47].                                               |       |       |              |
| budapest_connectome [2]     | Networks among neurons of both the adult male and adult hermaphrodite        | 1015  | 62552 | Biological   |
|                            | worms C. elegans, constructed from electron microscopy series, to include    |       |       |              |
|                            | directed edges (chemical) and undirected (gap junction), and spanning        |       |       |              |
|                            | including nodes for muscle and non-muscle end organs [13].                  | 514   | 2832  | Biological   |
|                            | Ten networks of protein-protein interactions in Caenorhabditis elegans       | 575   | 4500  | Biological   |
|                            | (nematode), from yeast two-hybrid experiments, biological process maps,      | 575   | 4500  | Biological   |
|                            | literature curation, orthologous interactions, and genetic interactions [13].| 469   | 1433  | Biological   |
|                            | Network of protein-protein interactions in Saccharomyces cerevisiae (budding | 2724  | 13064 | Biological   |
|                            | yeast), measured by co-complex associations identified by high-throughput affin-| 912   | 22738 | Biological   |
|                            | ity purification and mass spectrometry (AP/MS) [35].                         | 537   | 517   | Biological   |
|                            | Network of operons and their pairwise interactions for E. coli [30].         | 423   | 519   | Biological   |
|                            | Networks of carbon exchanges among species in the cypress wetlands of South  | 128   | 2075  | Biological   |
|                            | Florida. One network covers the wet and the other the dry season [37].       | 183   | 2434  | Biological   |
|                            | A food web among the species found in Little Rock Lake in Wisconsin [35].   | 183   | 2434  | Biological   |
|                            | Trophic-level species interactions in streams in New Zealand, Maine and North| 94    | 424   | Biological   |
|                            | Carolina [30].                                                              | 107   | 965   | Biological   |
Multiplex networks representing different types of genetic interactions, for different organisms. Layers represent (i) physical, (ii) association, (iii) co-localization, (iv) direct, and (v) suppressive, (vi) additive or synthetic genetic interaction [10].

Networks of neural interactions extracted from human patients using the Magnetic Resonance One-Click Pipeline (MROCP), where nodes are voxels of neural tissue and edges represent connections by single fibers [31].

Networks of recombinant antigen genes from the human malaria parasite P. falciparum [10].

Networks of protein-protein binding interactions among yeast proteins [43].

A global interaction web of interactions between nematodes and their host mammal species, extracted from the helmintR package and dataset [55].

A network from Chinese online encyclopedias Baidu [66].

Bipartite network of plants and pollinators from Kyoto University Forest of Ashi, Japan, from 1984 to 1987 [59].

A network representing the Western States Power Grid of the United States [59].

A network of recombinant antigen genes from the human malaria parasite P. falciparum [10].

A network of feeding links among taxa based on the 48 million years old uppermost Early Eocene Messel Shale [47].

A global interaction web of interactions between nematodes and their host mammal species, extracted from the helmintR package and dataset [55].

A network of word adjacencies of common adjectives and nouns in the novel Bible [68].

A network of noun phrases (places and names) in the King James Version of the Bible [48].

Several networks of software dependencies. Nodes represent libraries and a directed edge denotes a library dependency on another [58, 61].

An integrated snapshot of the structure of the Internet at the level of Autonomous Systems (ASs), reconstructed from BGP tables posted by the University of Oregon Route Views Project [55].
| First Two Bipartite Networks (Wikipedia) |
|----------------------------------------|
| | Citations among papers contained in the DBLP computer science bibliogra-
| | phy [71].|
| | A bipartite network of artistic works (movies, novels, etc.) and their tropes |
| | (stylistic conventions or devices), as extracted from tvtropes [57].|
| | Two bipartite networks of the affiliations between musical labels and other mu-
| | sical genres or musical "styles," as given in the discogs.com database [57].|
| | Two bipartite user-page networks extracted from Wikipedia, about books [74]. |
| | Two bipartite user-page networks extracted from Wikipedia, about news events [72]. |
| | | A bipartite user-page network extracted from Wikiquotes [72].|
| | | Three bipartite networks that make up the MovieLens 100K Dataset, a stable |
| | Three bipartite user-page networks extracted from Wiktionary, for French, |
| | | German, and English [72].|
| | | A web graph representing a crawl of a portion of the general WWW, from a 2002 |
| | | Google Programming contest [57].|
| | | Three bipartite networks that make up the MovieLens 100K Dataset, a stable |
| | | benchmark dataset of 100,000 ratings from 1,000 users on 1,700 movies [74]. |
| | | A directed network of hyperlinks among a large set of U.S [74].|
| | | A network of books about U.S [75].|
| | | Network of legal citations by the US Supreme Court (SCOTUS) [76, 77].|
| | | A web graph network originally constructed in 2003 as a testbed for informa-
| | | tion-retrieval techniques, including web search engines [75].|
| | | A bipartite network of languages and the countries in which they are spoken, as |
| | | estimated by Unicode [77].|
| | | Citations among patents in the United States, as found in the National Bureau |
| | | of Economic Research (NBER) database, from 1975 to 1999 [79].|
| | | Web graphs crawled from four Computer Science departments in 1998, with |
| | | each page manually classified into one of 7 categories: course, department, |
| | | faculty, project, staff, student, or other [50].|
| | | A network of scientific fields, extracted from the English Wikipedia in early |
| | | 2020 [81].|
| | | Network of word adjacency in texts of several languages including English, |
| | | French, Spanish and Japanese [82].|
| | | A network of word associations showing the count of such associations as collected |
| | | from subjects, from the Edinburgh Associative Thesaurus (EAT) [83].|
| | | A network of English words from the WordNet, denoting relationships between |
| | | words (synonymy, hyperonymy, meronymy, etc.) [84].|
| | | A network of words extracted from phrases on which advertisers bid, in Yahoo!|
| | | Google’s 653260 [85].|
| | | A transportation network of Chicago, USA, from an unknown date (probably |
| | | 12982 20627 Transportation late 20th century) [53].|
| | | A network of contiguous states in the USA, in which each state is a node and |
| | | two nodes are connected if they share a land-based geographic border [50].|
| | | A multiplex network of airline routes among European airports, where each of |
| | | the 37 edge types represents routes by a different airline [57].|
| | | A network of international "E-roads," mostly in Europe [56].|
| | | A network of air traffic routes, from the FAA (Federal Aviation Administration) |
| | | National Flight Data Center (NFDC) preferred routes database [59].|
| | | Multiplex network with 3 edge types representing links within the three layers |
| | | of London train stations: Underground, Overground and DLR [40].|
| | | A network of regularly occurring flights among airports worldwide, extracted |
| | | from the openflights.org dataset [59].|
| | | The road network for the entire United States, as extracted from the |
| | | OpenStreetMap project [51].|
| | | OpenStreetMap project [51].|
| | | Two roads in the contiguous United States, as extracted from the OpenStreetMap |
| | | project [51].|
| | | Road networks from three US states (CA, PA, TX), in which edges are |
| | | stretches of road and vertices are intersections of roads [57].|
| | | Urban street networks, corresponding to a 1-square-mile maps of 20 cities around |
| | | the urban centers of the United States [92, 93].|
Yearly snapshots of flights among all commercial airports in the United States from 1990 to today [94].

The road networks of the 50 US States and the District of Columbia based on UA Census 2000 TIGER/Line Files [95].

A network of friendships among 29 seventh grade students in Victoria, Australia [96].

Scientific collaborations between authors of papers submitted to arxiv.org [97].

A network of who-trusts-whom relationships among users of the Bitcoin Alpha platform [101].

A bipartite network of the affiliations between notable people and occupations, as extracted from Wikipedia by the DBpedia project [107].

A network of friendships among Facebook friends, which was made public as a result of legal action [111].

A small anonymized Facebook ego network, from April 2014. Nodes are Facebook profiles, and an edge exists if the two profiles are "friends" on Facebook [113].

An early snapshot of the federation network among web publishers using the ActivityPub protocol [116].

Friendship relationships and interactions (wall posts) for a subset of the Facebook social network in 2009, recorded over a 2 year period [118].
A network of American football games between Division IA colleges during reg-
ular season Fall 2000 [118, 119]. Two bipartite networks of users and restaurant locations in New York City on
Foursquare, from 24 October 2011 to 20 February 2012 [119]. A network of friendships among male students in a small high school in Illinois from 1988 to 2001 [120]. A set of networks of HIV transmissions between people through sexual, needle-
sharing, or social connections, based on combining 8 datasets collected from 1988 to 2001 [121]. A network of friendships among users of Hyves, an online social networking site in the Netherlands (comparable to Facebook at the time) [122]. The network of collaborations among jazz musicians, and among jazz bands, extracted from The Red Hot Jazz Archive digital database, covering bands that performed between 1912 and 1940 [124]. Network of friendships among members of a university karate club [125]. Bipartite network of members of the Abu Sayyaf Group in the Philippines, and the kidnapping events they were involved in [126]. User-band networks from the music website last.fm [127]. The network of scene coappearances of characters in Victor Hugo’s novel "Les Misérables". Edge weights denote the number of such occurrences [57]. A network of ratings given between users at Libimseti.cz, a Czech online dating website [128]. Network structure for four large online social networks [117]. A coauthorship network among scientists working on network science, from 2006. This network is a one-mode projection from the bipartite graph of authors and their scientific publications [59]. A network of scientific collaborations among institutions in New Zealand [129]. A network of friendships among users on catster.com and dogster.com [129]. A network of trust relationships among physicians in four midwestern (USA) cities in 1966 [130]. Coauthorships among the Pierre Auger Collaboration of physicists [131]. A network of human proximities among students at Massachusetts Institute of Technology (MIT), as measured by personal mobile phones [132]. A network of friendships between students living in a residence hall at Australian National University [133]. Contacts and friendship relations between students in a high school in Marseilles, France, in December 2013 [134]. Network of contacts between students in a high school in Marseilles, France [135]. Network of contacts among attendees of the ACM Hypertext 2009 confer-
ce [136]. A temporal network of contacts between individuals, measured in an office build-
ing in France, from June 24 to July 3, 2013 [136]. Network of contacts among students and teachers as of a primary school in Lyon, France, on consecutive days of in October 2009 [137]. Network of cooperation among students in the “Computer and Network Security” course at Ben-Gurion University, in 2013 [138]. A bipartite sexual affiliation network representing “swing unit” couples (one node 96 per couple) and the parties they attended [139]. A network of following relationships from Twitter, from a snowball sample crawl across “quality” users in 2009 [140]. A network representing follower-following relations among Twitter users asso-
ciated with the 15-M Movement or Anti-austerity movement in Spain, in the period April-May 2011 [141]. Tweet reply network related to the discovery of the Higgs boson [142]. Complete friendship and health advice social networks among households in 17 rural villages bordering Lake Victoria in Mayaguez District, Uganda in 2013 [143]. Web-based links between U.S. government agencies websites [144]. Networks of bill co-sponsorship tendencies among US Congresspeople, from 1973 (93rd Congress) to 2016 (114th Congress) [145, 146]. Interactions among users of 10 language-specific Wikipedias [57]. Nodes in the network represent (English) Wikipedia users and a directed edge from node i to node j represents that user i at least once edited a talk page of user j [147]. A bipartite network representing child-parent connections among family mem-
bers, collected in 2012 from WikiTree, an online genealogical website with 13+ million profiles [148]. A network of interpersonal contacts among windsurfers in southern California during the Fall of 1986. The edge weights indicate the perception of social af-
filiations majorized by the tasks in which each individual was asked to sort cards with other surfer’s name in the order of closeness [149]. Network of items for sale on amazon.com and the items they “recommend” [150]. A bipartite network of users and products on Amazon.com [151]. Bipartite network representing people and the books they have interacted with, from the BookCrossing website [152]. Bipartite network of directors and the companies on whose boards they sit, span-
ning 54 countries worldwide, constructed from data collected by the Financial Times [153].
A bipartite network of movies and the actors that played in them, as extracted from Wikipedia by the DBpedia project [107].

Bipartite network of the affiliations (employment relations) between professional athletes and their teams, as extracted from Wikipedia by the DBpedia project [107].

A large bipartite network of the affiliations (contractual relations) among musical artists and record labels, as given in the discogs.com database [57].

A bipartite network of users and the products they rated on the website Epinions.com [155].

A bipartite network of public EU procurement contracts, from 2008 to 2016, 839824 1841009 Economic between issuing buyers (public institutions such as a ministry or city hall) and supplying winners (a private firm) [155].

Networks representing the annual national public procurement markets of 26 European countries from 2008-2016, inclusive [155].

Multiplex network representing trade relationships between countries from the Food and Agricultural Organization of the United Nations [156].

The bipartite project-user membership network of the software development hosting site GitHub [157].

Two bipartite networks of users and jokes, extracted from the online joke recommender system Jester [155].

A bipartite network of users and the posts they have favorited, from the online Q&A site Stack Overflow [57].

A bipartite network between users and stories on digg.com from 2009 [159].