# Scalable detection of statistically significant communities and hierarchies, using message passing for modularity

**Pan Zhang and Cristopher Moore**[1]

Santa Fe Institute, Santa Fe, NM 87501

Edited by Giorgio Parisi, University of Rome, Rome, Italy, and approved October 31, 2014 (received for review May 28, 2014)

**Modularity is a popular measure of community structure. However, maximizing the modularity can lead to many competing partitions, with almost the same modularity, that are poorly correlated with each other. It can also produce illusory "communities" in random graphs where none exist. We address this problem by using the modularity as a Hamiltonian at finite temperature and using an efficient belief propagation algorithm to obtain the consensus of many partitions with high modularity, rather than looking for a single partition that maximizes it. We show analytically and numerically that the proposed algorithm works all of the way down to the detectability transition in networks generated by the stochastic block model. It also performs well on real-world networks, revealing large communities in some networks where previous work has claimed no communities exist. Finally we show that by applying our algorithm recursively, subdividing communities until no statistically significant subcommunities can be found, we can detect hierarchical structure in real-world networks more efficiently than previous methods.**

networks | community detection | message-passing algorithms | statistical significance | phase transitions

Community detection, or node clustering, is a key problem in network science, computer science, sociology, and biology. It aims to partition the nodes in a network into groups such that there are many edges connecting nodes within the same group and comparatively few edges connecting nodes in different groups.

Many methods have been proposed for this problem. These include spectral clustering, where we classify nodes according to the eigenvectors of a linear operator such as the adjacency matrix, the random walk matrix, the graph Laplacian, or other linear operators (1–3); statistical inference, where we fit the network with a generative model such as the stochastic block model (4–7); and a wide variety of other methods, e.g., refs. 8–10. See ref. 11 for a review.

We focus here on a popular measure of the quality of a partition, the modularity (e.g., refs. 8 and 12–14). We think of a partition $\{t\}$ into $q$ groups as a function $t: V \to \{1, \ldots, q\}$, where $t_i$ is the group to which node $i$ belongs. The modularity of a partition $\{t\}$ of a network with $n$ nodes and $m$ edges is defined as

$$Q(\{t\}) = \frac{1}{m}\left(\sum_{\langle ij\rangle \in \mathcal{E}} \delta_{t_i t_j} - \sum_{\langle ij\rangle} \frac{d_i d_j}{2m} \delta_{t_i t_j}\right). \quad [1]$$

Here $\mathcal{E}$ is the set of edges, $d_i$ is the degree of node $i$, and $\delta$ is the Kronecker delta function. The modularity is proportional to the number of edges connecting nodes in the same community minus the expected number of such edges if the graph were random conditioned on its degree distribution, that is, the expectation in a null model where $i$ and $j$ are connected with probability proportional to $d_i d_j$.

However, maximizing over all possible partitions often gives a large modularity even in random graphs with no community structure (15–18). Thus, maximizing the modularity can lead to overfitting, where the "optimal" partition simply reflects random noise. Even in real-world networks, the modularity often exhibits a large amount of degeneracy, with multiple local optima that are poorly correlated with each other and are not robust to small perturbations (19).

Thus, we need to add some notion of statistical significance to our algorithms. One approach is hypothesis testing, comparing various measures of community structure to the distribution we would see in a null model such as Erdős–Rényi (ER) graphs (20–22). However, even when communities really exist, the modularity of the true partition is often no higher than that of random graphs. In Fig. 1, we show partitions of two networks with the same size and degree distribution: an ER graph (*Left*) and a graph generated by the stochastic block model (*Right*), in the detectable regime where it is easy to find a partition correlated with the true one (5, 6). The true partition of the network in Fig. 1, *Right* has a smaller modularity than the partition found for the random graph in Fig. 1, *Left*. We can find a partition with higher modularity (and lower accuracy) in Fig. 1, *Right*, using, e.g., simulated annealing, but then the modularities we obtain for the two networks are similar. Thus, the usual approach of null distributions and *P* values for hypothesis testing does not appear to work.

We propose to solve this problem with the tools of statistical physics. As in ref. 16, we treat the modularity as the Hamiltonian of a spin system. We define the energy of a partition $\{t\}$ as $E(\{t\}) = -mQ(\{t\})$ and introduce a Gibbs distribution as a function of inverse temperature $\beta$, $P(\{t\}) \propto \mathrm{e}^{-\beta E(\{t\})}$. Rather than maximizing the modularity by searching for the ground state of

## Significance

**Most work on community detection does not address the issue of statistical significance, and many algorithms are prone to overfitting. We address this using tools from statistical physics. Rather than trying to find the partition of a network that maximizes the modularity, our approach seeks the consensus of many high-modularity partitions. We do this with a scalable message-passing algorithm, derived by treating the modularity as a Hamiltonian and applying the cavity method. We show analytically that our algorithm succeeds all the way down to the detectability transition in the stochastic block model; it also performs well on real-world networks. It also provides a principled method for determining the number of groups or hierarchies of communities and subcommunities.**

Author contributions: P.Z. and C.M. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: moore@santafe.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1409770111/-/DCSupplemental.

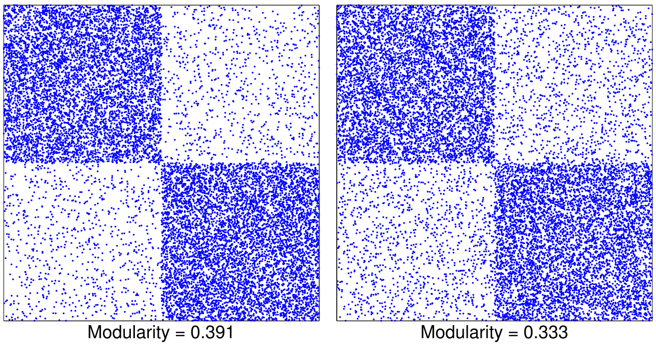


**Fig. 1.** The adjacency matrices of two networks, partitioned to show possible community structure. Each blue point is an edge. (*Left*) The network is an ER graph, with no real community structure; however, a search by simulated annealing finds a partition with modularity 0.391. (*Right*) The network has true communities and is generated by the stochastic block model, but the true partition has modularity of just 0.333. Thus, illusory communities in random graphs can have higher modularity than true communities in structured graphs. Both networks have size $n = 5{,}000$ and a Poisson degree distribution with mean $c = 3$; the network at *Right* has $c_{\mathrm{out}}/c_{\mathrm{in}} = 0.2$, in the easily detectable regime of the stochastic block model.

this system, we focus on its Gibbs distribution at a finite temperature, looking for many high-modularity partitions rather than a single one. In analogy with previous work on the stochastic block model (5, 6), we define a partition $\{\hat{t}\}$ by computing the marginals of the Gibbs distribution and assigning each node to its most likely community. Specifically, if $\psi_t^i$ is the marginal probability that $i$ belongs to group $t$, then $\hat{t}_i = \mathrm{argmax}_t\, \psi_t^i$, breaking ties randomly if more than one $t$ achieves the maximum. We call $\{\hat{t}\}$ the retrieval partition and call its modularity $Q(\{\hat{t}\})$ the retrieval modularity. We claim that $\{\hat{t}\}$ is a far better measure of significant community structure than the maximum-modularity partition. In the language of statistics, the maximum marginal prediction is better than the maximum a posteriori prediction (e.g., ref. 23). More informally, the consensus of many good solutions is better than the "best" single one (24, 25).

We give an efficient belief propagation (BP) algorithm to approximate these marginals, which is derived from the cavity method of statistical physics. This algorithm is highly scalable; each iteration takes linear time on sparse networks if the number of groups is fixed, and it converges rapidly in most cases. It is optimal in the sense that for synthetic graphs generated by the stochastic block model, it works all of the way down to the detectability transition. It provides a principled way to choose the number of communities, unlike other algorithms that tend to overfit. Finally, by applying this algorithm recursively, subdividing communities until no statistically significant subcommunities exist, we can uncover hierarchical structure.

We validate our approach with experiments on real and synthetic networks. In particular, we find significant large communities in some large networks where previous work claimed there were none. We also compare our algorithm with several others, finding that it obtains more accurate results, both in terms of determining the number of communities and in terms of matching their ground-truth structure.

## Results

**Results on the Stochastic Block Model.** Also called the planted partition model, the stochastic block model (SBM) is a popular ensemble of networks with community structure. There are $q$ groups of nodes, and each node $i$ has a group label $t_i^* \in \{1, \ldots, q\}$; thus $\{t^*\}$ is the true, or planted, partition. Edges are generated independently according to a $q \times q$ matrix $p$, by connecting each pair of nodes $\langle ij \rangle$ with probability $p_{t_i^*, t_j^*}$. Here for simplicity we discuss the commonly studied case where the $q$ groups have equal size and where $p$ has only two distinct entries, $p_{rs} = c_{\mathrm{in}}/n$ if $r = s$ and $c_{\mathrm{out}}/n$ if $r \neq s$. We use $\epsilon = c_{\mathrm{out}}/c_{\mathrm{in}}$ to denote the ratio between these two entries. In the assortative case, $c_{\mathrm{in}} > c_{\mathrm{out}}$ and $\epsilon < 1$. When $\varepsilon$ is small, the community structure is strong; when $\epsilon = 1$, the network becomes an ER graph.

For a given average degree $c = (c_{\mathrm{in}} + (q-1)c_{\mathrm{out}})/q$, there is a so-called detectability phase transition (5, 6), at a critical value

$$\epsilon^* = \frac{\sqrt{c} - 1}{\sqrt{c} - 1 + q}. \qquad [2]$$

For $\epsilon < \epsilon^*$, BP can label the nodes with high accuracy; for $\epsilon > \epsilon^*$, neither BP nor any other algorithm can label the nodes better than chance, and indeed no algorithm can distinguish the network from an ER graph with high probability. This transition was recently established rigorously in the case $q = 2$ (26–28).

For larger numbers of groups, the situation is more complicated. For $q \leq 4$, in the assortative case, this detectability transition coincides with the Kesten–Stigum bound (29, 30). For $q \geq 5$ the Kesten–Stigum bound marks a conjectured transition to a "hard but detectable" phase where community detection is still possible but takes exponential time, whereas the detectability transition is at a larger value of $\varepsilon$; that is, the thresholds for reconstruction and robust reconstruction become different. Our claim is that our algorithm succeeds down to the Kesten–Stigum bound, i.e., throughout the detectable regime for $q \leq 4$ and the easily detectable regime for $q \geq 5$.

In Fig. 2 we compare the behavior of our BP algorithm on ER graphs and a network generated by the SBM in the detectable regime. Both graphs have the same size and average degree $c = 3$. For the ER graph (Fig. 2, *Left*) there are just two phases, separated by a transition at $\beta^* = 1.317$: the paramagnetic phase, where BP converges to a factorized fixed point where every node is equally likely to be in every group, and the spin-glass phase, where replica symmetry is broken and BP fails to converge. The convergence time diverges at the transition. Note that in the spin-glass phase, the retrieval modularity returned by BP fluctuates wildly as BP jumps from one local optimum to another and has little meaning. In any case BP assumes replica symmetry, which is incorrect in this phase.

In contrast, the SBM network in Fig. 2, *Right* has strong community structure. In addition to the paramagnetic and spin-glass phases, there is now a retrieval phase in a range of $\beta$, where BP finds a retrieval state describing statistically significant community structure. The retrieval modularity jumps sharply at $\beta_{\mathrm{R}} = 1.072$,

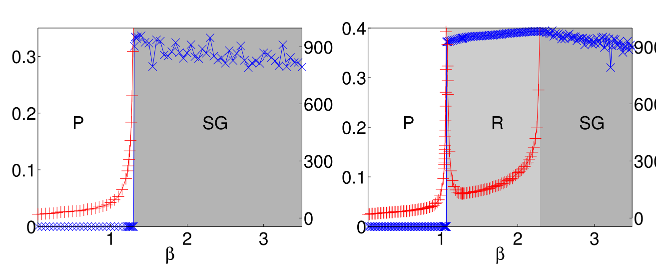


**Fig. 2.** Retrieval modularity (blue ×, left $y$ axis) and BP convergence time (red +, right $y$ axis) of an ER random graph (*Left*) and a network generated by the stochastic block model in the detectable regime (*Right*). Both networks have $n = 1{,}000$ and average degree $c = 3$, and the network on the right has $\epsilon = 0.2$. In both cases we ran BP with $q = 2$ groups. In the ER graph, which has no community structure, there are two phases, paramagnetic (P) and spin glass (SG), with a transition at $\beta^* = 1.317$. In the SBM network, there is an additional retrieval phase (R) between $\beta_{\mathrm{R}} = 1.072$ and $\beta_{\mathrm{SG}} = 2.27$, where BP finds a retrieval state with high modularity, indicating statistically significant community structure.

COMPUTER SCIENCES

SOCIAL SCIENCES

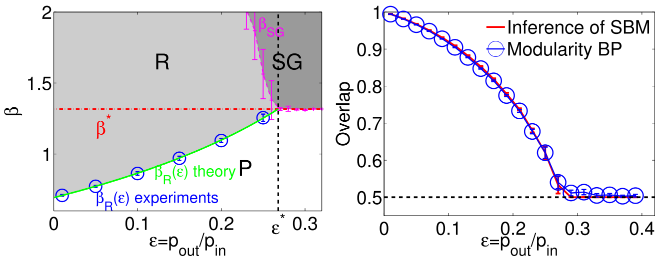


**Fig. 3.** (*Left*) Phase diagram for networks generated by the stochastic block model, showing the paramagnetic (P), retrieval (R), and spin-glass (SG) phases. Blue circles with error bars denote experimental estimates of $\beta_R$, the boundary between the paramagnetic and retrieval phases, and the solid green line shows our theoretical expression [**4**]. The spin-glass instability occurs for $\beta > \beta^*$ (2, 3) (red dashed line) and $\epsilon^*$ is the detectability transition (black dashed line). (*Right*) The overlap of the retrieval partition at $\beta = 1.315 \approx \beta^*$ (2, 3) (blue circles) and the partition obtained with the algorithm of ref. 5, which infers the parameters of the SBM with an additional EM learning algorithm. Each experiment is on the giant component of a network with $n = 10^5$, $q = 2$ groups, and average degree $c = 3$. We average over 10 random instances.

when we first enter this phase, and then increases gently to 0.393 as $\beta$ increases; for comparison, the modularity of the planted partition is $M_{\text{hidden}}(\epsilon) = 1/(1+\epsilon) - 1/2 = 0.33$. When we enter the spin-glass phase at $\beta_{SG} = 2.27$, the retrieval modularity fluctuates as in the ER graph. The convergence time diverges at both phase transitions.

We can compute two of these transition points analytically by analyzing the linear stability of the factorized fixed point (*Methods*). Stability against random perturbations gives

$$\beta^*(q, c) = \log\left(\frac{q}{\sqrt{c} - 1} + 1\right), \qquad [3]$$

and stability against correlated perturbations gives

$$\beta_R(q, c, \epsilon) = \log\left(\frac{q(1 + (q-1)\epsilon)}{c(1-\epsilon) - (1 + (q-1)\epsilon)} + 1\right). \qquad [4]$$

These cross at the Kesten–Stigum bound, where $\epsilon = \epsilon^*$. We do not currently have an analytic expression for $\beta_{SG}$.

In Fig. 3, *Left* we show the phase diagram of our algorithm on SBM networks, including the paramagnetic, retrieval, and spin-glass phases as a function of ε, with $q = 2$ and $c = 3$. The boundary $\beta_R$ between the paramagnetic and retrieval phases is in excellent agreement with our expression [**4**]. For $\epsilon < \epsilon^* \approx 0.267$, our algorithm finds a retrieval state for $\beta_R < \beta < \beta_{SG}$. In Fig. 3, *Right* we show the accuracy of the retrieval partition $\{\hat{t}\}$, defined as its overlap with the planted partition, i.e., the fraction of nodes labeled correctly.

We emphasize that $\beta^*$ is not the optimal value of $\beta$; i.e., it is not on the Nishimori line (23, 31, 32). However, the optimal $\beta$ depends on the parameters of the SBM (*SI Text*). Our claim is that setting $\beta = \beta^*$ in our algorithm succeeds throughout the easily detectable regime, even when the parameters are unknown. In Fig. 3, *Right* we compare our algorithm with that of refs. 5 and 6, which learns the SBM parameters using an expectation-maximization (EM) algorithm. Our algorithm provides nearly the same overlap, without the need for the EM loop.

**Results on Real-World Networks and Choosing the Number of Groups.** We tested our algorithm on a number of real-world networks. As for networks generated by the SBM in the detectable regime, we find a retrieval phase between the paramagnetic and spin-glass phases (*SI Text*). Rather than attempting to learn the optimal parameters or temperature for these networks, we simply set $\beta = \beta^*(q^*, c)$ as defined in [**3**], where $q^*$ is the ground-truth number of groups (if known) and $c$ is the average degree. Again, this value of $\beta$ is not optimal, and varying $\beta$ may improve the algorithm's performance; however, setting $\beta = \beta^*$ appears to work well in practice.

When the number of groups is not known, determining it is a classic model-selection problem. The maximum modularity typically grows with $q$. In contrast, the retrieval modularity stops growing when $q$ exceeds the correct value, giving us a principled method of choosing $q^*$ (*SI Text*). For those networks where $q^*$ is known, we found that this procedure agrees perfectly with the ground truth.

As shown in Table 1, our algorithm finds a retrieval state in all these networks, with high retrieval modularity and high overlap with the ground truth. For the Gnutella, Epinions, and web-Google networks, no ground truth is known; but in contrast with ref. 37, our algorithm finds significant large-scale communities.

Whereas most of these networks are assortative, one network in the table, the adjacency network of common adjectives and nouns in the novel *David Copperfield* (40) (see ref. 2), is disassortative, because nouns are more likely to be adjacent to adjectives than other nouns and vice versa. In this case, we found a retrieval state with negative modularity and high overlap with the ground truth, by setting $\beta$ to $-\beta^*(q^*, c)$.

**Results on Hierarchical Clustering.** Many networks appear to have hierarchical structure with communities and subcommunities on many scales (2, 8, 24, 38, 39). We can look for such structures by working recursively: We determine the optimal number $q^*$ of groups, divide the network into subgraphs, and apply the algorithm to each one. We stop dividing when there is no retrieval

**Table 1. Retrieval modularity, overlap between the retrieval partition and the ground truth, the number of groups $q^*$ as determined by our algorithm, the inverse temperature $\beta^*$ defined in [3], and the convergence time measured in seconds and iterations for several real-world networks (2, 33–37)**

| Network | $n$ | $m$ | $q^*$ | $\beta^*$ | $Q(\hat{t})$ | Overlap | Time, s | No. iterations |
|---|---|---|---|---|---|---|---|---|
| Zachary's karate club | 34 | 78 | 2 | 1.012 | 0.371 | 1 | 0.001 | 26 |
| Dolphin social network | 62 | 159 | 2 | 0.948 | 0.395 | 0.887 | 0.001 | 33 |
| Books about US politics | 105 | 441 | 3 | 0.948 | 0.521 | 0.829 | 0.002 | 23 |
| Word adjacencies | 112 | 425 | 2 | −0.761 | −0.275 | 0.848 | 0.003 | 35 |
| Political blogs | 1,222 | 16,714 | 2 | 0.387 | 0.426 | 0.948 | 0.043 | 18 |
| Gnutella | 62,586 | 147,892 | 7 | 0.995 | 0.517 | | 37.43 | 433 |
| Epinions | 75,888 | 405,740 | 4 | 0.632 | 0.429 | | 57.13 | 213 |
| Web-Google | 916,428 | 4,322,051 | 5 | 0.676 | 0.724 | | 2,331 | 505 |

For Gnutella, Epinions, and web-Google (37) no ground truth is known, but based on our results we claim, contrary to ref. 37, that these networks have statistically significant large-scale communities.

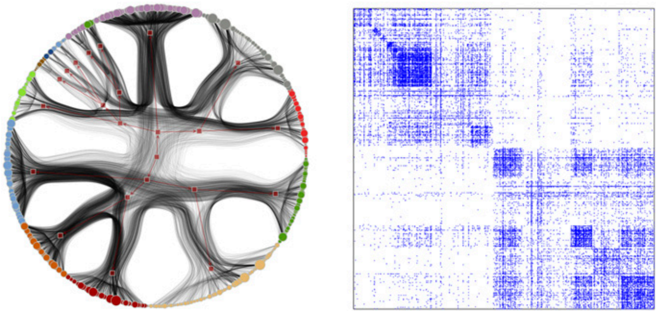

**Fig. 4.** (*Left*) A hierarchical division of the political blog network (34). We apply our technique recursively, looking for a retrieval state and optimizing the number of groups in which to split the community at each stage. We stop when no retrieval state is detected, indicating that the remaining groups have no statistically significant subcommunities. Each leaf denotes one node, the size indicates its degree, and the colors indicate different groups in the final division. (*Right*) The adjacency matrix of the network ordered according to this partition.

state, indicating that the remaining subgraphs have no significant internal structure.

For networks generated by the SBM, each subgraph is an ER graph. Our algorithm finds no retrieval state in the subgraphs, so it stops after one level of division. The same occurs in some small real-world networks, e.g., Zachary's karate club. In some larger real-world networks, on the other hand, our algorithm repeatedly finds a retrieval state in the subgraphs, suggesting a deep hierarchical structure.

An example is the network of political blogs (34). Our algorithm first finds two large communities corresponding to liberals and conservatives and agreeing with the ground-truth labels on 95% of the nodes. However, as shown in Fig. 4, it splits these into subcommunities, eventually finding a hierarchy five levels deep with a total of 14 subgroups. We show the adjacency matrix with nodes ordered by this final partition in Fig. 4, *Right* and the hierarchical structure is clearly visible. The modularity of the second through fifth levels is 0.426, 0.331, 0.285, and 0.282, respectively. This decreasing modularity may explain why the algorithm did not immediately split the network all of the way down to the subcommunities.

A nested SBM was used to explore hierarchical structure in ref. 39, where the blog network was also reported to have hierarchical structure. Our results are slightly different, giving 14 rather than 17 subgroups, but the first three levels of subdivision are similar.

**Comparison with Other Algorithms.** In this section we compare the performance of our algorithm with two popular algorithms: Louvain (9) and OSLOM (21). In particular, OSLOM tries to focus on statistically significant communities.

Louvain gives partitions with similar modularity to that of our algorithm, but with a much larger number of groups, particularly on large networks. For example, on the Gnutella and Epinions network (37), our algorithm finds $q^* = 7$ and $q^* = 4$ groups with modularity 0.517 and 0.429, respectively, whereas the Louvain method finds 66 and 949 groups with modularity 0.499 and 0.430, respectively. Thus, our algorithm finds large-scale communities, with a modularity similar to that of the smaller communities found by Louvain. Of course, we emphasize that maximizing the modularity is not our goal: Finding statistically significant communities is.

We show results on synthetic networks in Fig. 5. In Fig. 5, *Left* we apply Louvain, OSLOM, and our algorithm to SBM networks with $q = 6$. We compute the normalized mutual information (NMI) (41) between the inferred partition and the planted one. (We use the NMI rather than the overlap because the numbers of groups given by OSLOM and Louvain are very different from the planted partition.) For Louvain and OSLOM, the NMI drops off well below the detectability transition. In Fig. 5, *Right* we show the number of groups that each algorithm infers for an ER graph with $c = 4$. Our algorithm correctly chooses $q = 1$, recognizing that this network has no internal structure. The other algorithms overfit, inferring a number of communities that grow with $n$. In *SI Text* we report on experiments on benchmark networks with heavy-tailed degree distributions (42), with similar results.

## Discussion

We have presented a physics-based method for finding statistically significant communities. Rather than using an explicit generative or graphical model, it uses a popular measure of community structure, namely the modularity. It does not attempt to maximize the modularity, which is both computationally difficult and prone to overfitting. Instead it estimates the marginals of the Gibbs distribution, using a scalable BP algorithm derived from the cavity method (next section), and defines the retrieval partition by assigning each node to its most likely community according to these marginals.

In essence, the algorithm looks for the consensus of many partitions with high modularity. When this consensus exists, it indicates statistically significant community structure, as opposed to random fluctuations. Moreover, by testing for the existence of this retrieval state, as opposed to a spin-glass state where the algorithm fluctuates between many unrelated local optima, we can determine the correct number of groups and decompose a network hierarchically.

We note that this algorithm is related to BP for the degree-corrected stochastic block model (DCSBM). Specifically, for a fixed $\beta$, the modularity is linearly related to the log-likelihood of the DCSBM with particular parameters (*SI Text*). However, our algorithm does not have to learn the parameters of the block model with an EM algorithm or perform model selection between the stochastic block model and its degree-corrected variant (43). To be clear, $\beta$ is still a tunable parameter that can be optimized, but the heuristic value $\beta = \beta^*$ appears to work well for a wide range of networks.

In addition to the detectability transition in the SBM, another well-known barrier to community detection is the resolution limit (44) where communities become difficult to find when their size is $O(\sqrt{n})$ or less. In *SI Text*, we give some evidence that our hierarchical clustering algorithm overcomes this barrier. Namely, for the classic example of a ring of cliques, at the second level our algorithm divides the graph precisely into these cliques.

Another recent proposal for determining the number of groups is to use the number of real eigenvalues of the nonbacktracking matrix, outside the bulk of the spectrum (3). For some networks, such as the political blogs, this gives a larger number than the $q^*$

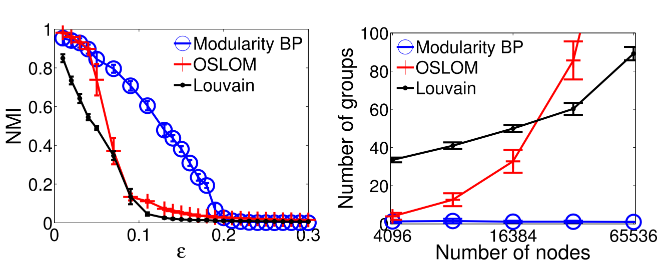


**Fig. 5.** Comparison of BP with Louvain and OSLOM on SBM networks with $n = 10^4$, $c = 6$, and $q = 6$. (*Left*) We show the normalized mutual information (NMI) between each algorithm's results and the true partition as a function of $\epsilon$; the other algorithms' NMI drops sharply well below the detectability transition at $\epsilon = 0.195$. (*Right*) We show the inferred number of groups on the giant component of an ER graph with $c = 4$. Whereas our algorithm correctly finds $q^* = 1$, the other algorithms overfit, finding a growing number of small communities as $n$ increases. Each point is averaged over 20 instances.

we found here; it may be that, in some sense, this method detects not just top-level communities, but also subcommunities deeper in the hierarchy. It would be interesting to perform a detailed comparison of the two methods.

Our approach can be extended to generalizations of the modularity, where the graph is weighted or where a parameter $\gamma$ represents the relative importance of the expected number of internal edges (16). Finally, it would be interesting to apply BP to other objective functions, such as normalized cut or conductance, devising Hamiltonians from them and considering the resulting Gibbs distributions.

Finally, we note that rather than running BP once and using the resulting marginals, we could use decimation (45) to fix the labels of the most biased nodes, run BP again to update the marginals, and so on. This would increase the running time of the algorithm, but it may improve its performance. Another approach would be reinforcement (45), where we add external fields that point toward the likely configuration. We leave this for future work.

## Methods

**Defining Statistical Significance.** As described above, an ER random graph has many partitions with high modularity. However, these partitions are nearly uncorrelated with each other. In the language of disordered materials, the landscape of partitions is glassy: Although the optimal one might be unique, there are many others whose modularity is almost as high, but have a large Hamming distance from the optimum and from each other. If we define a Gibbs distribution on the partitions, we encounter either a paramagnetic state where the marginals are uniform or a spin glass with replica symmetry breaking where we jump between local optima. In either case, focusing on any one of these optima is simply overfitting.

For networks such as Fig. 1, *Right* in contrast, there are many high-modularity partitions that are correlated with each other and with the ground truth. As a result, the landscape has a smooth valley surrounding the ground truth. At a suitable temperature, the Gibbs distribution is in a retrieval phase with both low energy (high modularity) and high entropy, giving it a lower free energy than that of the paramagnetic state, with its marginals biased toward the ground truth. When BP converges to a fixed point, it finds a (local) minimum of the Bethe free energy, approximating this lower free energy phase.

We propose the existence of this retrieval phase as a physics-based definition of statistical significance. When it exists, the retrieval partition defined by the maximum marginals is an optimal prediction of which nodes belong to which groups.

The idea of using the free energy to separate real community structure from random noise, and using the Gibbs marginals to define a partition, also appeared in refs. 5 and 6. However, that work is based on a specific generative model, namely the stochastic block model, and the energy is (minus) the log-likelihood of the observed network. In contrast, we avoid explicit generative models and focus directly on the modularity as a measure of community structure.

**The Cavity Method and Belief Propagation.** Our goal is to compute the marginal probability distribution that each node belongs to a given group and the free energy of the Gibbs distribution. We could do this using a Monte Carlo Markov chain algorithm. However, to obtain marginals we would need many independent samples, and to obtain the free energy we would need to sample at many different temperatures. Thus, the Monte Carlo Markov chain (MCMC) is prohibitively slow for our purposes.

Instead, for sparse networks, we can use belief propagation (46), known in statistical physics as the cavity method (47). BP makes a conditional independence assumption, which is exact only on trees; however, in the regimes we consider (the detectable regime of the stochastic block model and typical real-world graphs), its estimates of the marginals are quite accurate. It also provides an estimate of the free energy, called the Bethe free energy, which is a function of one- and two-point marginals.

BP works with "messages" $\psi_t^{i\to k}$: These are estimates, sent from node $i$ to node $k$, of the marginal probability that $t_i = t$ based on $i$'s interactions with nodes $j \neq k$. The update equations for these messages are as follows:

$$\psi_t^{i\to k} \propto \exp\left[-\frac{\beta d_i}{2m}\theta_t + \sum_{j\in\partial i\setminus k}\log\left(1+\psi_t^{j\to i}\left(e^{\beta}-1\right)\right)\right]. \quad [5]$$

Here $\partial i$ denotes the set of $i$'s neighbors, and $\theta_t = \sum_{j=1}^{n} d_j \psi_t^j$ denotes an external field acting on nodes in group $t$, which we update after each BP iteration. We refer to *SI Text* for detailed derivations of the BP update equations and Bethe free energy.

For $q$ groups and $m$ edges, each iteration of [**5**] takes time $O(qm)$. If $q$ is fixed, this is linear in the number of edges and linear in the number of nodes when the network is sparse (i.e., when the average degree is constant). Moreover, these updates can be easily parallelized. Empirically, the number of iterations required to converge appears to depend very weakly on the network size, although in some cases it must grow at least logarithmically.

**The Factorized Solution and Local Stability.** Observe that the factorized solution, $\psi_t^{j\to i} = 1/q$, where each node is equally likely to be in each possible group, is always a fixed point of [**5**]. If BP converges to this solution, we cannot label the nodes better than chance, and the retrieval modularity is zero. This is the paramagnetic state.

There are two other possibilities: BP fails to converge, or it converges to a nonfactorized fixed point, which we call the retrieval state. In the latter case, we can compute the marginals by

$$\psi_t^{i} \propto \exp\left[-\frac{\beta d_i}{2m}\theta_t + \sum_{j\in\partial i}\log\left(1+\psi_t^{j\to i}\left(e^{\beta}-1\right)\right)\right], \quad [6]$$

and define the retrieval partition $\hat{t}$ that assigns each node to its most likely community. This partition represents the consensus of the Gibbs distribution: It indicates that there are many high-modularity partitions that are correlated with each other. The retrieval modularity $Q(\{\hat{t}\})$ is then a good measure of the extent to which the network has statistically significant community structure.

On the other hand, if BP does not converge, this means that neither the factorized solution nor any other fixed point is locally stable; the spin-glass susceptibility diverges, and replica symmetry is broken. In other words, the space of partitions breaks into an exponential number of clusters, and BP jumps from one to another. The retrieval partition obtained using the current marginals will change to a very different partition if we run BP a bit longer or if we perturb the initial BP messages slightly. In the spin-glass phase, we are free to define a retrieval modularity from the current marginals, but it fluctuates rapidly and does not represent a consensus of many partitions.

The linear stability of the factorized solution can be characterized by computing the derivatives of messages with respect to each other at the factorized fixed point. Using [**5**], we find that $\partial\psi_t^{i\to k}/\partial\psi_s^{j\to i} = T_{st}$, where $T_{st}$ is the $q\times q$ matrix

$$T_{st} = \left.\frac{\partial\psi_t^{i\to k}}{\partial\psi_s^{j\to i}}\right|_{\frac{1}{q}} = \frac{e^{\beta}-1}{e^{\beta}-1+q}\left(\delta_{st}-\frac{1}{q}\right). \quad [7]$$

Its largest eigenvalue (in magnitude) is

$$\lambda = \frac{e^{\beta}-1}{e^{\beta}-1+q}. \quad [8]$$

On locally tree-like graphs with Poisson degree distributions and average degree $c$, the factorized fixed point is then unstable with respect to random noise whenever $c\lambda^2 > 1$. This is also known as the de Almeida–Thouless local stability condition (48), the Kesten–Stigum bound (29, 30), or the threshold for census or robust reconstruction (49, 50). In our case, it shows that $\beta$ must exceed a critical $\beta^*$ given by [**3**]. If the network has some other degree distribution but is otherwise random, [**3**] holds where $c$ is the average excess degree, i.e., the expected number of additional neighbors of the endpoint of a random edge.

If there is no statistically significant community structure, then BP has just two phases, the paramagnetic one and the spin glass: For $\beta < \beta^*$ it converges to the factorized fixed point, and for $\beta > \beta^*$ it does not converge at all. On the other hand, if there are statistically significant communities, then BP converges to a retrieval state in the range $\beta_R < \beta < \beta_{SG}$. Typically $\beta_R < \beta^*$ and $\beta^*$ is in the retrieval phase, because even if the factorized fixed point is locally stable, BP can still converge to a retrieval state if its free energy is lower than that of the paramagnetic solution. Thus, we can test for statistically significant communities by running BP at $\beta = \beta^*$. Note that our calculation of $\beta^*$ in [**3**] assumes that the network is random conditioned on its degree distribution; in principle $\beta^*$ could fall outside the retrieval phase for real-world networks. In that case, our heuristic method of setting $\beta = \beta^*$ fails, and it would be necessary to scan values of $\beta$ in the vicinity of $\beta^*$ for the retrieval state.

To estimate $\beta_R$, we again consider the linear stability of BP around the factorized fixed point; but now we consider arbitrary perturbations, as opposed to random noise. Let $T$ be the $q\times q$ matrix defined in [**7**]. The matrix of derivatives of all $2qm$ messages with respect to each other is a tensor product $T\otimes B$, where $B$ is the nonbacktracking matrix (3). The adaptive external field in the BP equations suppresses eigenvectors where every node is in the same community. As a result, the relevant eigenvalue is $\lambda\mu$, where $\lambda$ is the largest eigenvalue of $T$, and $\mu$ is the second-largest eigenvalue of $B$, and the factorized fixed point is unstable whenever $\lambda\mu > 1$. For networks generated by the SBM, we have (3)

$$\mu = \frac{c(1-\epsilon)}{1+(q-1)\epsilon}. \quad [9]$$

Combining this with [**8**] and setting $\lambda\mu = 1$ gives Eq. **4**.

However, this assumes that the corresponding eigenvector of $B$ is correlated with the community structure, so that perturbing BP away from the factorized fixed point will lead to the retrieval state. This is true as long as $\mu$ is outside the bulk of $B$'s eigenvalues, which are confined to a disk of radius $\sqrt{c}$ in the complex plane (3); if it is inside the bulk, then the community structure is washed out by isotropic eigenvectors and becomes hard to find. Thus, the communities are detectable as long as $\mu > \sqrt{c}$. This is equivalent to $\beta_R < \beta^*$ or equivalently $\epsilon < \epsilon^*$. Thus, the retrieval state exists all of the way down to the Kesten–Stigum transition where $\epsilon = \epsilon^*$, $\mu = \sqrt{c}$, and $\beta_R = \beta^*$. At that point, the relevant eigenvalue crosses into the bulk, and the retrieval phase disappears.

We note that the paramagnetic, retrieval, and spin-glass states were also studied in ref. 51, using a generalized Potts model and a heat bath MCMC algorithm. However, their Hamiltonian depends on a tunable cut-size parameter, rather than on a general measure of community structure such as the modularity. Moreover, it is difficult to obtain analytical results on phase transitions using MCMC algorithms, whereas the stability of BP fixed points is quite tractable.

**Defining the Spin-Glass Phase.** Although we have identified the spin-glass phase with the nonconvergence of belief propagation, the true phase diagram is potentially more complicated. The spin-glass phase is defined by the divergence of the spin-glass susceptibility. If this phase appears continuously, then in sparse problems this is equivalent to the sensitivity of the BP messages to noise, i.e., whether it converges to a stable fixed point. However, if the spin-glass phase appears discontinuously, it could be that BP converges even though the true susceptibility diverges (e.g., ref. 52).

We expect this to happen above the Nishimori line when the hard but detectable phase exists (5), when there is a retrieval state with lower free energy than the factorized fixed point but with an exponentially small basin of attraction, so that BP starting with random messages fails to converge to the true minimum of the free energy. Detecting this spin-glass phase would require us to go beyond the replica-symmetric BP equations used here to equations with one-step replica symmetry breaking (45). In the assortative case of the stochastic block model, the hard-but-detectable phase exists for $q \geq 5$. Happily, the corresponding range of parameters is quite narrow; nevertheless, more work on this needs to be done.

**ACKNOWLEDGMENTS.** We are grateful to Silvio Franz, Florent Krzakala, Mark Newman, Federico Ricci-Tersenghi, Christophe Schulke, and Lenka Zdeborová for helpful discussions and to Tiago de Paula Peixoto for drawing Fig. 4, *Left*, using his software at graph-tool.skewed.de/. This work was supported by the Air Force Office of Scientific Research and the Defense Advanced Research Planning Agency under Grant FA9550-12-1-0432.

1. Von Luxburg U (2007) A tutorial on spectral clustering. *Stat Comput* 17:395.
2. Newman MEJ (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys Rev E Stat Nonlin Soft Matter Phys* 74(3 Pt 2):036104.
3. Krzakala F, et al. (2013) Spectral redemption in clustering sparse networks. *Proc Natl Acad Sci USA* 110(52):20935–20940.
4. Hastings MB (2006) Community detection as an inference problem. *Phys Rev E Stat Nonlin Soft Matter Phys* 74(3 Pt 2):035102.
5. Decelle A, Krzakala F, Moore C, Zdeborová L (2011) Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys Rev E Stat Nonlin Soft Matter Phys* 84(6 Pt 2):066106.
6. Decelle A, Krzakala F, Moore C, Zdeborová L (2011) Inference and phase transitions in the detection of modules in sparse networks. *Phys Rev Lett* 107(6):065701.
7. Karrer B, Newman MEJ (2011) Stochastic block models and community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 83(1 Pt 2):016107.
8. Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 70(6 Pt 2):066111.
9. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech* 2008:P10008.
10. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci USA* 105(4):1118–1123.
11. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486:75.
12. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 69(2 Pt 2):026113.
13. Newman MEJ (2004) Fast algorithm for detecting community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 69(6 Pt 2):066133.
14. Duch J, Arenas A (2005) Community detection in complex networks using extremal optimization. *Phys Rev E Stat Nonlin Soft Matter Phys* 72(2 Pt 2):027104.
15. Guimerà R, Sales-Pardo M, Amaral LAN (2004) Modularity from fluctuations in random graphs and complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 70(2 Pt 2):025101.
16. Reichardt J, Bornholdt S (2006) Statistical mechanics of community detection. *Phys Rev E Stat Nonlin Soft Matter Phys* 74(1 Pt 2):016110.
17. Zdeborová L, Boettcher S (2010) A conjecture on the maximum cut and bisection width in random regular graphs. *J Stat Mech* 2010:P02020.
18. Sulc P, Zdeborová L (2010) Belief propagation for graph partitioning. *J Phys Math Gen* 43:B5003.
19. Good BH, de Montjoye YA, Clauset A (2010) Performance of modularity maximization in practical contexts. *Phys Rev E Stat Nonlin Soft Matter Phys* 81(4 Pt 2):046106.
20. Lancichinetti A, Radicchi F, Ramasco JJ (2010) Statistical significance of communities in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 81(4 Pt 2):046110.
21. Lancichinetti A, Radicchi F, Ramasco JJ, Fortunato S (2011) Finding statistically significant communities in networks. *PLoS ONE* 6(4):e18961.
22. Wilson JD, Wang S, Mucha PJ, Bhamidi S, Nobel AB (2009) *A Testing Based Extraction Algorithm for Identifying Significant Communities in Networks* (Oxford Univ Press) arXiv:1308.0777.
23. Iba Y (1999) The Nishimori line and Bayesian statistics. *J Phys Math Gen* 32:3875.
24. Clauset A, Moore C, Newman MEJ (2008) Hierarchical structure and the prediction of missing links in networks. *Nature* 453(7191):98–101.
25. Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. *Sci Rep* 2:336.
26. Mossel E, Neeman J, Sly A (2012) Stochastic block models and reconstruction. *arXiv*:1202.1499.
27. Massoulie L (2013) Community detection thresholds and the weak Ramanujan property. *arXiv*:1311.3085.
28. Mossel E, Neeman J, Sly A (2013) A proof of the block model threshold conjecture. *arXiv*:1311.4115.
29. Kesten H, Stigum BP (1966) A limit theorem for multidimensional Galton-Watson processes. *Ann Math Stat* 37:1211.
30. Kesten H, Stigum BP (1966) Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *Ann Math Stat* 37:1463.
31. Nishimori H (2012) *Statistical Physics of Spin Glasses and Information Processing* (Oxford Univ Press, Oxford).
32. Montanari A (2008) Estimating random variables from random sparse observations. *Eur Trans Telecomm* 19:385.
33. Zachary WW (1977) An information flow model for conflict and fission in small groups. *J Anthropol Res* 33(4):452–473.
34. Adamic LA, Glance N (2005) The political blogosphere and the 2004 US election: Divided they blog. *Proceedings of the Third International Workshop on Link Discovery* (ACM, New York), pp 36–43.
35. Lusseau D, et al. (2003) The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behav Ecol Sociobiol* 54:396.
36. Krebs V, Social Network Analysis software & services for organizations, communities, and their consultants. Available at www.orgnet.com/. Accessed September 26, 2014.
37. Leskovec J, Lang KJ, Dasgupta A, Mahoney MW (2009) Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Math* 6:29.
38. Sales-Pardo M, Guimerà R, Moreira AA, Amaral LAN (2007) Extracting the hierarchical organization of complex systems. *Proc Natl Acad Sci USA* 104(39):15224–15229.
39. Peixoto TP (2014) Hierarchical block structures and high-resolution model selection in large networks. *Phys Rev X* 4:011047.
40. Dickens C (1850) *The Personal History, Adventures, Experience and Observation of David Copperfield the Younger of Blunderstone Rookery (Which He Never Meant to Publish on Any Account)* (Bradbury and Evans).
41. Danon L, Diaz-Guilera A, Duch J, Arenas A (2005) Comparing community structure identification. *J Stat Mech* 2005:P09008.
42. Lancichinetti A, Fortunato S, Radicchi F (2008) Benchmark graphs for testing community detection algorithms. *Phys Rev E Stat Nonlin Soft Matter Phys* 78(4 Pt 2):046110.
43. Yan X, et al. (2014) Model selection for degree-corrected block models. *J Stat Mech* 2014:P05007.
44. Fortunato S, Barthélemy M (2007) Resolution limit in community detection. *Proc Natl Acad Sci USA* 104(1):36–41.
45. Mézard M, Montanari A (2009) *Information, Physics, and Computation* (Oxford Univ Press, Oxford).
46. Yedidia J, Freeman W, Weiss Y (2003) Understanding belief propagation and its generalizations. Exploring Artificial Intelligence in the New Millennium (Morgan Kaufmann Publishers Inc., San Francisco).
47. Mézard M, Parisi G (2001) The Bethe lattice spin glass revisited. *Eur Phys J B* 20:217.
48. De Almeida J, Thouless D (1978) Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *J Phys Math Gen* 11:983.
49. Mézard M, Montanari A (2006) Reconstruction on trees and spin glass transition. *J Stat Phys* 124:1317.
50. Janson S, Mossel E (2004) Robust reconstruction on trees is determined by the second eigenvalue. *Ann Probab* 32(3B):2630–2649.
51. Hu D, Ronhovde P, Nussinov Z (2012) Phase transitions in random Potts systems and the community detection problem. *Philos Mag* 92:406.
52. Zdeborová L (2009) Statistical physics of hard optimization problems. *Acta Phys Slov* 59:169.