## Supporting Text

In the supplementary material we first give additional details on the alignment algorithm and then discuss a simple procedure for identifying larger modules generating multiple smaller subgraphs.

The algorithm proceeds in four stages:

1. By enumeration all unique non-treelike subgraphs of size $n$ are found. We consider only subgraphs where each node carries at least two internal links, other than self-links ("exclusion of dangling links"). The reason is that including dangling links would generate from each subgraph an artificially inflated family of subgraphs generated by including all combinations of neighboring nodes into the subgraph. The enumeration is done by first finding all closed paths in the graph of length shorter than $2n-3$. (The maximum length derives from considering the non-treelike structure with the longest pathlength from the origin through all points of the subgraph back to the origin. The graph is considered as undirected at this stage.) The subgraphs are labeled by $\alpha = 1, \ldots, p^{\max}$.

2. The pairwise minimal mismatch $M^{\alpha\beta}$ for all *pairs of subgraphs* $\alpha, \beta$ is found by enumerating all $n!$ possible alignments of each pair of subgraphs. For each pair of subgraphs $\alpha$ and $\beta$ we determine whether they overlap by counting the number of nodes they have in common. The elements of the coupling matrix $\tilde{M}^{\alpha\beta}$ in the Hamiltonian **11** are given by $M^{\alpha\beta}$ if the subgraphs do not overlap, and by a large number, chosen to be 10, if they do.

3. The next task is to select a subset of the subgraphs such that the total score **10** is maximized at given values of the scoring parameters. To this end simulated annealing is used, with the (negative) score as the energy function, increasing the inverse temperature from 0 to 10 in 1000 Monte-Carlo sweeps. We assign each *subgraph* a spin variable: spin $s^\alpha = 1$ implies that the subgraph $\alpha$ is included in the alignment, spin $s^\alpha = 0$ that it is not. The contribution to Eq. **10** from the mismatch of two subgraphs acts as a coupling between their spins, the contribution of subgraph $\alpha$ to the total number of links $L$ in **10** acts as a local field, resulting in the Hamiltonian **11**. The evaluation of the last term in **10**, $\log(Z/Z_0)$, is discussed below.

The last step is repeated at different values of $\sigma$ and $\mu$ in order to perform the parametric optimization leading to Fig. 2*b*. The parameter $\sigma_0$ describing the null ensemble, on the other hand, is determined independently by considering the non-treelike subgraphs found in the randomized graph as described in the paper. $\sigma_0$ is chosen such that the average number of internal links in the ensemble of uncorrelated subgraphs with an enhanced number of links **5** equals that of the non-treelike subgraphs found in the randomized network,

$$\frac{1}{p}\sum_{\alpha=1}^{p}\langle L(\mathbf{c}^\alpha)\rangle_{\sigma_0,\mu=0} = \overline{L}_{\text{randomized}} \; .$$

Note that the ensemble **5** still depends on the connectivities of the nodes in each subgraph. The generalization to several *groups of subgraphs*, where only subgraphs from the same group are aligned with each other, can be done by admitting more states of the Potts-like spins $s^\alpha$. For $q$-state spins this would group the subgraphs into $q-1$ clusters much like in superparamagnetic clustering (1).

There are two approximations behind this algorithm. First, treelike subgraphs are excluded from the start. This step cuts down an enormous number of combinatorial possibilities associated with treelike subgraphs, which, different connectivities apart, are always locally similar.

Second, it uses the minimal mismatch obtained from the *pairwise* alignment of subgraphs (step 2), even though the minimal mismatch obtained by aligning a set of more than two subgraphs may be higher than that of the sum of pairwise alignments. This is easily seen by comparing the total number of alignments of all *pairs* chosen from $p$ subgraphs, $(n!)^{p(p-1)/2}$, with the number of alignments of $p$ subgraphs, $(n!)^p$. However, in the case of interest, where the graph contains multiple copies of a motif (possibly corrupted by noise), the sum of pairwise minimal mismatches will typically be very close to the minimal mismatch obtained from aligning all subgraphs simultaneously.

The maximal-score alignment $\mathcal{A}^\star(\sigma,\mu)$ turns out to be unique in most subgraphs. To see this, consider all alignments $\mathcal{A}^\alpha$ of a particular subgraph $\alpha$ with respect to the other subgraphs whose alignment is kept fixed. Two different alignments have the same score if and only if there is a permutation of the nodes $r_1^\alpha, \ldots, r_n^\alpha$ leaving both the adjacency matrix $c_{ij}^\alpha$ and the matrix $w_{ij}^\alpha$ defined above Eq. **4** invariant.

While symmetries of the adjacency matrices occur frequently, entries of the matrix $w_{ij}$ are unique in most subgraphs, since the connectivities in biological networks are broadly distributed.

We now discuss the normalizing constant **9** of the alignment ensemble **8**. We approximate the likelihood of given parameter values, which involves the sum over all alignments $\mathcal{A}$ as in **9**, by the corresponding maximum-score alignment. (In the literature for sequence alignment, this is known as the Viterbi approximation.) An improved likelihood estimate is possible using probabilistic graph alignment algorithms but is not expected to alter our results qualitatively. The optimal alignment has $p^\star \equiv p^\star(\sigma^\star, \mu^\star)$ subgraphs with average internal link number $\overline{L}^\star \equiv \overline{L}^\star(\sigma^\star, \mu^\star)$ and fuzziness $\overline{M}^\star \equiv \overline{M}^\star(\sigma^\star, \mu^\star)$. As may be seen by differentiation of **10** with respect to the scoring parameters, at $\sigma = \sigma^\star$ and $\mu = \mu^\star$ the $Q$ ensemble fits to the data set in the sense that the expectation values of the internal link number and the fuzziness equal the actual values,

$$\frac{1}{p}\sum_{\alpha=1}^{p}\langle L(\mathbf{c}^\alpha)\rangle_{\sigma^\star,\mu^\star} = \overline{L}^\star,$$

$$\frac{1}{p^2}\sum_{\alpha,\beta=1}^{p}\langle M(\mathbf{c}^\alpha,\mathbf{c}^\beta)\rangle_{\sigma^\star,\mu^\star} = \overline{M}^\star.$$

The normalizing constant **9** needs to be computed for two sets of parameters; for $\sigma$, $\mu$ characterizing the $Q$ ensemble, and for $\sigma_0$, $\mu_0$ characterizing the $Q_0$ ensemble. In both cases the normalizing constant consists of a trace over the link configuration $\{\mathbf{c}^\alpha\}$ in all subgraphs. Since the constant **9** factorizes in the link labels $i, j$, we consider only a single of these factors (a single "string"), drop the $i, j$ indices, and separate the bilinear form of the pairwise mismatch **3** into a quadratic and a linear part.

Formally, this expression is the partition function of a mean-field ferromagnet in a fluctuating field. The field depends on the local connectivities of each node along the "string" via the ensemble $P_0$, Eq. **4**. Using a Hubbard-Stratonovich transformation to linearize the quadratic term, the trace over $\{c^\alpha\}$ can be performed, giving

$$Z = \int \frac{dt}{\sqrt{2\pi/p}} \exp\left\{-pt^2/2 + \sum_{\alpha}^{p} g^\alpha(t)\right\} \ , [\mathbf{12}]$$

where

$$g^\alpha(t) = \log\left[(1-w^\alpha) + w^\alpha \exp\left\{\sqrt{2\mu}t + \sigma - \mu\right\}\right] \ .$$

For large $p$ this expression can be evaluated by a saddle point integral, giving

$$\log Z \approx -pt^{\star 2}/2 + \sum_{\alpha}^{p} g^\alpha(t^\star) + \mathcal{O}(\log p) \ ,$$

where $t^\star$ maximizes the exponent in Eq.**12**. The contribution to leading order of adding a new subgraph with index $\alpha$ is thus

$$\Delta \log Z \approx -t^{\star 2}/2 + g^\alpha(t^\star) \ .$$

The change of $t^\star$ as a finite number of subgraphs is added to or removed from the alignment alters the result only by terms of order $p^{-1}$. It thus turns out to be sufficient to update the saddle-point value $t^\star$ for each link $i, j$ once per Monte-Carlo sweep of the algorithm.

In order to compute for each pair $i, j$ in the Viterbi approximation, the one-to-one mapping between nodes in each subgraph $\mathcal{A}$ is needed, going beyond the *pairwise* alignment. This mapping is also needed to produce the plots of the consensus motifs in Fig. 4. It is produced by minimizing the fuzziness over the mapping between nodes in each subgraph, again using Monte-Carlo dynamics (100 Monte-Carlo sweeps while linearly increasing the inverse temperature from 0 to 10). The result of course depends on the subgraphs in the alignment, and thus the mapping ought to be updated each time a subgraph is added or removed from the alignment. In practice, however, one update of the mapping between nodes in each subgraph every 250 steps of the algorithm is sufficient. The reason for this is again that the mapping between nodes in subgraphs in the alignment is unchanged as motifs sufficiently close to the consensus motif enter/leave the alignment.

Finally, we discuss a simple procedure for identifying these structures *from the set of subgraphs at fixed (small)* $n$. First, for a given subgraph of size $n$, all neighbors with at least two links to the subgraph are enumerated. In this way, non-treelike subgraphs without dangling bonds with $n+1$ nodes are generated. This procedure is repeated for the entire list of $p$ subgraphs. Several subgraphs of size $n+1$ will occur repeatedly in this list: the more subgraphs of size $n$ can be be generated from the larger $n+1$ subgraph the more frequently it occurs on the list. Thus ranking the $n+1$ subgraphs according to the number of times they occur in the list, one obtains the $n+1$ subgraph from which the largest number of $n$

subgraphs derive. Clearly this procedure can be repeated iteratively, leading to subgraphs of increasing size. In fact, Fig. 4 $c$ is the result of applying this scheme to the non-treelike subgraphs with $n = 5$. Iterating twice, one finds two instances of the $n = 7$ layered structure of Fig. 4$c$.

[1] Blat,M., Wiseman,S., & Domany,E. (1996) *Phys. Rev. Lett.* **76** 3251-3255.