questioning the presence of (i, j) in each structure. By construction, the distribution of $X_{i,j}$ is given by the base pairing probability $\mathbb{P}(X_{i,j}(s) = 1) = p_{i,j}$. Any base pair, (i, j), has an *entropy*, defined by the information entropy of $X_{i,j}$, i.e.

$$H(X_{i,j}) = -p_{i,j} \log_2 p_{i,j} - (1 - p_{i,j}) \log_2(1 - p_{i,j}),$$

where the units of H are in bits. The entropy $H(X_{i,j})$ measures the uncertainty of the base pair (i, j) in Ω . When a base pair (i, j) is certain to either exist or not, its entropy $H(X_{i,j})$ is 0. However, in case $p_{i,j}$ is closer to $1/2$, $H(X_{i,j})$ becomes larger. The r.v. $X_{i,j}$ partitions the space Ω into two disjoint subspaces Ω_0 and Ω_1, where $\Omega_k = \{s \in \Omega : X_{i,j}(s) = k\}$ $(k = 0, 1)$, and the induced distributions are given by

$$p_0(s) = \frac{p(s)}{1 - p_{i,j}} \quad \text{for } s \in \Omega_0, \qquad p_1(s) = \frac{p(s)}{p_{i,j}} \quad \text{for } s \in \Omega_1.$$

Intuitively, $H(X_{i,j})$ quantifies the average bits of information we would expect to gain about the ensemble when querying a base pair (i, j) This motivates us to consider the *maximum entropy base pairs*, the base pair (i_0, j_0) having maximum entropy among all base pairs in Ω, i.e.

$$(i_0, j_0) = \operatorname*{argmax}_{(i,j)} H(X_{i,j}).$$

As we shall prove in Section, X_{i_0,j_0} produces maximally balanced splits.

0.1. The ensemble tree

Equipped with the notion of ensemble and bit query (i.e. the respective maximum entropy base pairs), we proceed by describing our strategy to identify the target structure as specified in Problem. The first step consists in having a closer look at the space of ensemble reductions. Each split obtained by partitioning the ensemble Ω using r.v. $X_{i,j}$, can in turn be bipartitioned itself via any of its maximum entropy base pairs. This recursive splitting induces the *ensemble tree*, $T(\Omega)$, whose vertices are sub-samples and in which its k-th layer represents a partition of the original ensemble into 2^k blocks. $T(\Omega)$, is a rooted binary tree, in which

$$\begin{aligned} \epsilon_{i,j}((x_1, \ldots, x_j), (y_1, \ldots, y_m)) &= (y_1, \ldots, y_{i-1}, x_1, \ldots, x_j, y_{i+1}, \ldots y_m) \\ \xi_{i,j}(x_1, \ldots, x_n) &= ((x_i, \ldots, x_j), (x_1, \ldots, x_{i-1}, x_{j+1}, \ldots, x_n)). \end{aligned}$$

To quantify to what extent modularity can discriminate base pairs, we perform computational experiments on random sequences via splittings. For each sequence, we consider its MFE structure s computed via ViennaRNA (**?**). Given two positions i and j, we cut the entire sequence $\mathbf{x}$ into two fragments, $\mathbf{x}_{i,j}$ and the remainder $\bar{\mathbf{x}}_{i,j}$, i.e., $\xi_{i,j}(\mathbf{x}) = (\mathbf{x}_{i,j}, \bar{\mathbf{x}}_{i,j})$. Subsequently, the two fragments $\mathbf{x}_{i,j}$ and $\bar{\mathbf{x}}_{i,j}$ refold into their MFE structures $s_{i,j}$ and $\bar{s}_{i,j}$, respectively, which are combined into a structure $\epsilon_{i,j}(s_{i,j}, \bar{s}_{i,j})$. If bases i and j are paired in s