S4 Appendix. Monte Carlo estimate of measures $F(C, S)$

One way to efficiently estimate the measure $F(C, S)$ of a given community $C$ within a larger multispecies system $S$ is to use a simple Monte Carlo approach. We can randomly sample a point on the unit sphere in $S = |S|$ dimensions by picking a random vector $x = \sum_{i \in S} \nu_i e_i$ where $\nu_i$ is chosen from a Gaussian distribution with a common variance for each $i$. We can efficiently test whether $x \in D_F(C, S)$ by checking $S$ linear conditions of the form $x \cdot b > 0$, as described below. We can then rapidly accumulate statistics on the fraction of random points in the sphere that lie in $D_F(C, S)$ to estimate $F(C, S)$.

In further detail, the linear conditions that determine which community $C \subset S$ is associated with a given point $x$ can be described in terms of the boundaries between a community $C$ and a community $C' = C \cup \{i\}, i \notin C$. A necessary condition for $x \in D_F(C, S)$ or $x \in D_F(C', S)$ is that

$$
(det[a'_{j_1} \ldots a'_{j_C} x]) = \mp(det[a'_{j_1} \ldots a'_{j_C} e_i]), \ j_k \in C,
$$

(S6)

where $a'$ denotes the restriction of the vector $a$ to the subspace associated with community $C'$, and the negative/positive sign is associated with the community $C/C'$. This follows because the linear hypersurface that separates $C, C'$ is spanned by the vectors $a'_{j_k}$ and $e_i, i \notin C'$. (Note that these conditions continue to be valid in the degenerate case $C = \{\}$, which is needed to fix the boundaries of communities $C' = \{i\}$ containing only a single species.) To check whether $x \in D_F(C, S)$, we thus simply check each of the $S$ conditions associated with the boundaries of $C$ with communities that differ by a single species $i$, where $i$ is added or subtracted from the community depending on whether $i \in C$.

If all these conditions are satisfied then $x \in D_F(C, S)$.

Since each of the determinants on the LHS of Eq (S6) can be written in the form

$$
det[a'_{j_1} \ldots a'_{j_C} x] = b_{C,i} \cdot x,
$$

by computing the vectors $b_{C,i}$ in advance for each combination $C, i$ we can efficiently sample and test many points to give a good Monte Carlo estimate of $F(C, S)$. A simple mathematica code implementing this Monte Carlo algorithm is available with the other code supporting the results of this work.