2 Largest common substructures in probabilistic combinatorics

Consider the following general setting. There is a set of n labeled elements $[n] := \{1, 2, \ldots, n\}$. There is an instance $\mathcal{S}$ of a "combinatorial structure" built over these elements. The type of structure is such that for any subset $A \subset [n]$ there is an induced substructure of the same type on A. Three examples of types:

- graphs on vertex-set $[n]$
- partial orders on the set $[n]$
- cladograms (leaf-labeled trees – see below) on leaf-set $[n]$.

Given two distinct instances $\mathcal{S}_1, \mathcal{S}_2$ of the same type of structure on $[n]$, we can ask for each $A \subset [n]$ whether the two induced substructures on A are identical; and so we can define

$$c(\mathcal{S}_1, \mathcal{S}_2) = \max\{\#A : \text{ induced substructures are identical}\}$$

where $\#A$ denotes cardinality. Finally, given a probability distribution μ_n on the set of all structures of a particular type, we can consider the random variable

$$C_n = c(\mathcal{S}_1, \mathcal{S}_2) \text{ where } \mathcal{S}_1, \mathcal{S}_2 \text{ are independent picks from } \mu_n.$$

This general framework includes the following two well-known examples.

Example 1. Suppose the type is "graph" and the distribution μ_n is the usual random graph $G(n, p)$ in which possible edges are independently present with probability p. Given two instances G_1, G_2 of graphs we can define the "similarity" graph G to have an edge (i, j) iff both or neither of G_1, G_2 has the edge (i, j). Then

$$c(G_1, G_2) = \text{cl}(G) := \text{ maximal clique size of } G.$$

Moreover if $\mathcal{G}_1, \mathcal{G}_2$ are independent picks from $G(n, p)$ then their "similarity" is distributed as $G(n, q)$ for $q = p^2 + (1 - p)^2$. Thus C_n is just the maximal clique size of a random graph, a well-understood quantity ([12] section 11.1).

Example 2. Suppose the type is "total order" and μ_n is the uniform distribution on all $n!$ total orders on $[n]$. A few moments thought shows that here C_n is distributed as the longest increasing subsequence of a (single) uniform random permutation. This is again a well-studied quantity, of recent interest because of its connection with extreme eigenvalues of random matrices [8, 11, 28].

Of course these two examples are atypical, in that "by symmetry" a problem about two independent random structures reduces to a problem about one random structure, but they suggest that investigation of other examples may be interesting. Here are two new examples.

Example 3. Figure 1 shows a *cladogram* on $[n]$ (rooted unordered binary tree with non-root leaves labeled by $[n]$) for $n = 11$, together with the sub-cladogram on $A = \{1, 2, 3, 4\}$.

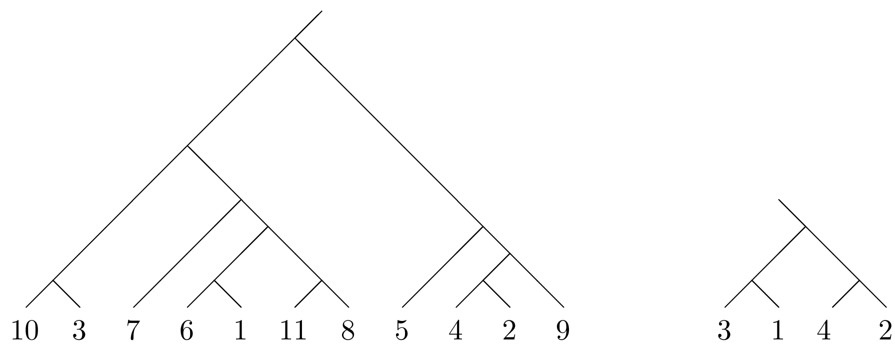

Figure 1. A cladogram on [11] and the induced sub-cladogram on [4].

There are two natural probability measures on n-cladograms:
(a) uniform on all $(2n-3)!!$ cladograms;
(b) the *coalescent*, starting with n lineages and successively joining two randomly-chosen lineages into one lineage.
We conjecture that in both cases

$$EC_n = n^{\gamma+o(1)}$$

for different constants $\gamma_a, \gamma_b < 1/2$. We do not have conjectures for numerical values, but one can consider continuous limits of the relevant structures and seek to define candidate constants γ in terms of the limit random structures.

Example 4. Amongst several models for random partial orders [13], consider the random two-dimensional partial order on $[n]$. This is the partial order obtained by taking n points (x_i, y_i), $1 \le i \le n$ uniformly randomly in the unit square $[0,1]^2$ and using the induced "coordinatewise" partial order [29]. Here the natural conjecture is

$$EC_n \sim cn^{1/3}, \text{ for some } 0 < c < \infty. \tag{2}$$

Remarkably, there are two quite different ways to obtain subsets $A \subset [n]$ of size $\approx n^{1/3}$ such that the partial orders agree on A.

(i) Partition $[0,1]^2$ into subsquares of side $n^{-1/3}$. Take B as the set of i such that the i'th point in both processes falls into the same subsquare, so $E\#B = n \times n^{-2/3} = n^{1/3}$. Then take A as a maximal subset of B such that no two of the corresponding subsquares are in the same row or column.
(ii) Take C as the set of i such that in both processes the i'th point is within $n^{-1/3}$ of the reverse diagonal in $[0,1]^2$. Again $\#C$ is order $n^{1/3}$. And one can choose $A \subset C$ with $\#A/\#C$ non-vanishing such that each partial order on A is the trivial partial order.
It is not hard (Graham Brightwell, personal communication) to prove an $O(n^{1/3})$ upper bound using the first moment method. But establishing a value for, or existence of, the presumed limit constant c in (2) may be genuinely hard.