Algorithm 18.1 Data perturbation search

Procedure Search-with-Data-Perturbation (
  $\mathcal{G}_\emptyset$, // initial network structure
  $\mathcal{D}$ // Fully observed data set
  score, // Score
  $\mathcal{O}$, // A set of search operator
  Search, // Search procedure
  t_0, // Initial perturbation size
  γ, // Reduction in perturbation size
)

  $\mathcal{G} \leftarrow$ Search$(\mathcal{G}_\emptyset, \mathcal{D}, \text{score}, \mathcal{O})$
  $\mathcal{G}_{\text{best}} \leftarrow \mathcal{G}$
  $t \leftarrow t_0$
  for $i = 1, \ldots$ until convergence
    $\mathcal{D}' \leftarrow$ Perturb$(\mathcal{D}, t)$
    $\mathcal{G} \leftarrow$ Search$(\mathcal{G}, \mathcal{D}', \text{score}, \mathcal{O})$
    if $\text{score}(\mathcal{G} : \mathcal{D}) > \text{score}(\mathcal{G}_{\text{best}} : \mathcal{D})$ **then**
      $\mathcal{G}_{\text{best}} \leftarrow \mathcal{G}$
    $t \leftarrow \gamma \cdot t$

  return $\mathcal{G}_{\text{best}}$

to the values in the original data. Thus, we expect that big differences between networks are preserved. That is, if $\text{score}(\mathcal{G}_{X \to Y} : \mathcal{D}) \gg \text{score}(\mathcal{G}_2 : \mathcal{D})$, then we expect that $\text{score}(\mathcal{G}_{X \to Y} : \mathcal{D}') \gg \text{score}(\mathcal{G}_2 : \mathcal{D}')$. On the other hand, the perturbation does change the comparison between networks that are similar. The basic intuition is that the score using $\mathcal{D}'$ has the same broad outline as the score using $\mathcal{D}$, yet might have different fine-grained topology. This suggests that a structure $\mathcal{G}$ that is a local maximum when using the score on $\mathcal{D}$ is no longer a local maximum when using $\mathcal{D}'$. The magnitude of perturbation determines the level of details that are preserved after the perturbation.

weighted data instances

We note that instead of duplicating and removing instances, we can achieve perturbation by *weighting* data instances. Much of the discussion on scoring networks and related topics applies without change if we assign weight to each instance. Formally, the only difference is the computation of sufficient statistics. If we have weights $w[m]$ for the m'th instance, then the sufficient statistics are redefined as:

$$M[\boldsymbol{z}] = \sum_m \mathbf{1}\{\boldsymbol{Z}[m] = \boldsymbol{z}\} \cdot w[m].$$

Note that when $w[m] = 1$, this reduces to the standard definition of sufficient statistics. Instance duplication and deletion lead to integer weights. However, we can easily consider perturbation that results in fractional weights. This leads to a continuous spectrum of data perturbations that