1 Maximum degree of $4\Delta_{avg} + 5$

1.1 Algorithm

For each node i, let's denote α_i its initial degree before running the algorithm. We separate the nodes in two subsets L and H which respectively contain low-degree nodes and the high-degree nodes. The low-degree nodes are the $\frac{n}{4}$ nodes with the lowest degree, and the high-degree nodes are the other ones. Amongst the high-degree nodes, we call high-out-degree (resp. high-in-degree) an high-degree node whose out-degree (resp. in-degree) is greater than $2\Delta_{avg}$. We note HO and HI the subsets of high-out and high-in-degree nodes. Note that high-degree nodes can be neither high-out nor high-in-degree.

We want to remove all the initial connexions between high-degree nodes. To remove an edge, we use the help of another node which will reroute it, by adding a first edge from the source to him and a second one from him to the destination. Both low and high-degree nodes can help, and we note β_i the number of edges a node i can help.

Let's choose β_i that way:

- if $i \in L$, then $\beta_i = \lceil \frac{2}{3}\Delta_{avg} - \frac{1}{2}\alpha_i \rceil \in \mathbb{N}$
- if $i \in H$, then $\beta_i = max(\lceil \frac{2}{3}\Delta_{avg} - \frac{1}{6}\alpha_i \rceil, 0) \in \mathbb{N}$

Thanks to helping nodes, we remove all the initial connexions between high-degree nodes.

Before going further in the algorithm, let's justify that for a low-degree node i, β_i is indeed positive because it may not be intuitive.

We can show it by doing a observation on low-degree nodes. Let's study a general case where the low-degree nodes are the n_l lowest ones and the high-degrees nodes the other ones ($0 < n_l < n$). Let's sort the degrees. [i, ..., nl] are the low-degree nodes and [nl+1, ..., n] are the high-degree ones. We take a look at the total degree:

$$n\Delta_{avg} = \sum_{i<nl} \alpha_i + \alpha_{nl} + \sum_{i>nl} \alpha_i$$

$$n\Delta_{avg} \geq \alpha_{nl} + \sum_{i>nl} \alpha_i$$

and if $i > nl$ then $\alpha_i > \alpha_n l$ because the nodes are sorted. So

$$n\Delta_{avg} \geq \alpha_{nl}(n - nl + 1)$$

$$\forall i \in L, \alpha_i \leq \alpha_{nl} \leq \frac{n\Delta_{avg}}{n - nl + 1} \leq \frac{n\Delta_{avg}}{n - nl}$$

By replacing n_l with $\frac{n}{4}$, we have $\forall i \in L, \alpha_i \leq \frac{4}{3}\Delta_{avg}$ and so $\beta_i >= 0$

Once we have removed the initial edges between high-degree nodes, we do melhorn trees with the incoming edges of high-in-degree nodes and with outgoing edges of high-out-degree nodes. Some high-in and high-out-degree nodes may have helped nodes. In that case we do not include those helped nodes in the melhorn tree of the helper.

1.2 Analysis

The first thing to know is if we are able, with our defintion of $beta_i$ to help enough edges. Let's note B the number of edges to remove and A the number of edges in which a low degree node is involved. B exactly equals to the number of initial edges between high-degree nodes because those are the edges we want to remove. So we have :

$$m = A + B$$

Furthermore we have

$$\sum_{i \in L} \alpha_i \leq 2A$$

Because for an extreme case, there is no edges between low and high-degree nodes so $\sum_{i \in L} \alpha_i = 2A$. Otherwise we have connexions with high-degree nodes which will increase A for a given total degree $\sum_{i \in L} \alpha_i$. Then it leads to

$$B \leq m - \frac{1}{2} \sum_{i \in L} \alpha_i$$

It means that beeing able to help more than $m - \frac{1}{2}\sum_{i \in L} \alpha_i$ is sufficient to remove all the initial edges between high-degree nodes. Now we show that our choice of β_i allows us to help a sufficient number of edges. To this end we sum the β_i.

- if $i \in L$, $\beta_i \geq \frac{2}{3}\Delta_{avg} - \frac{1}{2}\alpha_i$

$$\sum_{i \in L} \beta_i \geq \frac{2}{3} * \frac{n}{4}\Delta_{avg} - \frac{1}{2}\sum_{i \in L} \alpha_i$$

$$\sum_{i \in L} \beta_i \geq \frac{m}{3} - \frac{1}{2}\sum_{i \in L} \alpha_i$$

- if $i \in H$, $\beta_i \geq \frac{2}{3}\Delta_{avg} - \frac{1}{6}\alpha_i$

$$\sum_{i \in L} \beta_i \geq \frac{2}{3} * \frac{3n}{4}\Delta_{avg} - \frac{1}{6}\sum_{i \in H} \alpha_i$$

$$\sum_{i \in L} \beta_i \geq m - \frac{1}{6}\sum_{i \in H} \alpha_i$$

And because $\sum_{i\in H} \alpha_i \leq 2m$ We have

$$\sum_{i\in L} \beta_i \geq m - \frac{1}{3}m$$

We sum our two previous results in order to have a lower bound on $\sum_i \beta_i$

$$\sum_{i} \beta_i \geq m - \frac{1}{2}\sum_{i\in L} \alpha_i$$

And that way we are able to help enough edges.

Finally we evaluate the maximal final degree of the nodes. We note γ_i the final degree of the node i. We study the different subsets of nodes to find an upper bound for each case. For any node, to be involved in a melhorn tree can increase a lot its initial degree. For one initial connexion which will be replaced in a melhorn tree, it can lead at the end to three connexions.

- if $i \in L$, in the worst case, the initial edges from i are going to high-in degree node and the initial edges in destination to i are coming from high-out degree node. So for each initial edge, i will be involved in a melhorn tree. So the degree increases from α_i to at most $3\alpha_i$. But this node may also have helped edges and in the worst case, it has only helped edges from high-out to high-in-degree nodes. This adds $3 \times 2\beta_i$ more to the final degree. At the end we have

  $$\gamma_i \leq 3(2\beta_i + \alpha_i)$$

  and

  $$\beta_i < \frac{2}{3}\Delta_{avg} - \frac{1}{2}\alpha_i + 1$$

  So

  $$\gamma_i \leq 4\Delta_{avg} + 5$$

- if $i \in HO \cap HI$, the intial connexions of i will not lead to melhorn trees because they have been replaced by connexions with an helper node. And because we forced it in the description of the algorithm, if this helper creates a melhorn tree, i will not be involved in. So the degree of i due to initial connexions will not increase. Even better because i is both high-out and high-in degree, and making so two melhorn trees on intial in and out edges, this degree will decrease to 2. Like before we need to add the contribution in the degree of helping other nodes.

  $$\gamma_i \leq 2 + 6\beta_i$$

  if $\beta_i = 0$, $\gamma_i \leq 2$ else $\beta_i = \lceil \frac{2}{3}\Delta_{avg} - \frac{1}{6}\alpha_i \rceil > 0$

  $$\gamma_i < 2 + 6(\frac{2}{3}\Delta_{avg} - \frac{1}{6}\alpha_i + 1) = 8 + 4\Delta_{avg} - \alpha_i$$

  But $\alpha_i > 4\Delta_{avg}$ so $\gamma_i \leq 7$

- if $i \in HO \setminus HI$ (same reasoning for $i \in HI \setminus HO$)

  Let's note α_i^+ the initial out-degree of i and α_i^- the initial in-degree We have :

  1. $\alpha_i = \alpha_i^+ + \alpha_i^-$
  2. $\alpha_i^+ \geq 2\Delta_{avg}$ because $i \in HO$
  3. $\alpha_i^- < 2\Delta_{avg}$ because $i \notin HI$

  By doing the same reasoning as previously, the contribution of initial out-edges will be 1, because we do a melhorn from them and the contribution of intial in-edges will be α_i^-. Again we add the contribution of helping other nodes.

  $$\gamma_i \leq 1 + \alpha_i^- + 6\beta_i$$

  if $\beta_i = 0$, $\gamma_i \leq 2 + \alpha_i^- < 2 + 2\Delta_{avg}$ else

  $$\gamma_i < 1 + \alpha_i^- + 6(\frac{2}{3}\Delta_{avg} - \frac{1}{6}\alpha_i + 1) = 8 + 4\Delta_{avg} - \alpha_i^+$$

  But $\alpha_i^+ > 2\Delta_{avg}$ so $\gamma_i \leq 6 + 2\Delta_{avg}$

- if $i \in H \setminus (HI \cup HO)$, their will be no melhorn trees from the initial connexions because the out and in-degrees are less than $2\Delta_{avg}$. So the contribution of initial connexion will be equals to α_i. One last time we add the contribution of helping nodes.

  $$\gamma_i \leq \alpha_i + 6\beta_i$$

  because $\alpha_i < 4\Delta_i$, $\beta_i = \lceil \frac{2}{3}\Delta_{avg} - \frac{1}{6}\alpha_i \rceil > 0$ so

  $$\gamma_i < \alpha_i + 6(\frac{2}{3}\Delta_{avg} - \frac{1}{6}\alpha_i + 1)$$

  $$\gamma_i \leq 4\Delta_{avg} + 5$$

Finally if we sum up these results, we can say that this algorithm will generate a DNA N with maximum degree of $4\Delta_{avg} + 5$ instead of $12\Delta_{avg}$ in the previous version.