construct an optimum phylogeny for M_j. Therefore the proof follows by induction. ■

C. Initial Bounds

In this sub-section we bound the probability of correct guesses, analyze the running time and show how to derandomize the algorithm. We perform two guesses at Steps 2a and 2c. Lemmas 4.2 and 4.6 bound the probability that all the guesses performed at these steps are correct throughout the execution of the algorithm.

Lemma 4.2: The probability that all guesses performed at Step 2a are correct is at least 4^{-q}.

Proof: Implementation: The guess at Step 2a is implemented by selecting v uniformly at random from $\cup_i N(M_i)$.

To prove the lemma, we first show that the number of iterations of the `while` loop (step 2) is at most q. Consider any one iteration of the while loop. Since v is a non-isolated vertex of the conflict graph, $c(v)$ shares all four gametes with some other character c' in some M_j. Therefore, in every optimum phylogeny $T^*_{M_j}$ that mutates $c(v)$ exactly once, there exists a path P starting with edge e_1 and ending with e_3 both mutating c', and containing edge e_2 mutating $c(v)$. Furthermore, the path P contains no other mutations of $c(v)$ or c'. At the end of the current iteration, M_j is replaced with $M0$ and $M1$. Both subtrees of $T^*_{M_j}$ containing $M0$ and $M1$ contain (at least) one mutation of c' each. Therefore, $\texttt{penalty}(M0) + \texttt{penalty}(M1) < \texttt{penalty}(M_j)$. Since $\texttt{penalty}(I) \leq q$, there can be at most q iterations of the while loop.

We now bound the probability. Intuitively, if $| \cup_i N(M_i)|$ is very large, then the probability of a correct guess is large, since at most q out of $| \cup_i N(M_i)|$ characters can mutate multiple times in $T^*_{M_j}$. On the other hand if $| \cup_i N(M_i)| = q$ then we terminate the loop. Formally, at each iteration $| \cup_i N(M_i)|$ reduces by at least 1 (guessed vertex v is no longer in $\cup_i N(M_i)$). Therefore, in the worst case (to minimize the probability of correct guesses), we can have q iterations of the loop, with $q+1$ non-isolated vertices in the last iteration and $2q$ in the first iteration. The probability in such a case that all guesses are correct is at least

$$(\frac{q}{2q}) \times (\frac{q-1}{2q-1}) \times \ldots \times (\frac{1}{q+1}) = \frac{1}{\binom{2q}{q}} \geq 2^{-2q}$$

■