Exercise: Repeat Resolution

June 16, 2010

1 Algorithm of Tammi

Given the following multiple alignment and corresponding error rates find all pairs of DNPs with $p^{corr} \leq p^{corr}_{max} = 0.25$. Why is this threshold not appropriate?

Note: Since there are many calculation steps, it is recommended to implement an algorithm or use a spreadsheet e.g. Excel.

A	T	G	C	T	.01	.01	.001	.01	.1
C	T	A	A	A	.01	.01	.01	.01	.01
A	T	G	C	A	.001	.001	.001	.001	.01
A	T	G	C	A	.001	.001	.001	.001	.01
C	T	T	A	A	.001	.001	.001	.001	.001
C	T	C	A	A	.001	.01	.01	.01	.001
A	A	T	C	A	.01	.1	.1	.01	.01

2 Algorithm of Kececioglu

Given the following multiple alignment: The columns 2 and 5 are identified as DNPs and

C	A	T	C	A
C	A	T	C	C
C	T	G	C	T
C	A	T	C	A
C	T	T	G	T

k=2, that means the DNPs can be split into 2 groups.
Build a K_n graph and formulate an ILP for the problem. The ILP is going to have a lot of constraints. Solve it with a LP solver (e.g. soplex).