## A Class Representative Model for Pure Parsimony Haplotyping under Uncertain Data

D. Catanzaro ULB, Department of Computer Science dcatanz@ulb.ac.be

M. Labbé ULB, Department of Computer Science mlabbe@ulb.ac.be

L. Porretta ULB, Department of Computer Science lporrett@ulb.ac.be

The Pure Parsimony Haplotyping (PPH) problem is a NP-hard combinatorial optimization problem that consists of finding the minimum number of haplotypes necessary to explain a given set of genotypes. PPH has attracted more and more attention in recent years due to its importance in analysis of many fine-scale genetic data. Its application fields range from mapping complex disease genes to inferring population histories, passing through designing drugs, functional genomics and pharmacogenetics. In this article we investigate, for the first time, a recent version of PPH called the Pure Parsimony Haplotype problem under Uncertain Data (PPH-UD). This version mainly arises when the input genotypes are not accurate, i.e., when some single nucleotide polymorphisms are missing or affected by errors. We propose an exact approach to solution of PPH-UD based on an extended version of class representative model for PPH presented by Catanzaro D, Godi A, Labbé M (2009), currently the state-of-the-art integer programming model for PPH. The model is efficient, accurate, compact, polynomial-sized, easy to implement, solvable with any solver for mixed integer programming, and usable in all those cases for which the parsimony criterion is well suited for haplotype estimation.