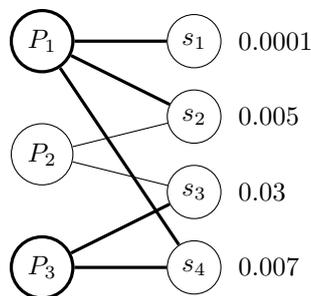


# Identification de protéines et prise en compte des peptides partagés

Sylvain Bouveret  
LIG – Grenoble INP

Thomas Burger  
CNRS – CEA /iRTSV

L'objectif de ce travail est d'aborder la problématique de l'identification de protéines à partir de peptides sous l'angle de l'optimisation combinatoire. Plus précisément, nous cherchons à proposer une modélisation de ce problème, fondée sur un ensemble d'hypothèses simplificatrices mais réalistes, permettant la prise en compte des peptides partagés entre plusieurs protéines, ainsi que les incertitudes liées à l'identification des peptides eux-mêmes. Formellement, nous modélisons dans un premier temps le problème sous la forme d'un graphe bipartite associant chaque protéine à une liste de peptides qu'elle est supposée contenir, et associant à chaque peptide une probabilité critique traduisant la confiance en l'observation et l'identification de ce peptide dans le mélange à analyser (voir la figure 1). Dans notre modèle, une *explication* est un ensemble de protéines supposées expliquer la présence de peptides observés dans le mélange. Nous associons à une explication un *score*, supposé refléter d'une part la complétude de l'explication (*i.e* le nombre de peptides expliqués par rapport au nombre de peptides observés), et d'autre part le degré de confiance de cette explication, calculé par agrégation des probabilités critiques élémentaires des peptides concernés. Selon l'opérateur d'agrégation utilisé, ce problème peut s'exprimer formellement comme un problème linéaire en nombre entiers, ce qui laisse entrevoir une possibilité de résolution efficace par des algorithmes généralistes.



Les données de l'instance :

- Protéines :  $\{P_1, P_2, P_3\}$
- Peptides :  $\{s_1, s_2, s_3, s_4\}$

Une explication possible apparaît en gras sur la figure.

FIGURE 1 – Le graphe bipartite et les probabilités critiques correspondant à un exemple d'instance de problème d'identification à trois protéines et quatre peptides,