

Apprentissage automatique basé sur l'agrégation pour la prédiction de liens dans les réseaux d'interactions protéine-protéine

Hajer AKID^{*,**}, Kirsley CHENNEN^{*}, Gabriel FREY^{*},
Julie THOMPSON^{*}, Mounir BEN AYED^{**}, Nicolas LACHICHE^{*}

^{*}ICube - UMR 7357, Université de Strasbourg

^{**}REGIM-Lab - LR11ES48, Université de Sfax

1 Problématique

Les interactions protéine-protéine (IPP) sont souvent modélisées intuitivement à l'aide d'un graphe non orienté dans lequel les noeuds représentent les protéines et les arcs reliant ces protéines représentent les interactions. Ainsi, afin de prédire de nouveaux arcs, plusieurs mesures de similarité basées sur la topologie du graphe ont été utilisées telles que l'indice des voisins communs, Adamic-Adar, l'allocation des ressources et l'attachement préférentiel. Récemment, une nouvelle mesure L3 (Kovács et al., 2019) basée sur l'existence ou non de chemins de longueur 3 entre les protéines non reliées a été introduite. Cette mesure s'est avérée plus performante en précision que les mesures de similarité basées sur la topologie. Toutefois, L3 aussi n'est pas définie pour des graphes étiquetés.

2 Approche proposée

Nous proposons une nouvelle mesure AVGMIN pour agréger les informations portées par les étiquettes de liens intermédiaires entre les protéines non directement reliés, comme dans la base de données STRING (Szklarczyk et al., 2021) où les arcs portent des scores allant de 0 à 1000 indiquant la confiance en chaque arc. Nous commençons par une première agrégation à l'aide de la fonction minimum (MIN) pour trouver le score minimal sur chaque chemin intermédiaire. Ensuite, nous appliquons la fonction moyenne (AVG) sur les minimums de tous les chemins indirects comme illustré dans la figure 1.

Nous proposons aussi d'utiliser AVGMIN, le nombre de liens intermédiaires entre les protéines non reliées, L3 et les mesures de similarité classiques comme entrée d'un algorithme d'apprentissage tel que l'arbre de modèle pour estimer les scores manquants.