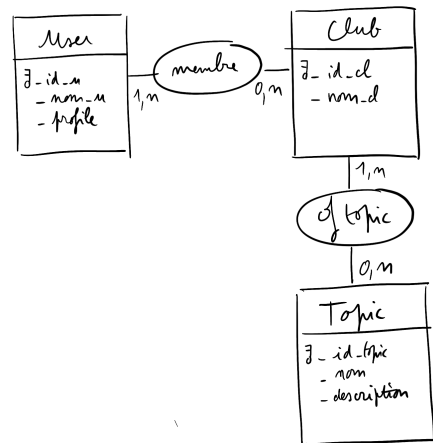


# TD Network analysis for information retrieval

Soit un réseau social, que l'on nommera  $\mathcal{Y}$ , qui permet de :

- mettre en relation des utilisateurs  $U$  via leur appartenance à des *clubs*  $C$  (autrement dit un utilisateur  $u$  appartient à au moins un club mais possiblement plusieurs),
- spécialiser les clubs en fonction d'un certain nombre de thématiques (un club peut être associé à plusieurs thématiques),
- poster des messages au format textuel sur les forums associés aux clubs.

Le modèle relationnel de la base de données ressemble à ceci (on omet ici la table des messages) :



On considère que les relations **membre** et **oftopic** sont des relations binaires (par ex., un utilisateur fait partie d'un club ou bien non). Elles sont encodées dans deux matrices :

- $UC$ , telle que  $UC[u,c]=1$  si l'utilisateur  $u \in U$  est membre du club  $c \in C$ ,
- $CT$ , telle que  $CT[c,t]=1$  si le club  $c$  est associé à la thématique  $t \in T$ .

A partir de ce modèle des données, on vous demande d'élaborer un prototype d'algorithme de recommandation automatique afin de mettre en place un réseau social qui n'existe pas encore, encodé par une nouvelle relation **ami**. Vous répondrez aux questions suivantes qui doivent vous guider vers l'élaboration de ce prototype.

1. Tout d'abord, on vous demande de proposer un algorithme **reco1** qui permet de recommander à un utilisateur  $u \in U$  de se connecter comme ami avec  $v \in U$  s'ils suivent des clubs en commun. Cet algorithme devra utiliser une matrice carrée d'affinité  $S_c$  dont vous préciserez le calcul en vous inspirant des mesures de proximité vues en cours.
2. On vous précise à présent que les clubs qui possèdent trop d'abonnés doivent être pénalisés dans la mesure de l'affinité entre deux utilisateurs. Indiquez comment calculer le nombre d'abonnés des clubs à partir de  $UC$ , puis comment modifier la mesure de l'algorithme précédent afin de prendre cette nouvelle contrainte en compte.
3. On souhaite enrichir l'algorithme de recommandation en rapprochant les utilisateurs sur la base des thématiques qui les intéressent le plus. Comment pouvez-vous calculer automatiquement une nouvelle matrice d'affinité  $UT$ , telle que  $UT[u,t]$  reflète l'intérêt qu'un utilisateur  $u$  porte à la thématique  $t$ , bien sûr via les clubs auxquels il est abonné? Vous préciserez bien le calcul et comment interpréter la valeur  $UT[u,t]$ . A partir de cette nouvelle matrice, donnez l'algorithme de recommandation **reco2** permettant de rapprocher un utilisateur  $u$  à un utilisateur  $v$ , basé sur cette affinité thématique.
4. Indiquez comment, à partir des exercices précédents, vous pouvez construire un graphe d'utilisateurs  $G_T$  qui encode l'information structurelle (i.e.,  $u$  et  $v$  sont connectés s'ils sont proches en terme de thématique, via  $UT$ ) et ajoute l'information textuelle attachée au profil des utilisateurs (champ enregistré dans **profile** de la table **User**).

5. A partir du graphe  $G_T$ , écrivez un algorithme qui permet de calculer des nouvelles représentations  $z_u$  des utilisateurs, représentations qui prendront en compte à la fois la proximité thématique mais également les profils similaires. Cet algorithme doit suivre le paradigme des Graph Convolutional Networks (GCN) et calculer des représentations d'ordre 3 (càd prenant en compte un voisinage de profondeur 3). Vous préciserez bien comment vous encodez chacune des informations et comment vous les utilisez pour calculer les  $z$ .
6. Expliquez brièvement comment vous pouvez mettre au point un algorithme de recommandation **reco3** qui prend en compte ces représentations latentes pour résoudre le problème.