

# Flixster data

Raphaël Fournier-S'niehotta, Tiphaine Viard

Links graph  $L = (V_L, E_L)$  : social network between users

- $|V_L| = 786936$ ,
- $|E_L| = 5897324$ ,
- $\delta(L) = 1.9 \cdot 10^{-5}$ .
- Une composante connexe géante

Bipartite ratings graph  $R = (\top_R, \perp_R, E_R)$  : ratings of movies by users

- $|\top_R| = 147612$  (users),
- $|\perp_R| = 48794$  (movies),
- $|E_R| = 8196065$ ,
- $\delta(R) = 2.2 \cdot 10^{-3}$ ,
- Une composante connexe géante.

Number of users in social network : 786936

Number of users in ratings : 147612

Number of users in friends but not ratings  $|V_L \cap \top_R|$  : 147335

En retirant les nœuds qui n'ont aucun rating, il reste 137925 nœuds; Il y a environ 10000 nœuds qui ont des ratings mais sont connectés à des nœuds qui n'ont aucun rating.

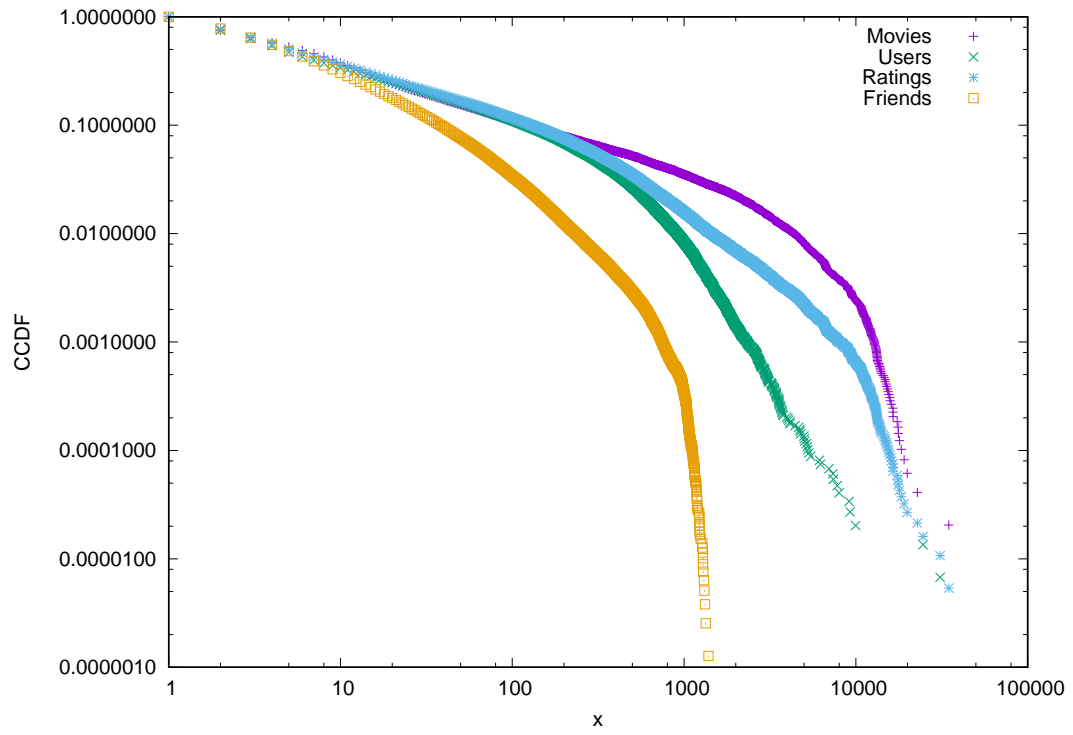


Figure 1: Inverse cumulative degree distributions.

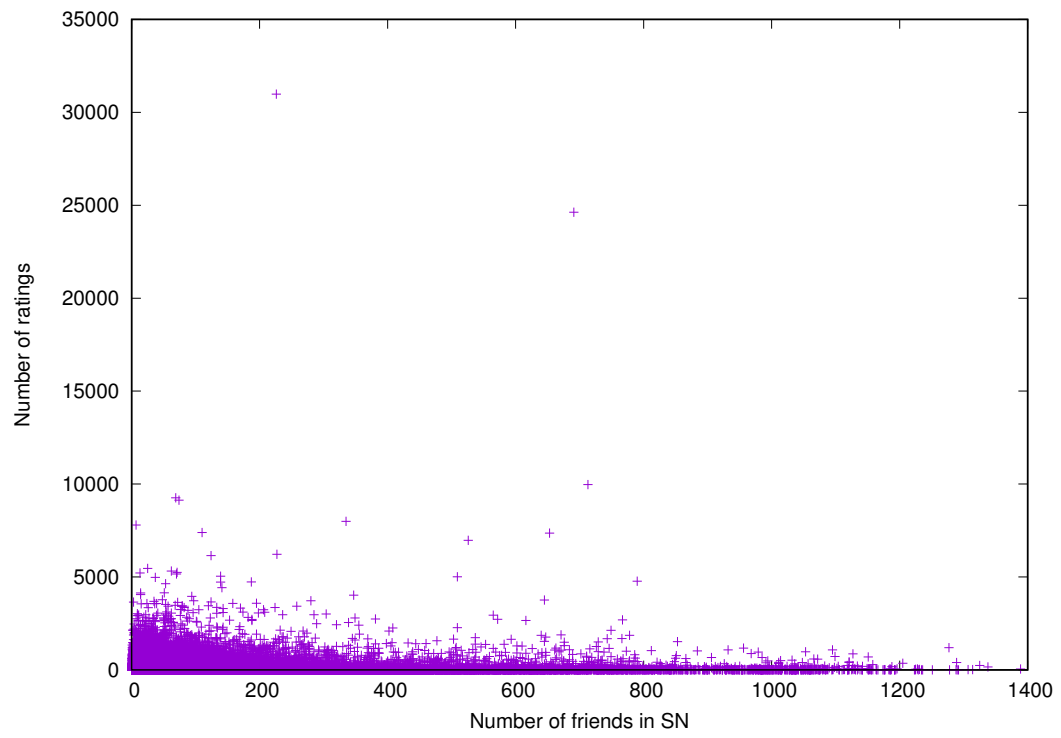


Figure 2: Inverse cumulative degree distributions.

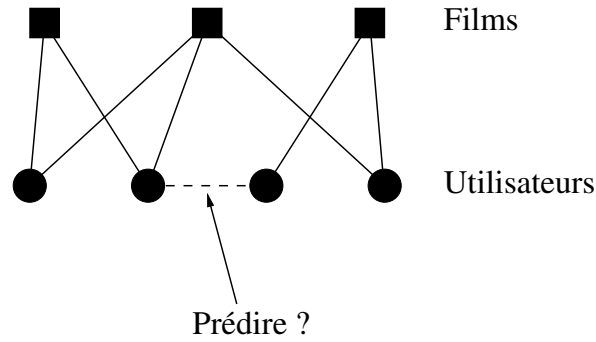


Figure 3:

## 1 Prédiction de liens

Peut-on prédire les liens du réseau social en utilisant des méthodes de prédiction de liens sur le biparti (voir figure 1).

Ou alors, peut-on faire de la prédiction de liens sur le biparti en pondérant par l'existence ou non d'un lien dans le réseau social, voir figure 1 ?

On cherche à prédire la note que  $u$  donnerait à  $F$ ,  $\hat{S}_F(u)$ .

$$\hat{S}_F(u) = \sum_{v \in N(F)} \alpha_{uv} \cdot \beta_{uv} \cdot S_F(v)$$

où  $S_F(v)$  est la note donnée par  $v$  pour le film  $F$ ,  $N(F)$  est le voisinage du film  $F$ ,

$$\alpha_{uv} = \begin{cases} 1 & \text{si } (u, v) \in E_L \\ 0 & \text{sinon} \end{cases}$$

et

$$\beta_{u,v} = \text{Jaccard?}$$

Quelles sont les conditions limites ? On veut peut-être plutôt  $\text{argmax}(\alpha_{uv} \cdot \beta_{uv})$ , pour éviter de recommander des films très mal notés par  $v$  ? On peut voir  $\alpha_{uv} \cdot \beta_{uv}$  comme un terme de confiance de la recommandation.