

DumboAsAService

PredictLikes: Blog like-profil klaszterezés

Nótai István (notaiistv@gmail.com)

Szakállas Dávid(david.szakallas@gmail.com)

Mátyás-Barta Csongor(mbcsongor@gmail.com)

2015. december 8.

PredictLikes - blogok "like-profil" alapján klaszterezése és a csoportosítás vizualizációja:

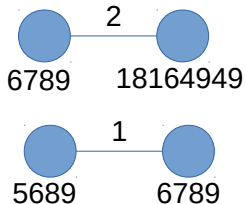
Like profil a mi értelmezésünkben = Ha két blognak van közös likeolója, akkor ők hasonló blogok.

```
{ "uid": "34168956",  
  "likes":  
    [ { "blog": "18164949" },  
      { "blog": "6789" } ] }
```

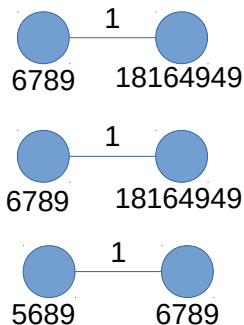


MapReduce - Éllista építés

Reduce



Map

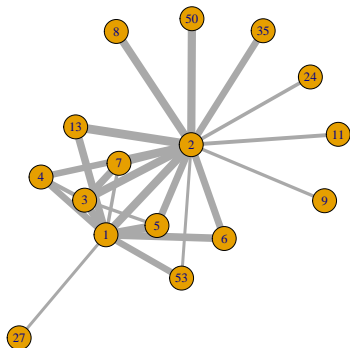


Community Search

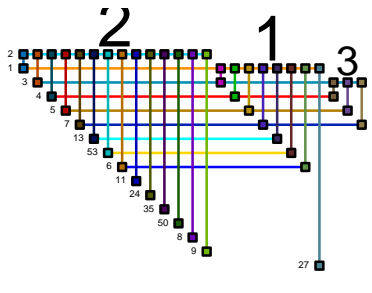
Végeredmény és vizualizáció

Közel 7000 csomópont és 700k él

igraph



BioFabric



Éllista építés

- local Hadoop
MapReduce(Java)
- 15GB RAM

Gráf feldolgozás

- R
- igraph R csomag [1]
- BioFabric R port [2]

Köszönjük a figyelmet!

Hivatkozások

- 1 <http://igraph.org/r/>
- 2 <https://github.com/wjrl/RBioFabric/blob/master/R/bioFabric.R>