

Viikkoraportti 3

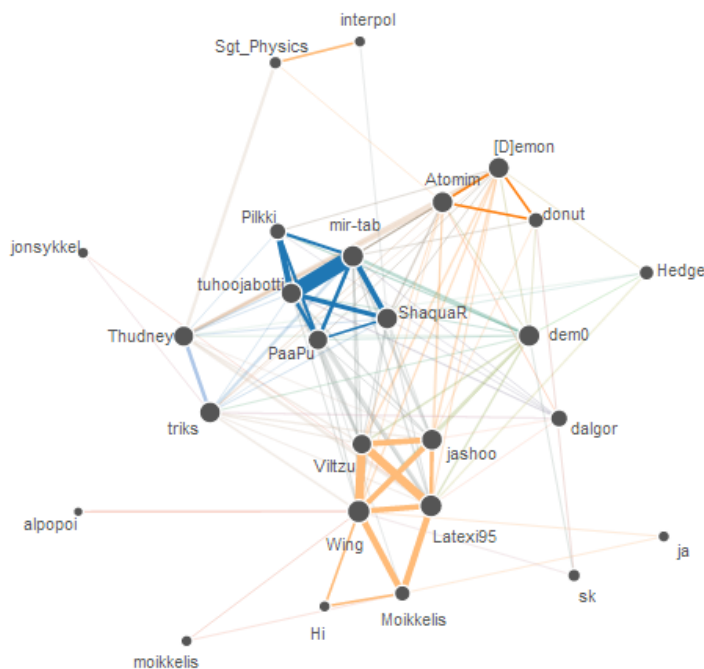
Ryvästäminen (eh, klusterointi)

Päätin toteuttaa Markov Cluster -algoritmin (MCL). Lähde: https://www.cs.ucsb.edu/~xyan/classes/CS595D-2009winter/MCL_Presentation2.pdf

Algoritmi perustuu siihen, että tehdään satunnaisia reittejä verkossa käyttäen Markovin ketjuja. Tällöin on todennäköistä, että reitti pysyy klusterin sisällä, koska klusterin sisällä on enemmän kaaria, kuin klustereiden välillä: *“By doing random walks upon the graph, it may be possible to discover where the flow tends to gather, and therefore, where clusters are”*

Algoritmi käyttää verkon matriisiesitystä ja laskee tilasiirtymien todennäköisyydet siihen. Kunnes lopulta päädytään lopputilaan.

Aluksi algoritmin tuloksen tulkinnassa oli hieman epäselvyyttä, joka aiheutti vääriä visualisaatioita (Ei tullut mieleen ottaa kuvakaappausta). Lopputulos: <http://lahdenvuo.info/social/clusters/cb2/10.json>



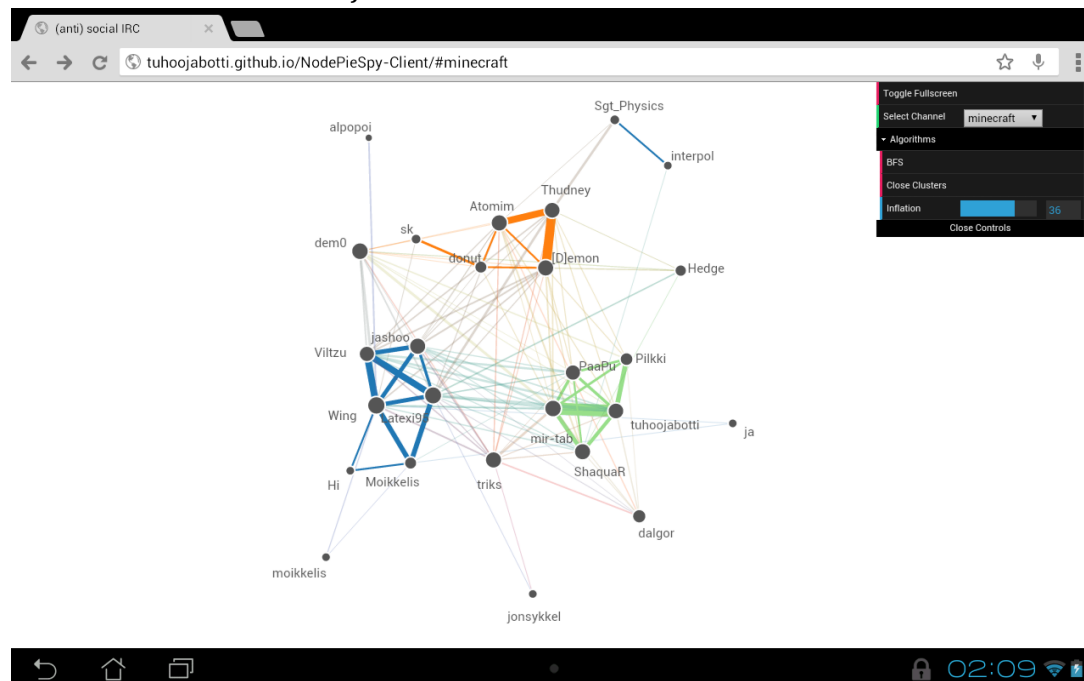
12.1.2014
Ville Lahdenvuo
Viikkoraportti

Refaktorointia

Refaktoroin Analyser-luokkaa ja irroitin itse verkon siitä omaksi luokaksi. Tämä selkeyttää algoritmejä, koska niille voi antaa suoraan verkon parametrina.

Mobiilitesti

Päätin kokeilla miten ohjelma toimii mobiiliselaimella:



Käyttöliittymää lukuunottamatta näyttää toimivan vallon mainiosti, vaikka mitään mobiilioptimoitinta ei ole tehty.

Optimointia

Ryvästämisessä kesti verkon koosta riippuen hyvin pitkä aika. Huomasin vian algoritmin loppumisen tarkistamisessa, jonka takia algoritmia suoritettiin aina maksimi-iteraatiomäärä, vaikka tulos olikin jo valmis. Tämän korjauksen jälkeen klusterointi toimii käytännössä silmänräpäyksessä. Suurimallakin syötteellä suoritusaika pysyy alle yhden sekunnin.

Epärealistisellakin 10 000 rivin syötteellä suurimman kanavan klusterointi onnistui noin 15 sekunnissa.

12.1.2014
Ville Lahdenvuo
Viikkoraportti

Esimerkkidata

Kopioin hieman esimerkkidataa reposiion example-kansioon ja lisäsin käyttöohjeeseen ohjeet, miten esimerkkipalvelin käynnistetään ja miten GUI asennetaan.