Visualizing Wikipedia Communities using Gephi

A closer look at our communities

Haitham Shammaa
April, 16, 2014
Wiki communities

What we do know…
- Number of edits by users
- Total number of users/editors/active editors
- Number of users by countries/language
… etc (basically individual stats)

What we don’t know …
- Community leaders (user centrality)
- Interaction between users (Community health)
How do we do that?

“Simplicity is the ultimate sophistication.”
– Leonardo da Vinci
A random Research Community
(that exists in one of the parallel universes)

Dario  Aaron  Leila  Oliver
Jaime  Erik  Haitham  Jonathan  Bizzario
Example: emails, in and out!

* most of our data mining work so far stopped at this dimension
Who emails whom?

Cool, but not very informative!
Layout using graph energy

Erik
in: 70
out: 45

Leila
in: 15
out: 7

Aaron
in: 25
out: 28

Dario
in: 100
out: 3

Oliver
in: 60
out: 130

Jonathan
in: 45
out: 40

Jaime
in: 30
out: 25

Haitham
in: 40
out: 20

in: 100
out: 3
How many emails exchanged between each two?
Coloring by # of in & out emails

- Dario
  - in: 100
  - out: 20

- Aaron
  - in: 25
  - out: 28

- Oliver
  - in: 60
  - out: 130

- Erik
  - in: 70
  - out: 45

- Leila
  - in: 15
  - out: 7

- Jonathan
  - in: 45
  - out: 40

- Jaime
  - in: 30
  - out: 25

- Haitham
  - in: 40
  - out: 20
How do we do this on Wikipedia?