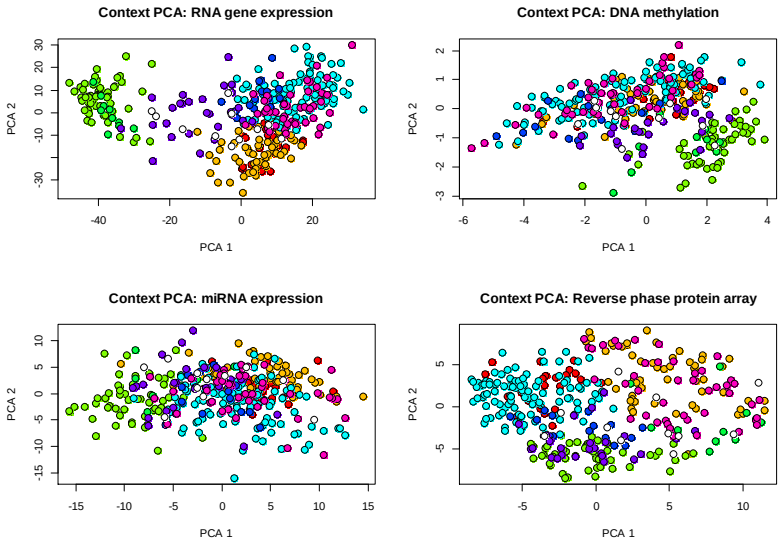
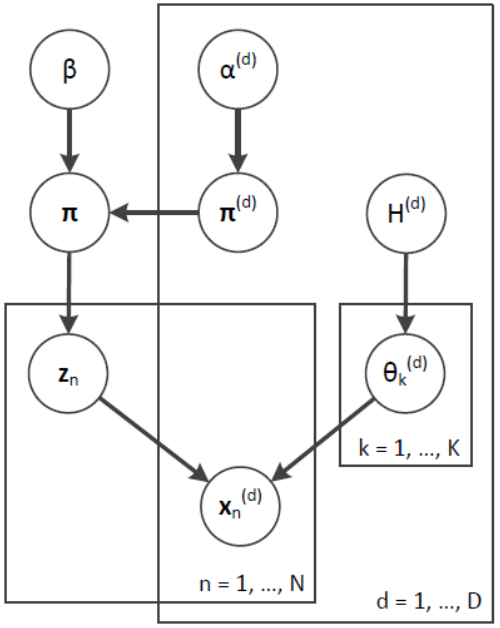
Understanding social networks with F#

Evelina Gabasova
@evelgab

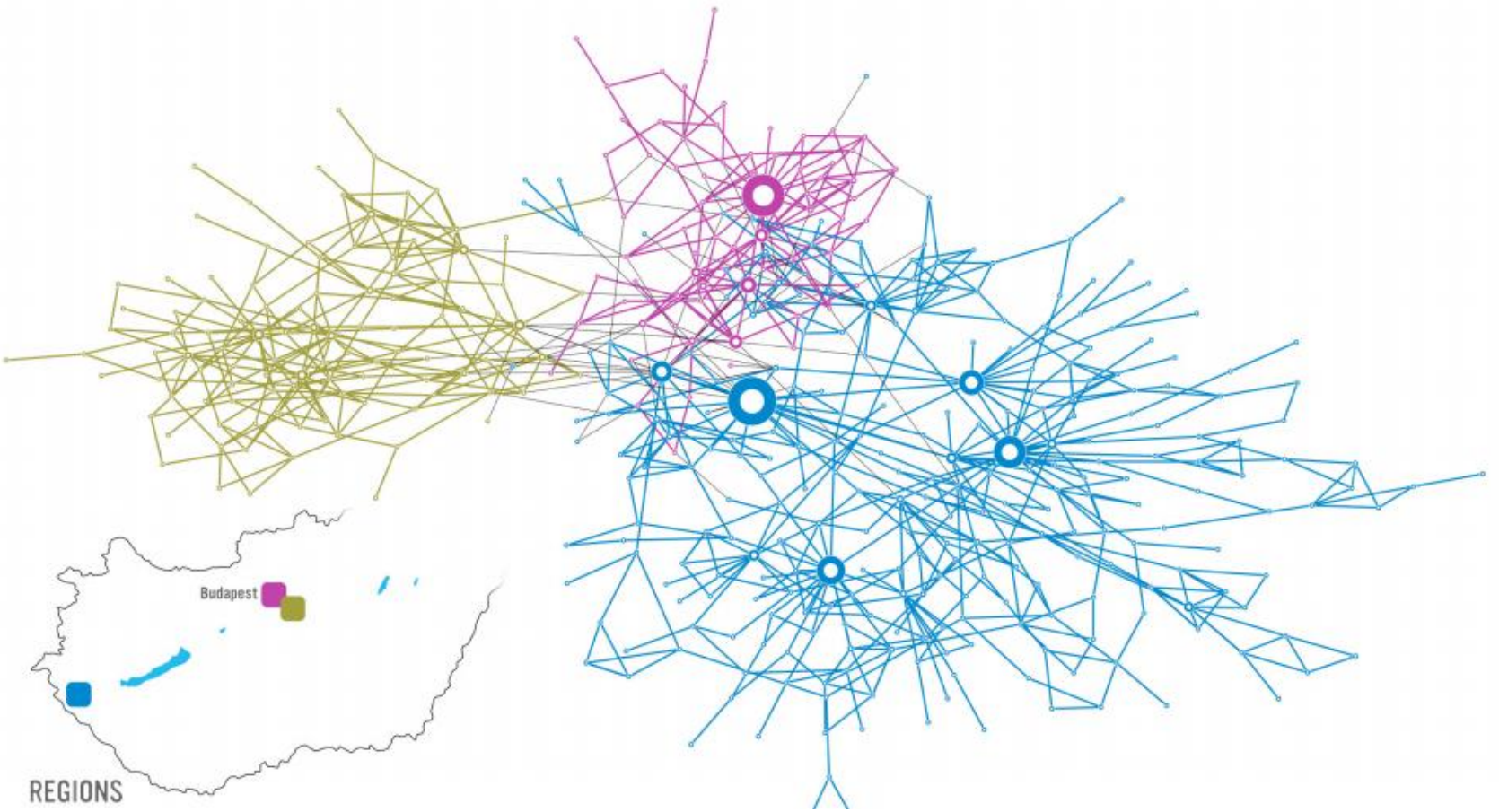
evelinag.com
@evelgab



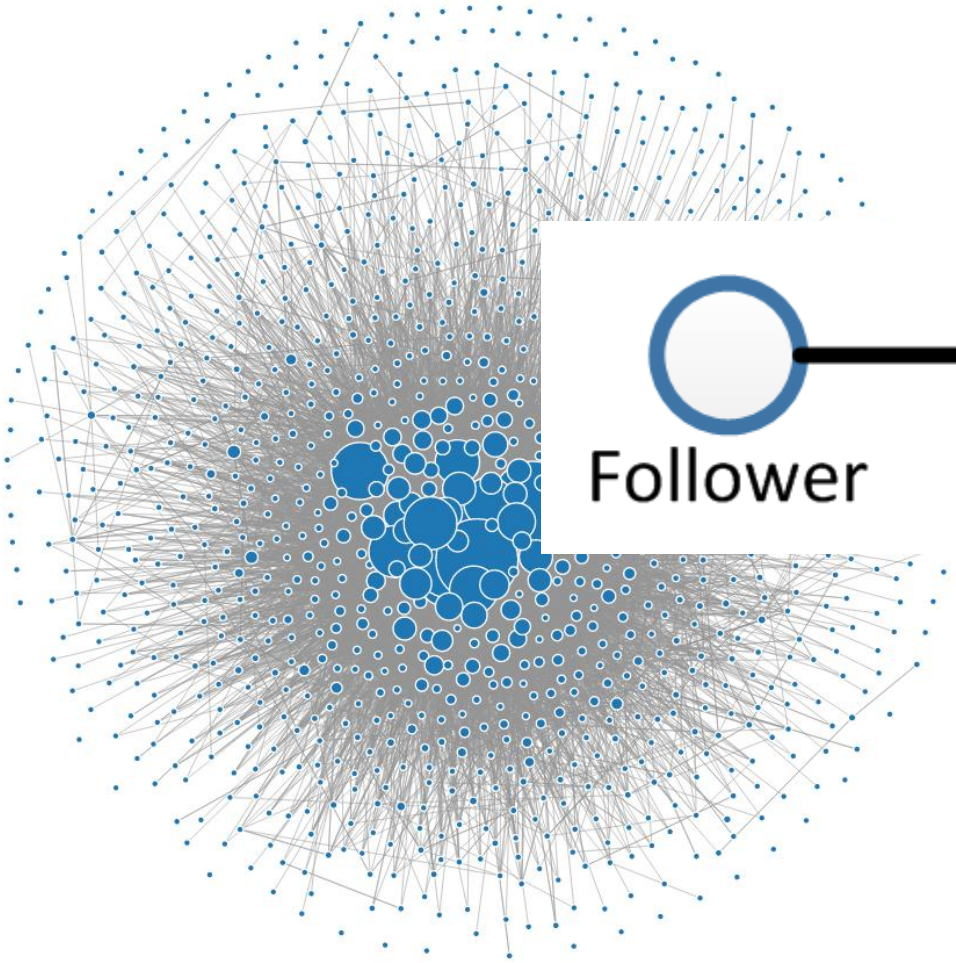
$$
\begin{aligned}
& p\left(\boldsymbol{\pi}^{(d)} \middle| \alpha_d, \beta, \boldsymbol{\pi}\right) \\
& \quad \propto p\left(\boldsymbol{\pi}^{(d)} \middle| \alpha_d\right)\ p\left(\boldsymbol{\pi} \middle| \boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(D)}, \beta\right) \\
& \quad = \left\{ \frac{\Gamma\left(K^{(d)}\alpha_d\right)}{\left(\Gamma(\alpha_d)\right)^{K^{(d)}}} \prod_{k=1}^{K^{(d)}} \left(\pi_k^{(d)}\right)^{\alpha_d - 1} \right\} \times \\
& \qquad \times \left\{ \frac{\Gamma\left(\sum_{i1=1}^{K^{(1)}} \cdots \sum_{iD=1}^{K^{(D)}} \left[\beta \times \pi_{i_1}^{(1)} \times \cdots \times \pi_{i_D}^{(D)}\right]\right)}{\prod_{i1=1}^{K^{(1)}} \cdots \prod_{iD=1}^{K^{(D)}} \Gamma\left(\beta \times \pi_{i_1}^{(1)} \times \cdots \times \pi_{i_D}^{(D)}\right)} \prod_{i_1=1}^{K^{(1)}} \cdots \prod_{i_D=1}^{K^{(D)}} \left(\boldsymbol{\pi}_{i_1,\ldots,i_D}\right)^{\beta \pi_{i_1}^{(1)} \ldots \pi_{i_D}^{(D)} - 1} \right\} \\
& \quad = \frac{\Gamma\left(K^{(d)}\alpha_d\right)}{\left(\Gamma(\alpha_d)\right)^{K^{(d)}}} \times \frac{\Gamma(\beta)}{\prod_{i1=1}^{K^{(1)}} \cdots \prod_{iD=1}^{K^{(D)}} \Gamma\left(\beta \times \pi_{i_1}^{(1)} \times \cdots \times \pi_{i_D}^{(D)}\right)} \times \\
& \qquad \times \left\{ \prod_{k=1}^{K^{(d)}} \left(\pi_k^{(d)}\right)^{\alpha_d - 1} \right\} \left\{ \prod_{i_1=1}^{K^{(1)}} \cdots \prod_{i_D=1}^{K^{(D)}} \left(\boldsymbol{\pi}_{i_1,\ldots,i_D}\right)^{\beta \pi_{i_1}^{(1)} \ldots \pi_{i_D}^{(D)} - 1} \right\} \\
& \quad \propto \frac{\Gamma\left(K^{(d)}\alpha_d\right)}{\left(\Gamma(\alpha_d)\right)^{K^{(d)}}} \left\{ \prod_{k=1}^{K^{(d)}} \left(\pi_k^{(d)}\right)^{\alpha_d - 1} \right\} \left\{ \prod_{i_1=1}^{K^{(1)}} \cdots \prod_{i_D=1}^{K^{(D)}} \frac{1}{\Gamma\left(\beta \times \pi_{i_1}^{(1)} \times \cdots \times \pi_{i_D}^{(D)}\right)} \left(\boldsymbol{\pi}_{i_1,\ldots,i_D}\right)^{\beta \pi_{i_1}^{(1)} \ldots \pi_{i_D}^{(D)} - 1} \right\}
\end{aligned}
\tag{18}
$$

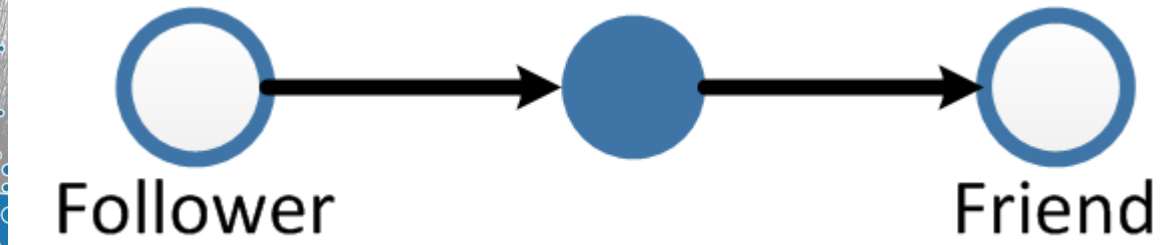


Insights from social networks

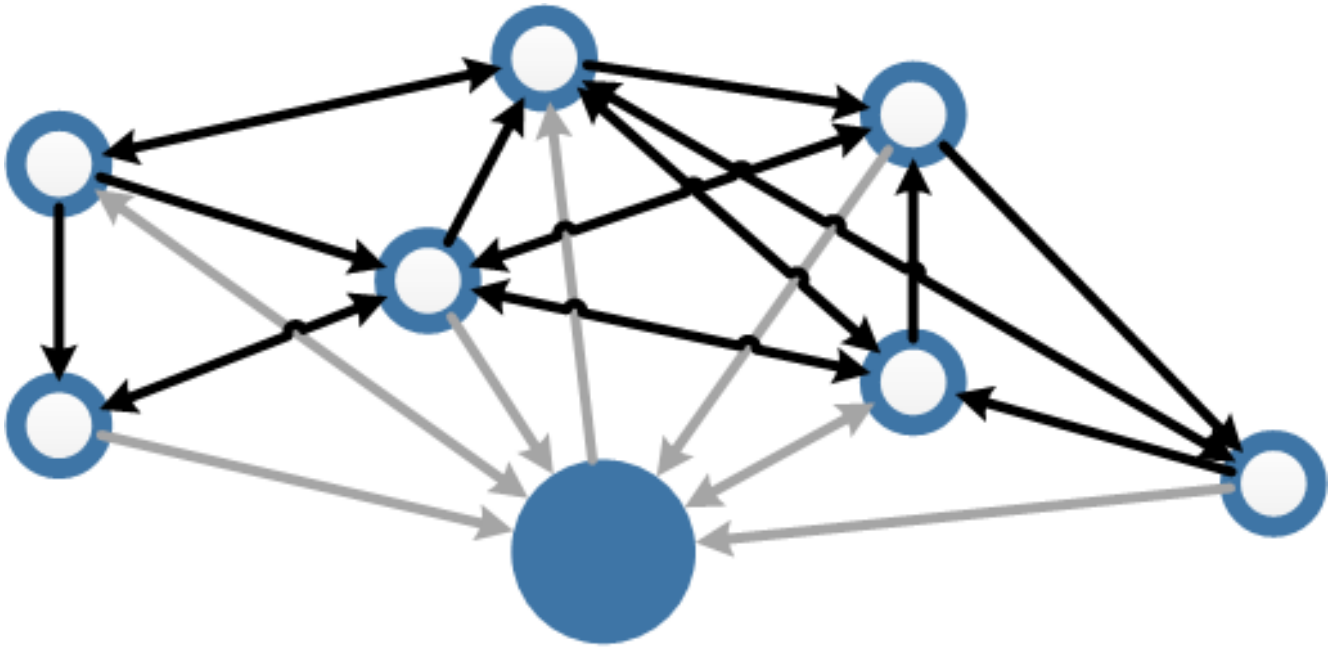


Twitter network





How large is your ego?



@fsharporg

Connecting to twitter

Interaction rates

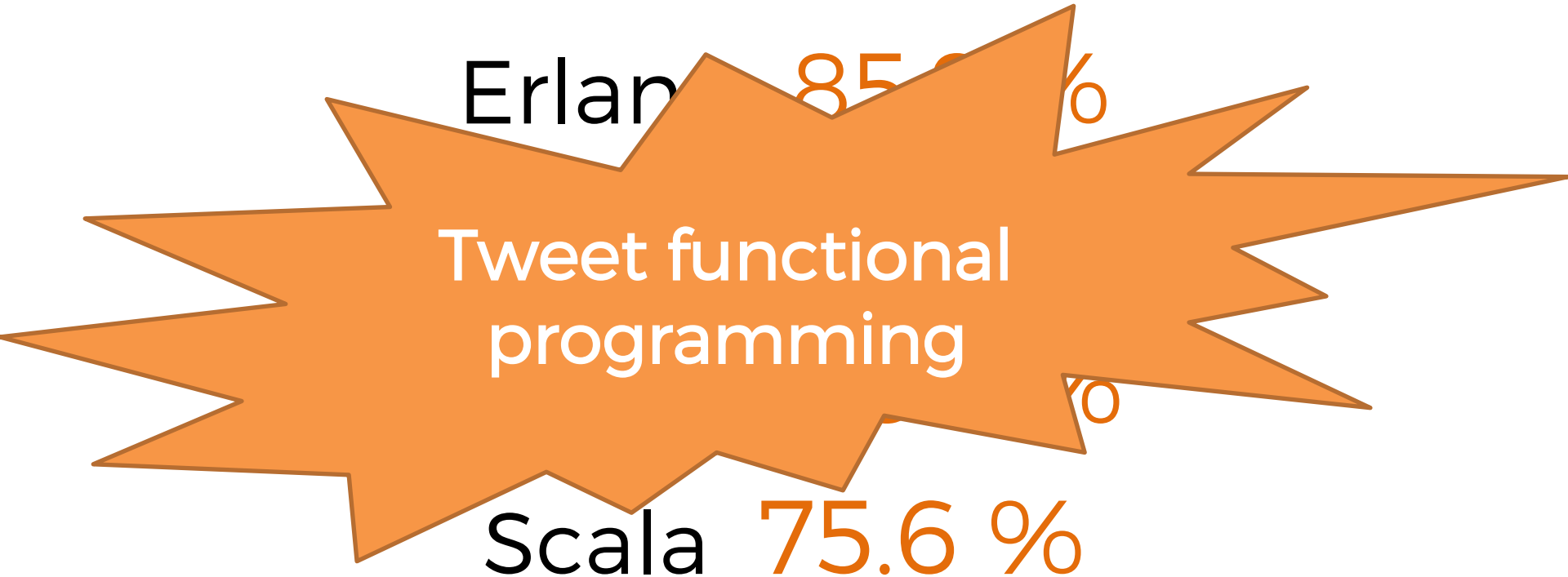
F#	86.1 %
Erlang	85.9 %
Clojure	82.7 %
Haskell	75.7 %
Scala	75.6 %

Interaction rates

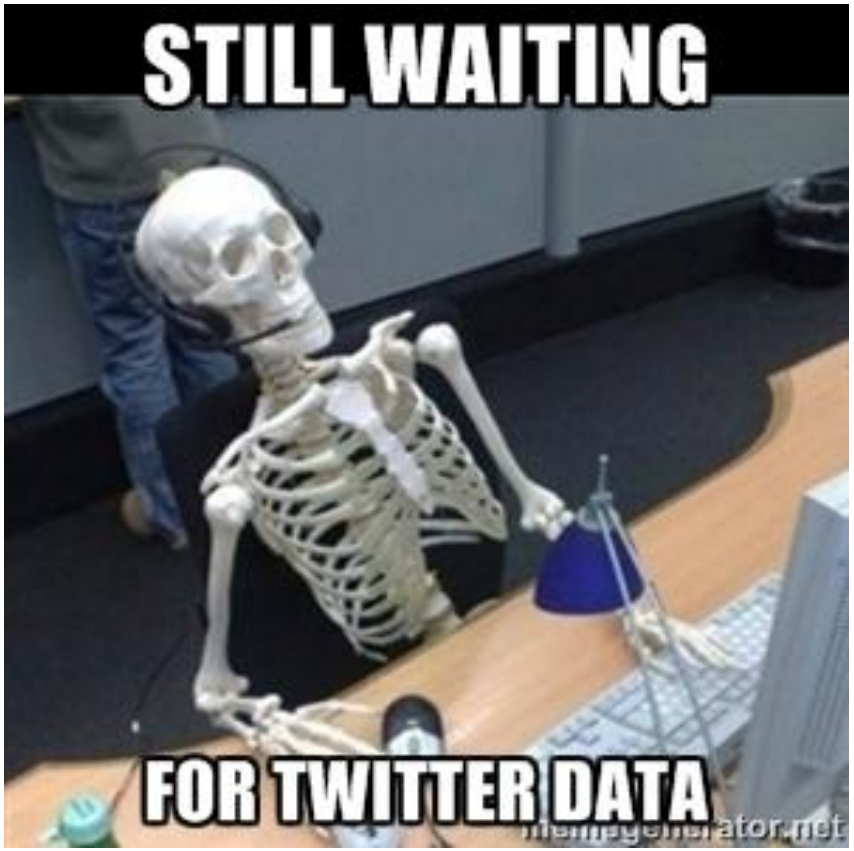
F# 86.1 %

Erlan 85.[illegible] %

Scala 75.6 %



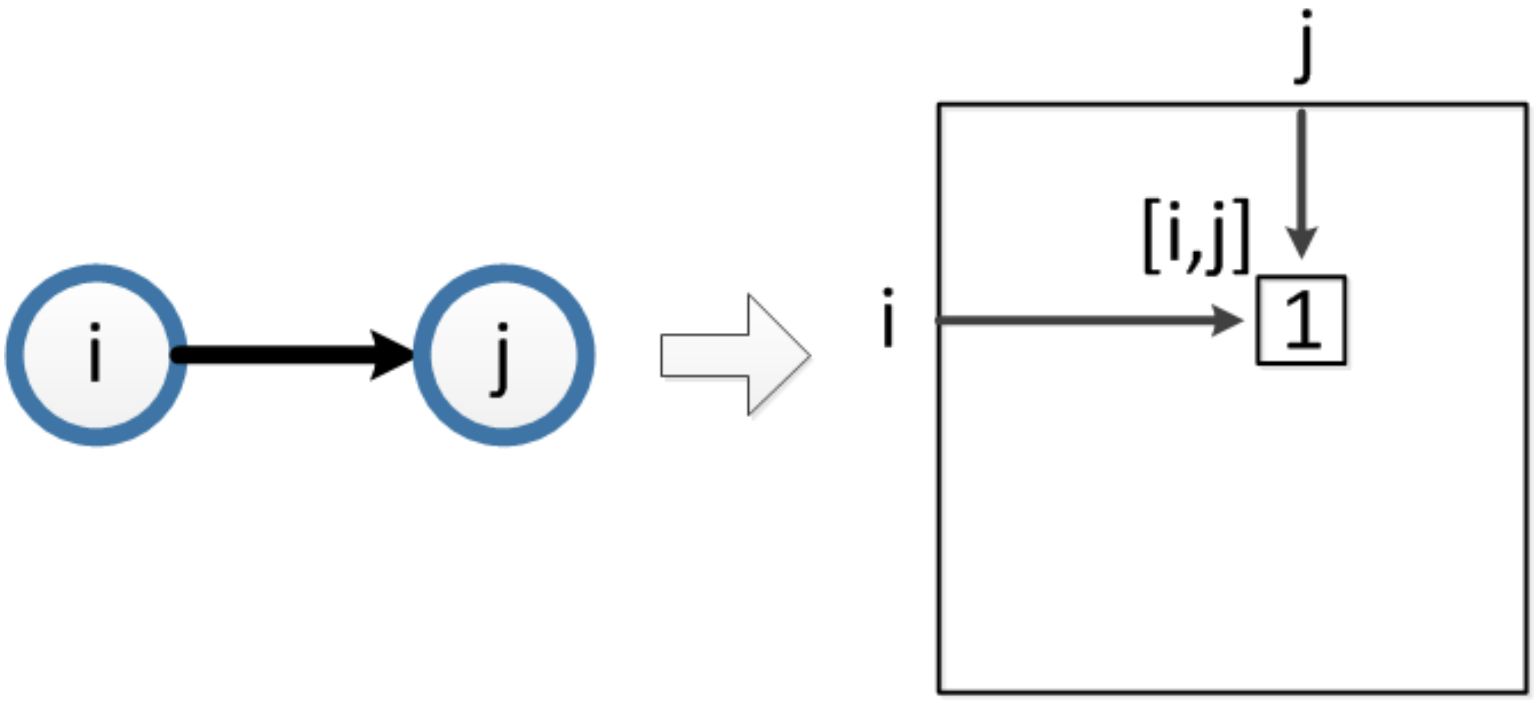
Downloading data



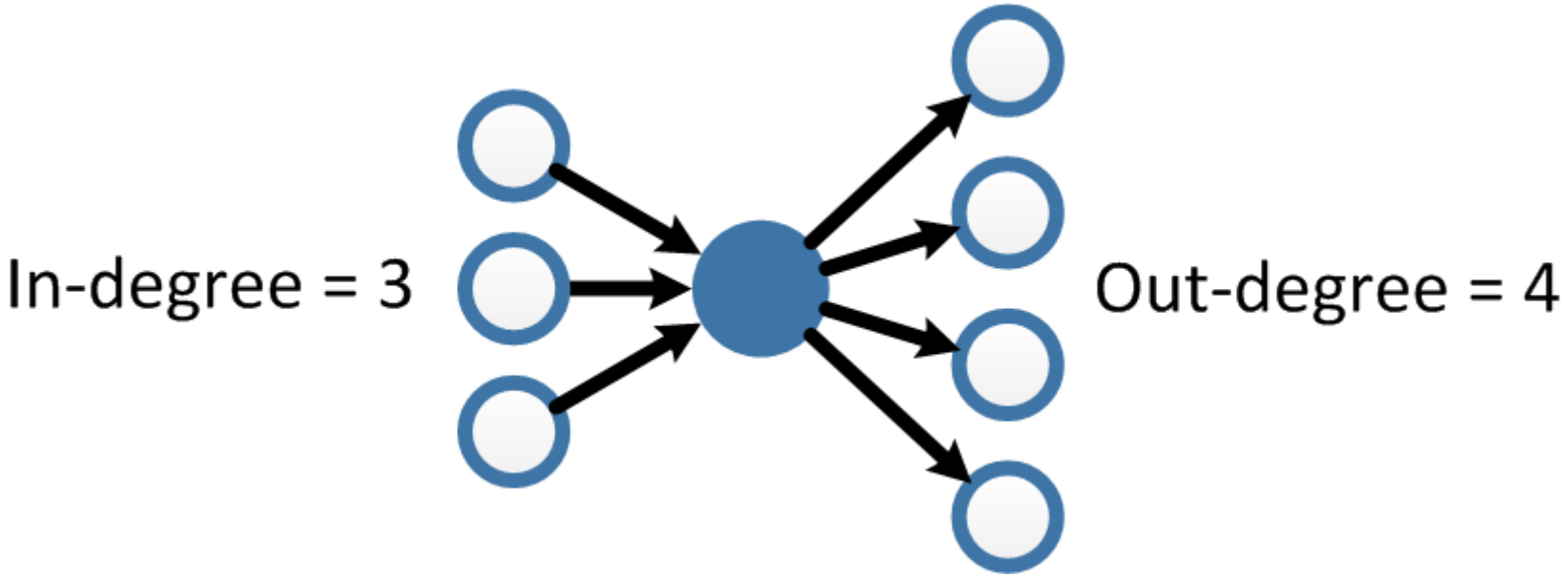
1) List of nodes
2) Connections between nodes

Twitter API allows only 15 requests every 15 minutes to list connections.

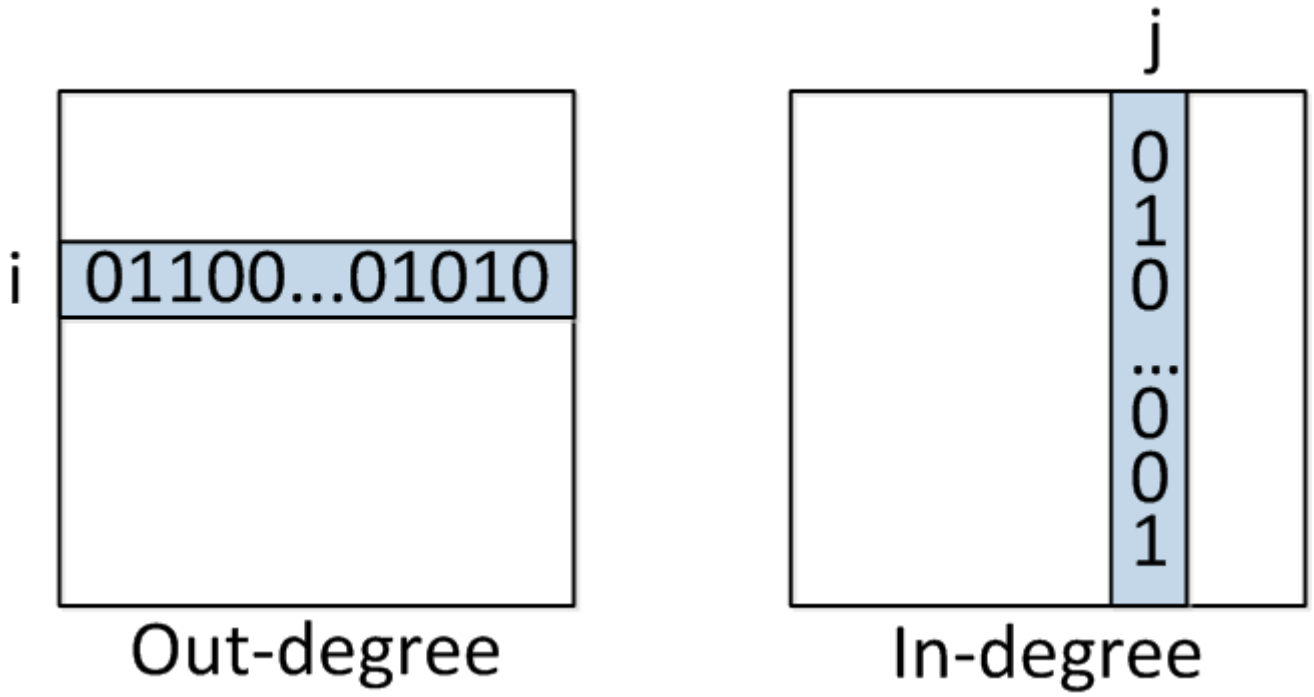
Adjacency matrix



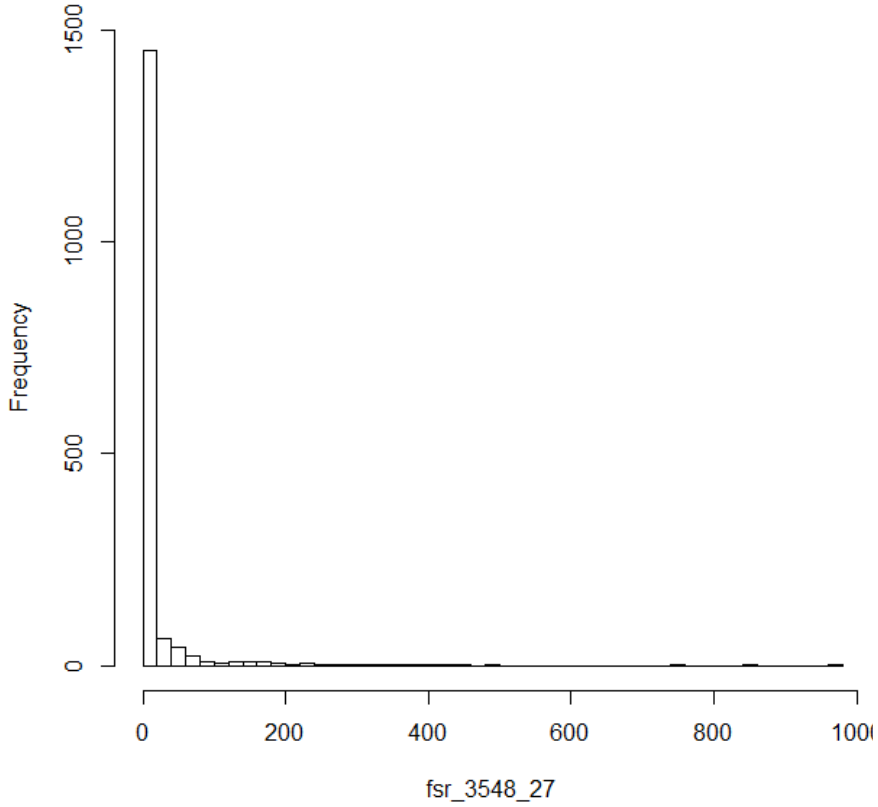
Degrees

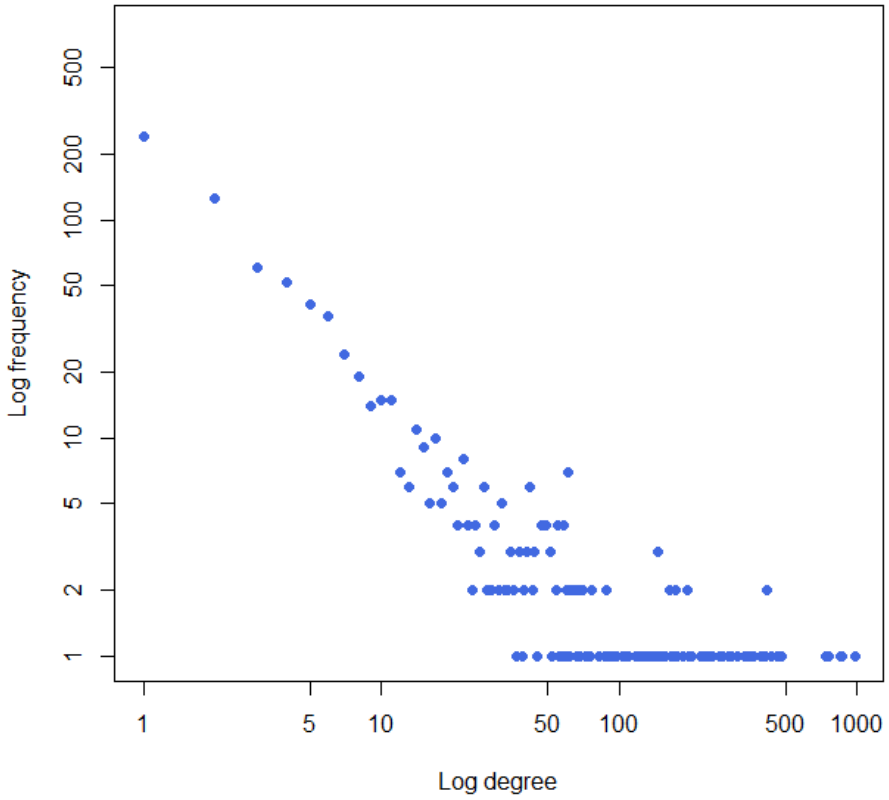


Degrees



Degree distribution





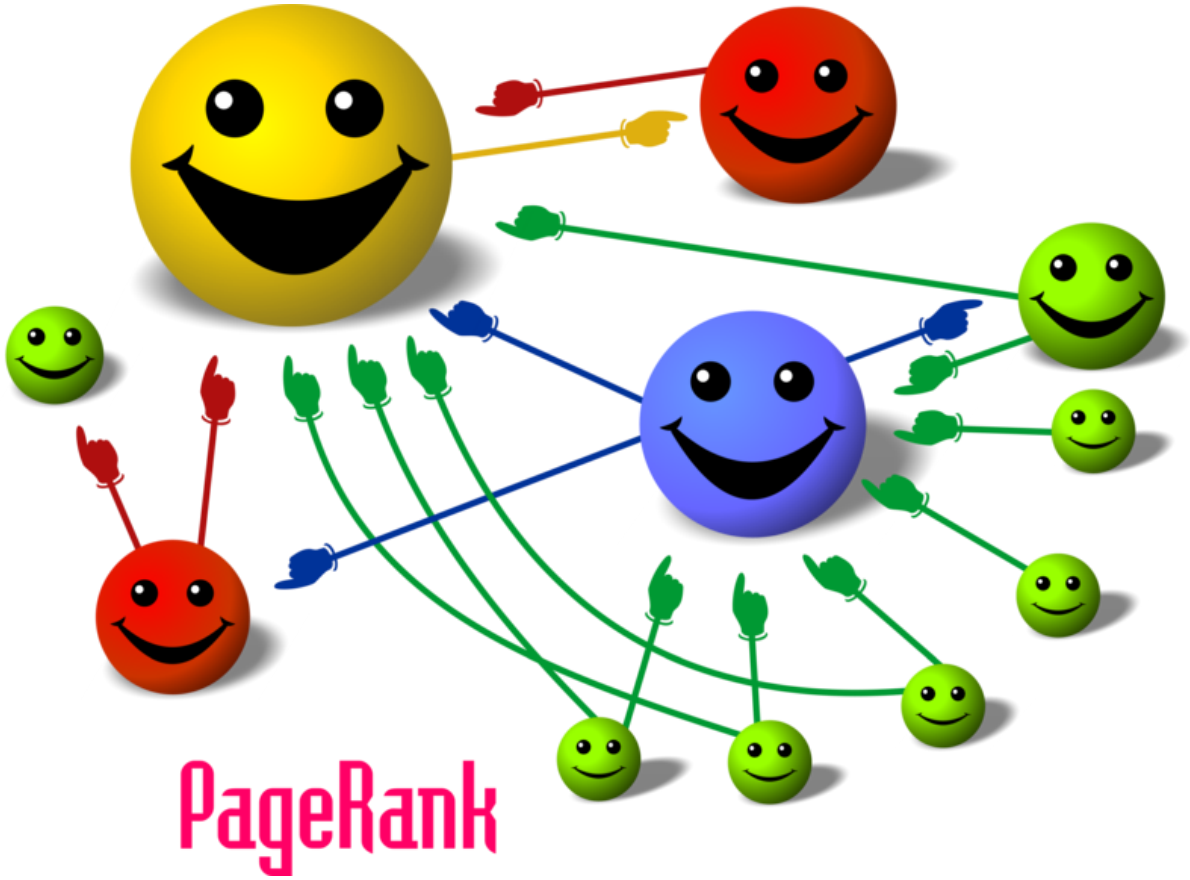
Hubs in networks

How to identify most

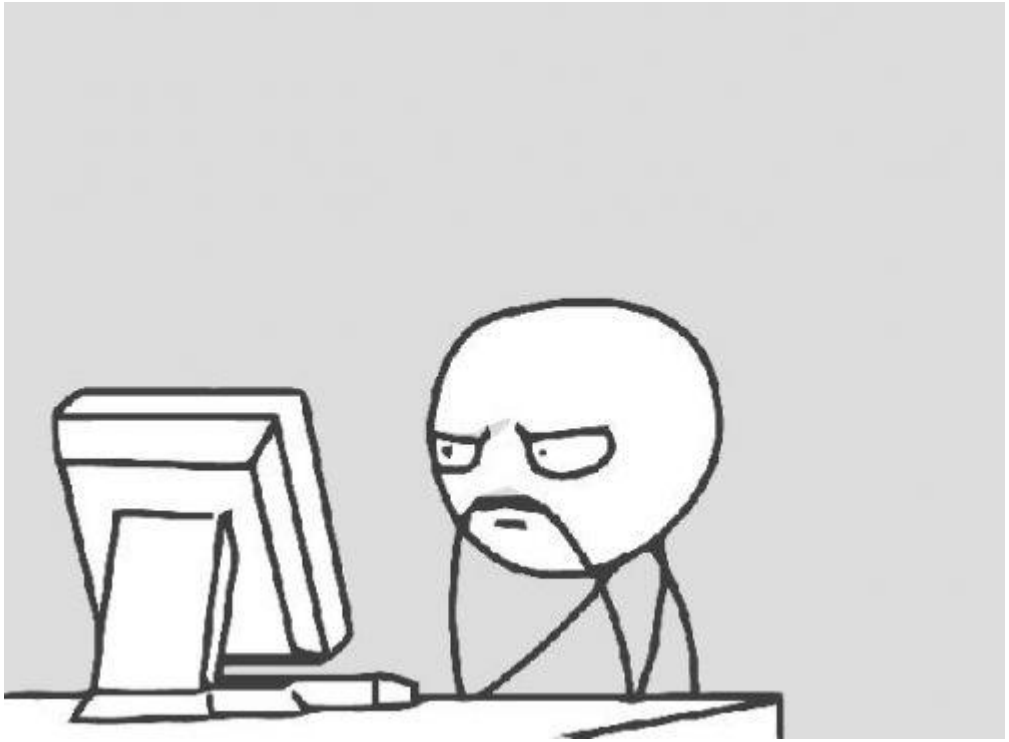
important nodes in a network?

Centrality with PageRank

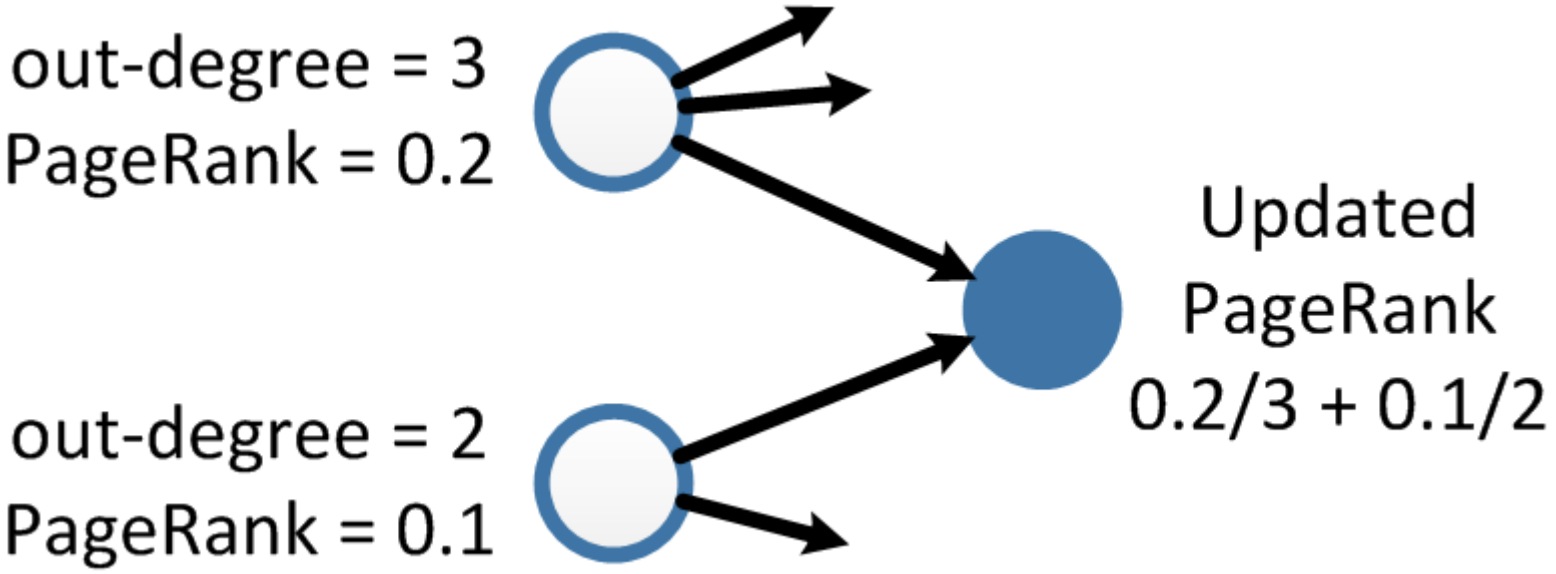
Your followers are not created equal.



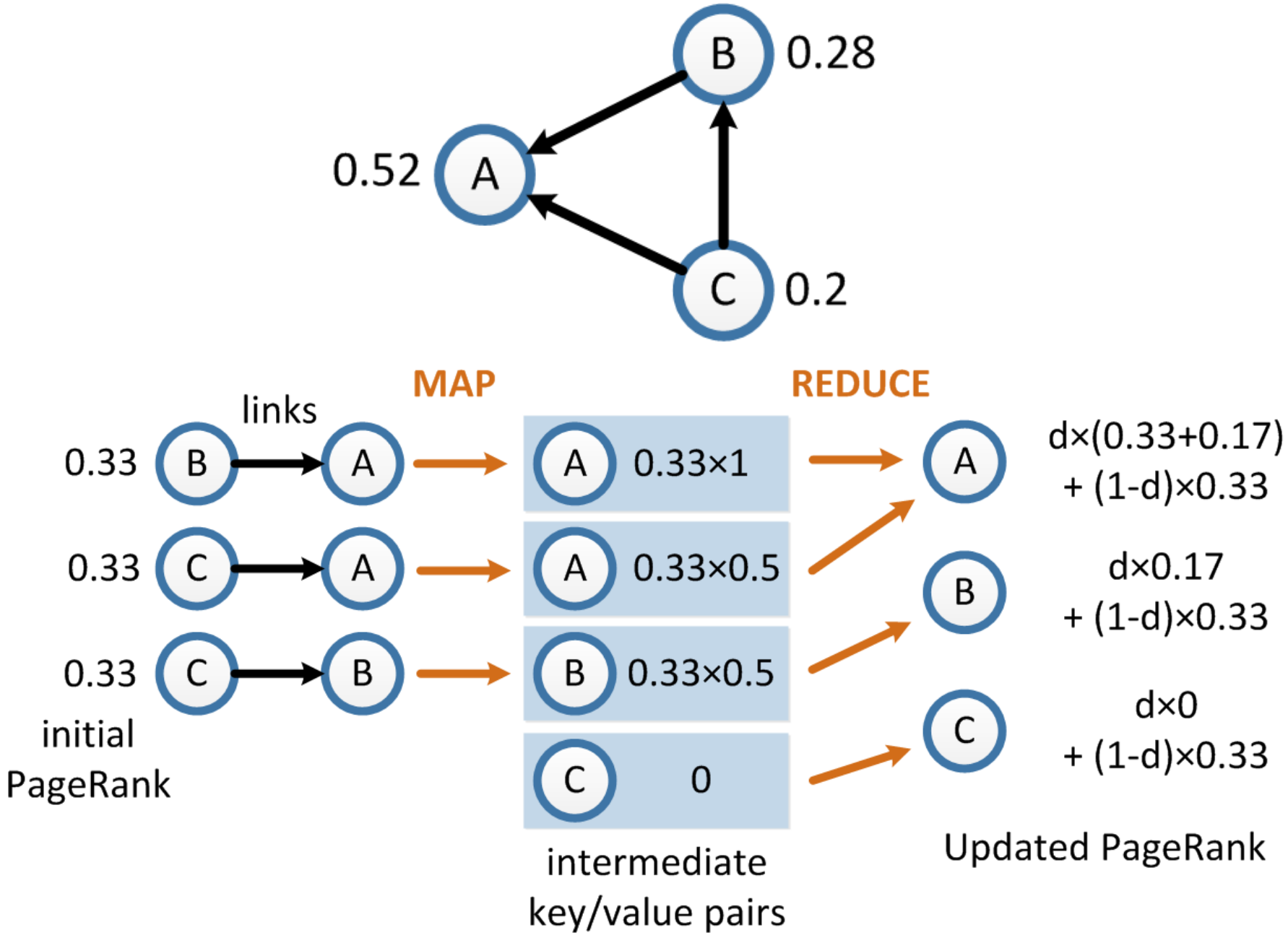
Random surfer model



Centrality with PageRank



+ random jumps


B
0.28
0.52
A
C
0.2
MAP
REDUCE
links
0.33
B
A
A
0.33×1
A
d×(0.33+0.17)
+ (1-d)×0.33
0.33
C
A
A
0.33×0.5
B
d×0.17
+ (1-d)×0.33
0.33
C
B
B
0.33×0.5
C
d×0
+ (1-d)×0.33
initial
PageRank
C
0
intermediate
key/value pairs
Updated PageRank

PageRank changes

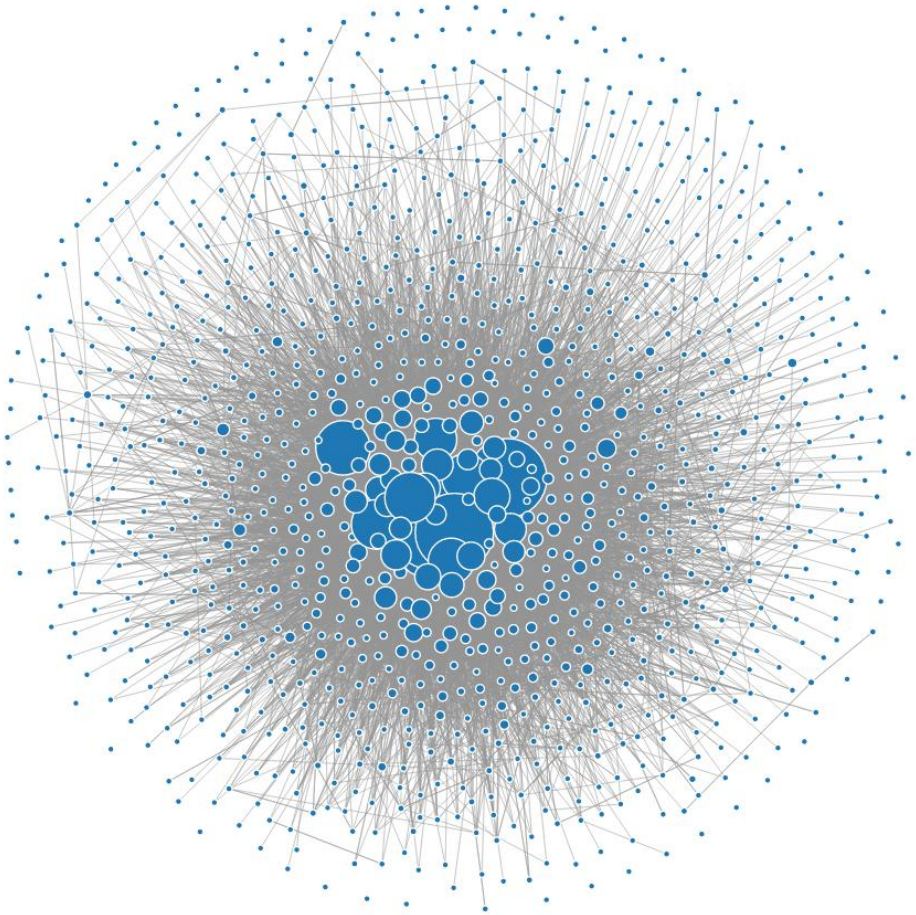
February

1. migueldeicaza (0.033130)
2. dsyme (0.032783)
3. tomaspetricek (0.027756)
4. LincolnAtkinson (0.021993)
5. VisualFSharp (0.020233)
6. c4fsharp (0.019720)
7. rickasaurus (0.019189)
8. ptrelford (0.018099)
9. 1tgr (0.016525)
10. sforkmann (0.014970)

September

1. dsyme (0.028640)
2. migueldeicaza (0.024808)
3. VisualFSharp (0.024479)
4. tomaspetricek (0.021066)
5. c4fsharp (0.019612)
6. rickasaurus (0.014272)
7. sforkmann (0.013471)
8. 1tgr (0.012768)
9. ptrelford (0.012669)
10. FSPowerTools (0.012113)

Visualization with D3.js



Go play
with data!

Thank you

@evelgab
evelina@evelinag.com
https://github.com/evelinag

fsharp.org

F# eXchange 2015
17 April, London