Bachelor Thesis No. 3743
Information propagation in online social networks

Iva Miholić
Advisor: Asst. Prof. Mile Šikić

Department of Electronic Systems and Information Processing
Faculty of Electrical Engineering and Computing
University of Zagreb

9 Jul 2014

Networks

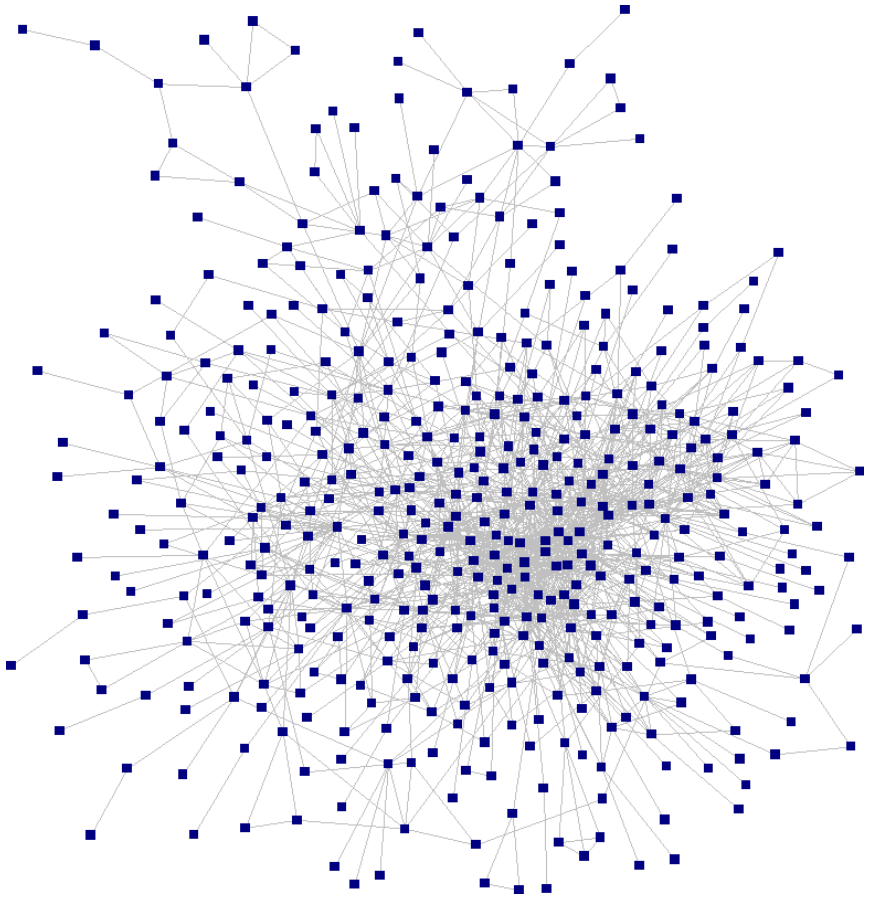


Figure: A network graph of Paul Erdõs and his collaborators. The nodes represent mathematicians and the edges represent the relationship "wrote a paper with".

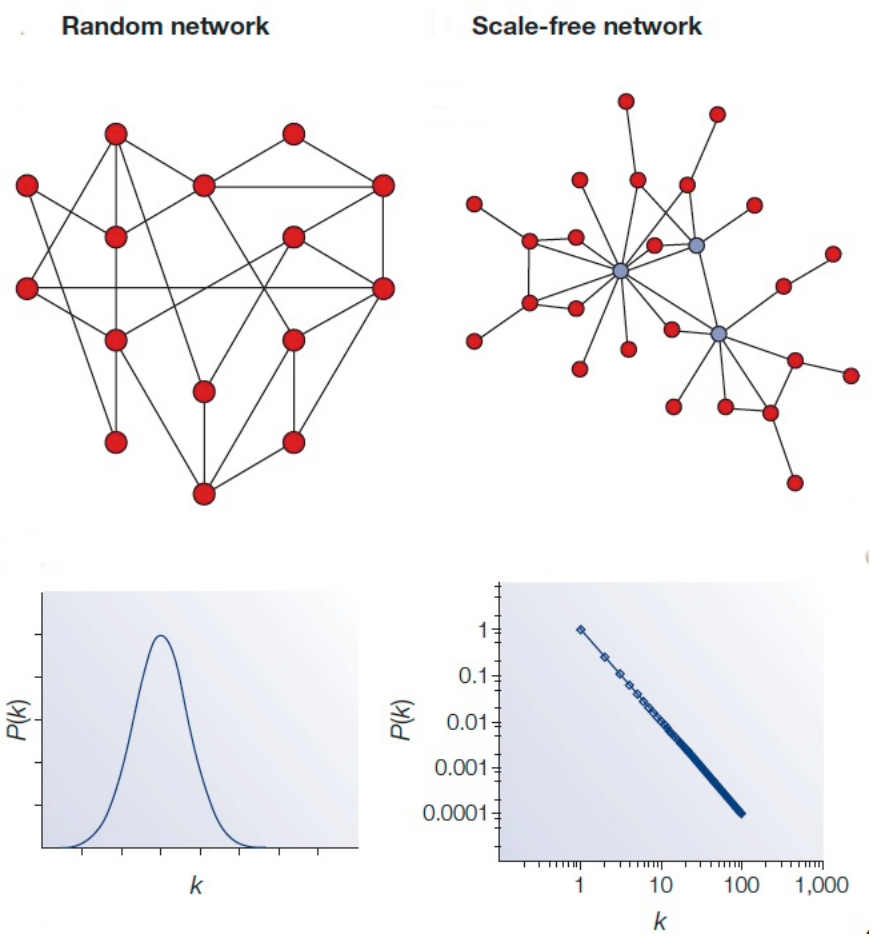


Figure: Comparison of a random and a scale-free network. $p(k)$ stands for the probabilty that a vertex chosen unformly at random has degree k.

- propagation of the information item amongst the nodes of the social network
- closed world + social influence => information propagation similar to diffusion
- In the real world, external influence is present. That is visible on the propagation dynamic.
- Model of propagation that takes into account both internal and external influence.


referendum2013.hr
referendum 2013.hr
Referendum 2013
Prijavite se putem Facebooka, glasajte i saznajte što vaši prijatelji i ostali misle o ovoj temi.
Prijavite se putem Facebooka
Sviđa mi se
Tebi, Sandra Trkulja i 7.354 drugih se ovo sviđa.
Podijeli
7,3 tisuća
Ukupni broj glasova: 11701 (osvježavanje svakih 60 sekundi)
Rezultate mogu vidjeti samo registrirani korisnici.
U svrhu ovoga istraživanja prikupljamo podatke o vašem stavu o referendumskom pitanju, kao i određene podatke s Facebooka (Facebook identifikacijski broj, godinu rođenja, lokaciju, spol i popis vaših prijatelja), čime želimo dobiti uvid u načine kako međusobna poznanstva utječu na stavove korisnika Facebooka. Vaši individualni odgovor na referendumsko pitanje i podaci o vašem profilu neće biti vidljivi drugim korisnicima, već će samo biti vidljiv anonimni prosjek odgovora. Jedino želite vi vidjeti prosjek odgovora vaših prijatelja. Istraživači garantiraju da prikupljeni podaci neće biti korišteni ni u koje druge svrhe osim znanstveno-istraživačke. Prijavom na ovaj upitnik potvrđujete da ste suglasni s ovim pravilima korištenja.
O projektu
Ova web stranica predstavlja dio znanstvenoga istraživanja u području društvenih mreža. Cilj ovog istraživanja je utvrditi kako stav okoline utječe na stav pojedinca te kako se širi vijest u društvenim mrežama.
Više
Autori projekta
Ovaj projekt je nastao kao rezultat znanstvenog istraživanja na temu širenja informacija na društvenim mrežama Nine Antulova-Fantulina i Ive Miholić pod vodstvom doc.dr.sc. Mile Šikića. U cilju pomoći u prikupljanju podataka i izgradnje web sustava, projektu su se pridružili: Bruno Rahle, Matija Piškorec, Tomislav Lipić, Vedran Ivanac i Matej Mihelčić. Projekt je financiran iz vlastitih sredstava autora projekta. Kontakt osoba: Mile Šikić, referendum2013.hr@gmail.com
©referendum2013.hr

Registration dynamic

- 1.695 million nodes, 4.461 million edges
- 11606 registered users, 11 recorded articles

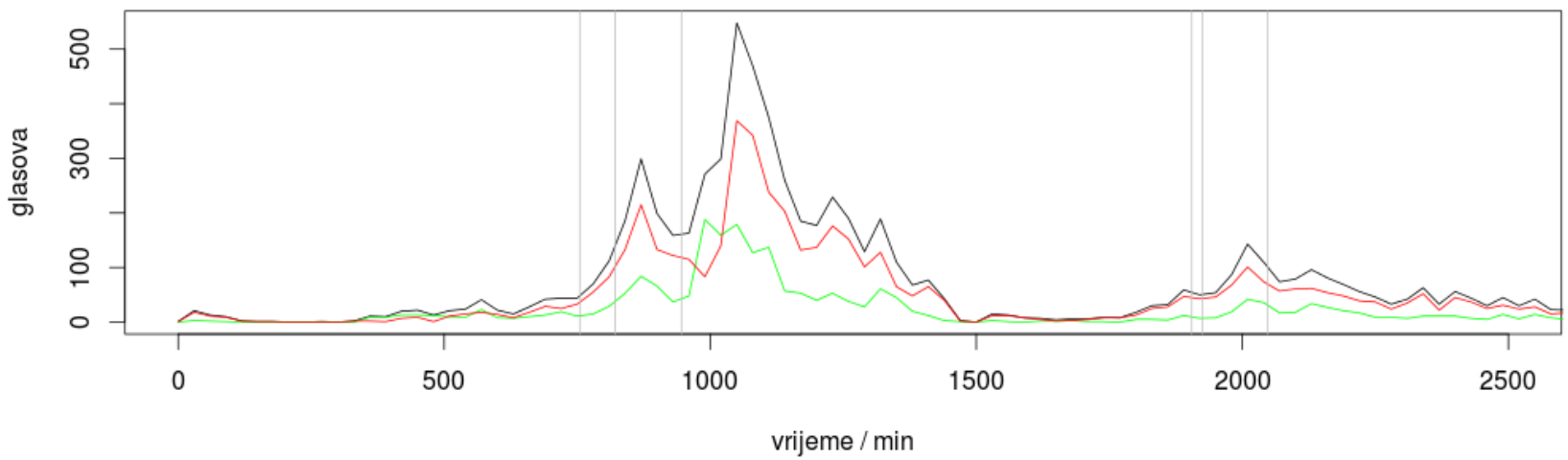
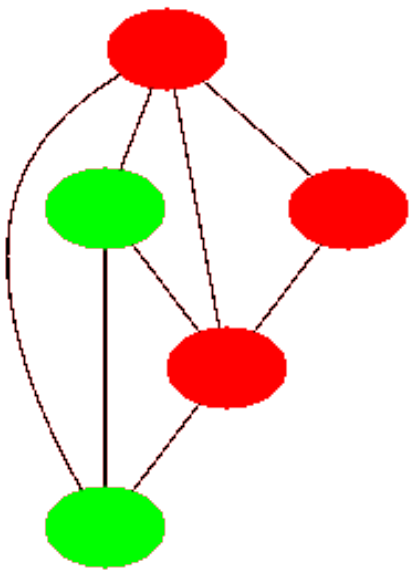


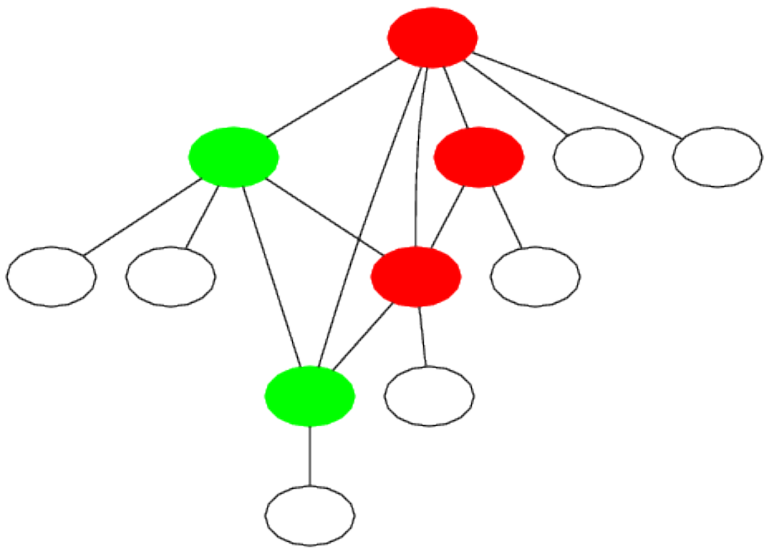
Figure: Dynamic of registrations for the first two days of application being active.

Datasets

Restricted Referendum Dataset



Complete Referendum Dataset



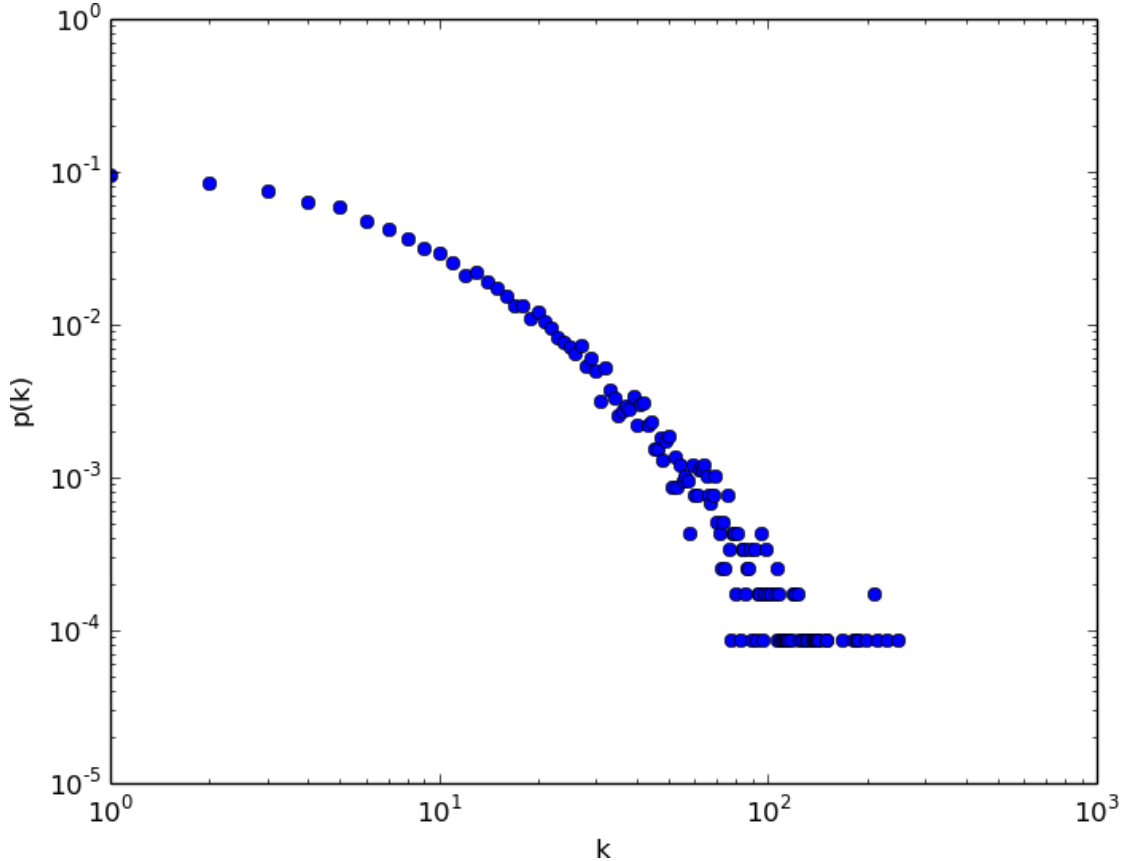


Figure: The degree distribution for the network of registered users. $p(k)$ stands for the probabilty that a vertex chosen unformly at random has degree k.

- **Q: Is the adoption of an information item for the given network mainly internal or external influence driven?**
- information item = information on the existance of the application
- network = Restricted and Complete Referendum Dataset

The structure of the network

- nodes of social network + external nodes
- every external node can influence every peer node (a node in social network)
- set of external nodes form a complete graph
- peer nodes do not influence the external nodes

Probability of activation

$$P(x, y; \alpha, \beta, \gamma) = \frac{1}{1 + e^{-\alpha \ln(x+1) - \beta \ln(y+1) - \gamma}}, \tag{1}$$

- x - the number of already active peer neighbours
- y - the number of already active authorities
- α - *peer coefficient*, measure of internal influence
- β - *authority coefficient*, measure of external influence
- γ - *externality coefficient*, the impact of other factors

Assumptions

- Propagation happens in discrete time steps, from timestamp 1 to timestamp T.
- For every time stamp and every currently inactive node, success of the node trying to become active is observed as a Bernoulli random variable.
- All observed events are independant.
- From the time a node gets activated, its influence is constant and does not fade away with time.
- Every external node has an equal probability of influencing any peer node.

Model estimation

$$P(x, y; \alpha, \beta, \gamma) = \frac{1}{1 + e^{-\alpha \ln(x+1) - \beta \ln(y+1) - \gamma}}, \tag{2}$$

- Estimate parameters α, β and γ with observed events of activation tryouts at every timestep for every currently unactive node using the principle of maximal likelihood.

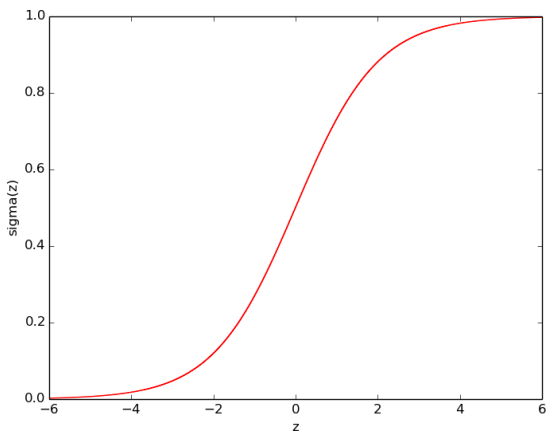


Figure: Function $\sigma(z) = \frac{1}{1+e^{-z}}$.

Principle of maximal likelihood

Take α, β and γ that maximize

$$L(\alpha, \beta, \gamma) = \prod_{x,y} P(x,y)^{A(x,y)}(1 - P(x,y))^{N(x,y)} \tag{3}$$

or minimize

$$-\ln L(\alpha, \beta, \gamma) = -\sum_{x,y} A(x,y) \ln P(x,y) - \sum_{x,y} N(x,y) \ln(1 - P(x,y)). \tag{4}$$

- $A(x,y)$ - number of observed succesful activations.
- $N(x,y)$ - number of observed unsuccesful activations.

Randomization Test

- Given α, β, γ can we just conclude $(\alpha > \beta) => $ internal influence is dominant?
- *time-shuffle* test is used to verify this conclusion.

time-shuffle test

- randomize activation time of all eventually active nodes in the dataset
- obtain $\alpha(D')$ and $\beta(D)'$ for the new dataset D'
- update S_α and S_β
- repeat

time-shuffle test

- randomize activation time of all eventually active nodes in the dataset
- obtain $\alpha(D')$ and $\beta(D)'$ for the new dataset D'
- update S_α and S_β
- repeat

The *strength of peer influence*, S_α, is defined as

$$S_\alpha = P_{D' \in \mathcal{D}}(\alpha > \alpha(D')). \tag{5}$$

and similarly for the *strength of authority influence*, S_β:

$$S_\beta = P_{D' \in \mathcal{D}}(\beta > \beta(D')). \tag{6}$$

Restricted Referendum Dataset

$\alpha = 0.16432$, $\beta = 0.41826$, $\gamma = -8.84820$

$S_\alpha = 0.996$, $S_\beta = 0.633$

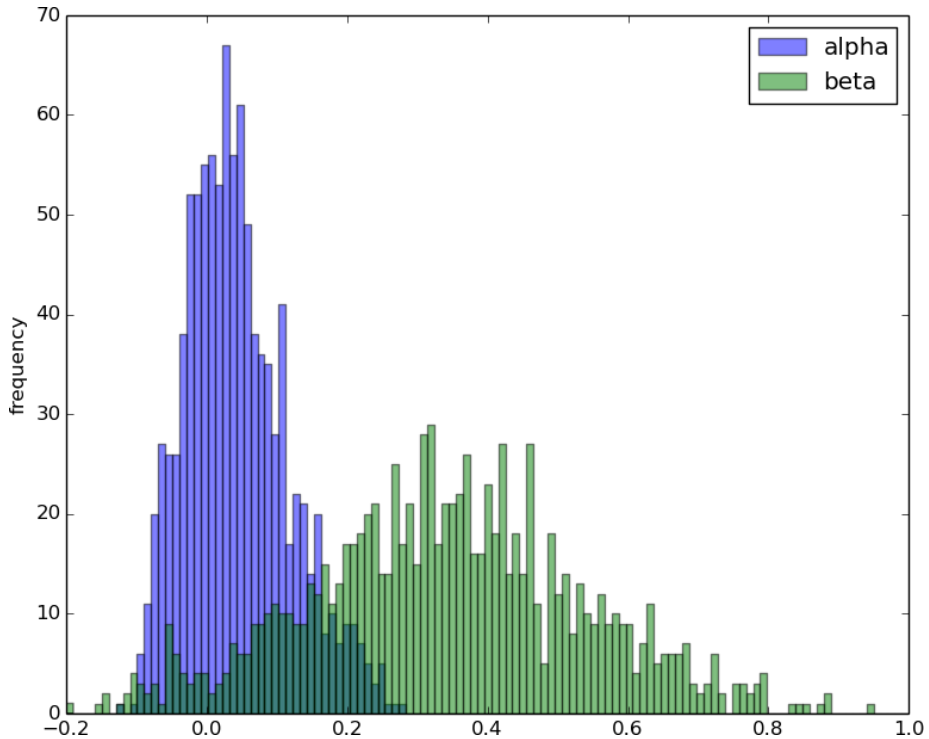


Figure: Frequency histograms of estimated values of α (blue) and β (green) obtained for the Restricted Referendum Dataset with time-shuffle test based on 1000 randomized instances.

Complete Referendum Dataset

$\alpha = 1.44990$, $\beta = -1.02360$, $\gamma = -13.52627$

y	0	1	2	3	4	5
successes	0.77791	9.00000	7.34400	3.91023	4.52381	3.74797
trials	1697340	1694519	1693785	1690362	1689491	1689169
success rate (10^{-6})	0.458	5.311	4.336	2.313	2.678	2.219
y	6	7	8	9	10	11
successes	1.26634	0.89345	0.67893	1.41053	3.40826	1.62857
trials	1687883	1685728	1684609	1684391	1684161	1639898
success rate (10^{-6})	0.750	0.530	0.403	0.837	2.024	0.993

Table: Average frequencies of observed successful activations per time step, average frequencies of trials per time step and success rates for the time period when y authorities were active on the Complete Referendum Datasets.

Redefining the model of external influence

$$P(x, y; \alpha, \beta, \gamma) = \frac{1}{1 + e^{-\alpha \ln(x+1) - \beta \ln(y+1) - \gamma}}, \tag{7}$$

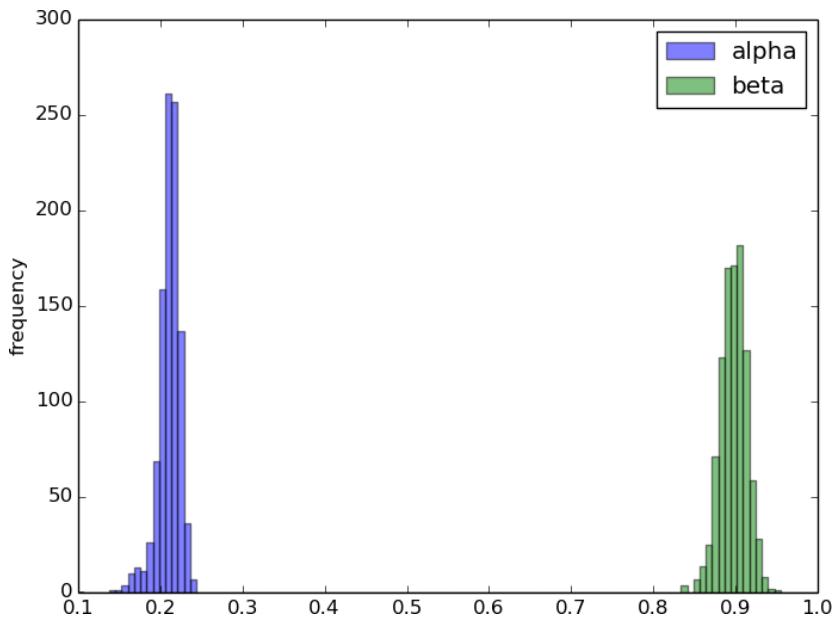
- x - the number of already active peer neighbours
- y - **the number of activations of nodes with** 0 **friends active at the time of the event**

- now y doesn't necessarily have positive correlation with time

Restricted Referendum Dataset

$\alpha = 0.16697$, $\beta = 0.92495$,
$\gamma = -8.60957$, $S_\alpha = 0.018$,
$S_\beta = 0.959$

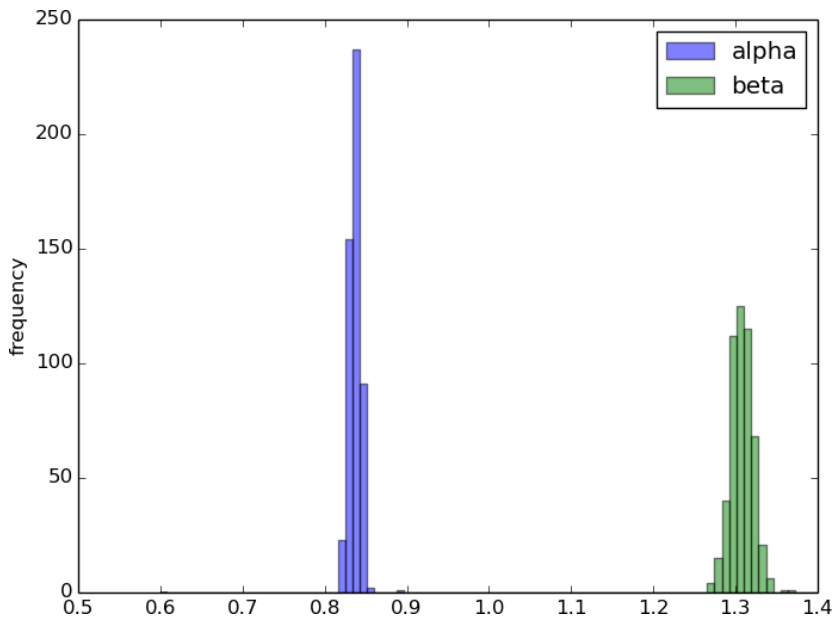
- based on 1000 randomized instances



Complete Referendum Dataset

$\alpha = 0.89215$, $\beta = 1.36409$,
$\gamma = -15.7230$, $S_\alpha = 1.000$,
$S_\beta = 0.998$

- based on 500 randomized instances



- **Q: Is the adoption of an information item for the given network mainly internal or external influence driven?**
- **A: The adoption is mainly external influence driven.**