Social networks bot detection using Benford's law

M. Kolomeets[12], D. Levshun[12], S. Soloviev[23], A. Chechulin[1], I. Kotenko[1]
[1]St. Petersburg Federal Research Center of the Russian Academy of Sciences
[2]Université Paul Sabatier Toulouse III
[3]Toulouse Institute of Computer Science Research (IRIT)

Bot types in social networks

- Bot - a social network account that **cheats metrics** and **does not express the real opinion** (if any) of its creator
- Controlled by:
  - Software (automated)
  - Human (human animated bots)
- Created by:
  - Software
  - Human
  - Hacking / Buying / **Renting** an account from a real user

What we can analyze to detect bot?

- Account metrics
- Distributions of friend’s metrics
- Network centrality measures
- Text
- Meaning/Emotion/Information content
- Timeline

What methods can be used to detect bot?

- Analytical
- **Statistical**
- Network Science (calculation of centrality measures on graphs)
- Machine Learning

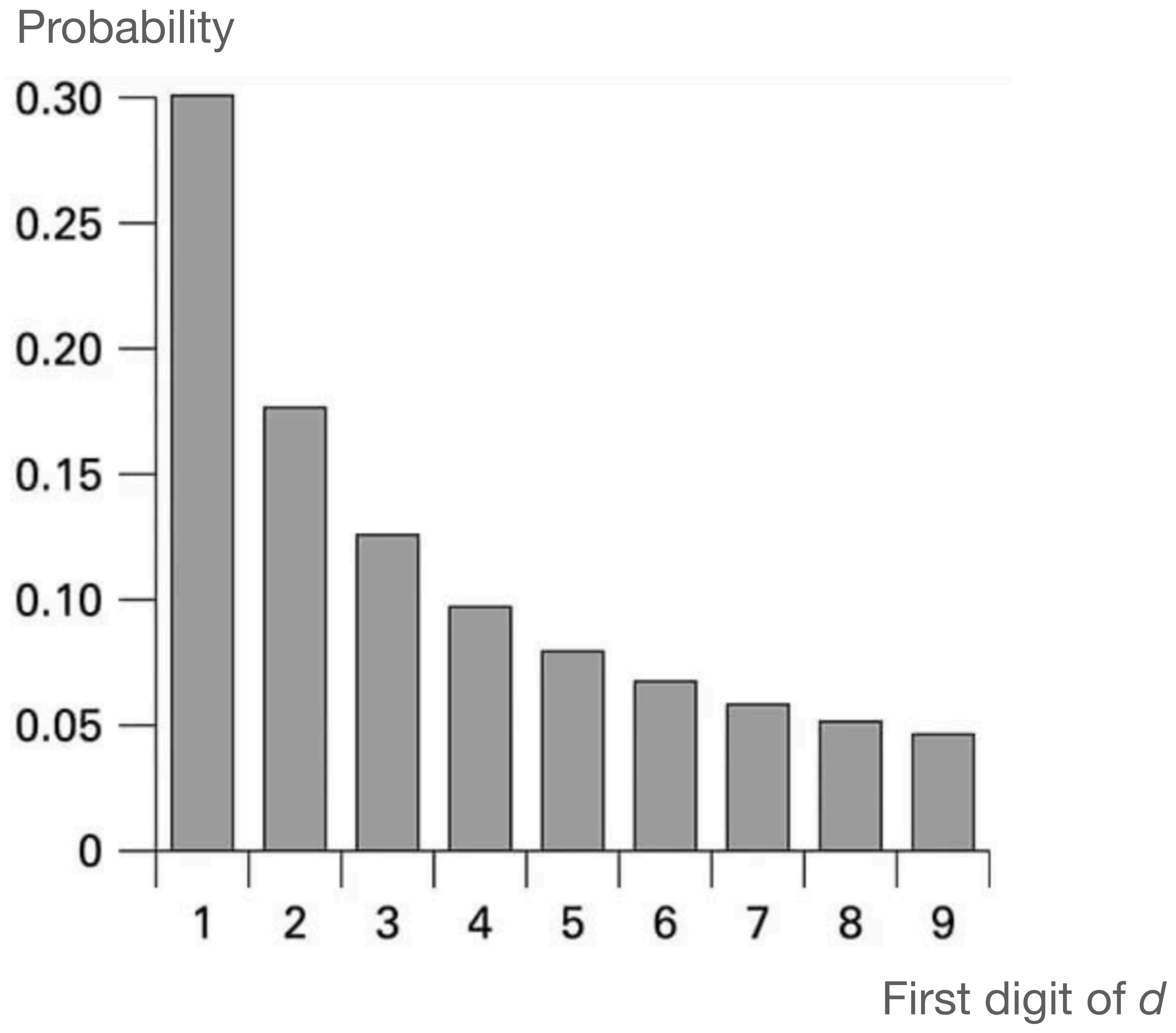
Benford’s law

- A dataset **satisfies** Benford's law

  if the probability of observing a **first digit of *d***

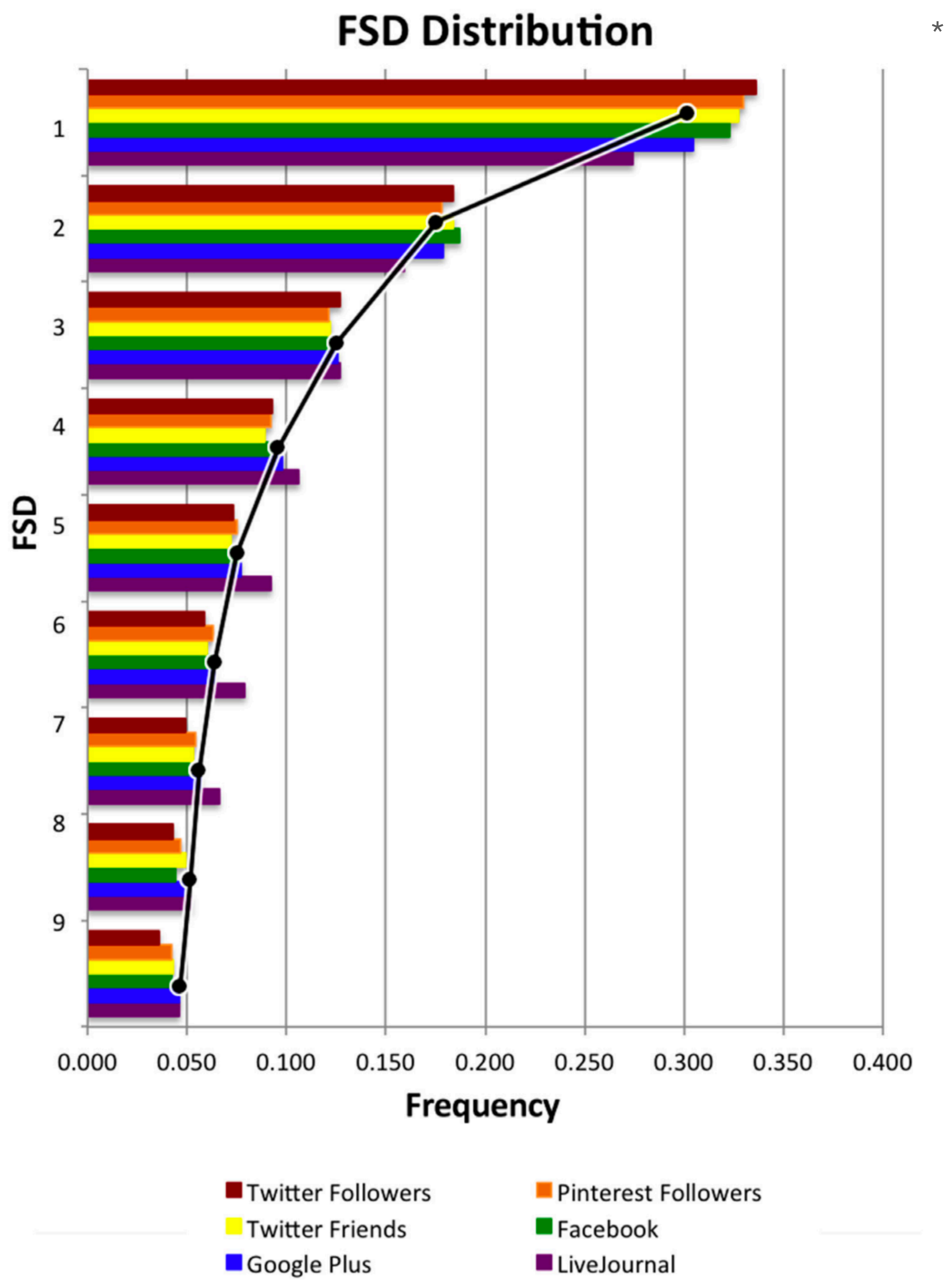
  is approximately $log_{10}(\frac{d+1}{d})$.

749, **3**9, **2**4, **7**2, **4**74, **6**293, **1**0, **2**, **6**11, **2**8, …



Benford’s law in social network

- Successfully used to detect fraud
- It is proved that it is applicable to the social networks (w.r.t. various metrics parameters)
- Hypothesis:
    - Real **users - obey** Benford’s law
    - **Bot - do not obey** Benford’s law



*Golbeck J. Benford’s law applies to online social networks // PLoS One. 2015. T. 10. № 8.

Experiment

Dataset	Type	Number of profiles
bot_1	software bots	301
bot_2	software bots	295
bot_3	live bots	298
bot_4	software bots	301
bot_5	software bots	303
bot_6	live bots	304
bot_7	live bots	302
bot_8	live bots	357
user_1	activists	385
user_2	mass media	298
user_3	developers	332
user_4	sport	224
user_5	mass media	420
user_6	blog	251
user_7	commerce	284
user_8	festival	259
user_9	sport	181
user_10	developers	397

	metric 1	…	metric *M*
account 1	**194**	…	**63**
…	…	…	…
account *N*	**76**	…	**24**

metric 1	1, …, 7
…	…
metric *M*	6, …, 2

	p-value
metric 1	0.97
…	…
metric *M*	0.64

Kolmogorov–Smirnov test

Table 2. Table with p-values for real user datasets

Dataset	friends	groups	followers	subscriptions	albums	photos	posts
user_1	0.9895	0.9999	0.9999	0.9895	0.9895	0.9999	0.9895
user_2	0.9895	0.9794	0.3364	0.9999	0.3364	0.3364	0.9999
user_3	0.9794	0.9895	0.9999	0.9895	0.4647	0.6994	0.9794
user_4	0.6994	0.3364	0.3364	0.9794	0.3364	0.6994	0.9794
user_5	0.9794	0.9794	0.9895	0.9895	0.1243	0.7301	0.9999
user_6	0.9895	0.6994	0.9999	0.9794	0.3364	0.6994	0.7301
user_7	0.9794	0.9794	0.7301	0.6994	0.7301	0.1243	0.6994
user_8	0.9895	0.9895	0.9999	0.9895	0.3364	0.9794	0.9794
user_9	0.6994	0.9895	0.9794	0.9794	0.9539	0.9296	0.9794
user_10	0.9999	0.9794	0.9999	0.9794	0.3364	0.9895	0.9794

Table 3. Table with p-values for bot datasets

Dataset	friends	groups	followers	subscriptions	albums	photos	posts
bot_1	0.1243	0.0016	0.9894	0.3364	0.9439	0.5136	0.9793
bot_2	0.6993	0.0366	0.9894	0.3364	0.9776	0.6171	0.6993
bot_3	0.6993	0.0630	0.9793	0.1243	0.9907	0.7344	0.9793
bot_4	0.5436	0.5727	0.8239	0.6993	0.7090	0.7740	0.5436
bot_5	0.3364	0.3364	0.6993	0.3364	0.5454	0.8994	0.6993
bot_6	0.9793	0.9999	0.6993	0.9999	0.6993	0.6993	0.9999
bot_7	0.9793	0.9999	0.9793	0.7301	0.5907	0.7229	0.9894
bot_8	0.9999	0.9894	0.6993	0.9793	0.9793	0.9793	0.9999

Results and Discussion

- Albums, photos, posts, and followers are usually hidden with privacy settings
- Probably, it is possible to detect software bots
- Probably, usage of Benford's law is not enough to detect human animated bots

Table 2. Table with p-values for real user datasets

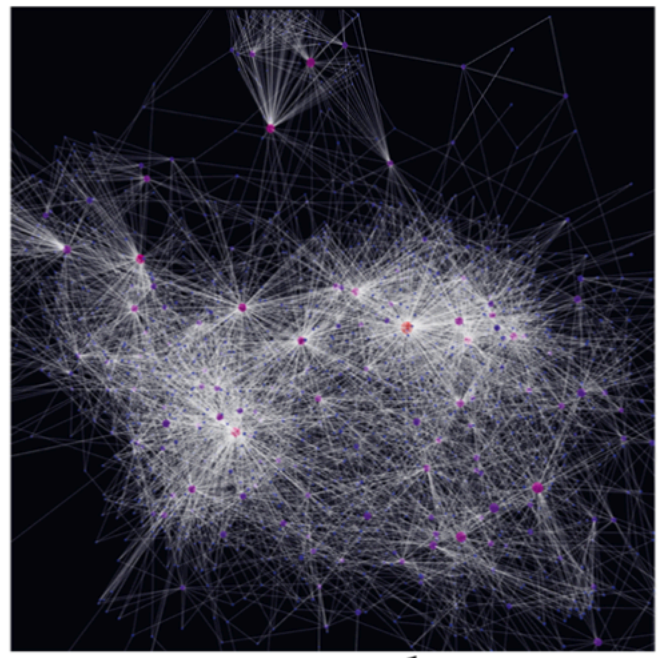
Dataset	friends	groups	followers	subscriptions	albums	photos	posts
user_1	0.9895	0.9999	0.9999	0.9895	0.9895	0.9999	0.9895
user_2	0.9895	0.9794	0.3364	0.9999	0.3364	0.3364	0.9999
user_3	0.9794	0.9895	0.9999	0.9895	0.4647	0.6994	0.9794
user_4	0.6994	0.3364	0.3364	0.9794	0.3364	0.6994	0.9794
user_5	0.9794	0.9794	0.9895	0.9895	0.1243	0.7301	0.9999
user_6	0.9895	0.6994	0.9999	0.9794	0.3364	0.6994	0.7301
user_7	0.9794	0.9794	0.7301	0.6994	0.7301	0.1243	0.6994
user_8	0.9895	0.9895	0.9999	0.9895	0.3364	0.9794	0.9794
user_9	0.6994	0.9895	0.9794	0.9794	0.9539	0.9296	0.9794
user_10	0.9999	0.9794	0.9999	0.9794	0.3364	0.9895	0.9794

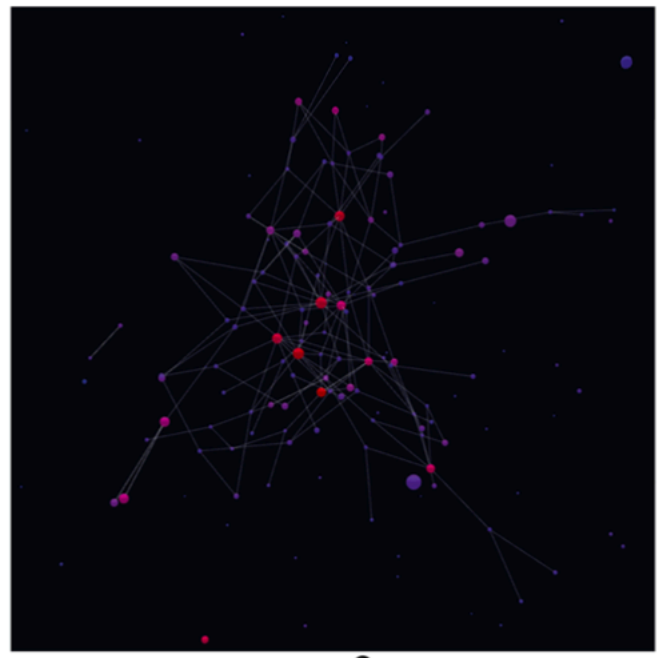
Table 3. Table with p-values for bot datasets

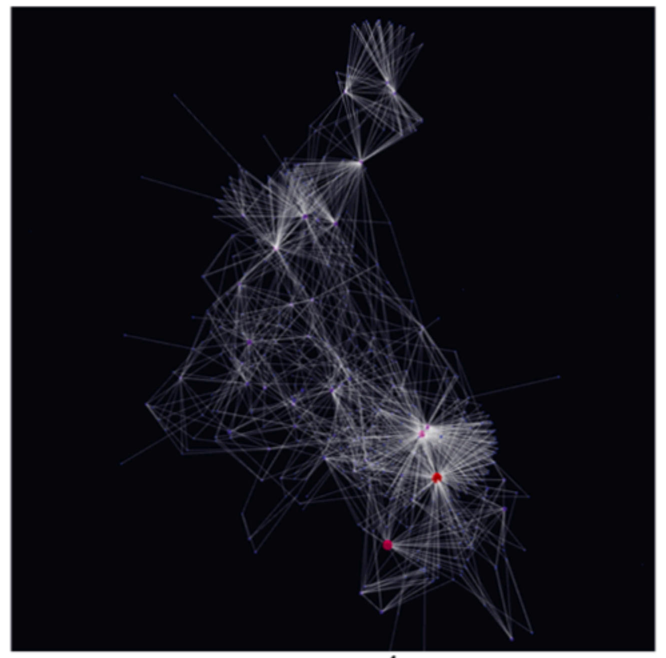
Dataset	friends	groups	followers	subscriptions	albums	photos	posts
bot_1	0.1243	0.0016	0.9894	0.3364	0.9439	0.5136	0.9793
bot_2	0.6993	0.0366	0.9894	0.3364	0.9776	0.6171	0.6993
bot_3	0.6993	0.0630	0.9793	0.1243	0.9907	0.7344	0.9793
bot_4	0.5436	0.5727	0.8239	0.6993	0.7090	0.7740	0.5436
bot_5	0.3364	0.3364	0.6993	0.3364	0.5454	0.8994	0.6993
bot_6	0.9793	0.9999	0.6993	0.9999	0.6993	0.6993	0.9999
bot_7	0.9793	0.9999	0.9793	0.7301	0.5907	0.7229	0.9894
bot_8	0.9999	0.9894	0.6993	0.9793	0.9793	0.9793	0.9999

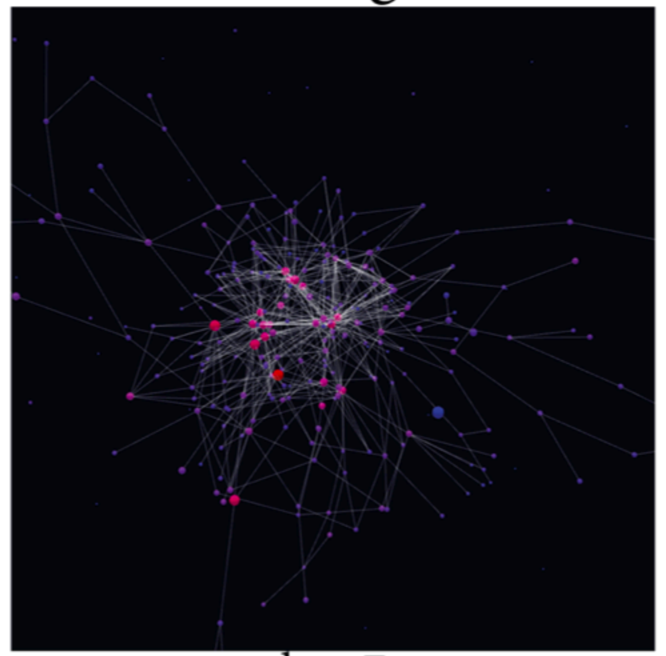
Discussion

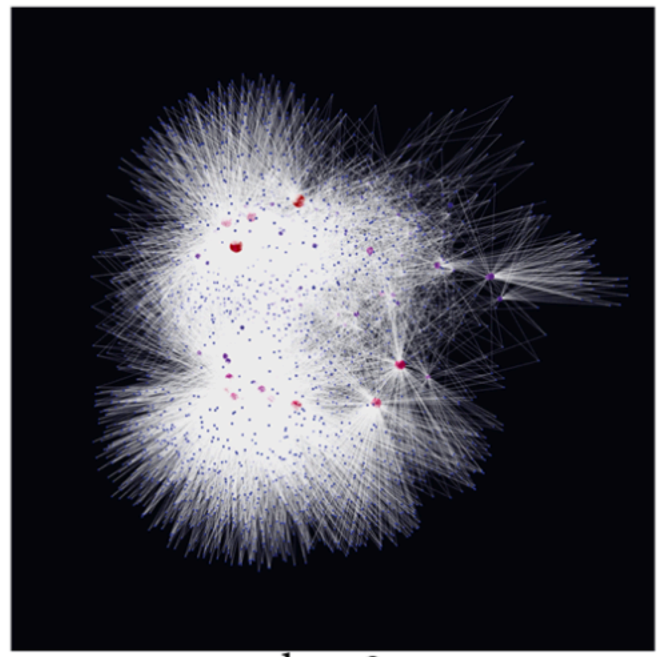
- bot_3 - is an exception
  but (probably) the seller was cheating


user 1


user 2


user 4


bot 7


bot 3

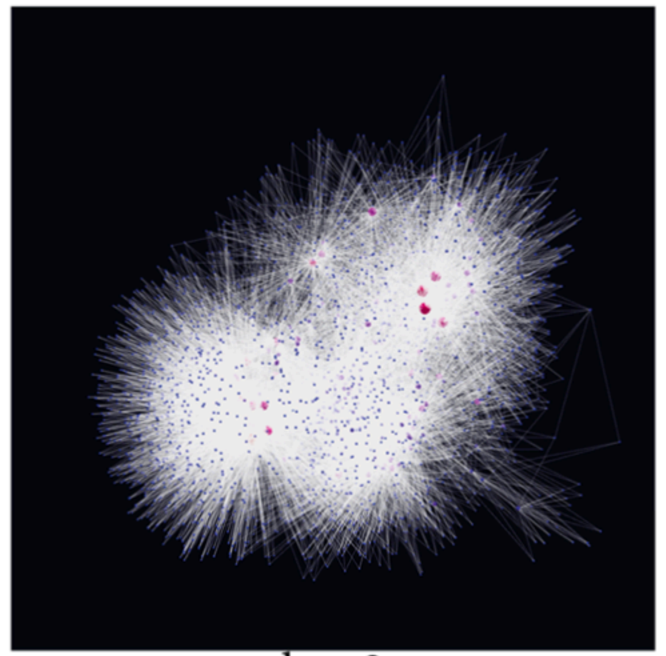

bot 2

Table 2. Table with p-values for real user datasets

Dataset	friends	groups	followers	subscriptions	albums	photos	posts
user_1	0.9895	0.9999	0.9999	0.9895	0.9895	0.9999	0.9895
user_2	0.9895	0.9794	0.3364	0.9999	0.3364	0.3364	0.9999
user_3	0.9794	0.9895	0.9999	0.9895	0.4647	0.6994	0.9794
user_4	0.6994	0.3364	0.3364	0.9794	0.3364	0.6994	0.9794
user_5	0.9794	0.9794	0.9895	0.9895	0.1243	0.7301	0.9999
user_6	0.9895	0.6994	0.9999	0.9794	0.3364	0.6994	0.7301
user_7	0.9794	0.9794	0.7301	0.6994	0.7301	0.1243	0.6994
user_8	0.9895	0.9895	0.9999	0.9895	0.3364	0.9794	0.9794
user_9	0.6994	0.9895	0.9794	0.9794	0.9539	0.9296	0.9794
user_10	0.9999	0.9794	0.9999	0.9794	0.3364	0.9895	0.9794

Table 3. Table with p-values for bot datasets

Dataset	friends	groups	followers	subscriptions	albums	photos	posts
bot_1	0.1243	0.0016	0.9894	0.3364	0.9439	0.5136	0.9793
bot_2	0.6993	0.0366	0.9894	0.3364	0.9776	0.6171	0.6993
bot_3	0.6993	0.0630	0.9793	0.1243	0.9907	0.7344	0.9793
bot_4	0.5436	0.5727	0.8239	0.6993	0.7090	0.7740	0.5436
bot_5	0.3364	0.3364	0.6993	0.3364	0.5454	0.8994	0.6993
bot_6	0.9793	0.9999	0.6993	0.9999	0.6993	0.6993	0.9999
bot_7	0.9793	0.9999	0.9793	0.7301	0.5907	0.7229	0.9894
bot_8	0.9999	0.9894	0.6993	0.9793	0.9793	0.9793	0.9999

Dataset	Type	Number of profiles
bot_1	software bots	301
bot_2	software bots	295
bot_3	live bots	298
bot_4	software bots	301
bot_5	software bots	303
bot_6	live bots	304
bot_7	live bots	302
bot_8	live bots	357
user_1	activists	385
user_2	mass media	298
user_3	developers	332
user_4	sport	224
user_5	mass media	420
user_6	blog	251
user_7	commerce	284
user_8	festival	259
user_9	sport	181
user_10	developers	397

Results

and plans for future research

*

- The method is not able to identify bots individually, since it analyzes distributions.
- Perhaps, bots add to their friend list other bots. So, applying this method to the friend list can identify bots individually.
  This hypothesis needs to be tested.
- Perhaps, the p-value will be useful as a feature in ML.
  This hypothesis needs to be tested.

M. Kolomeets[12], D. Levshun[12], S. Soloviev[23], A. Chechulin[1], I. Kotenko[1]
[1]St. Petersburg Federal Research Center of the Russian Academy of Sciences
[2]Université Paul Sabatier Toulouse III
[3]Toulouse Institute of Computer Science Research (IRIT)