Reputation systems against social engineering and manipulation in online environments

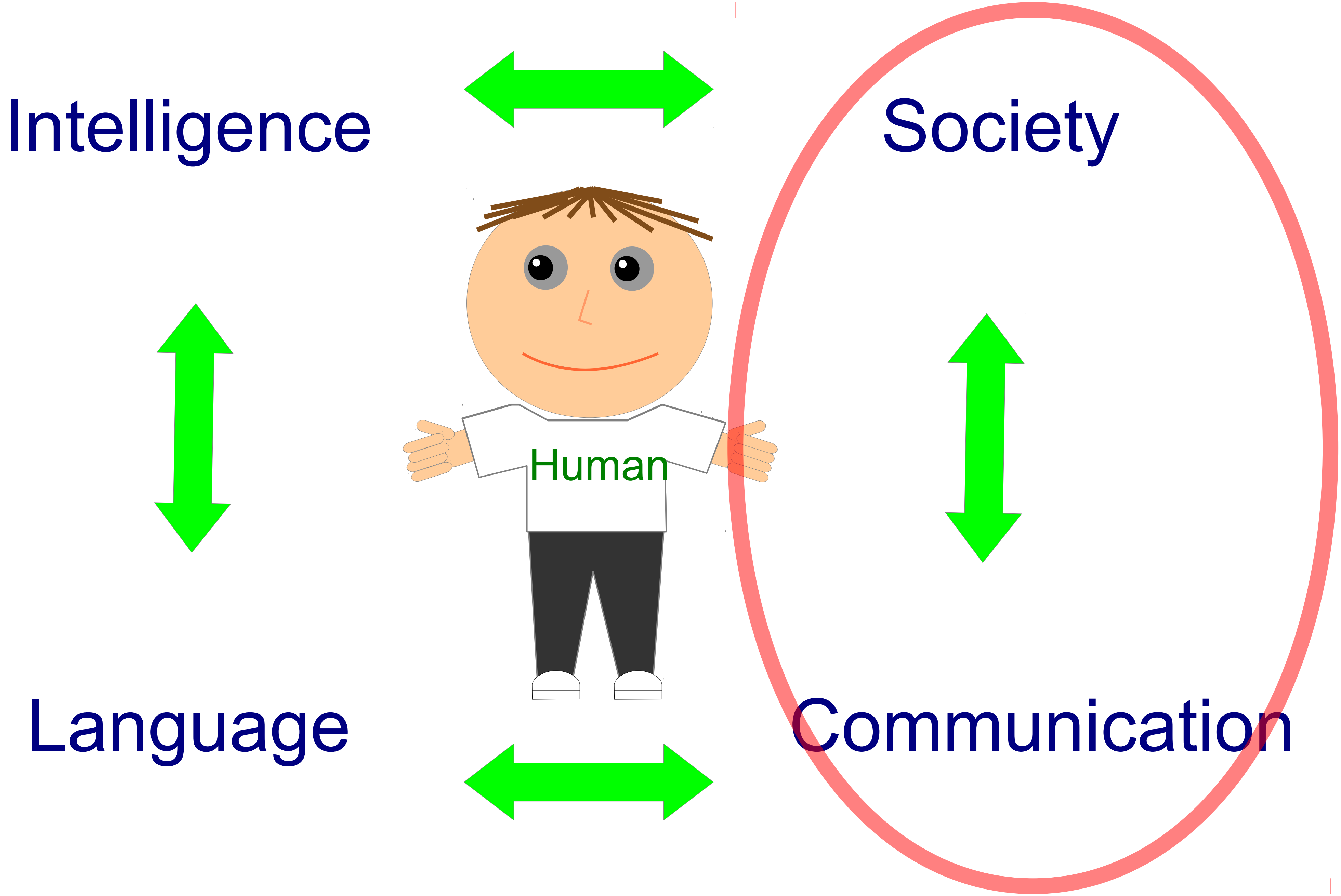
Anton Kolonin
akolonin@aigents.com

Novosibirsk State University
*THE REAL SCIENCE

AIGENTS
https://aigents.com

SingularityNET
https://singularitynet.io

Evolution of Social Complexity



Social Communication Challenges

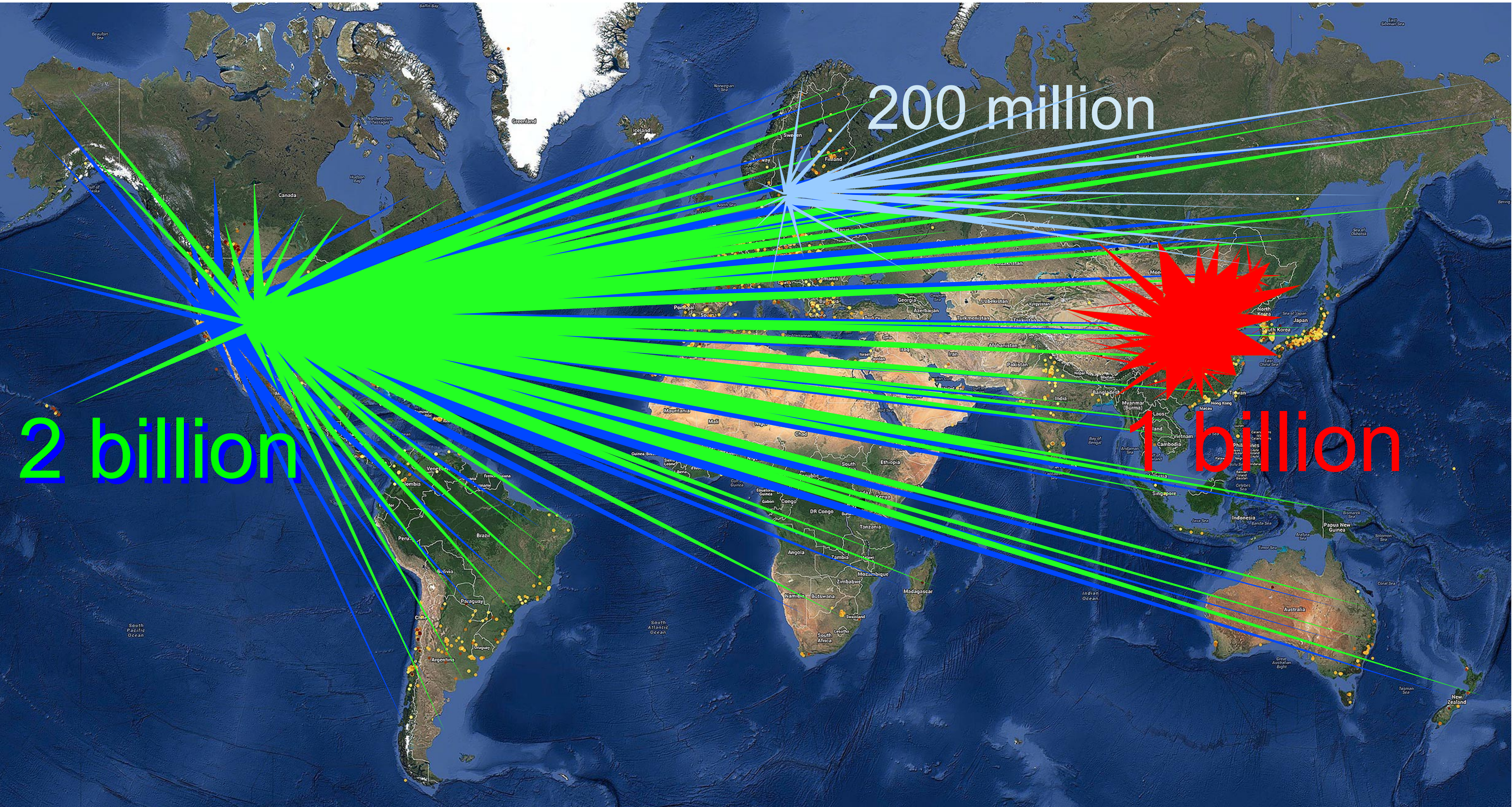
What have changed in last 50 years?

Connectivity – ~~tens~~ *millions* of people

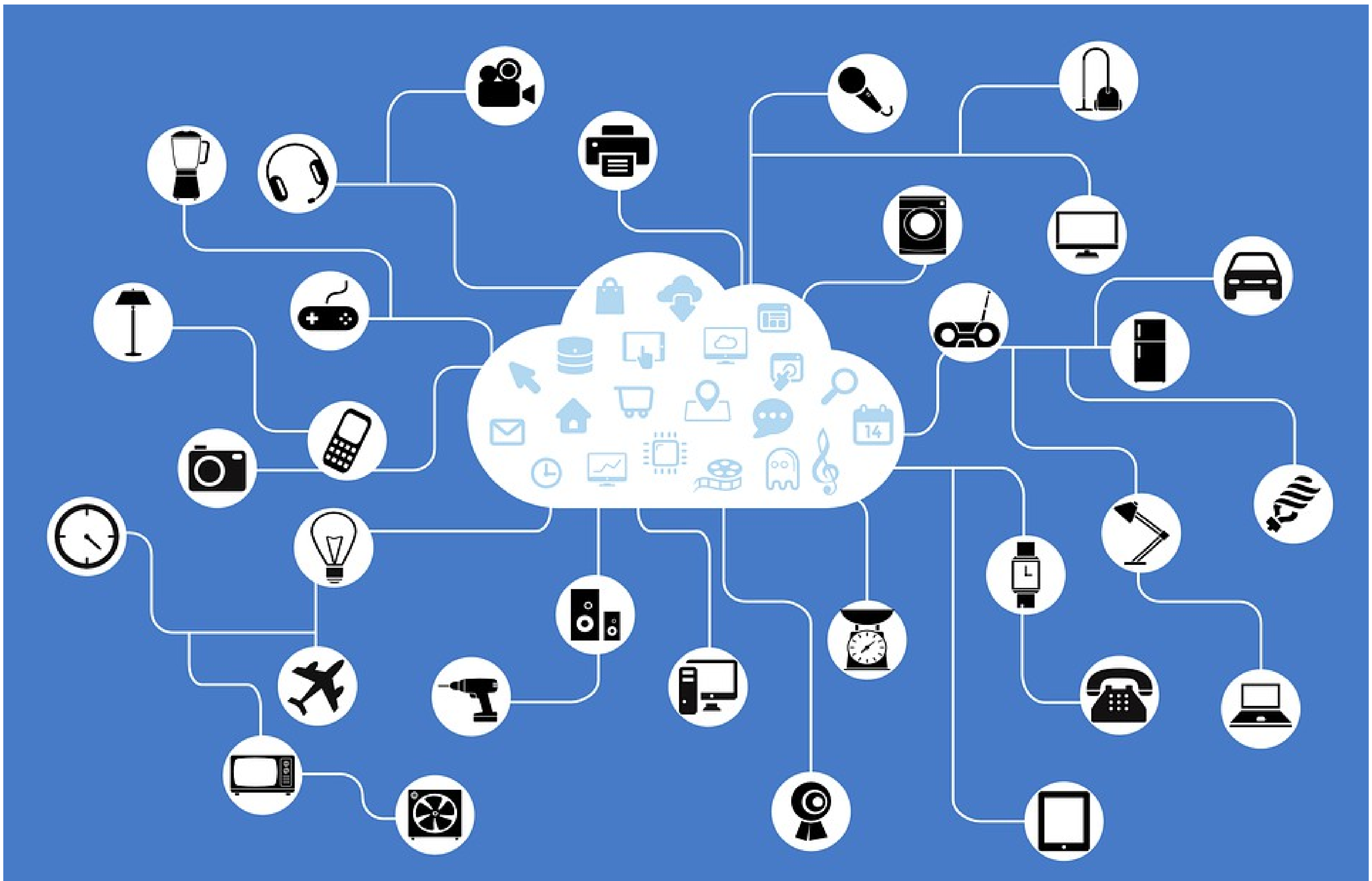
Speed – of ~~speaking and writing~~ *light*

Reliability – ~~relatives/neighbors~~ *strangers*

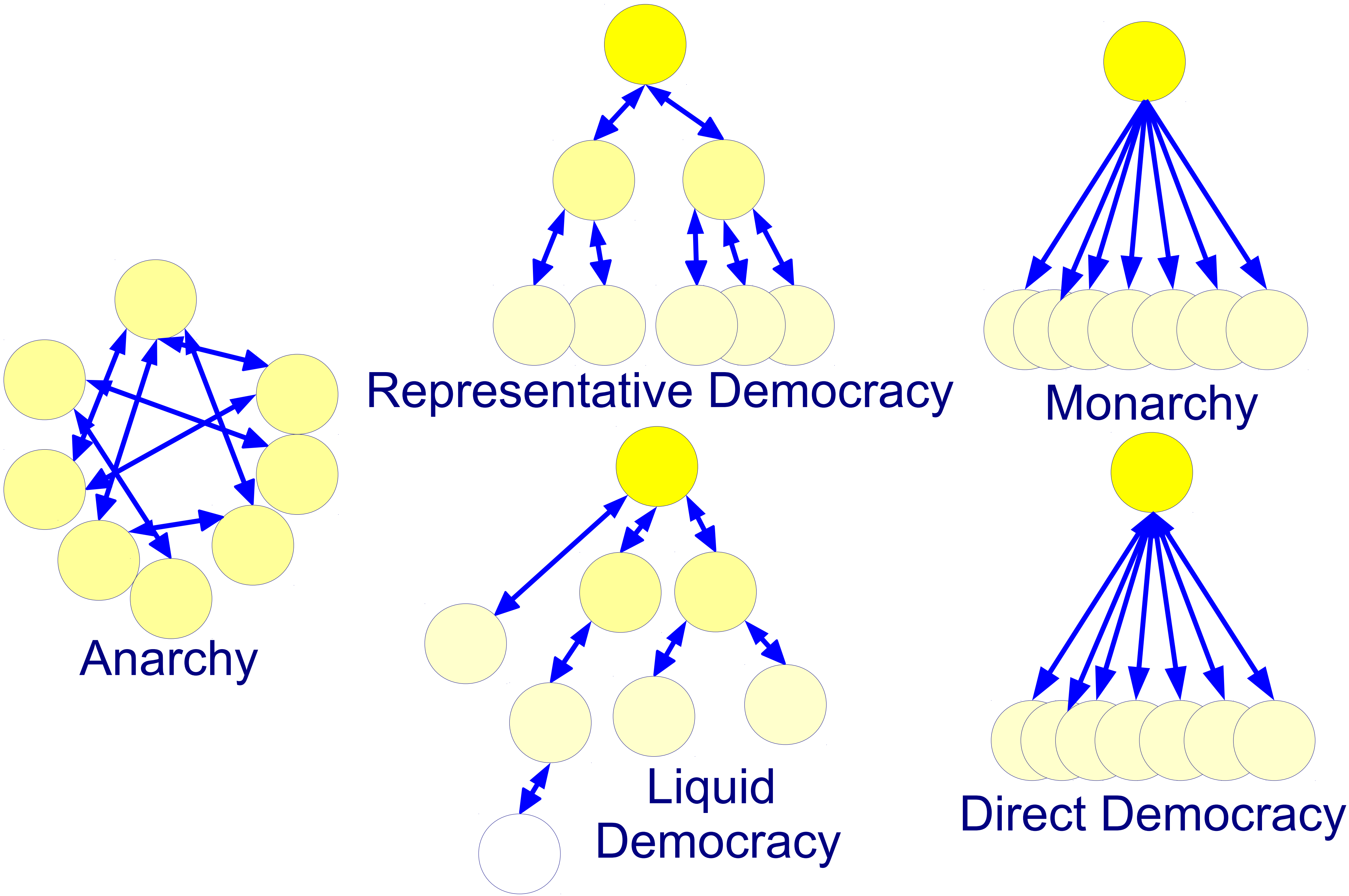
People involved in “social computing” monthly:
Google+Facebook – worldwide, Telegram – worldwide
WeChat+Baidu+QQ - in China



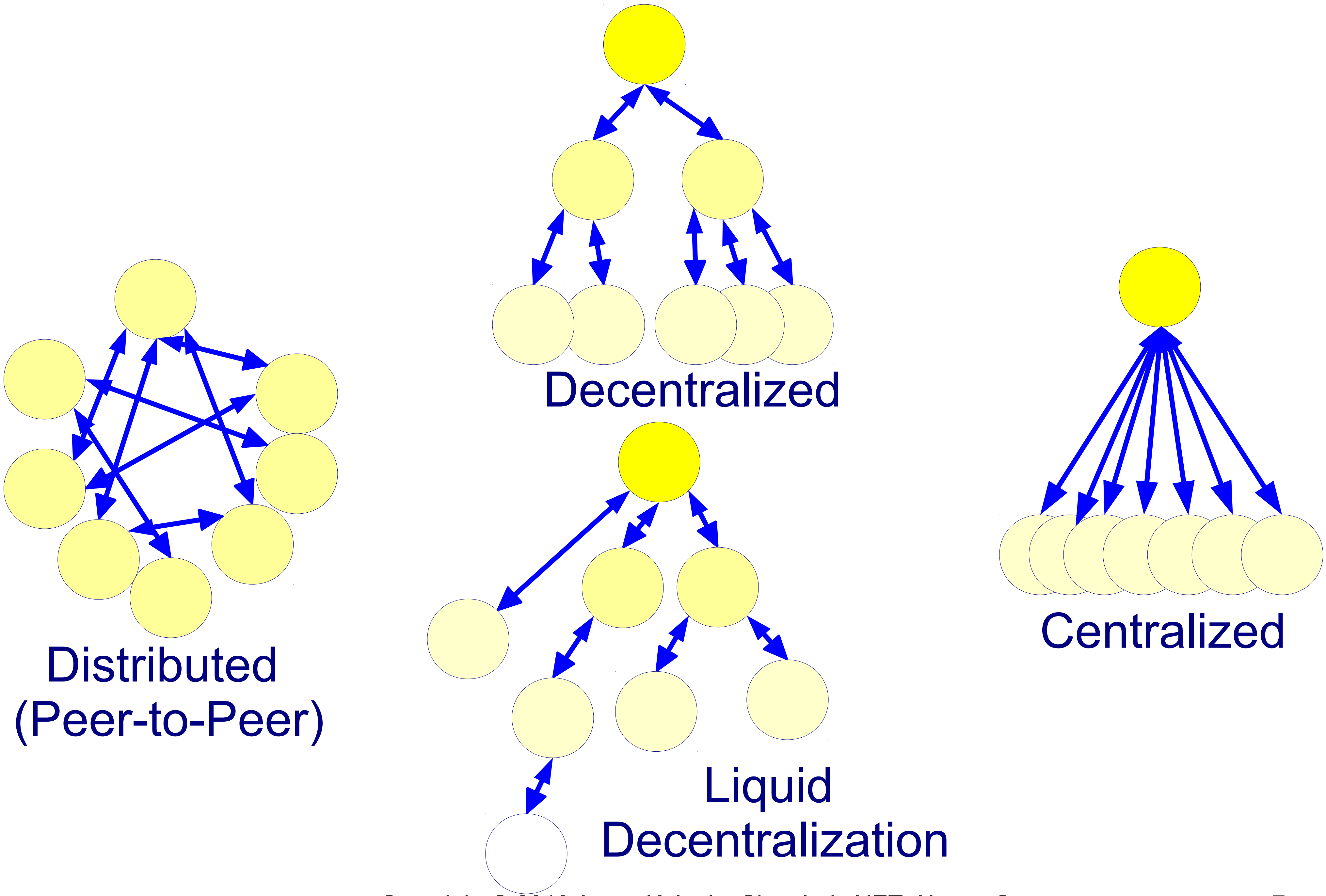
World-wide social network of 7.5 billion humans, is accompanied with 15 billion IoT devices in 2018 with many of them supplied with AI in the next years



Governance and Reputation in Human Societies

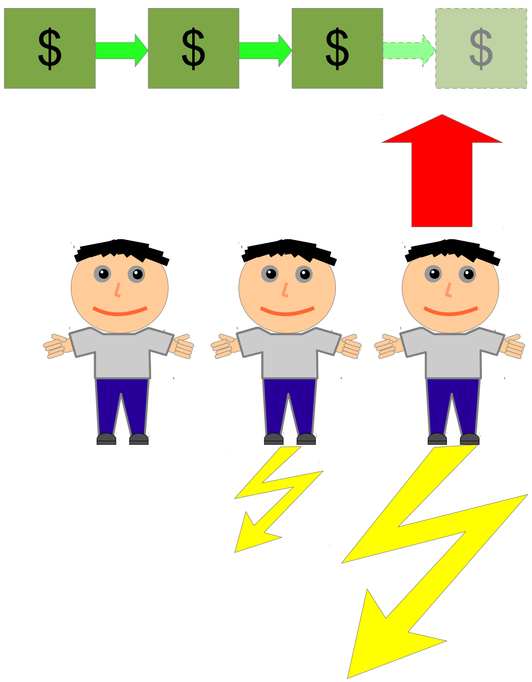


Governance and Control in Computing Environments

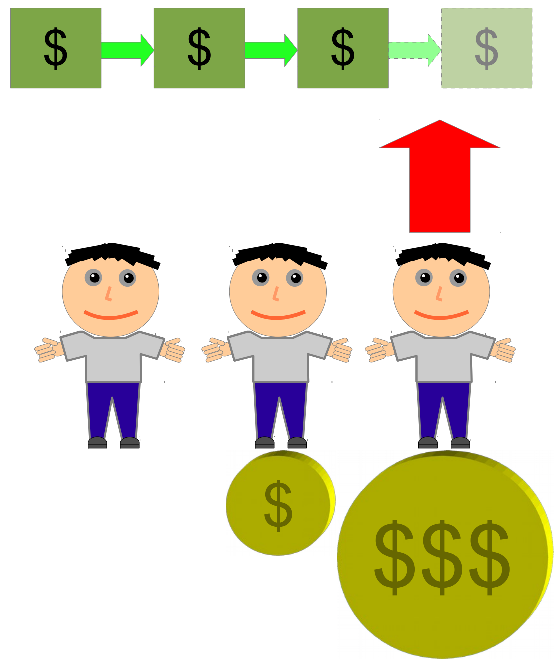


Consensus – technology to govern distributed multi-agent systems such as blockchains or societies, resistant to takeover and scam.

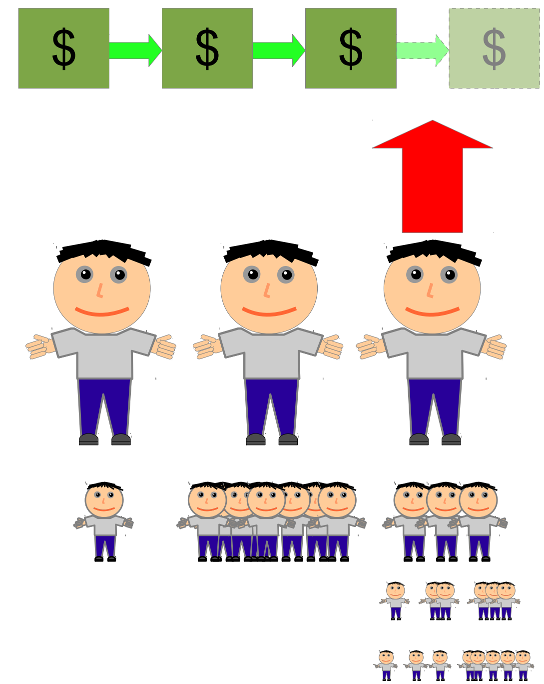
Proof-Of-Work



Proof-Of-Stake



Proof-Of-Reputation



$$R_i = \sum_t \sum_j (R_j * V_{ijt})$$

Force is Power:
Those who own more computing resources govern the network.

Money is Power:
Those who have more money govern the network.

Reputation is Power:
Those who earn a better reputation and a greater long-term audience base govern the network.

Reputation Systems – Solving Problems

Marketplaces	Unfair competition, gaming ratings
News filtering	Fake news, information wars
Social Networking	Spam, abuse, harassment
Socio-psychological security	Broken relationships
Financial security	Scam
Blockchain consensuses	Consensus takeover
Democratic Governance	State instability

Reputation Systems Ingredients

Data:

- Ratings
- Stakes
- Payments
- Spendings
- Reviews
- Mentions
- Tips
- etc.

Principles:

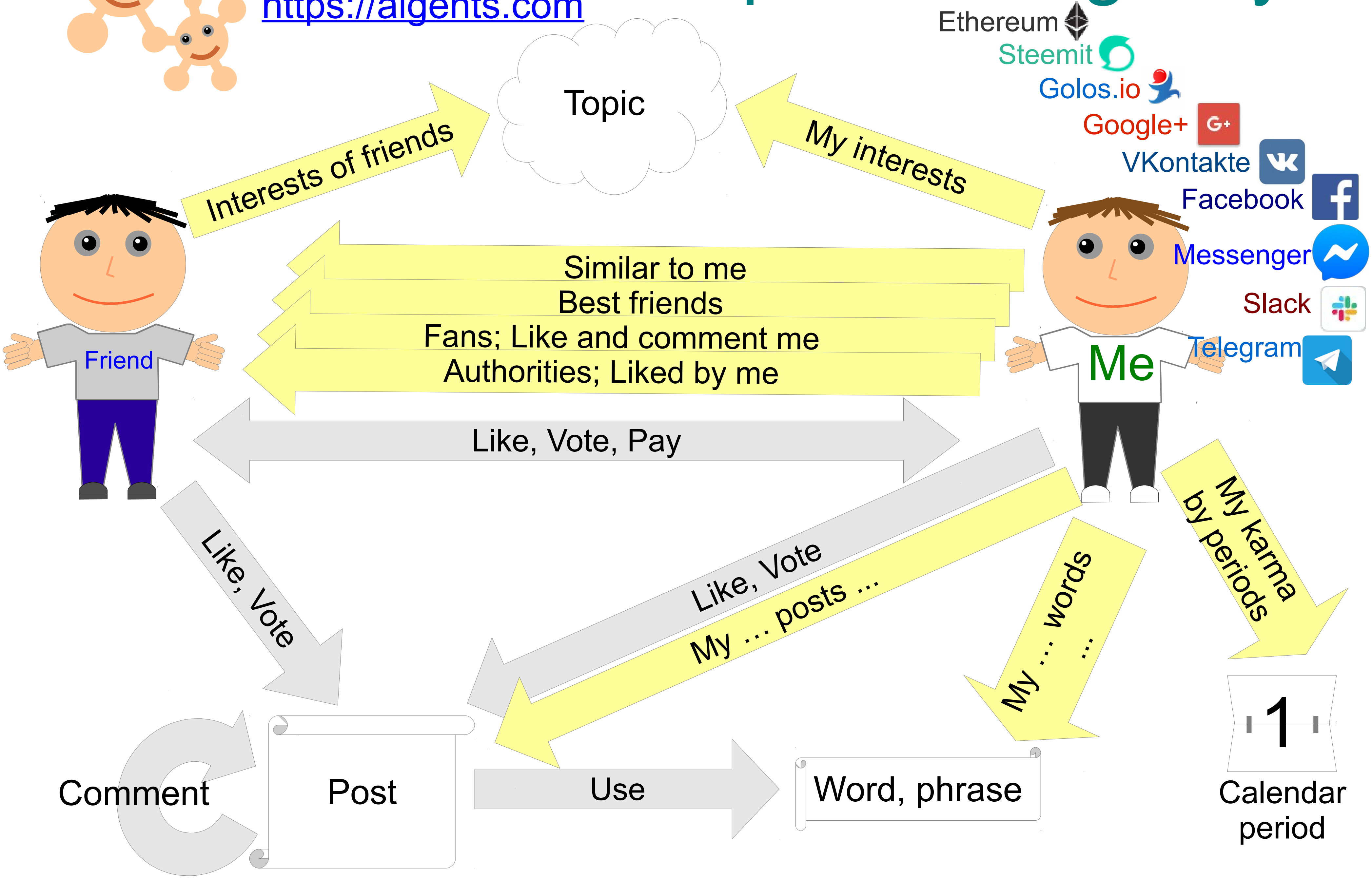
- Liquid ranking!
- Weighted ranking!
- Time scoping!
- Data openness!
- Code openness?
- Human precedence?
- Non-anonymity?
- No right to oblivion?

Results:

- Rank
- Reputation
- Karma
- Social capital

Reputation Agency

AIGENTS
https://aigents.com



AIGENTS

https://aigents.com

Social Computing

Ethereum
Steemit
Golos.io
Google+
VKontakte
Facebook
Messenger
Slack
Telegram

Best friends

$$B_{ij} = (L_{ij}+C_{ij})*(L_{ji}+C_{ji}) \,/\, Max_{j=1,J}((L_{ij}+C_{ij})*(L_{ji}+C_{ji}))$$

Fans and followers

$$F_{ij} = ((L_{ji}+C_{ji})/(1+L_{ij}+C_{ij}))/Max_{j=1,J}((L_{ji}+C_{ji})/(1+L_{ij}+C_{ij}))$$

Like and comment me

$$F'_{ij} = (L_{ji}+C_{ji}) \,/\, \mathrm{Max}_{j=1,J}(L_{ji}+C_{ji})$$

Authorities and opinion leaders

$$A_{ij} = ((L_{ij}+C_{ij})/(1+L_{ji}+C_{ji})) \,/\, Max_{j=1,J}((L_{ij}+C_{ij})/(1+L_{ji}+C_{ji}))$$

Liked by me

$$A'_{ij} = (L_{ij}+C_{ij}) \,/\, Max_{j=1,J}(L_{ij}+C_{ij})$$

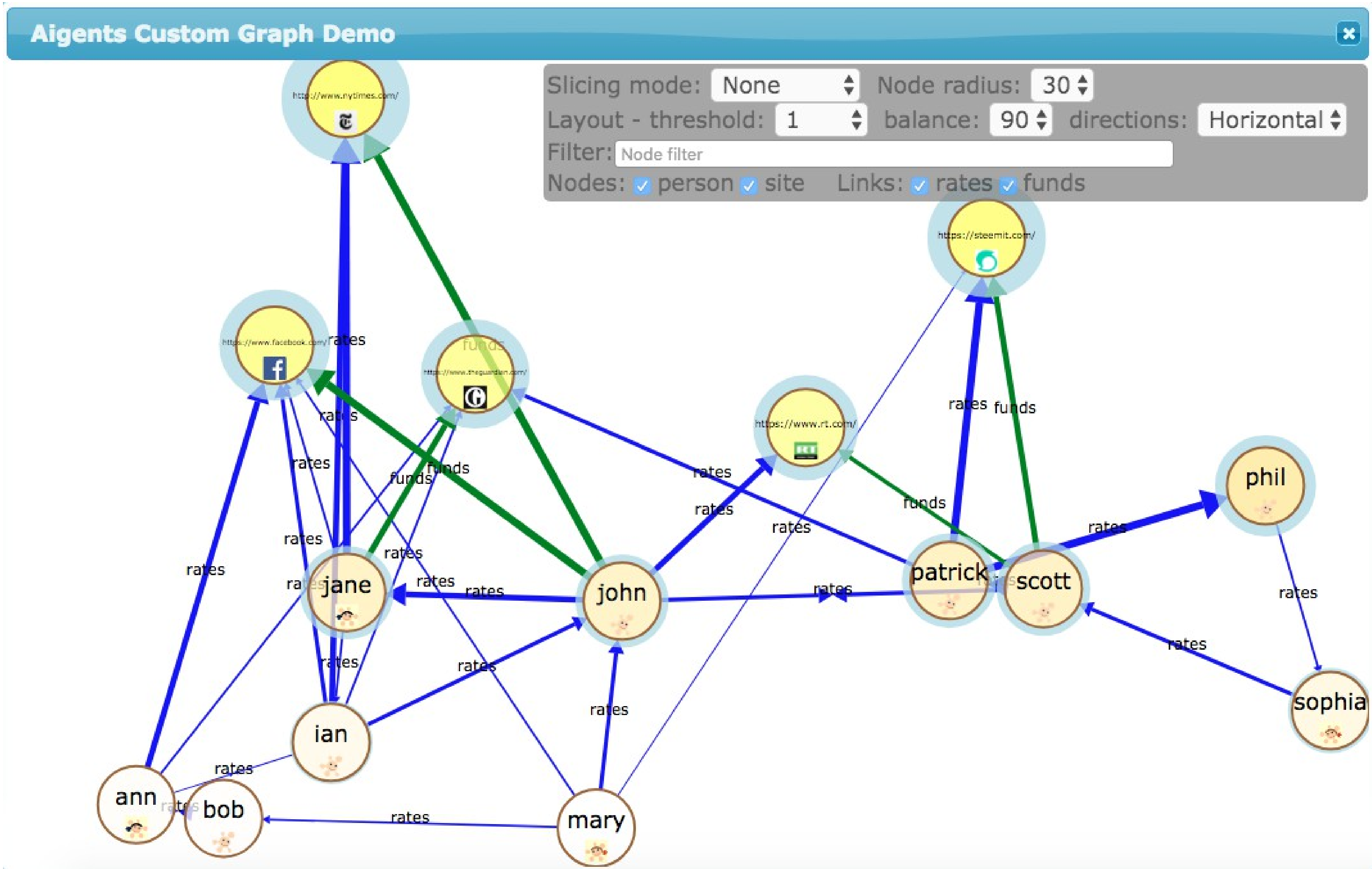
My karma by periods

$$K_{it} = \sum_{j,t}(L_{ji}+C_{ji}) \,/\, Max_{t=1,T}\sum_{j,t}(L_{ji}+C_{ji})$$

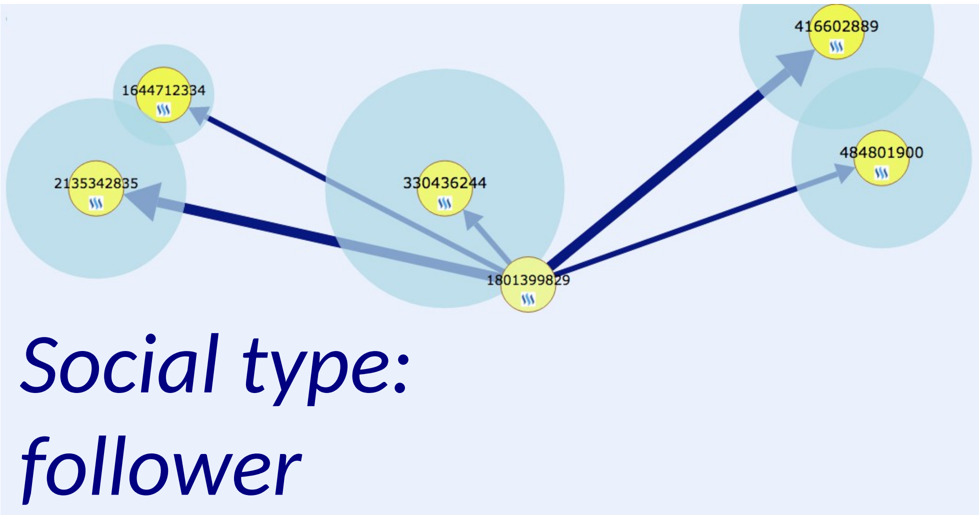
Explanation:

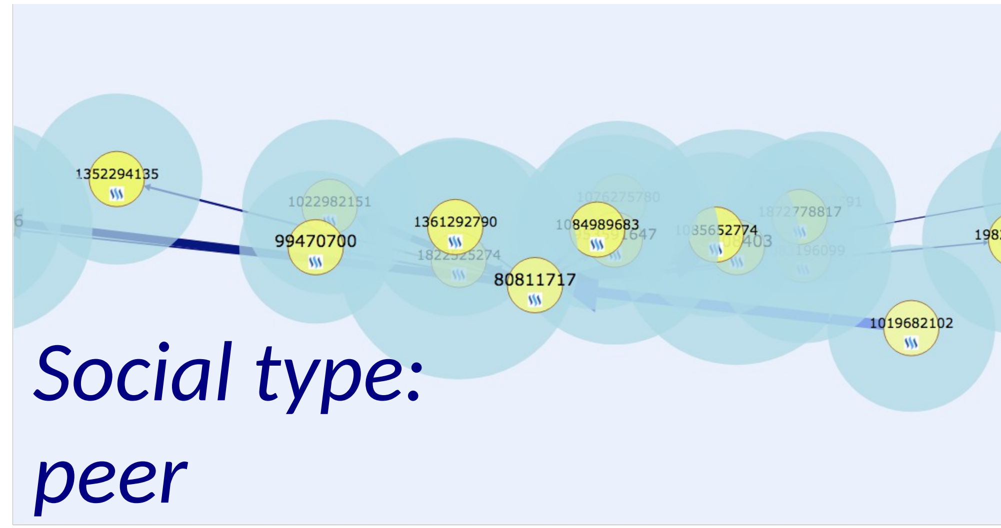
i – user being primarily explored
j – other user in context of i
L_{ji} – “likes” by user j to posts by i
C_{ji} – comments by user j to posts by i
1 – avoiding division by zero

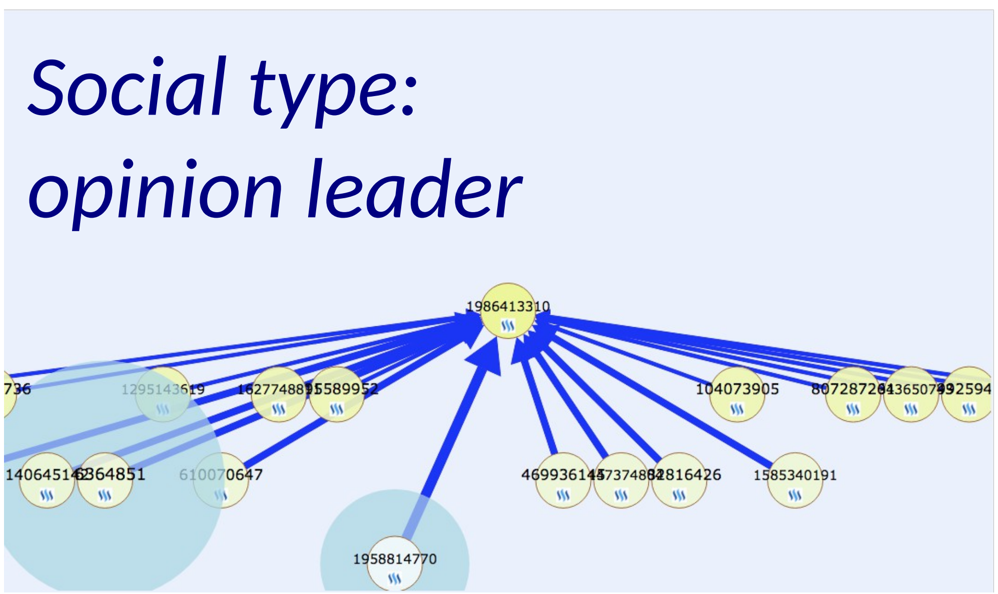
Social Networking: Helping community to understand opinion leaders and news agenda makers, helping leaders to understand audience (demonstration example, not real data).

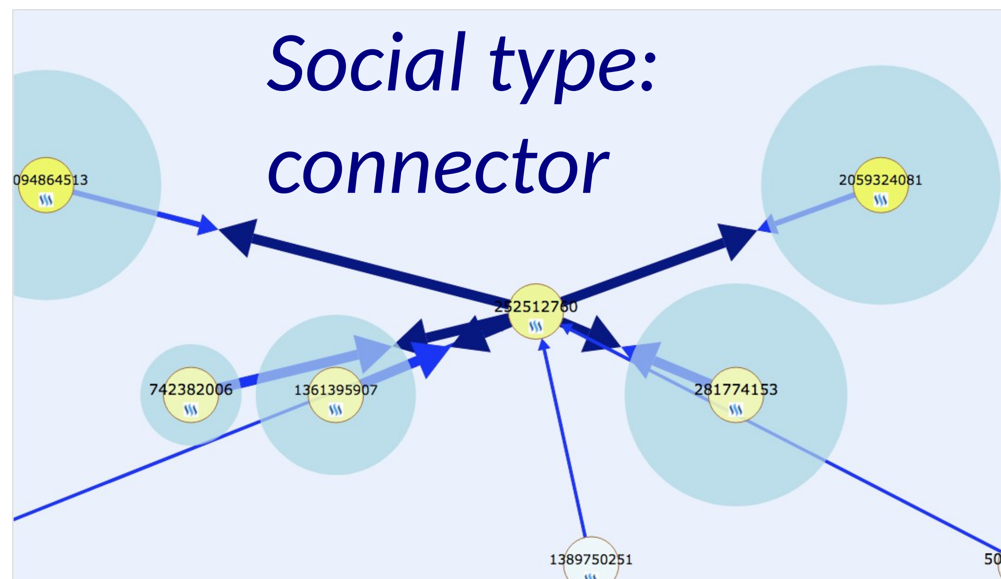


Social Networking: Helping community members to understand themselves better and perform more efficiently online – using tracks in social networks and online resources, capture interests, relationships, communication patterns and social structures.

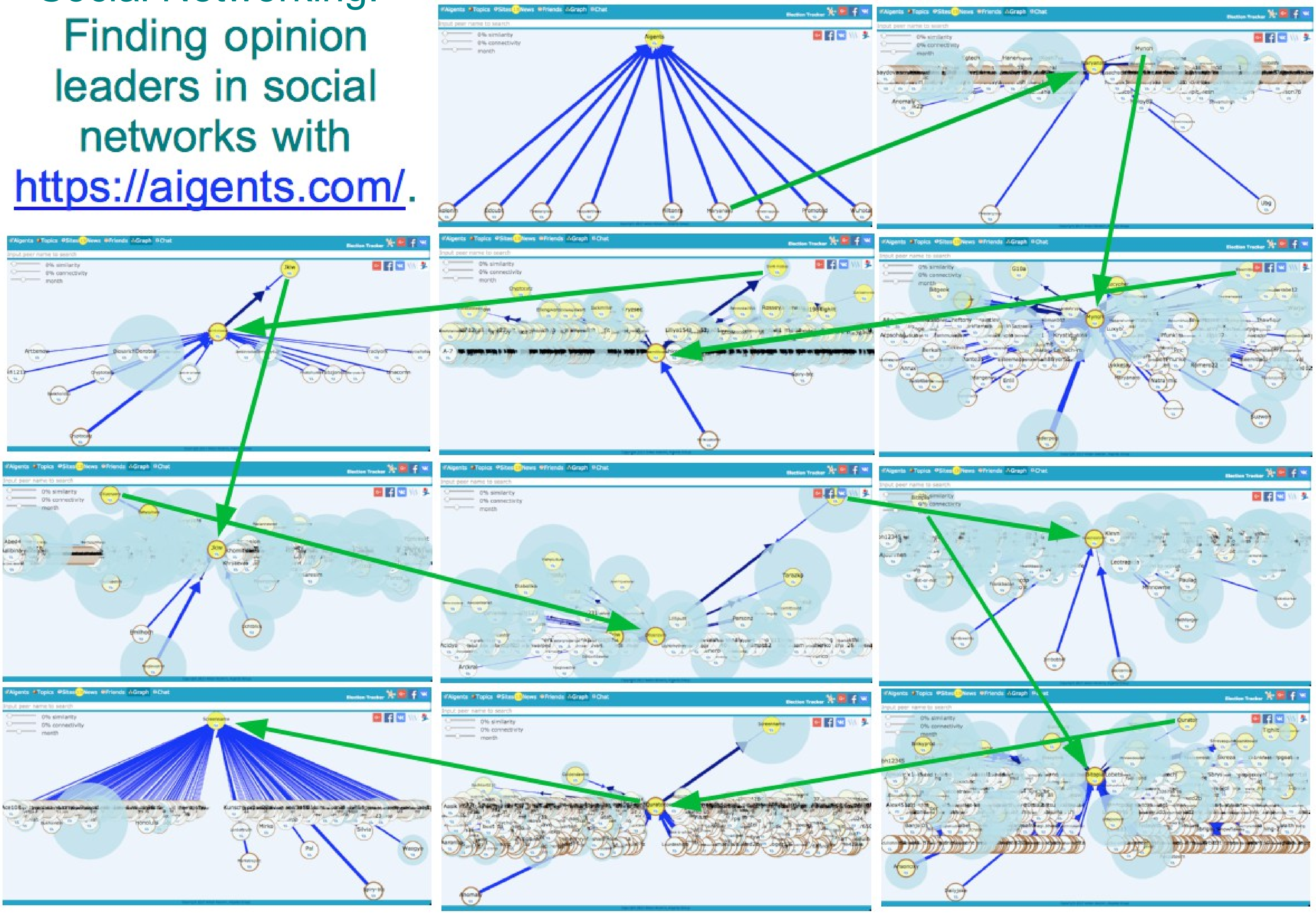






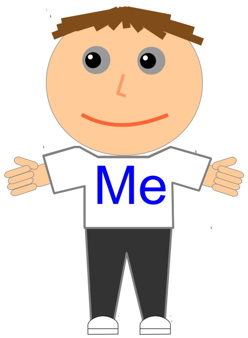


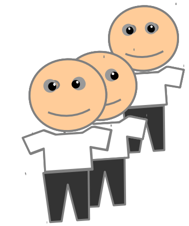
Social Networking: Finding opinion leaders in social networks with https://aigents.com/.



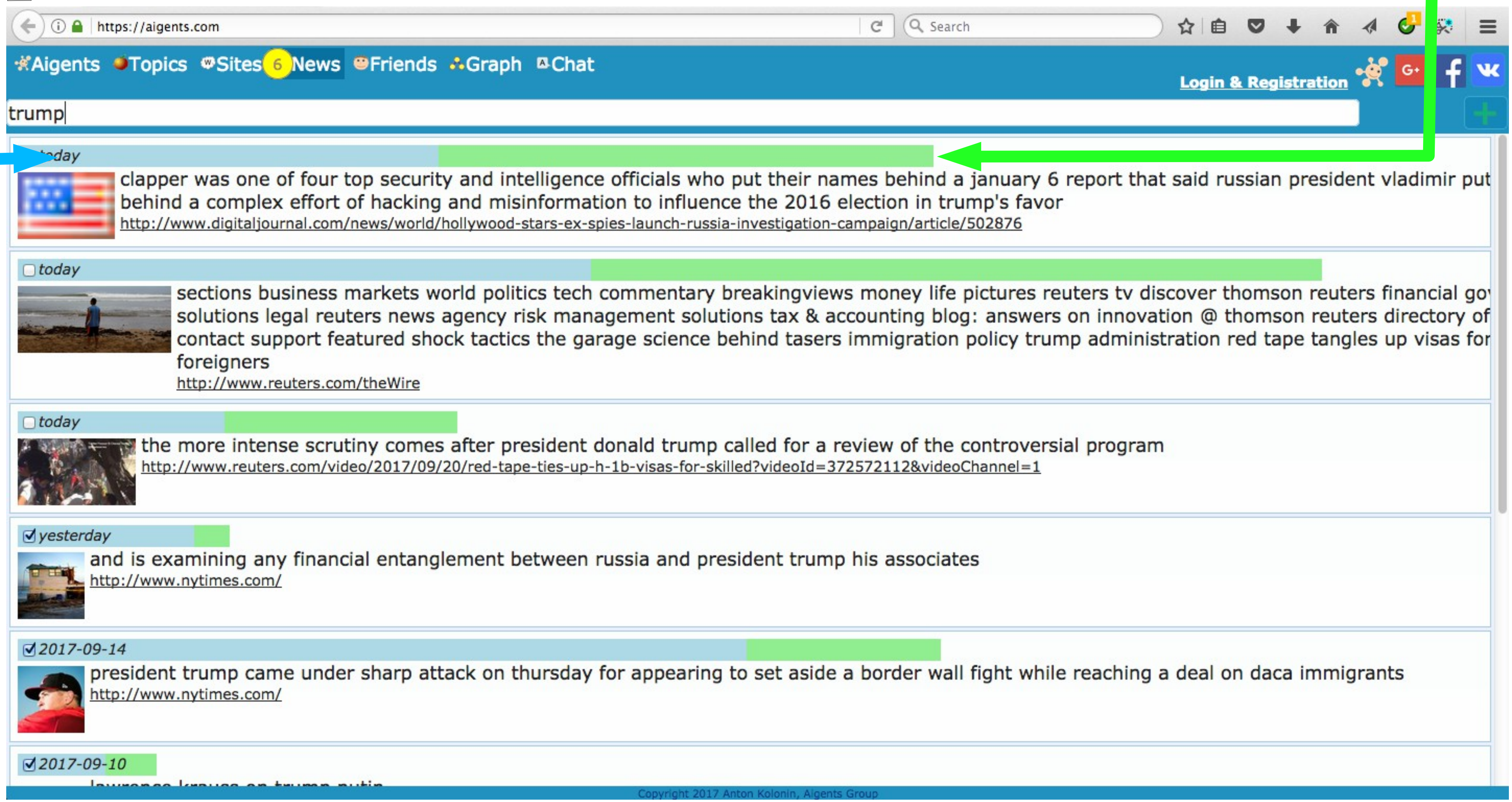
Collaborative News Filtering with Aigents:

Monitoring web pages and extracting textual information with account to Personal and Social relevances





Friends

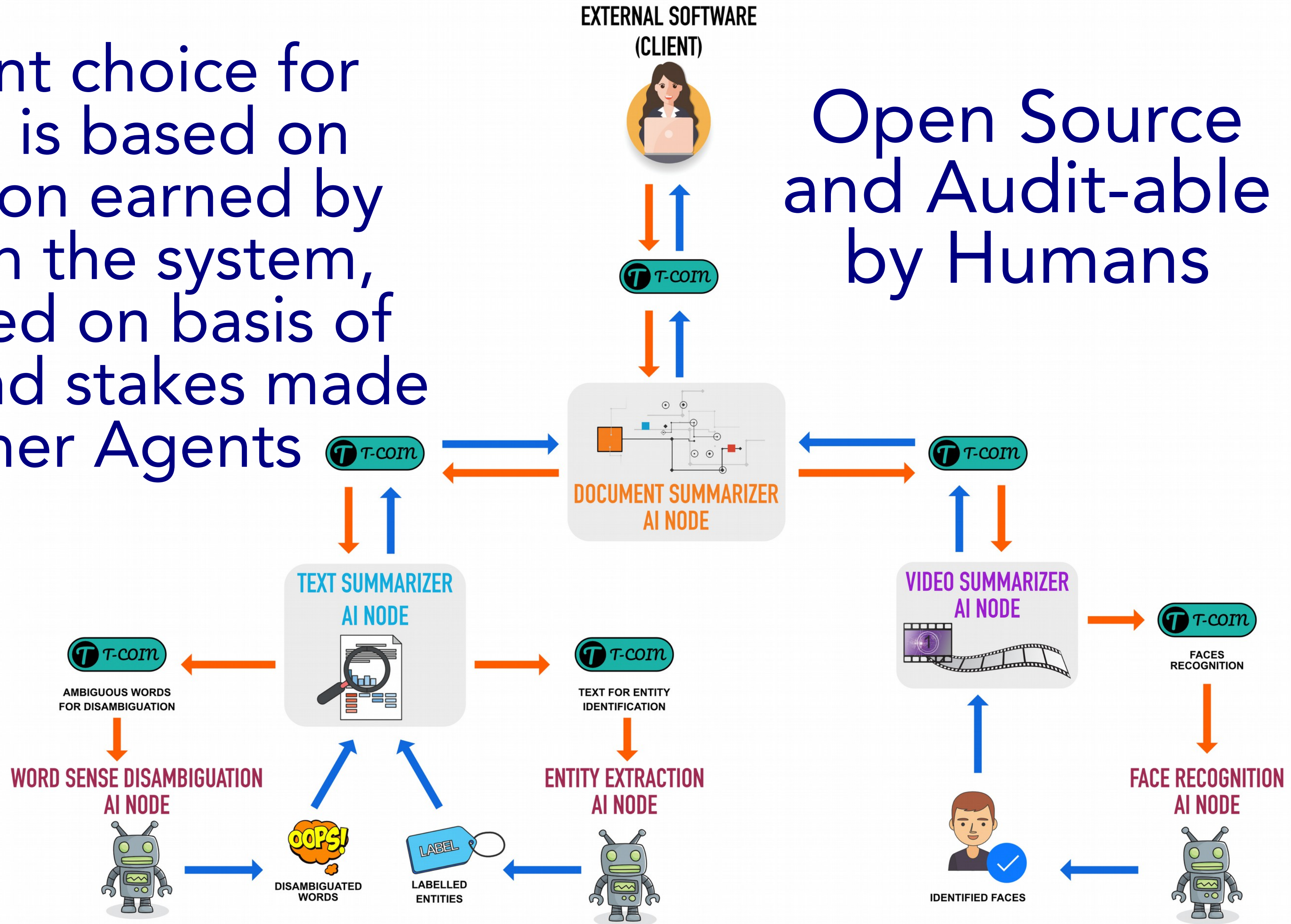


SingularityNET
https://singularitynet.io

Marketplaces and for Products and Services:

AI Agent choice for service is based on reputation earned by Agent in the system, computed on basis of ratings and stakes made by other Agents

Open Source and Audit-able by Humans



SingularityNET
https://singularitynet.io

Weighted Liquid Rank

Algorithm 1 Weighted Liquid Rank (simplified version)

Inputs:
1) Volume of rated transactions each with financial value of the purchased product or service and rating value evaluating quality of the product/service, covering specified period of time;
2) Reputation ranks for every participant at the end of the previous time period.

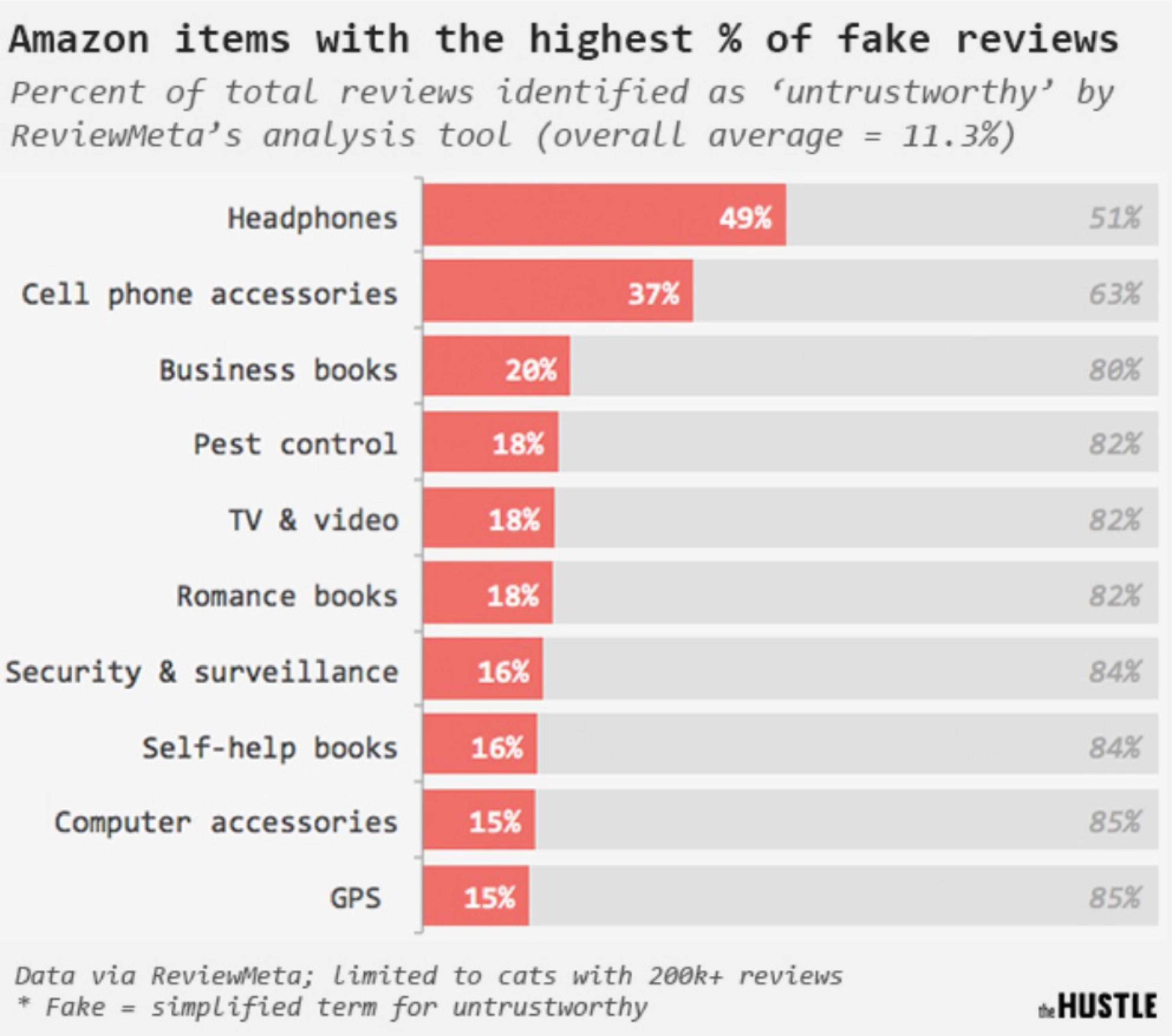
Parameters: List of parmeters, affecting computations - default value, logarithmic ratings, conservatism, decayed value, etc.

Outputs: Reputation ranks for every participant at the end of the previous time period.

```
foreach of transactions do
    let rater_value be rank of the rater at the end of previous period of default value
    let rating_value be rating supplied by trasaction rater (consumer) to ratee (supplier)
    let rating_weight be financial value of the transaction of its logarithm, if logarithmic ratings parameter is set to true
    sum rater_value*rating_value*rating_weight for every ratee
end foreach
do normalization of the sum of the muliplications per ratee to range 0.0-1.0, get differential_ranks
do blending of the old_ranks known at the end of previous peiod with differential_ranks based on parameter of conservatism, so that new_ranks = (old_ranks*conservatism+N*(1-differential_ranks)), using decayed value if no rating are given to ratee during the period
do normalization of new_ranks to range 0.0-1.0
return new_ranks
```

- R_d - default initial reputation rank;
- R_c - decayed reputation in range to be approached by inactive agents eventually;
- *C* - conservatism as a blending "alpha" factor between the previous reputation rank recorded at the beginning of the observed period and the differential one obtained during the observation period;
- *FullNorm* – when this boolean option is set to *True* the reputation system performs a full-scale normalization of incremental ratings;
- *LogRatings* - when this boolean option is set to *True* the reputation system applies *log10(1+value)* to financial values used for weighting explicit ratings;
- *Aggregation* - when this boolean option is set to *True* the reputation system aggregates all explicit ratings between each unique combination of two agents with computes a weighted average of ratings across the observation period;
- *Downrating* - when this boolean option is set to *True* the reputation system translates original explicit rating values in range *0.0-0.25* to negative values in range *-1.0* to *0.0* and original values in range *0.25-1.0* to the interval *0.0-1.0*.
- *UpdatePeriod* – the number of days to update reputation state, considered as observation period for computing incremental reputations.

Real Pain: Resist Reputation Gaming



SingularityNET
https://singularitynet.io

Marketplaces and for Products and Services:

Reputation System	AR	Good	Bad	Good2Bad	MVR	Bad/Good2Bad	LTS	PFS
None	2	42845	5164	2036	8.3	2.54	4.8%	39%
Regular RS	2	43994	5692	2291	7.7	2.48	5.2%	40%
Weighted Rank	2	42884	5391	2291	8.0	2.35	5.3%	42%
Weighted Denominated	2	42332	6100	2333	6.9	2.61	5.5%	38%
No RS	10	42763	1129	2036	37.9	0.55	4.8%	180%
Regular RS	10	45705	991	2291	46.1	0.43	5.0%	231%
Weighted Rank	10	42425	1242	204	34.2	6.09	0.5%	16%
Weighted Denominated	10	42338	1022	2296	41.4	0.45	5.4%	225%
No RS	20	42763	561	2036	76.2	0.28	4.8%	363%
Regular RS	20	45705	491	2291	93.1	0.21	5.0%	467%
Weighted Rank	20	45672	570	204	80.1	2.79	0.4%	36%
Weighted Denominated	20	42338	505	2296	83.8	0.22	5.4%	455%

Expected summary for Reputation System usability with no Liquid Rank, where reputaion of the raters can not be accessed (based on "10 agents operating during 10 days with FR=4 (fairness ratio), TR=1, AR=2,10,20, supliers=50%, consumers=50%"):

1) MVR below 10 - better not use any reputation system at all
2) MVR above 10 - A MUST to use "Weighted Rank" based reputaion system
3) For MVR below 10 - need to find way to access reputation of the raters

SingularityNET
https://singularitynet.io

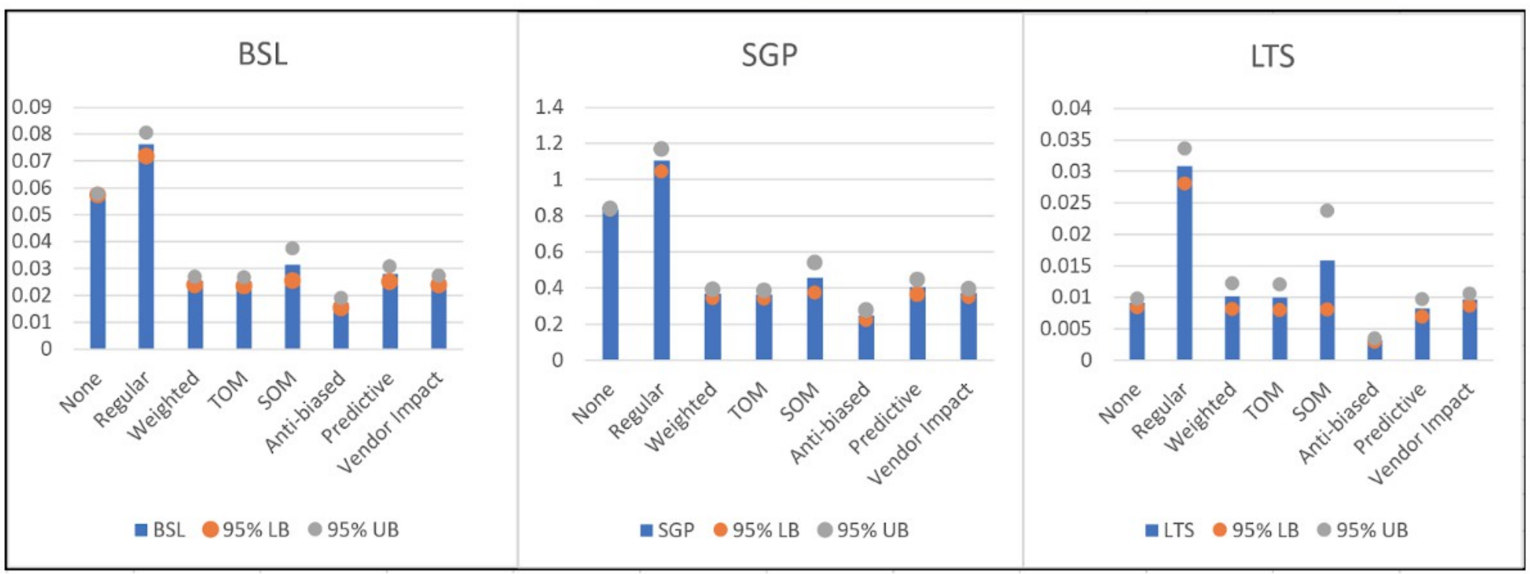
Reputation System for Marketplaces:

Scam Period	Reputation System	Loss to Scam (LTS)	Profit from Scam (PFS)	LTS Relative Decrease	PFS Relative Decrease
182	No	2.4%	44%		
182	Regular	2.7%	49%	-13%	-13%
182	Weighted	2.3%	42%	2%	3%
182	TOM-based	1.4%	30%	41%	31%
182	SOM-based	2.2%	40%	8%	7%
92	No	3.0%	54%		
92	Regular	3.5%	65%	-19%	-20%
92	Weighted	2.8%	52%	5%	4%
92	TOM-based	1.7%	36%	43%	33%
92	SOM-based	2.6%	47%	13%	12%
30	No	3.9%	73%		
30	Regular	4.7%	86%	-19%	-18%
30	Weighted	3.3%	59%	17%	19%
30	TOM-based	1.5%	31%	63%	58%
30	SOM-based	1.5%	27%	63%	63%
10	No	4.4%	81%		
10	Regular	4.7%	88%	-7%	-8%
10	Weighted	3.0%	54%	33%	33%
10	TOM-based	0.2%	3%	**96%**	**96%**
10	SOM-based	0.3%	6%	**93%**	**93%**

- No reputation system: participants are making decisions relying only on their own memories and not referring to any reputation system.
- Regular reputation system: standard version of reputation system. Does not take into account any factors other than values of ratings that consumers make to suppliers.
- Weighted reputation system: When considering ratings as regular reputation system does, accounts to financial values of transactions between participants so that rating values are weighted by costs of transactions that are rated.
- TOM-based reputation system: In addition to weighting ratings with financial values per-transaction, weights the ratings based on the rater's time on the market (TOM) as a "proof-of-time". That is, the raters (buyers) are implicitly rated based on how long have they been on the market. So, rating by buyer with a longer history influences reputation of a seller more than the one made by rater with shorter history.
- SOM-based reputation system: In addition to weighting ratings with financial values per-transaction, weights the ratings based on rater's spendings on the market (SOM) as a "proof-of-burn" value. That is, the raters (buyers) are implicitly rated based on how much they spend on this market. So, rating by buyer with a lot of spendings influences reputation more than the one made by rater with smaller spendings.

Reputation System for Marketplaces against Reputation Gaming

Reputation System Type	OMU	LTS	BSL	SGP
None	0.99	0.01	0.06	0.83
Regular	0.97	0.03	0.08	1.11
Weighted	0.99	0.01	0.03	0.37
TOM	0.99	0.01	0.03	0.36
SOM	0.99	0.02	0.03	0.46
Anti-biased	1.00	0.00	0.02	0.25
Predictive	0.99	0.01	0.03	0.40
Vendor Impact	0.99	0.01	0.03	0.37



- Table and charts presenting performance of financial metrics for different reputation systems using adaptive simulation. The charts show a 95% confidence interval for the highest and lowest the true values could be (had we repeated the simulations indefinitely).
- Compared results between "Regular" and "Weighted" reputation system, TOM/SOM (time/spendings on the market) based ones, "Anti-biased", "Predictive" and "Vendor Impact" reputation system. The optimisation was targeting to make OMU (Organic Market Utility) higher and making the other metrics such as LTS (Loss to Scam), BSL (Buyers Satisfaction Loss), SGP (Seller Gaming Profit) lower.
- Use of "Regular" reputation system makes all financial metrics instantly worse than in the case when no reputation system is used at all - just because of the reputation gaming redirecting the market to the dishonest providers increasing their profits (SGP), decreasing the volume of honest market (OMU) and causing losses for buyers (LTS and BSL). We can also see than most ot the reputation system configurations, such as "Anti-biased", Weighted, TOM, "Predictive", and "Vendor Impact" improve the financial metrics. The LTS column shows that the best "Anti-biased" reputation system configuration reduced the total market volume spent on scams to zero making the OMU approached 1.00, rounding to the first two decimal places.

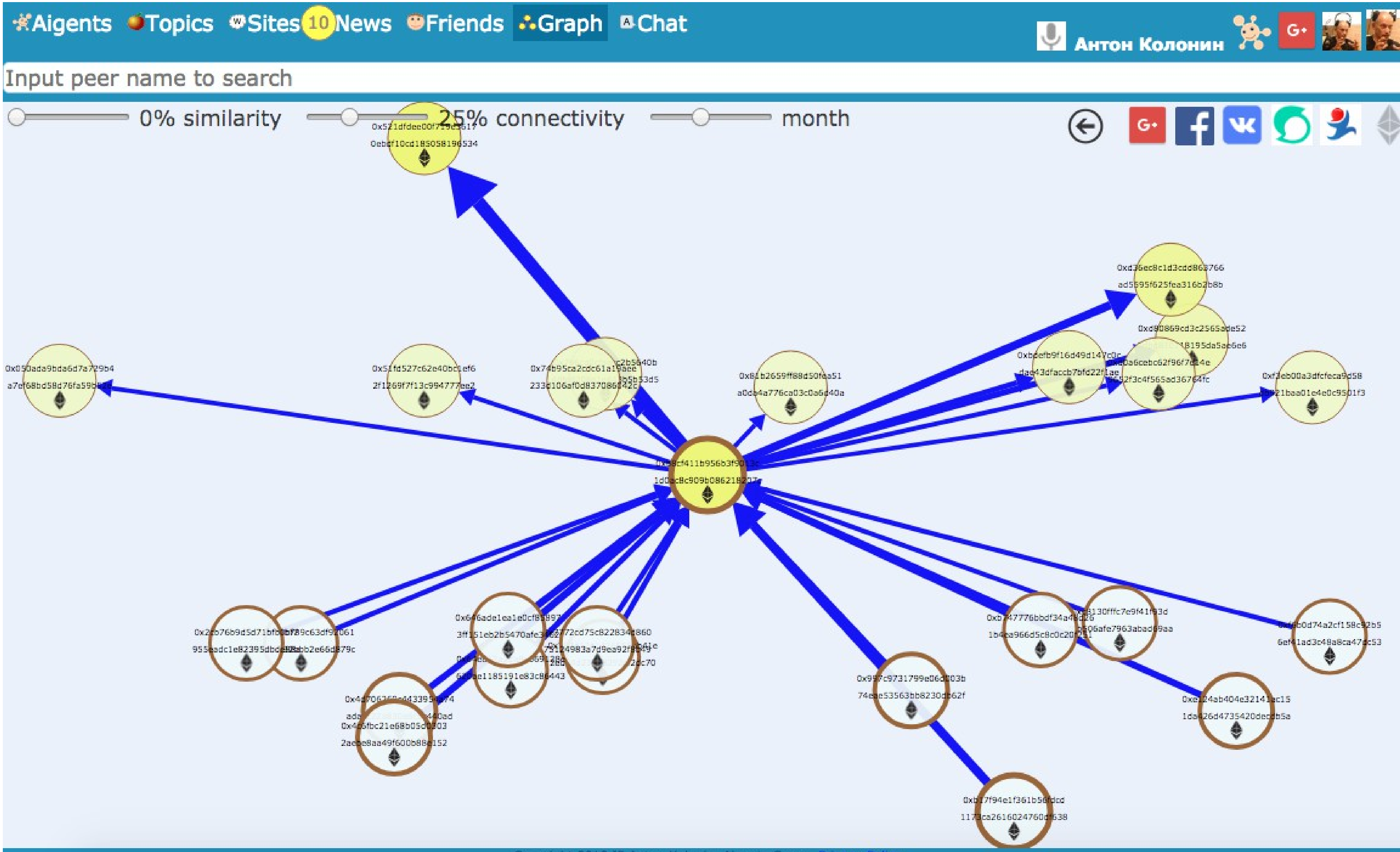
Reputation System for Marketplaces:

SingularityNET
https://singularitynet.io

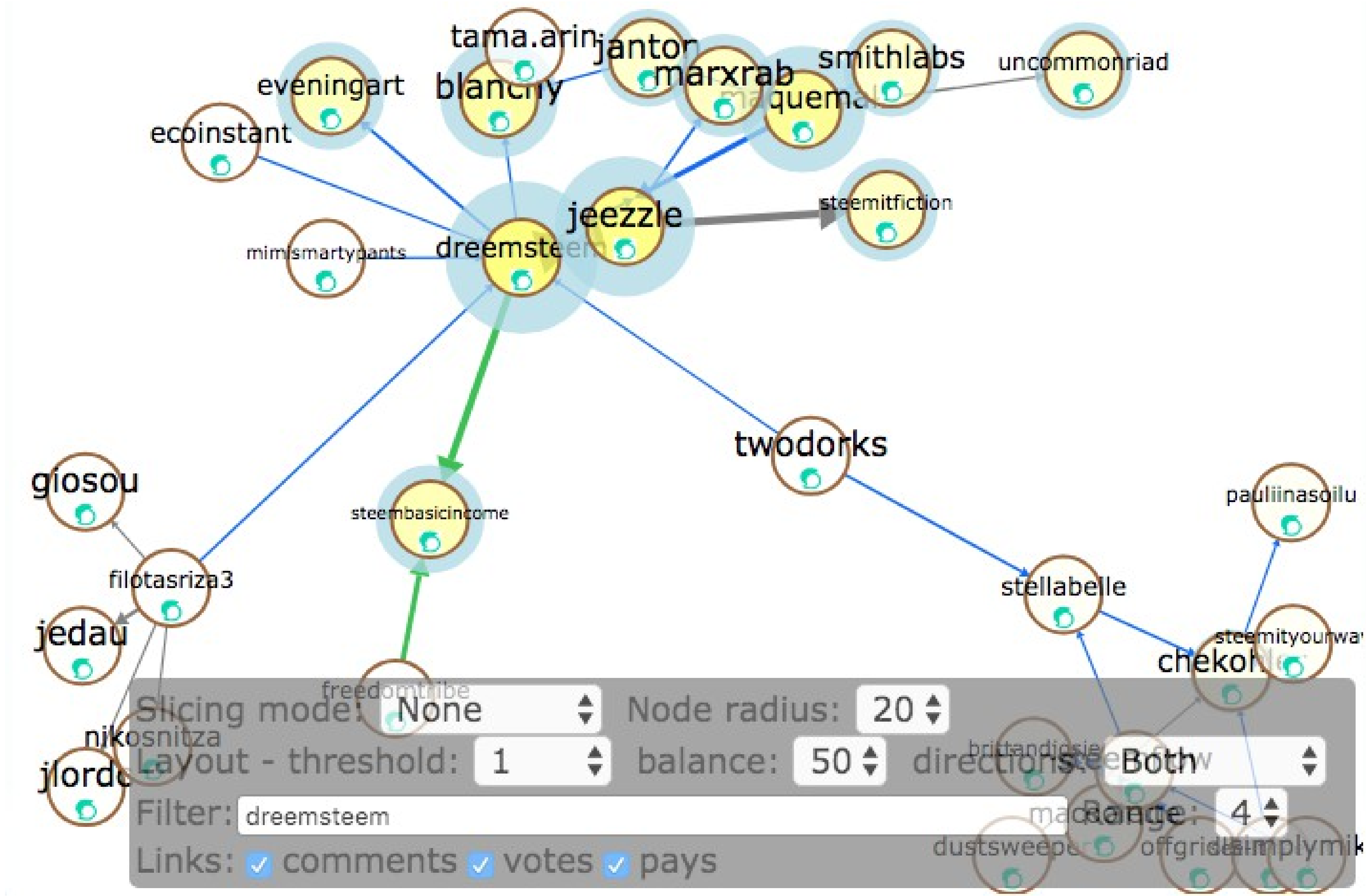
Using Reputation System for protection from scam identifying dishonest suppliers.

Ranks of Suppliers, dishonest Supplier (including alias) in red and honest suppliers in blue

Financial Security: Making sense of financial ecosystem, cash flows and transaction patterns in blockchains such as Ethereum.



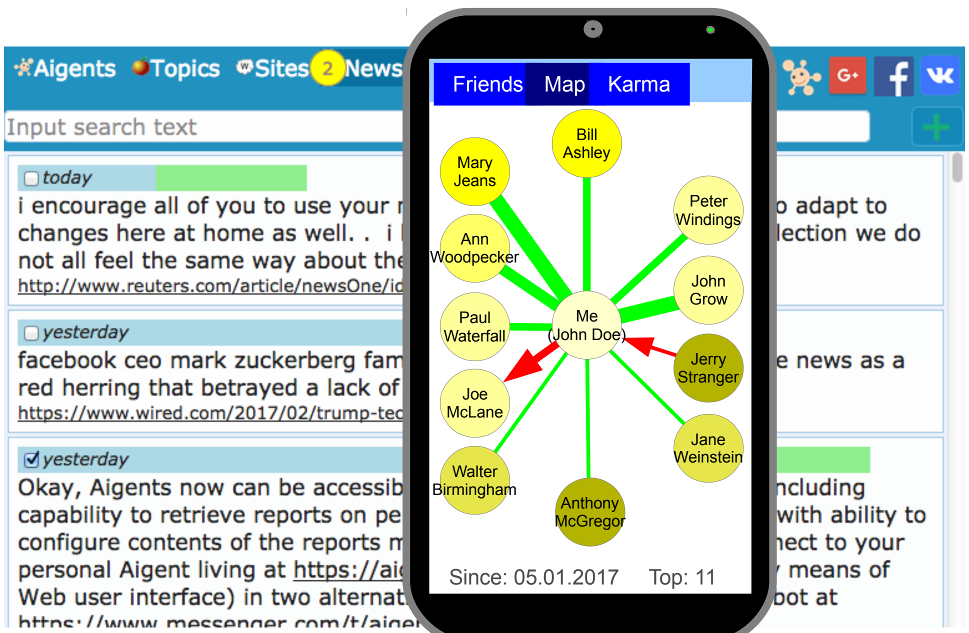
Financial Security: Making sense of complex socio-financial network dynamics based on synergy of financial, textual and emotional interactions in distributed online platform such as Steemit.



Financial Security: Evaluate trustworthiness and its dynamics for anonymous accounts in open public networks based on reputations computed on explicit and implicit rating data.

Socio-psychological Security: Encouraging users to conduct positive and effective communications with partners while guarding users from being manipulated themselves or being offensive to others.

I connect my “virtual agent” to my social networks and communication channels and let it learn about my partners and preferences.



“Agent” extracts information from networks and online communications automatically, analyses all posts, comments and messages and alerts once there are important messages coming in or out – encouraging and positive or manipulative and offensive.

Reputation systems and liquid democracy may become key elements in human-computer environments

Anton Kolonin
akolonin@aigents.com

iHeart

Novosibirsk State University
*THE REAL SCIENCE

NEXUS

ocean

zkylos

AIGENTS
https://aigents.com

SingularityNET
https://singularitynet.io

TravelChain

CISCO

HUAWEI

HANSON ROBOTICS

