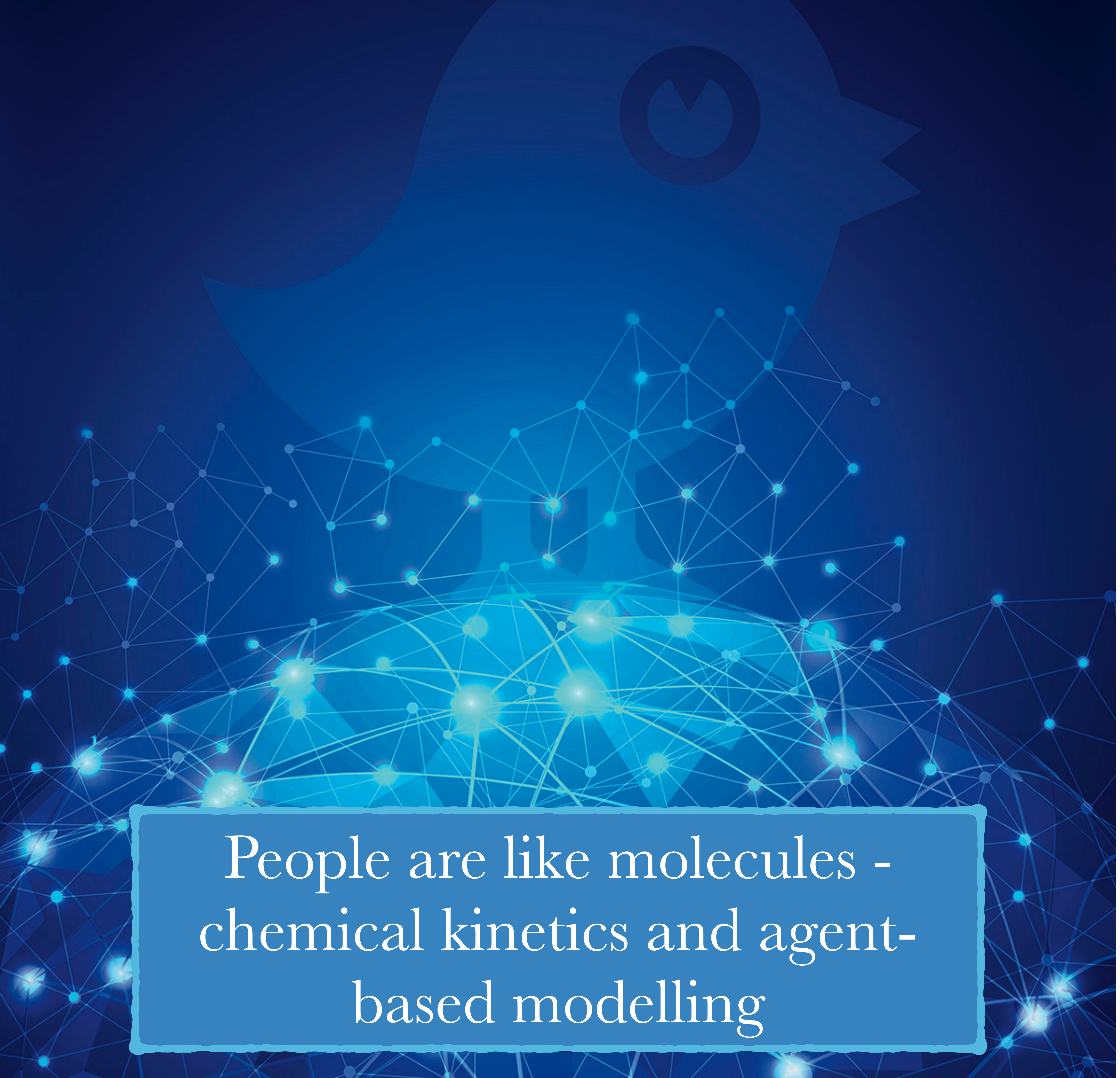
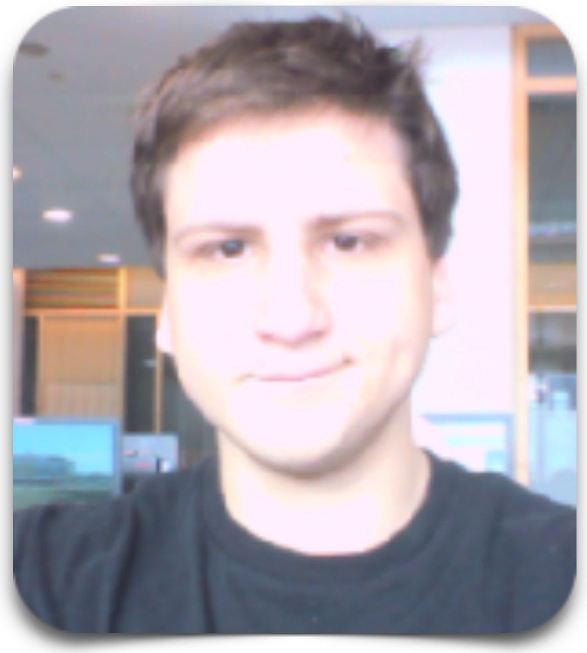
I. Tamblyn



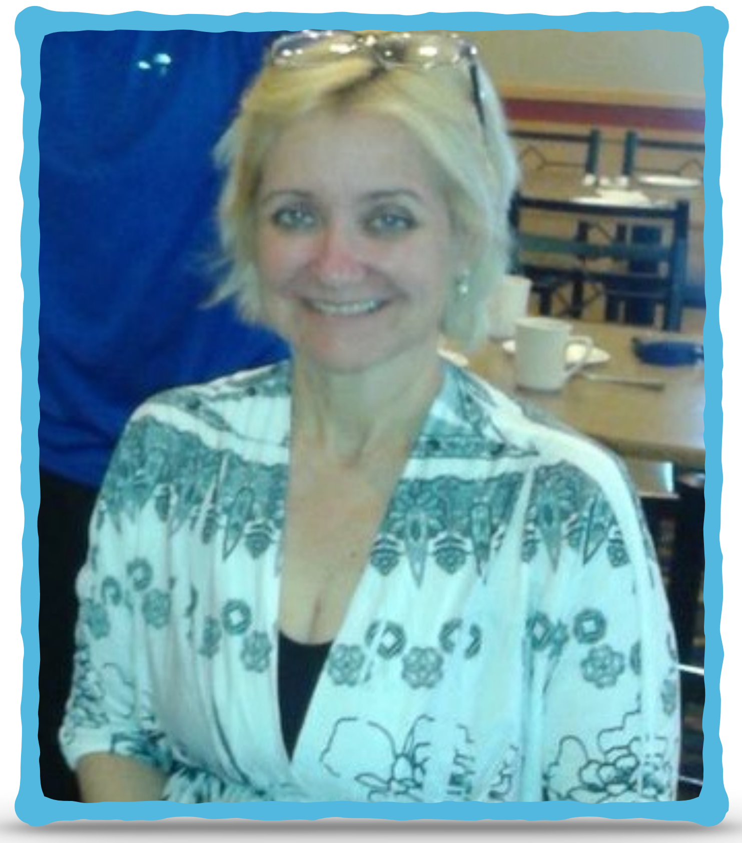
People are like molecules - chemical kinetics and agent-based modelling

K. Ryczko



A. Domurad

J. Grover



L. Kettle

N. Buhagier

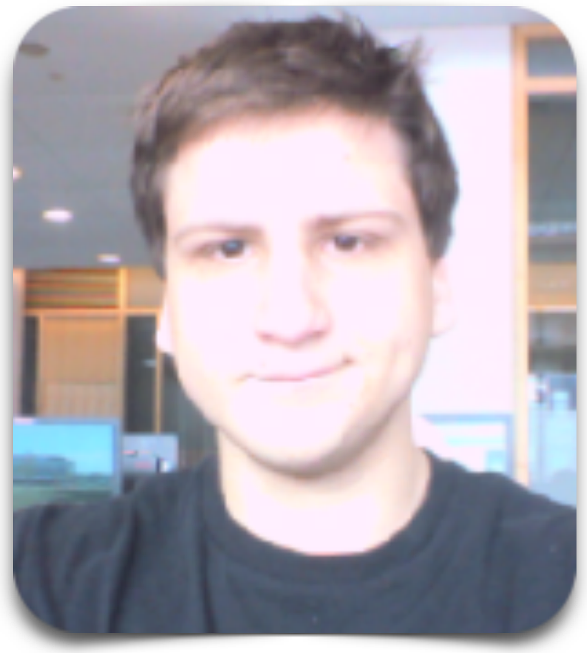
http://clean.energyscience.ca

http://hashkat.org

I. Tamblyn

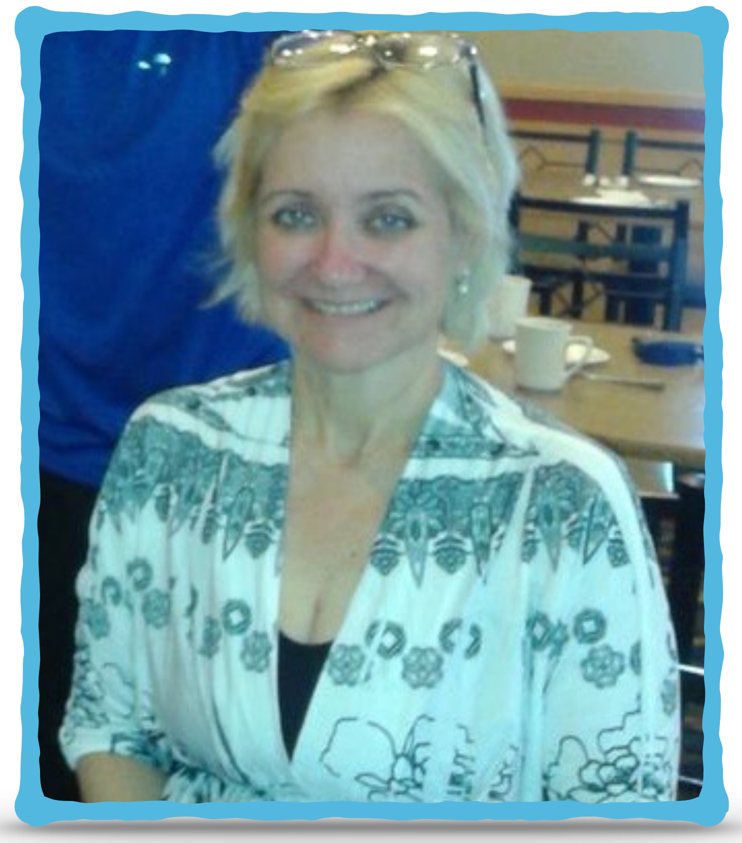
People are like molecules - chemical kinetics and agent-based modelling

K. Ryczko



A. Domurad

J. Grover



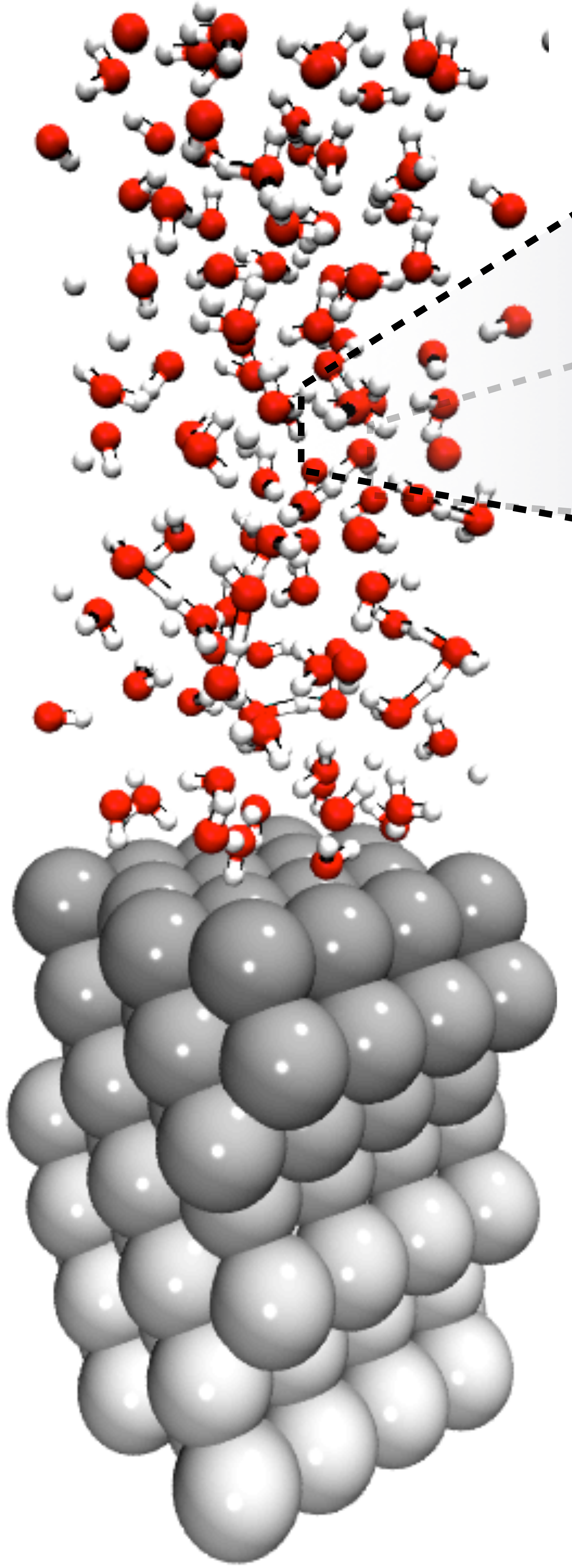
L. Kettle

N. Buhagier

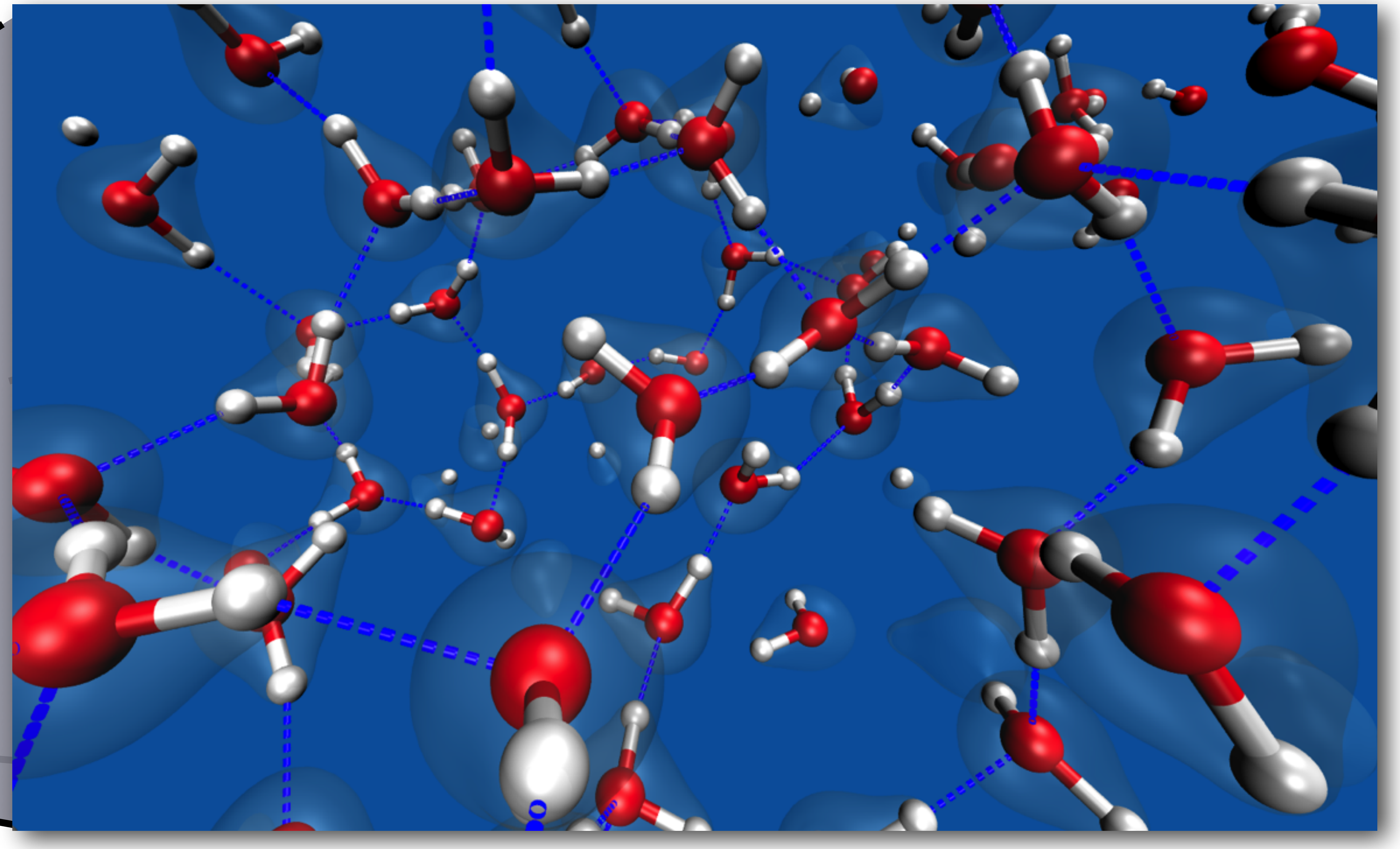
http://clean.energyscience.ca

http://hashkat.org

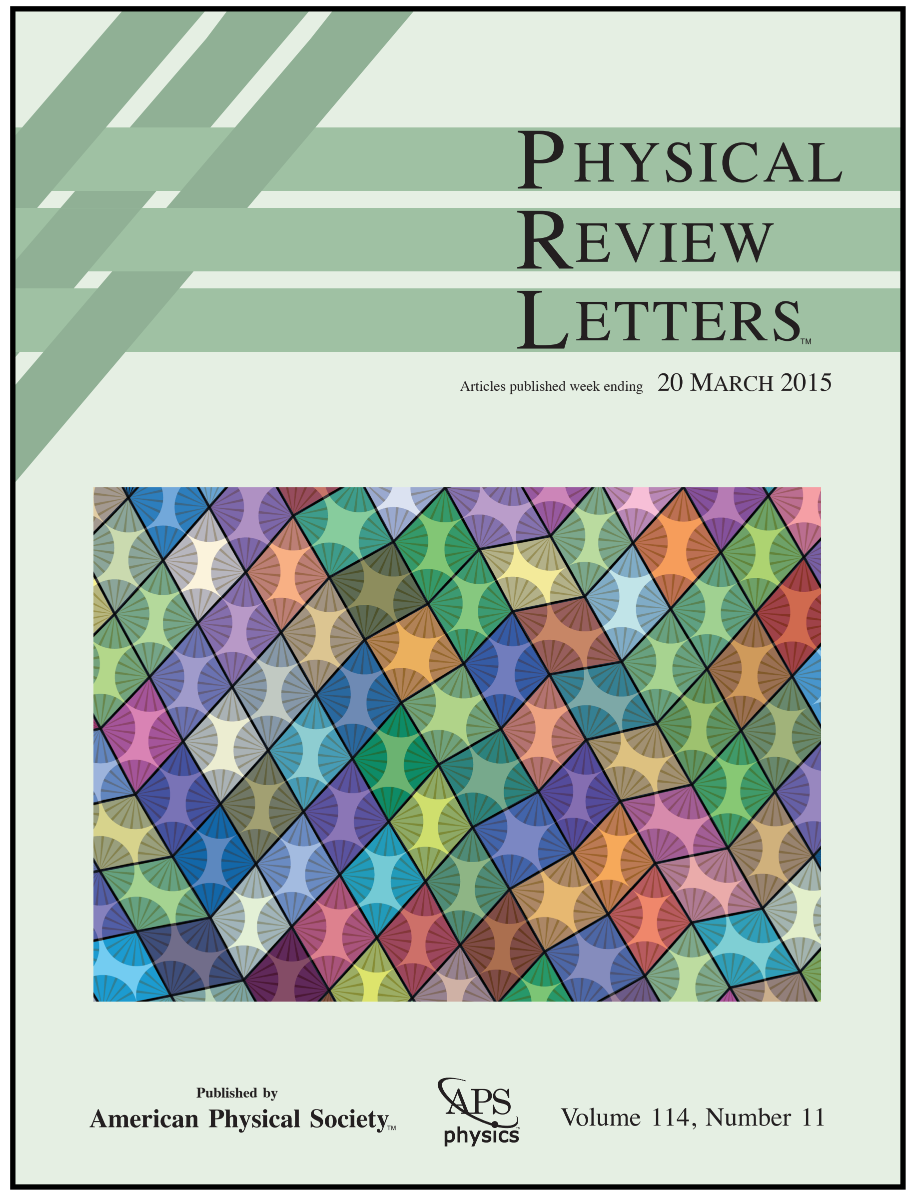
liquid water



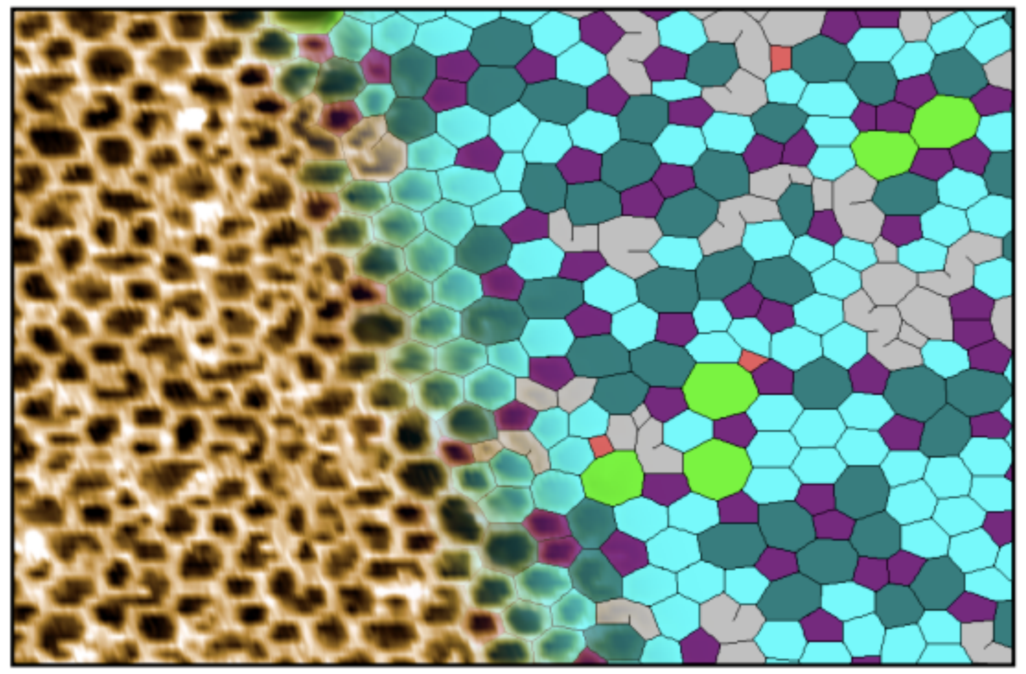
Platinum electrode



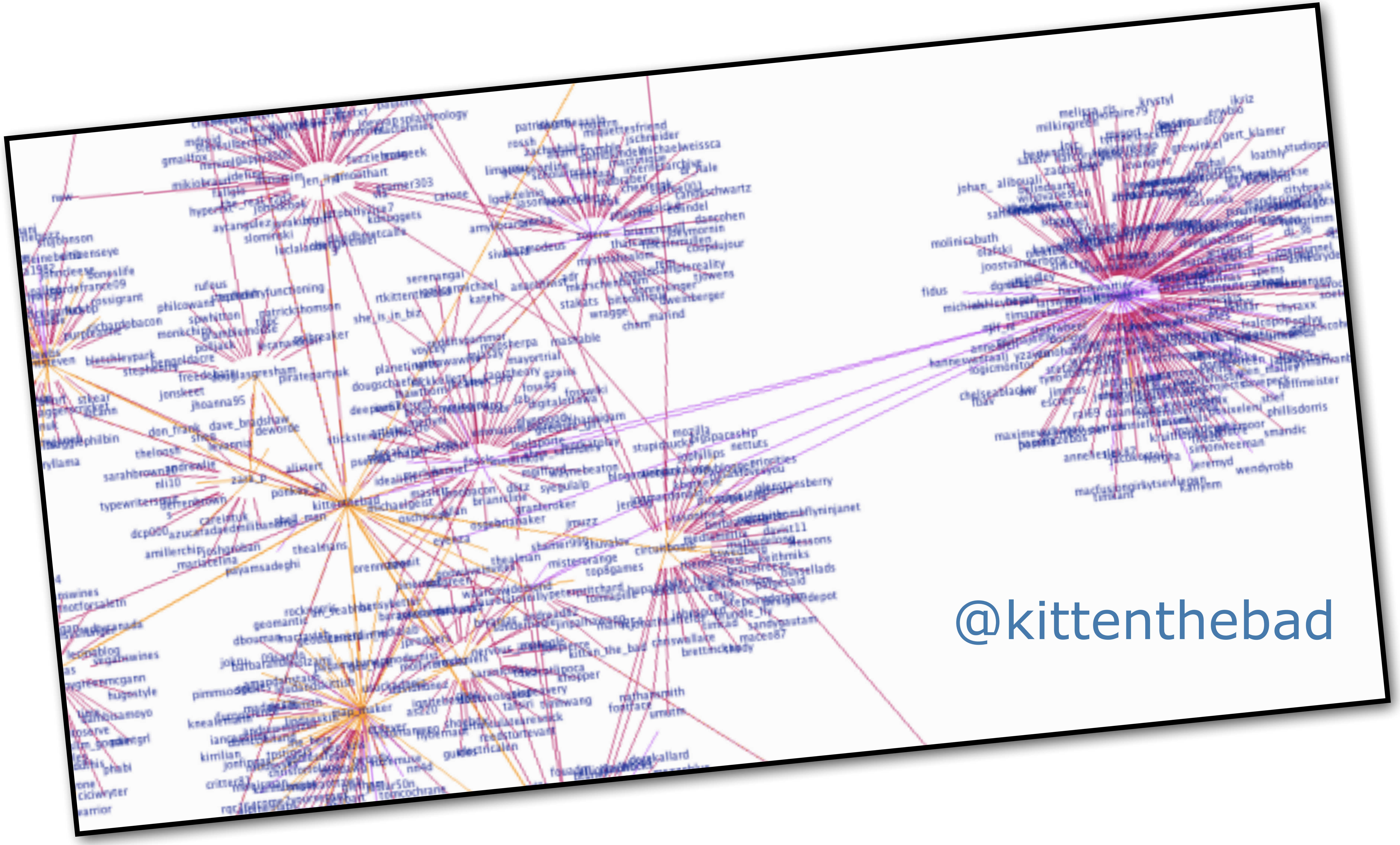
H-bonding



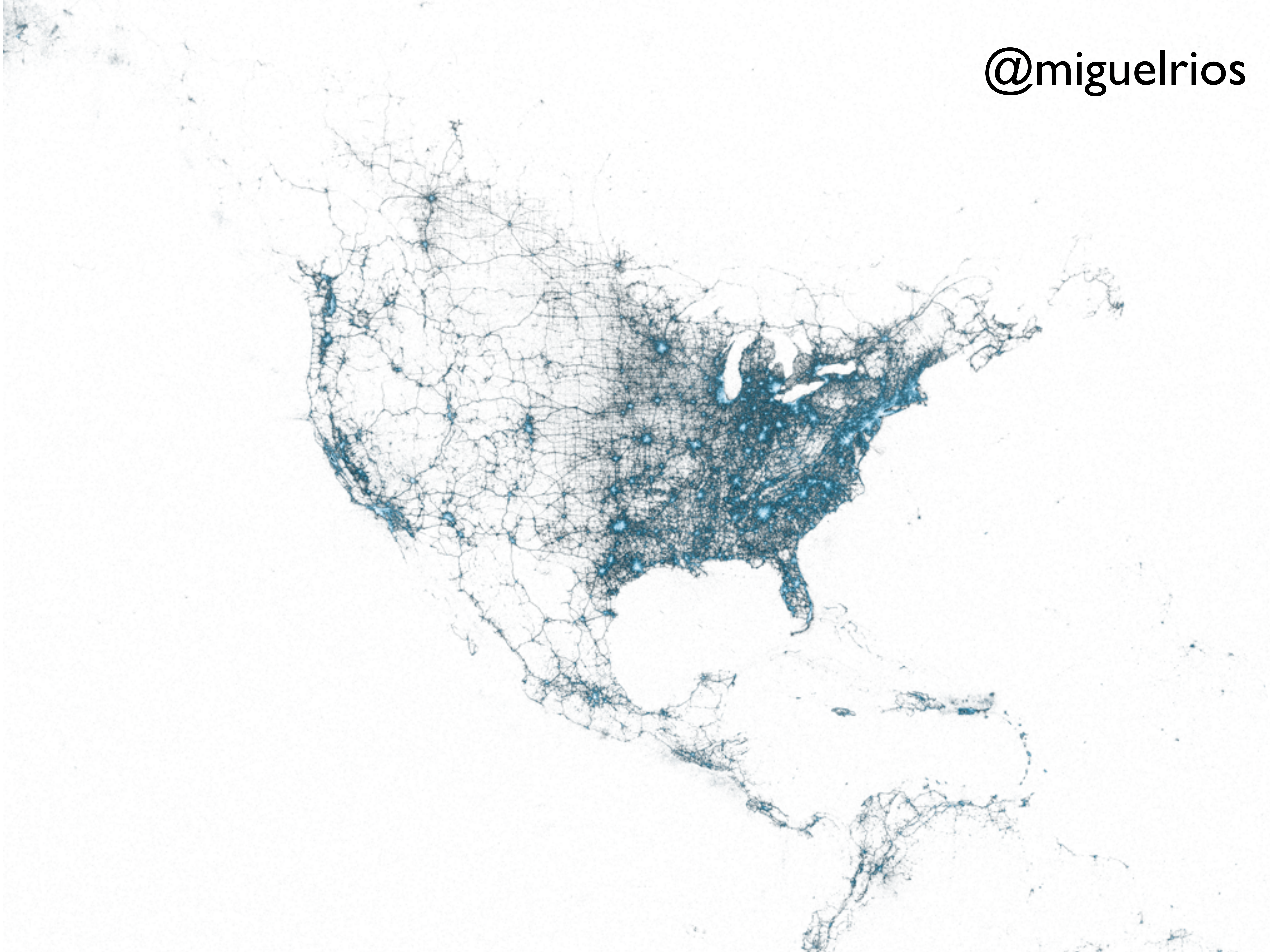
self assembled networks - OPV



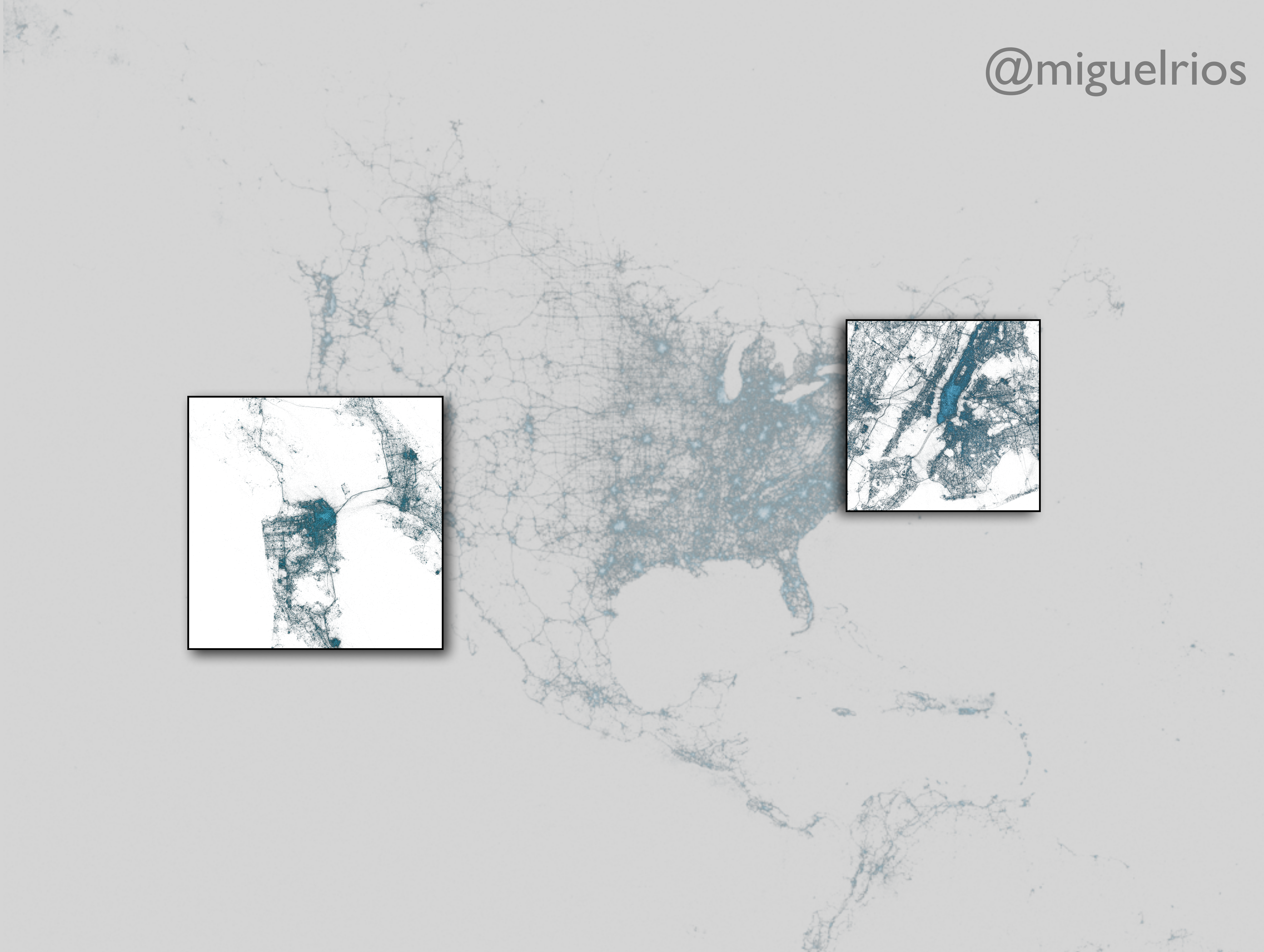
Whitelam, <u>Tamblyn</u>, Beton, and Garrahan, *Phys. Rev. Lett.* 114, 115702 (2015)
Whitelam, <u>Tamblyn</u>, *et al.*, *Phys. Rev. X*, 3, (2014)
Whitelam, <u>Tamblyn</u>, Beton, and Garrahan, *Phys. Rev. Lett.* 108, 035702 (2012)


@kittenthebad

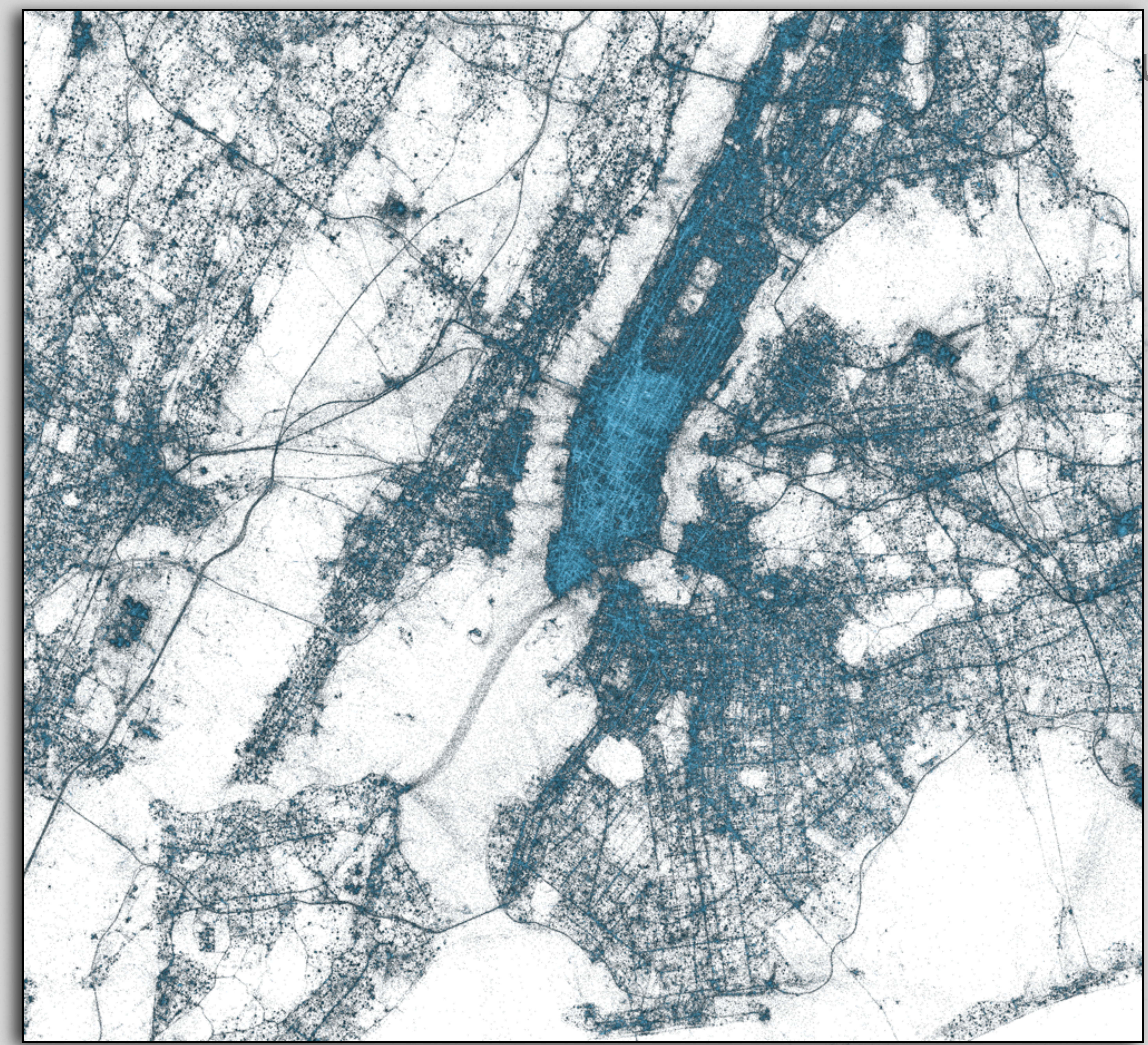
@miguelrios

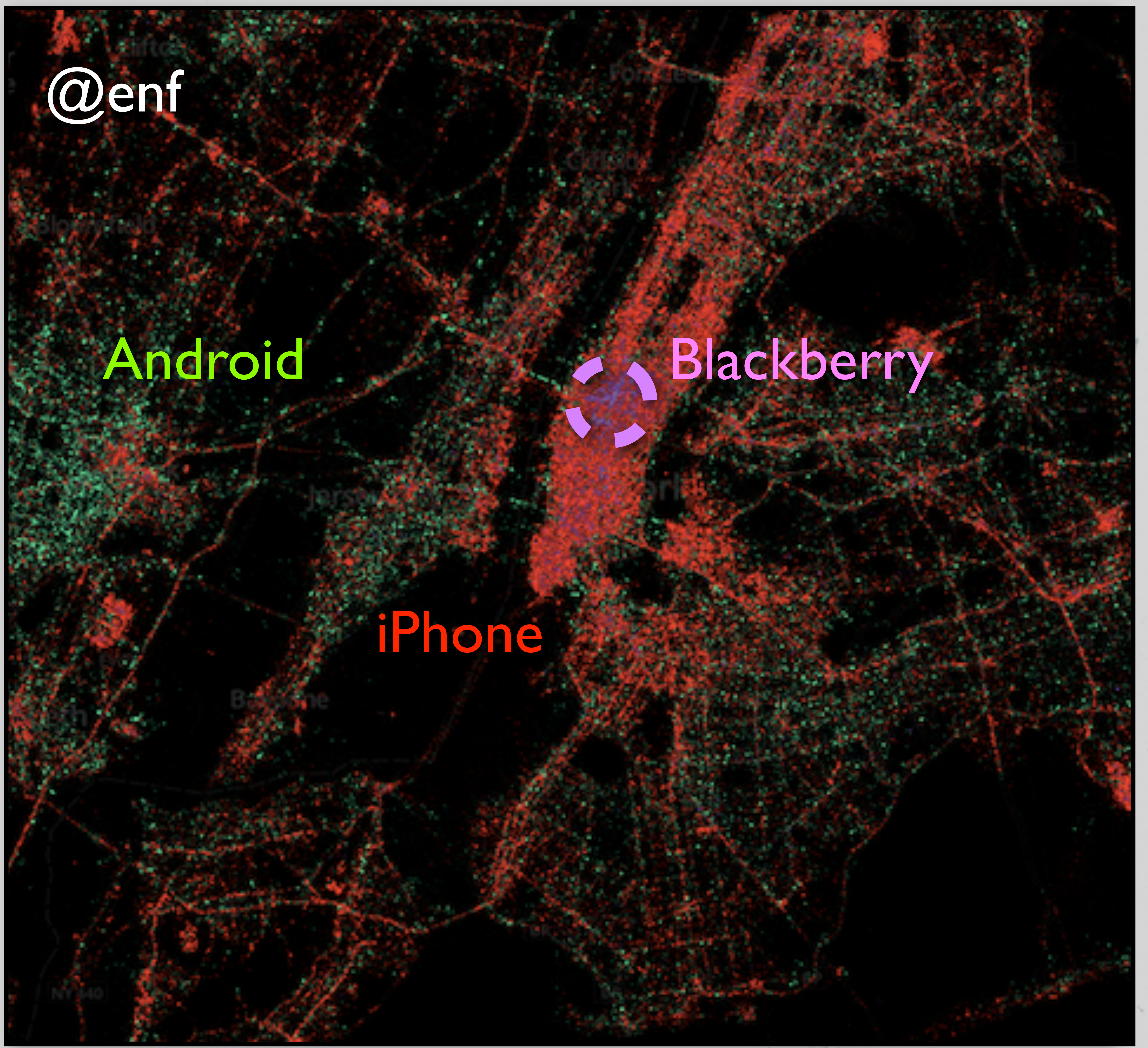


@miguelrios

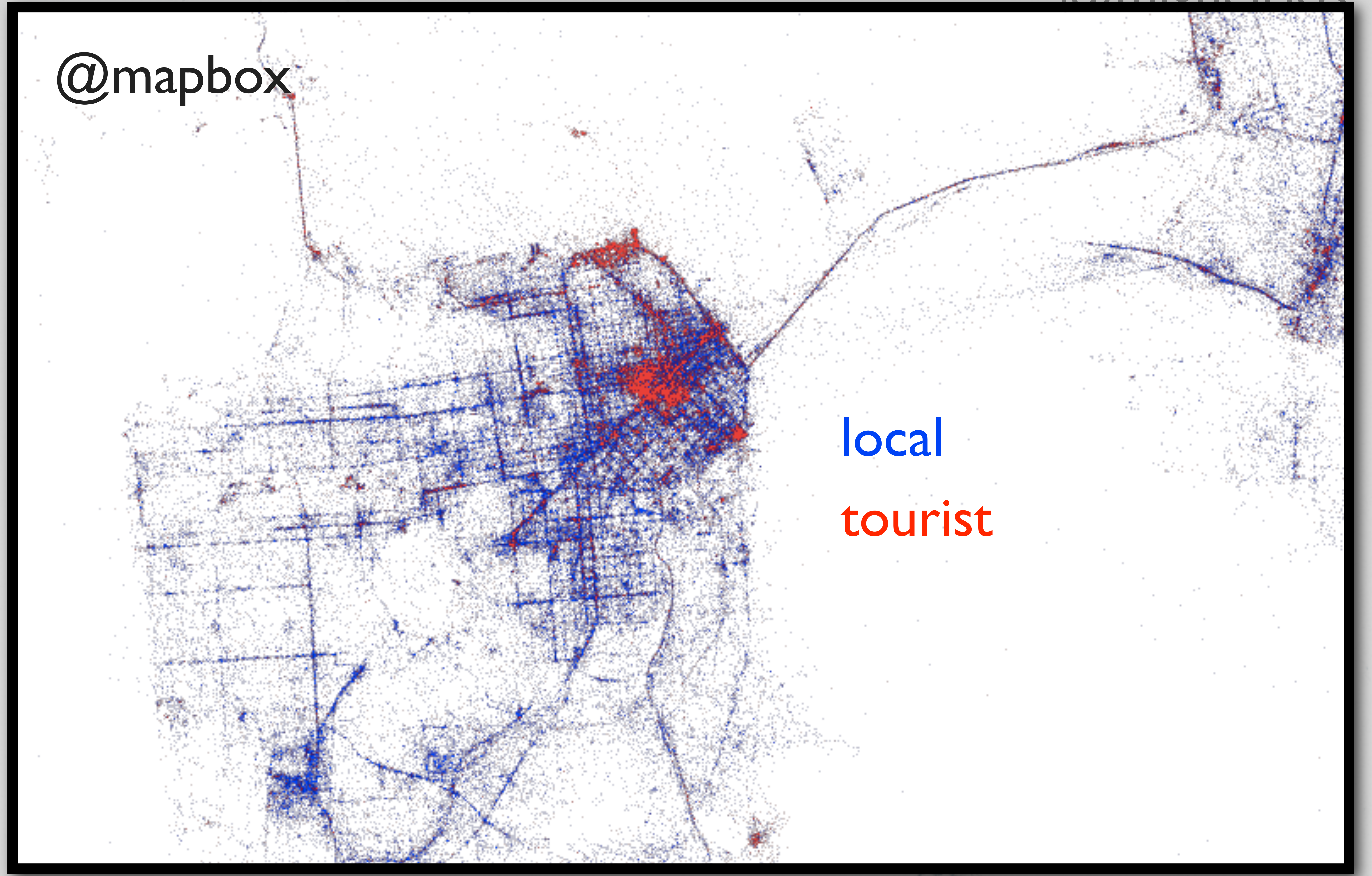


uelrios



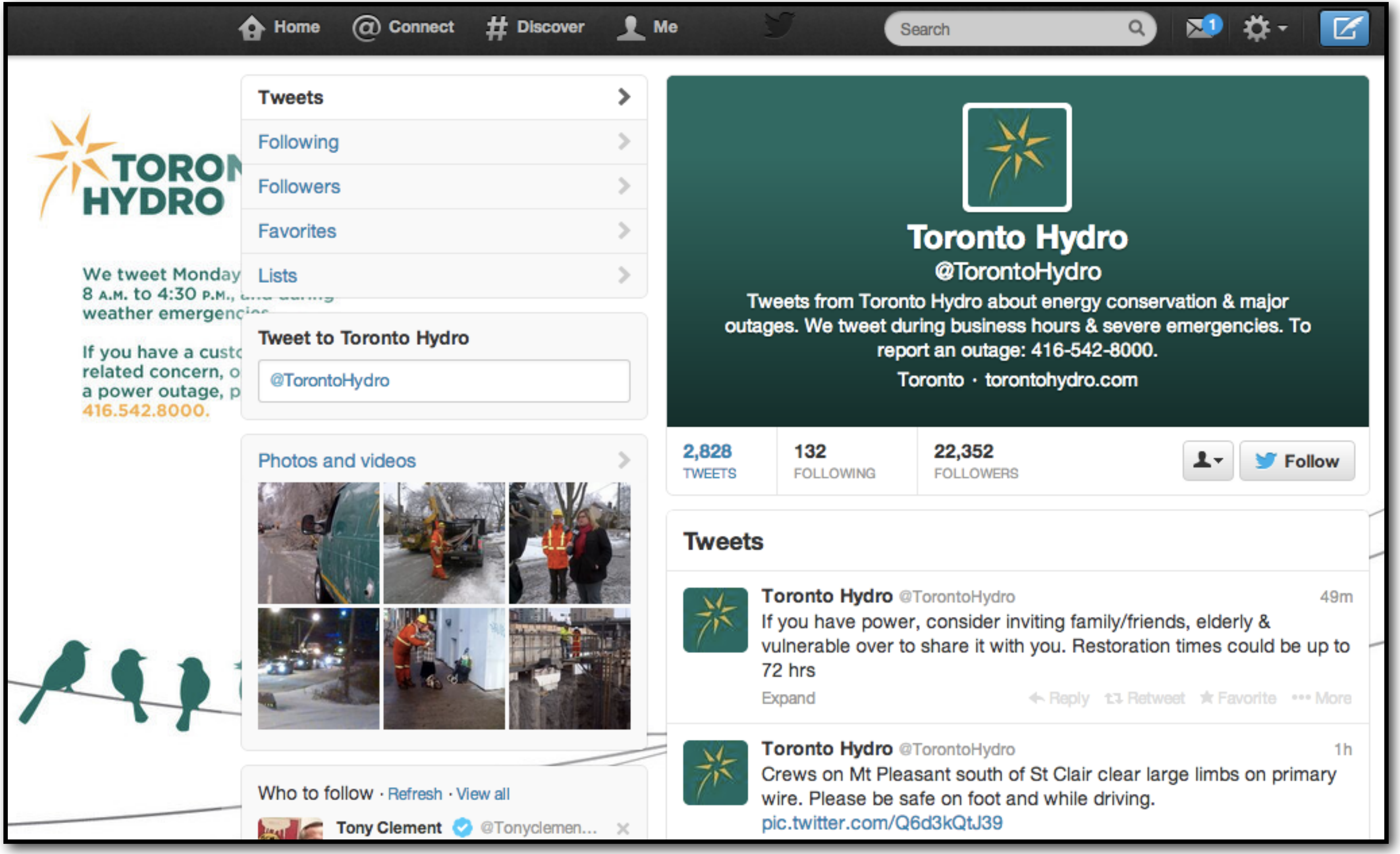

@enf
Android
Blackberry
iPhone

@miguelrios



@mapbox

Visitors land at Pearson airport and head downtown



December 2013 Power Outage


@jkrums
US AIRWAYS
2009

@UkrProgress
@StateDept

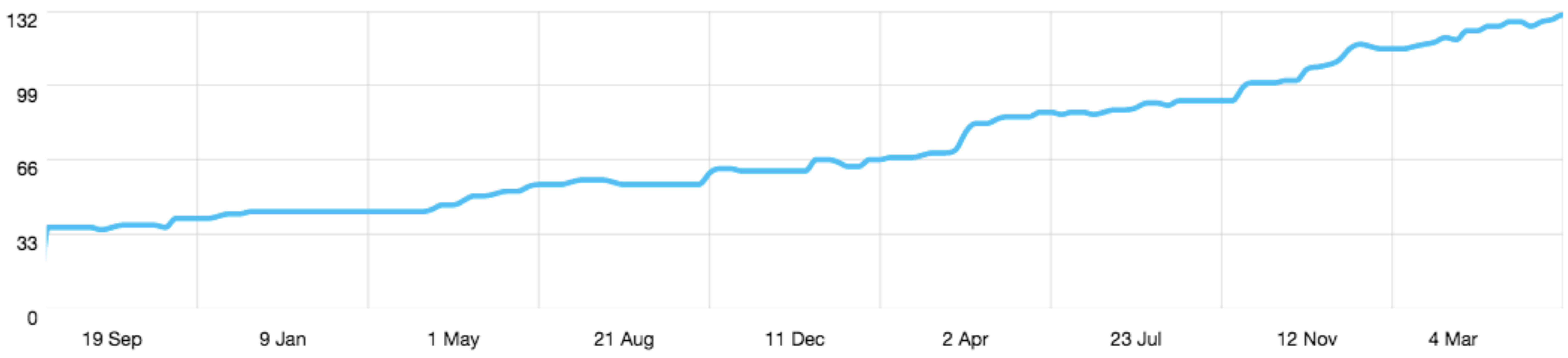
Прогресс для Украины
@UkrProgress

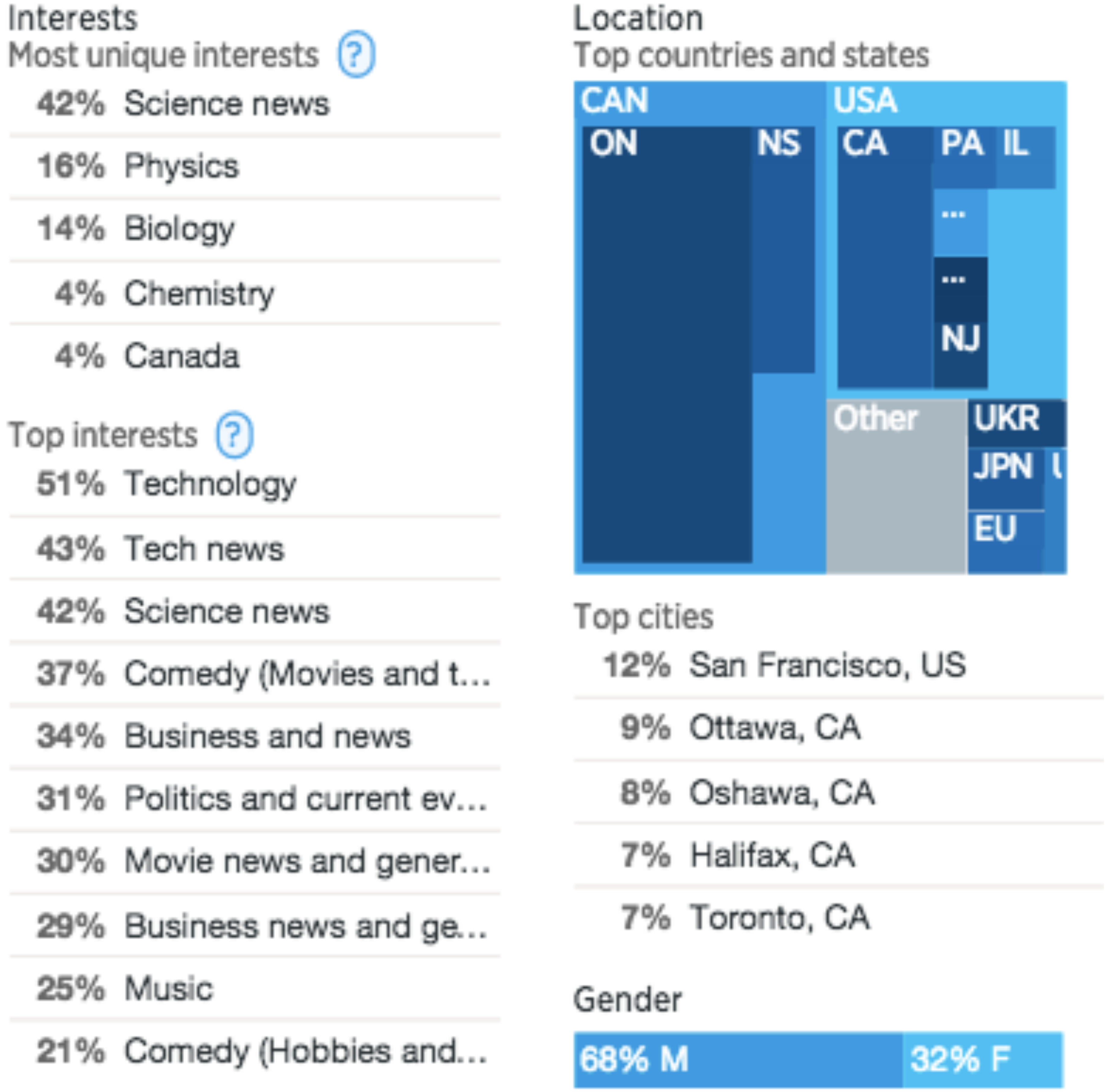
Follow

MT@Interpreter_Mag Поддерживаемые Россией сепаратисты заняли 28 городов и деревень с февраля bit.ly/1JqJ2Qp

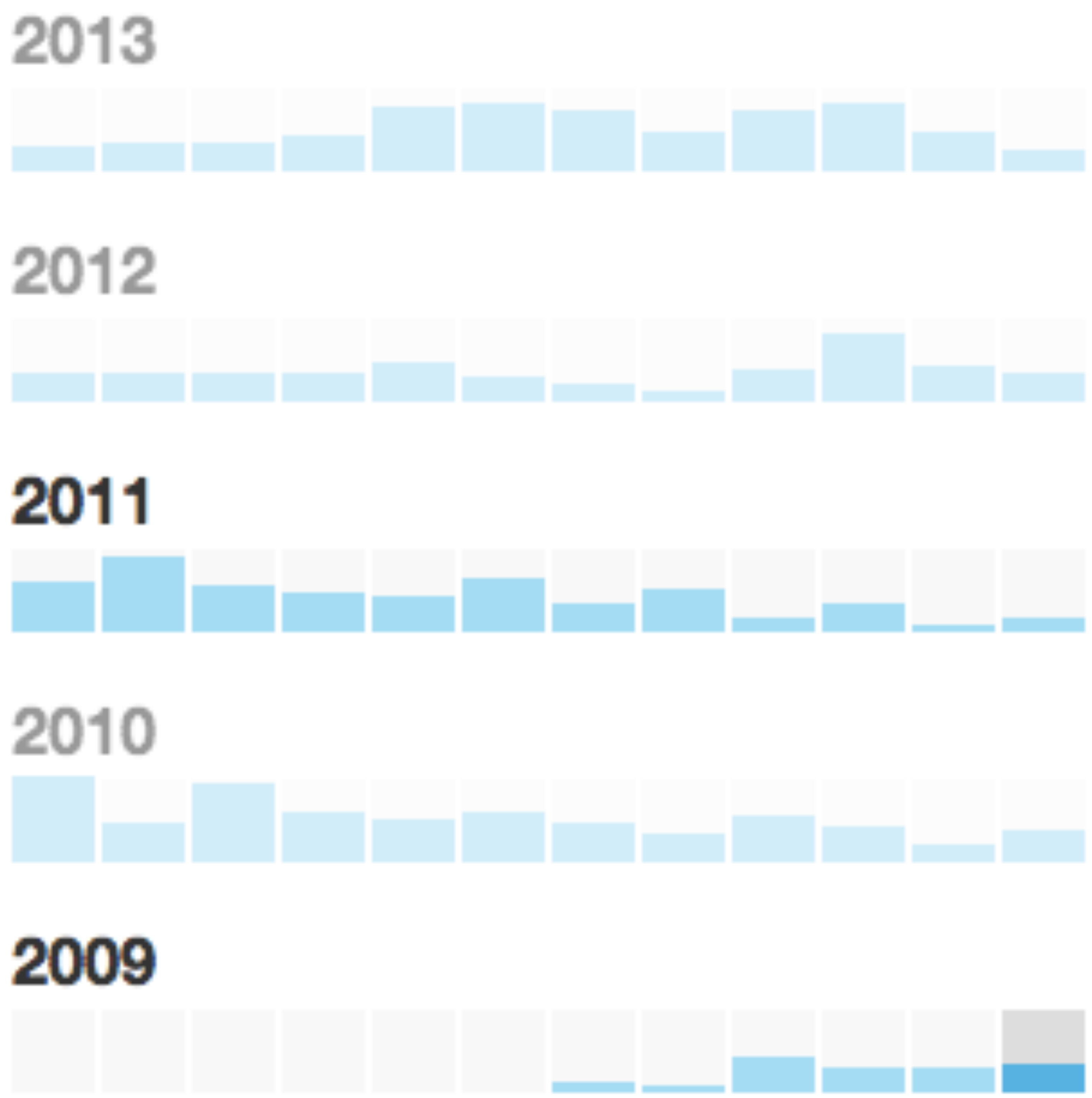
View translation

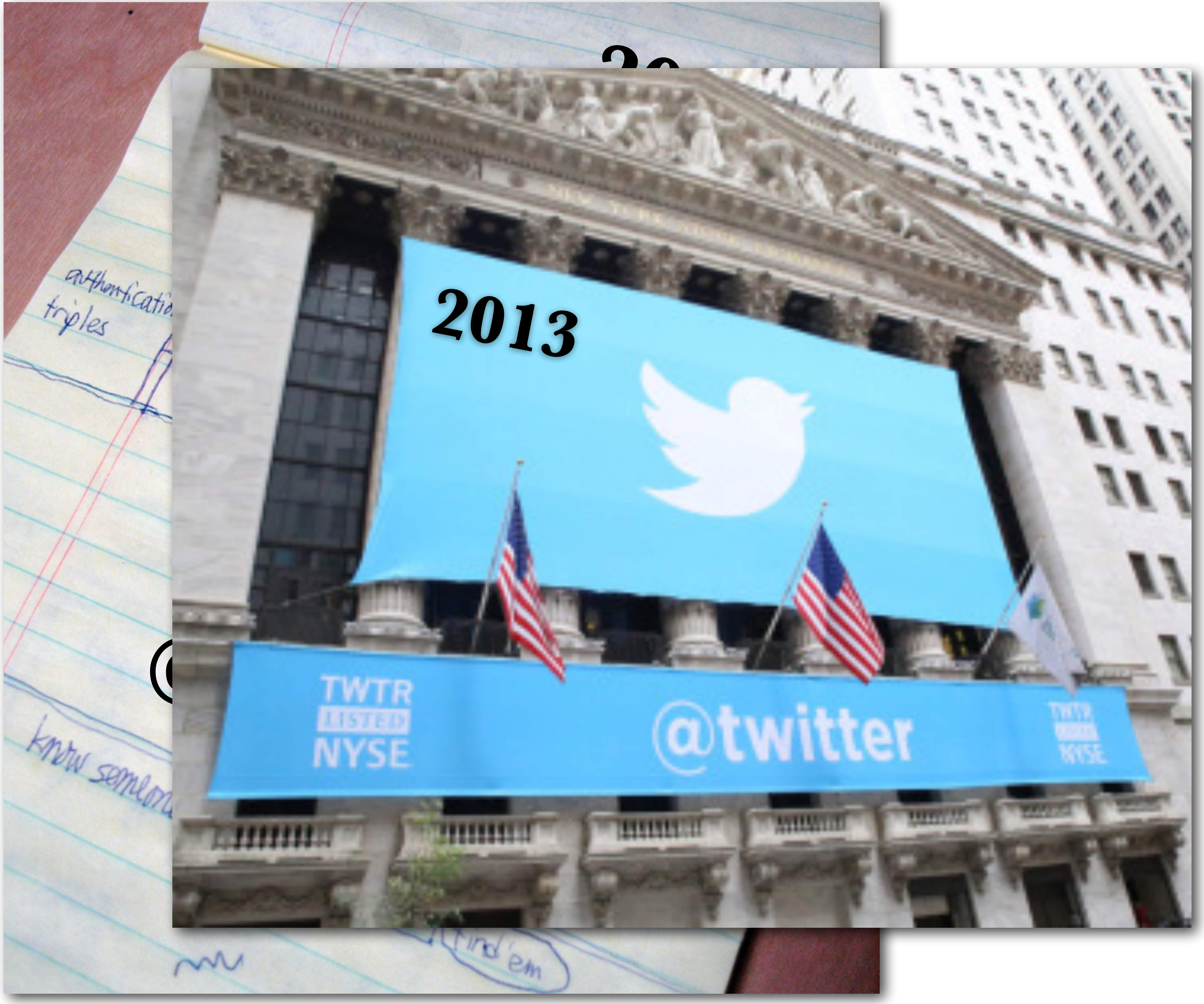
RETWEETS 8 FAVORITES 2



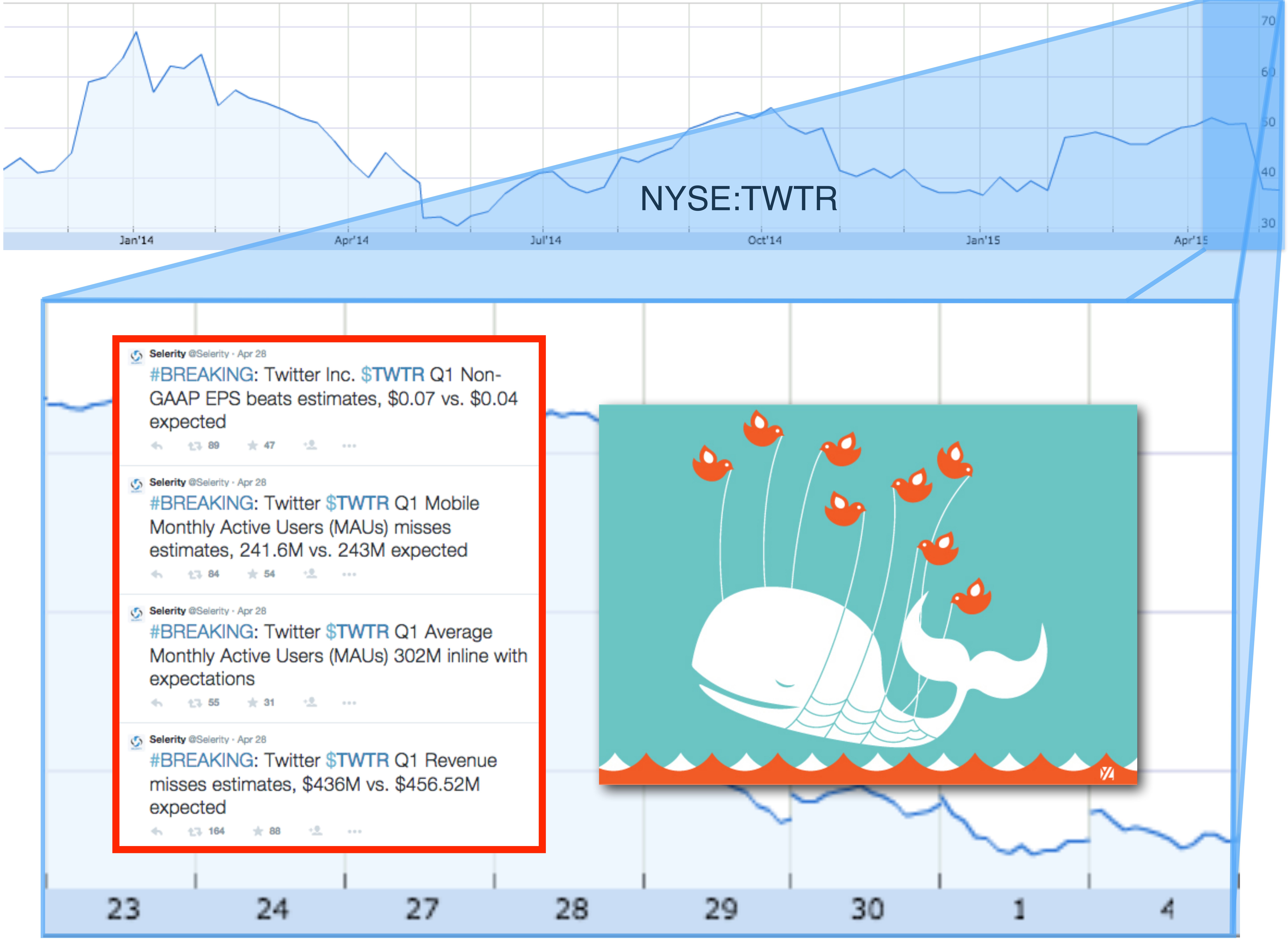


User behaviour changes over time




triples
find 'em
2013
TWTR
LISTED
NYSE
@twitter


70
60
50
40
30
NYSE:TWTR
Jan'14
Apr'14
Jul'14
Oct'14
Jan'15
Apr'15


NYSE:TWTR
70
60
50
40
30
Jan'14
Apr'14
Jul'14
Oct'14
Jan'15
Apr'15
Selerity @Selerity · Apr 28
#BREAKING: Twitter Inc. $TWTR Q1 Non-GAAP EPS beats estimates, $0.07 vs. $0.04 expected
89 47
Selerity @Selerity · Apr 28
#BREAKING: Twitter $TWTR Q1 Mobile Monthly Active Users (MAUs) misses estimates, 241.6M vs. 243M expected
84 54
Selerity @Selerity · Apr 28
#BREAKING: Twitter $TWTR Q1 Average Monthly Active Users (MAUs) 302M inline with expectations
55 31
Selerity @Selerity · Apr 28
#BREAKING: Twitter $TWTR Q1 Revenue misses estimates, $436M vs. $456.52M expected
164 88
23
24
27
28
29
30
1
4

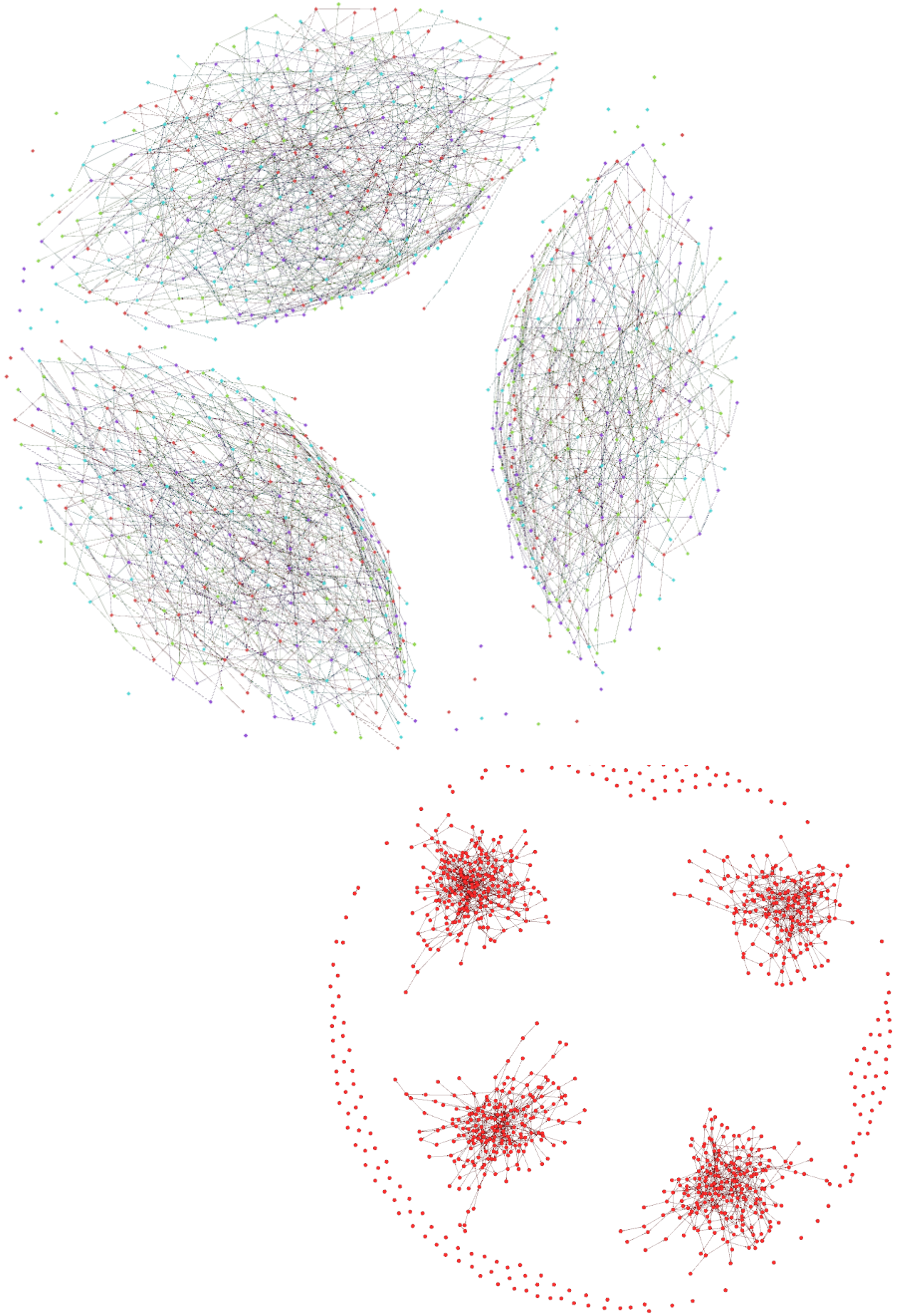
NYSE:TWTR
70
60
50
40
30
Jan'14
Apr'14
Jul'14
Oct'14
Jan'15
Apr'15

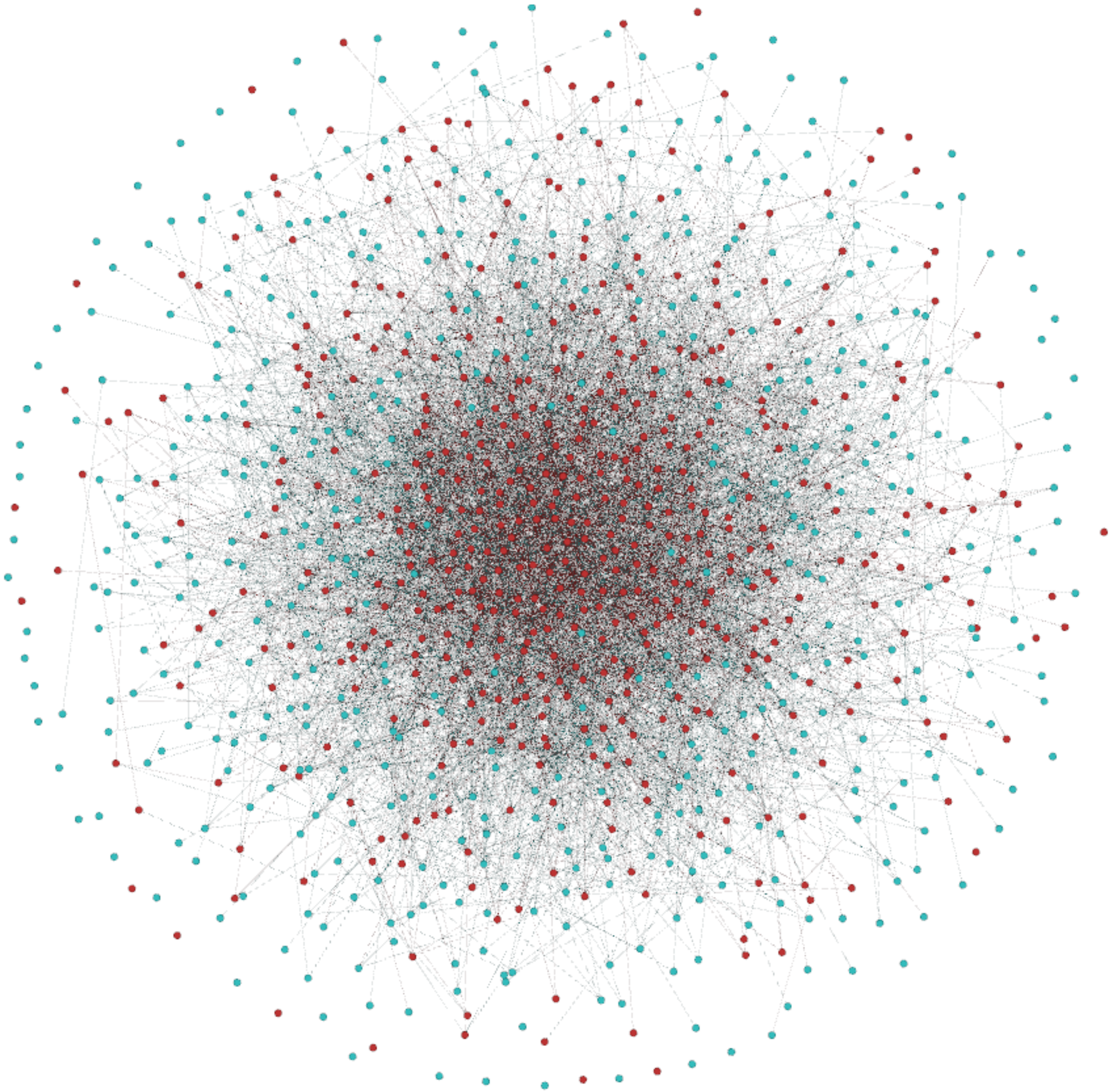
NYSE:LNKD
250
200
150
100
2012
2013
2014

NASDAQ:FB
80
70
60
50
40
30
20
2013
2014

Why simulate an online social network?

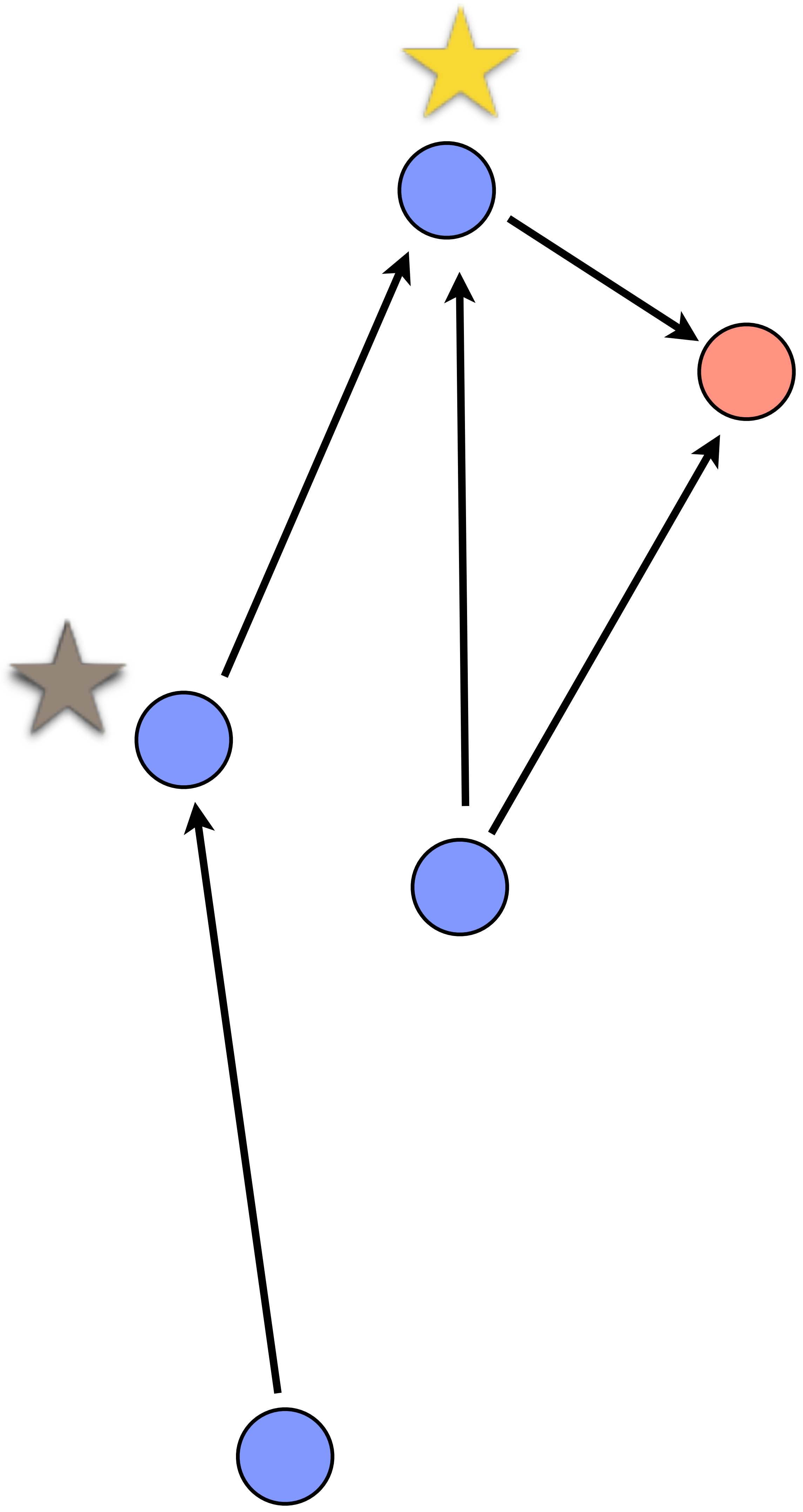
- Test hypotheses (e.g. growth mechanism, information flow, etc)
- Provides "ground truth" for stream based data aggregation & analysis (e.g. phonebook)
- Optimize pricing strategies (e.g. advertising)
- Predict future user behaviour - rule changes
- Validate agent based modelling approaches
- Conduct safe experiments (e.g. Facebook)





What events are possible?

- join
- tweet
- follow
- retweet / mention



- Upon follow: you can view and re-transmit information
- Your followers can do the same
- hashtags in the tweets, ‘#physics’ is an example

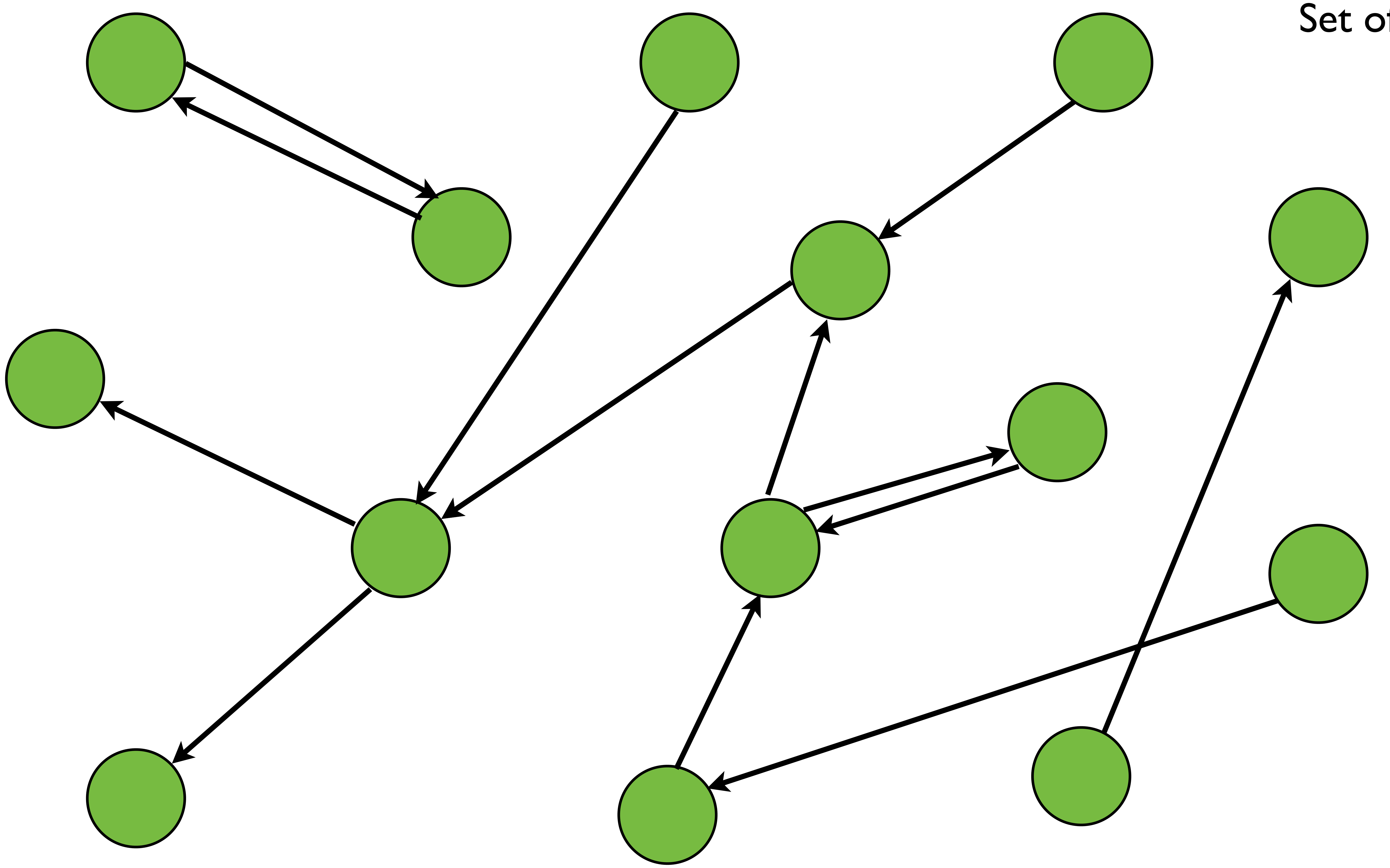
- Different online social networks have different rules

Introduction to networks/graphs

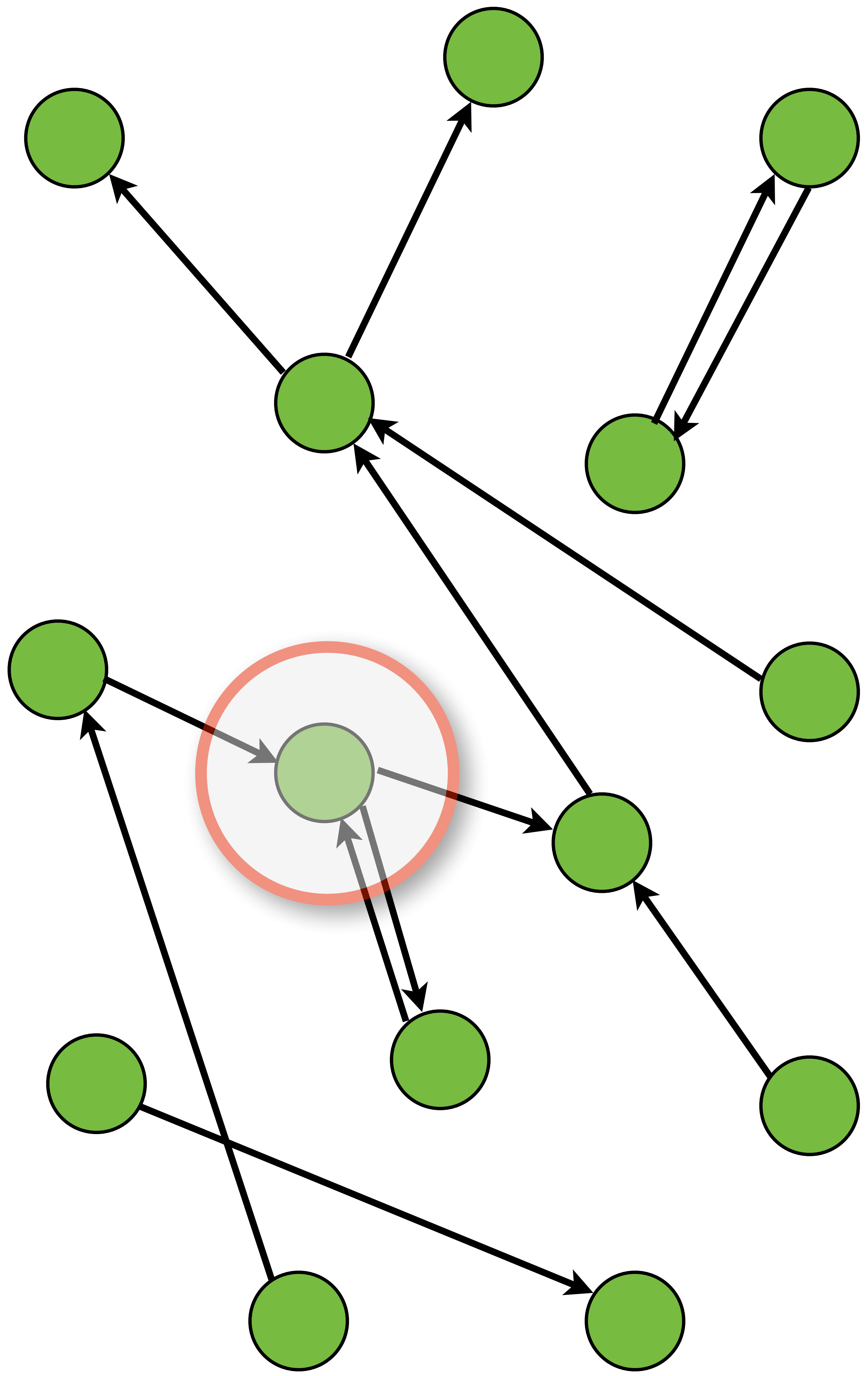
$$G(N, E) = \{N\} + \{E\}$$

Set of nodes -

Set of edges -



Degree of a node

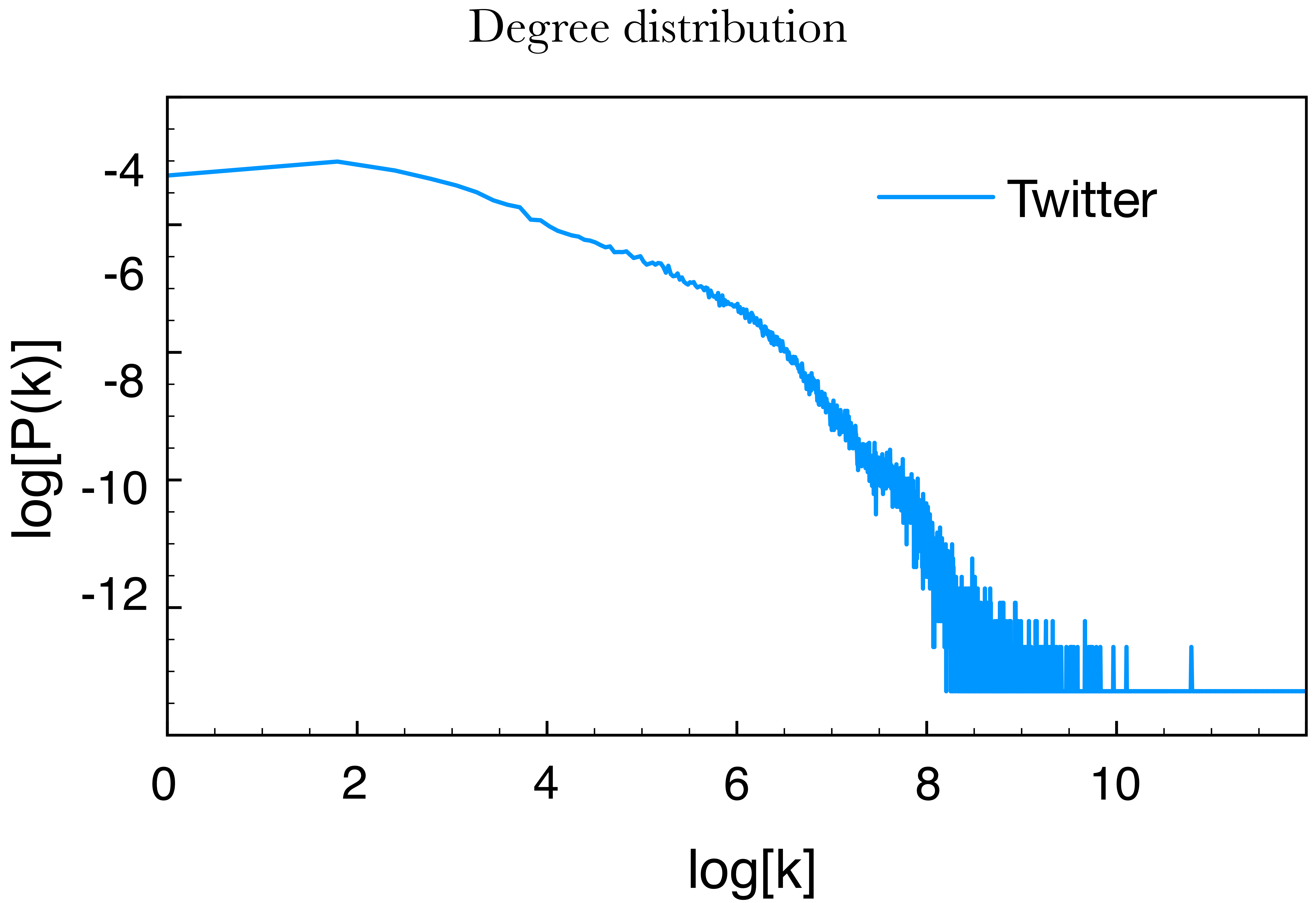


- How many connections does a selected node have?

Example:

- In-degree: 2
- Out-degree: 2
- Cumulative degree: 4

- If we do this for all nodes we obtain a degree distribution.


Degree distribution
Twitter
-4
-6
-8
-10
-12
log[P(k)]
0
2
4
6
8
10
log[k]

Userid math

32 bit unsigned int = 0 to 4,294,967,295

4.3 billion accounts

1 userid = 4 bytes - 8 bytes

Twitter monthly active users ~ 300 million $\longrightarrow$ 1.2 GB

(daily ~ 100 million)

Facebook monthly active users ~ 1.4 billion $\longrightarrow$ 5.6 GB

(daily ~ 940 million)

users

followers

4471	7882	6594	8931	3513	5348	4694	4236	7492	5837	7290	3280	7715	6392	7117	1580	7304	8932	9161	1039	4769	7995	7819	3678	2714	510	7286	6369	1509	1854	7579	8190	1742
1470	4373	6365	3681	172	2321	722	4881	9923	3163	4106	7670	9110	8590	8578	9572	6254	651	1166	5138	7489	2595	3743	9251	1434	7640	5246	4098	5686	3343	7921	4809	6550
3776	8393	7925	629	7898	4098	6529	102	4731	2849	5499	6825	1489	2620	9472	138	4997	3602	1402	7591	5027	3323	232	4559	7658	8111	1232	1817	1557	3298	1600	7532	5085
7463	3319	7308	7926	3029	6379	4345	3043	5539	170	9061	1969	5375	3591	6485	6761	7312	7407	5845	8565	2398	1368	4714	9243	4306	6122	6197	7860	5254	5932	7150	906	9276
8482	734	2414	8519	9445	2921	2420	2112	7713	781	2946	7565	1954	5829	4580	8733	9512	4161	645	7706	5919	5738	4500	61	8254	3268	9672	3201	7817	6007	6293	7502	9545
5350	6479	4688	1772	9971	8711	8429	7575	6256	6546	3643	4473	432	5069	7184	9878	7732	7306	8775	3517	9412	1028	854	9944	1936	1916	372	2289	615	6728	9834	7482	4381
3445	5649	1336	579	4038	7273	3496	8053	7376	8611	1689	7274	8587	9787	6527	7971	6989	4377	1685	1635	9989	561	3749	7146	1896	8203	9173	3653	3583	1956	1220	7561	4961
2288	4648	2949	21	459	8987	4161	2527	2634	1867	4525	3186	9657	2478	6905	1830	7199	7842	6411	7462	952	2400	7676	6575	9149	3587	9863	6614	2180	8366	14	8539	7994
6342	2259	3855	5504	7985	1733	3168	7719	826	3183	2653	4361	6951	3078	1795	9321	4607	4759	1644	3405	3508	7524	2407	3146	9914	3098	458	8106	1371	294	8802	6978	4473
3659	8398	6364	3908	4614	4948	6203	5435	4553	3539	4459	2192	4370	3584	9574	9628	5175	4464	6164	5321	1858	8271	2787	1288	5756	7250	5719	9309	7780	987	9059	745	4181
5808	2905	9551	933	6188	3240	1239	6567	9465	1099	1127	1559	4421	2788	2162	6862	2214	2684	6468	3865	6619	8065	2869	248	5959	7768	5194	3595	8304	8294	5815	5513	8552
140	146	7291	6316	1814	3521	7667	514	5550	8966	8945	6806	3041	1671	5083	4504	7581	4008	3535	341	151	2737	4996	4867	4154	4598	3458	3913	653	9020	8221	2263	1728
9412	4203	5772	9177	3464	3535	8798	3546	3519	8226	4923	2848	6813	8910	8949	1013	3203	8633	1395	3940	7103	5898	7803	4831	6095	7298	7219	1268	19	9701	3506	6015	5075
6812	9157	8711	7563	3889	8619	2957	8767	4247	3827	1154	2976	1756	8404	9075	9830	239	8134	8594	9172	1847	4460	6505	8162	7489	1022	7840	8699	2825	9755	9354	2326	3653
8156	843	9792	2183	3261	4581	5195	1518	8758	6857	3204	4041	3728	7512	6449	3452	4618	4910	2186	3519	2481	7831	6342	3736	6278	204	49	8930	6388	9602	2443	8167	2755
5207	5446	554	6221	7146	5365	219	9792	3298	6592	7456	1610	9049	9677	5693	9389	5544	7945	9345	4911	3604	3389	885	4122	6900	9974	4369	2824	8306	779	3113	5486	91
8969	2617	7978	4734	5490	2789	7860	4722	5016	3474	1644	8739	767	8689	9662	5670	8778	408	7651	2179	1507	1866	4237	4706	2013	3153	3210	5596	9723	4458	1369	800	3112
9219	2099	4056	7878	7920	2122	9920	1994	5538	1807	3915	7535	1123	8674	9517	2489	9105	3250	9311	6309	92	2098	2865	6196	5385	8637	574	2743	1733	8115	3210	9378	3991
3359	3105	106	4329	4637	6479	7145	9596	1909	4517	2054	6775	4355	1574	4584	7731	9157	4145	5523	4503	6793	8184	9124	3287	499	4131	6709	8503	7237	3536	5227	5140	6933
494	6607	7742	476	2661	8903	1807	1820	1677	7001	5736	4855	6464	6569	1909	1446	5042	3907	6626	4899	4194	6332	3341	298	8394	5936	210	6816	4287	4013	8642	2337	875
8035	4232	6018	2142	2638	2804	2340	1430	220	9951	7580	3712	9106	853	4251	8702	5584	9183	5679	4169	2640	256	6967	2390	3228	4551	4986	1619	1695	2365	2765	9210	4694

Userid math

$$32 \text{ bit unsigned int} = 0 \text{ to } 4,\underbrace{294,967}_{4.3 \text{ billion accounts}},295$$

1 userid = 4 bytes - 8 bytes

Twitter monthly active users ~ 300 million $\longrightarrow$ 1.2 GB

(daily ~ 100 million)

$\times 210 = 250 \text{ GB}$

Facebook monthly active users ~ 1.4 billion $\longrightarrow$ 5.6 GB

(daily ~ 940 million)

$\times \dfrac{340}{2} = 980 \text{ GB}$

Our code #k@
http://hashkat.org

- Written mostly in C++, with some Python and Lua.
- Source ~ 7000 lines of code
- Repository consists of ~80 files
- Full documentation and tutorials are available online
- Obtain a random network with 5 M nodes in about an hour and 30 M (128 Gb RAM) nodes in about 42 hours.
- Free & Open Source (GPLv3)
- Agent based model

Network =

Agent 0	Agent 1	Agent 2	Agent 3	...	Agent N

Array of structs

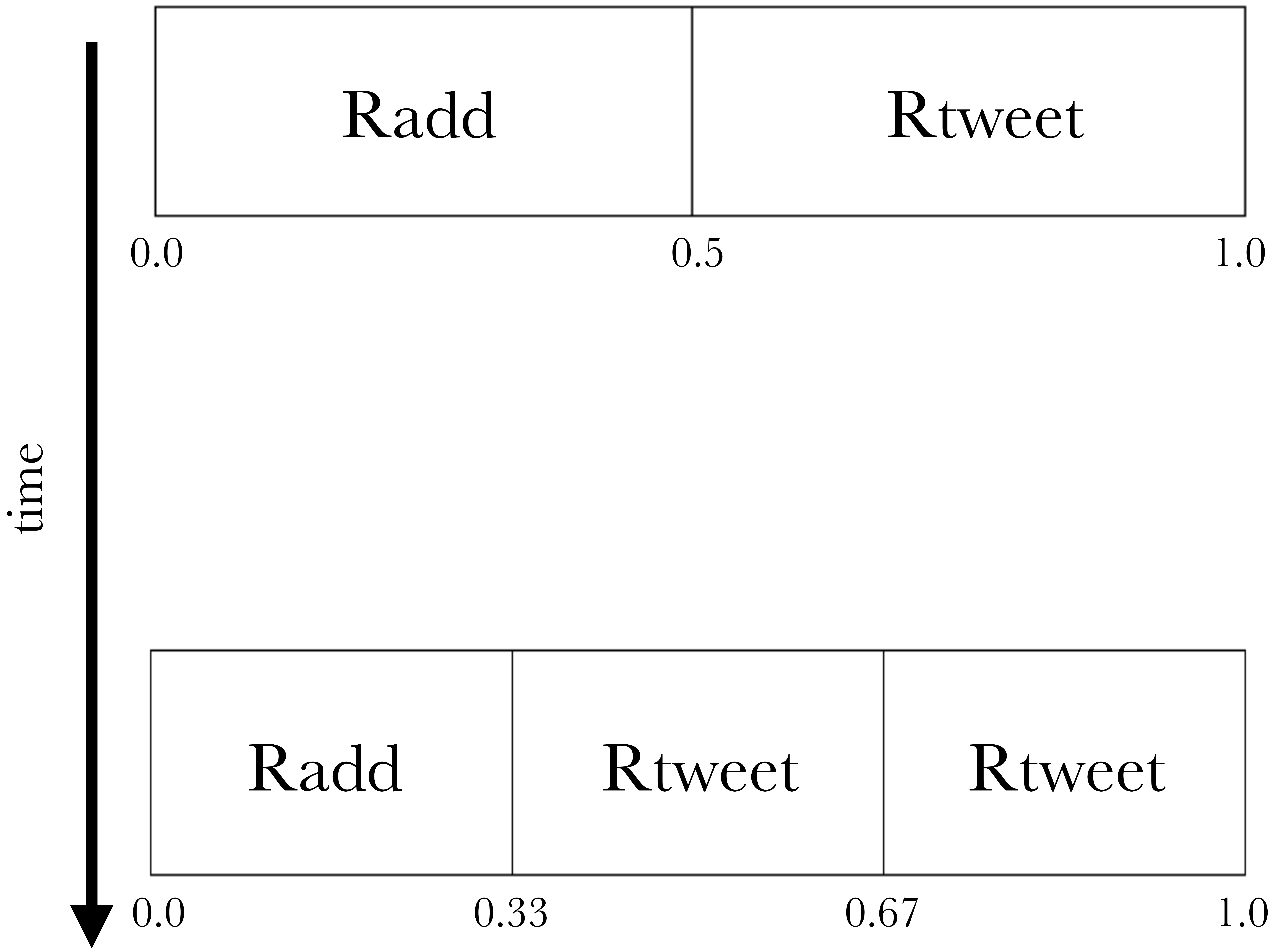
- id = 1
- Agent type
- Number of tweets, retweets
- Region
- Creation time
- Language
- Ideology
- Humour preference
- List of chatty people
- Following and follower set

Kinetic Monte Carlo

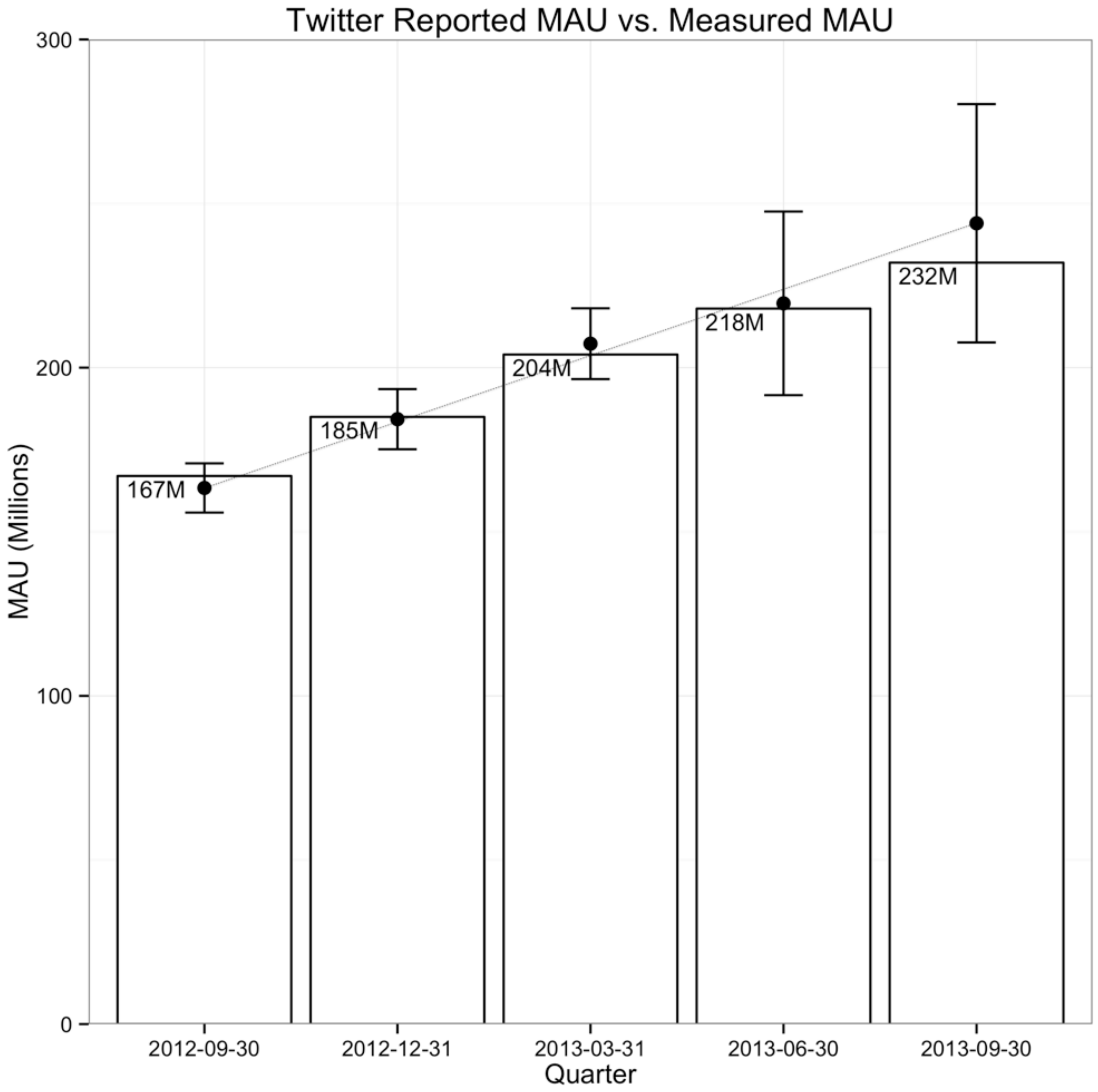
- An event based Monte Carlo method
- Progresses time in the simulation
- The only thing you need is:
  - a random number generator
  - the rates for every event in your system
- kMC is NOT Metropolis Hastings

Kinetic Monte Carlo (algorithm)

1. Set the time t = 0
2. Form a list of all possible rates r_i in the system.
3. Calculate the cumulative function or the total rate: $$R_{\text{tot}} = \sum_{i=1}^{N} r_i$$
4. Generate a uniform number u_1 on the interval (0,1].
5. Carry out an event based on the random number.
6. Get a new random number u_2 on the interval (0,1].
7. Move forward in time by Δt where $\Delta t = -\ln(u_2)/R_{tot}$.
8. Return to step 2.


time
Radd
Rtweet
0.0
0.5
1.0
Radd
Rtweet
Rtweet
0.0
0.33
0.67
1.0

Experimental guidance



sampling error reflects collection method

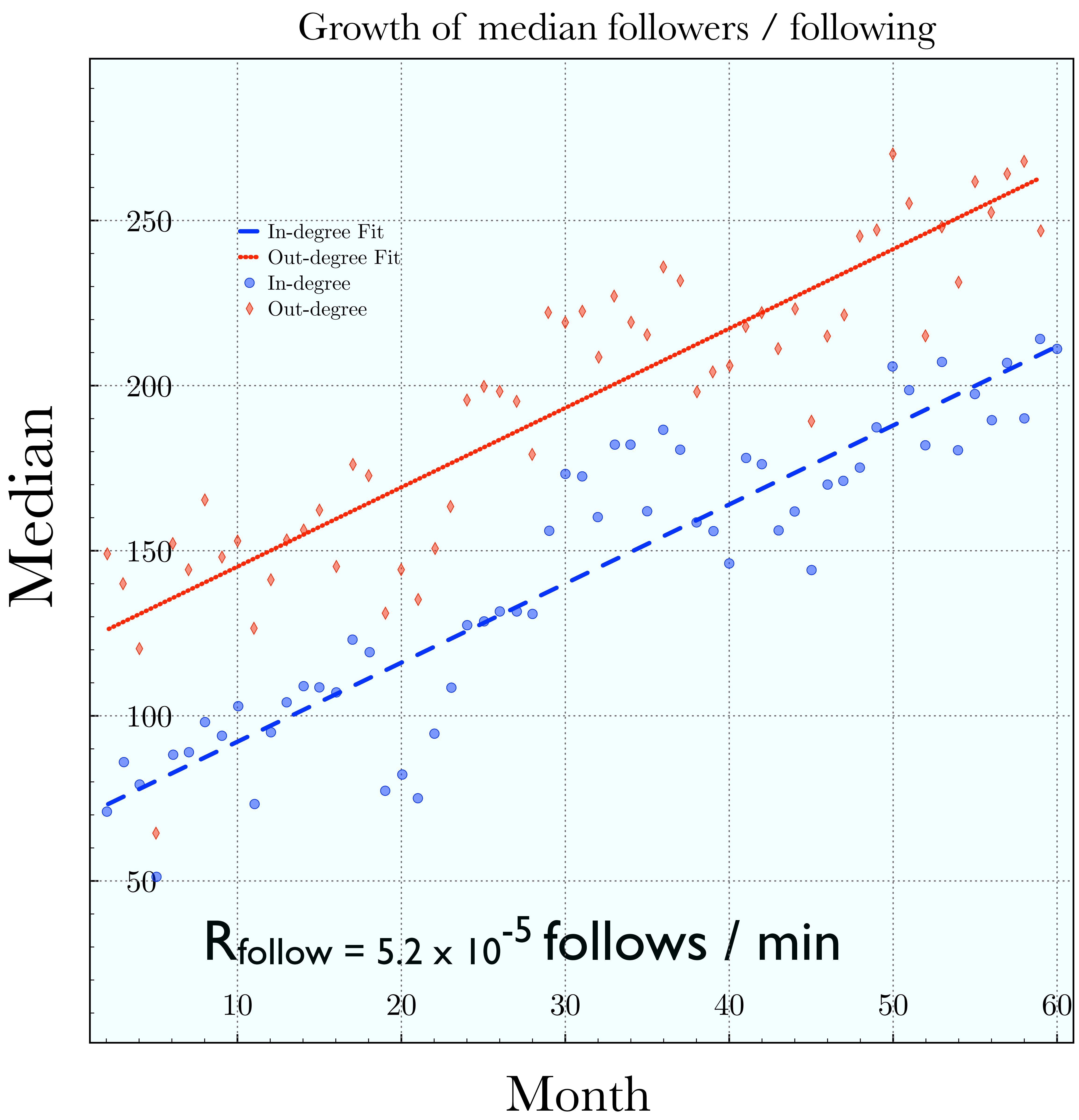
R_{add} = 123.7 users/min

Size of the graph (IPO)
- 300 million

Tweets / day (IPO)
- 500 million

K. White, private comm

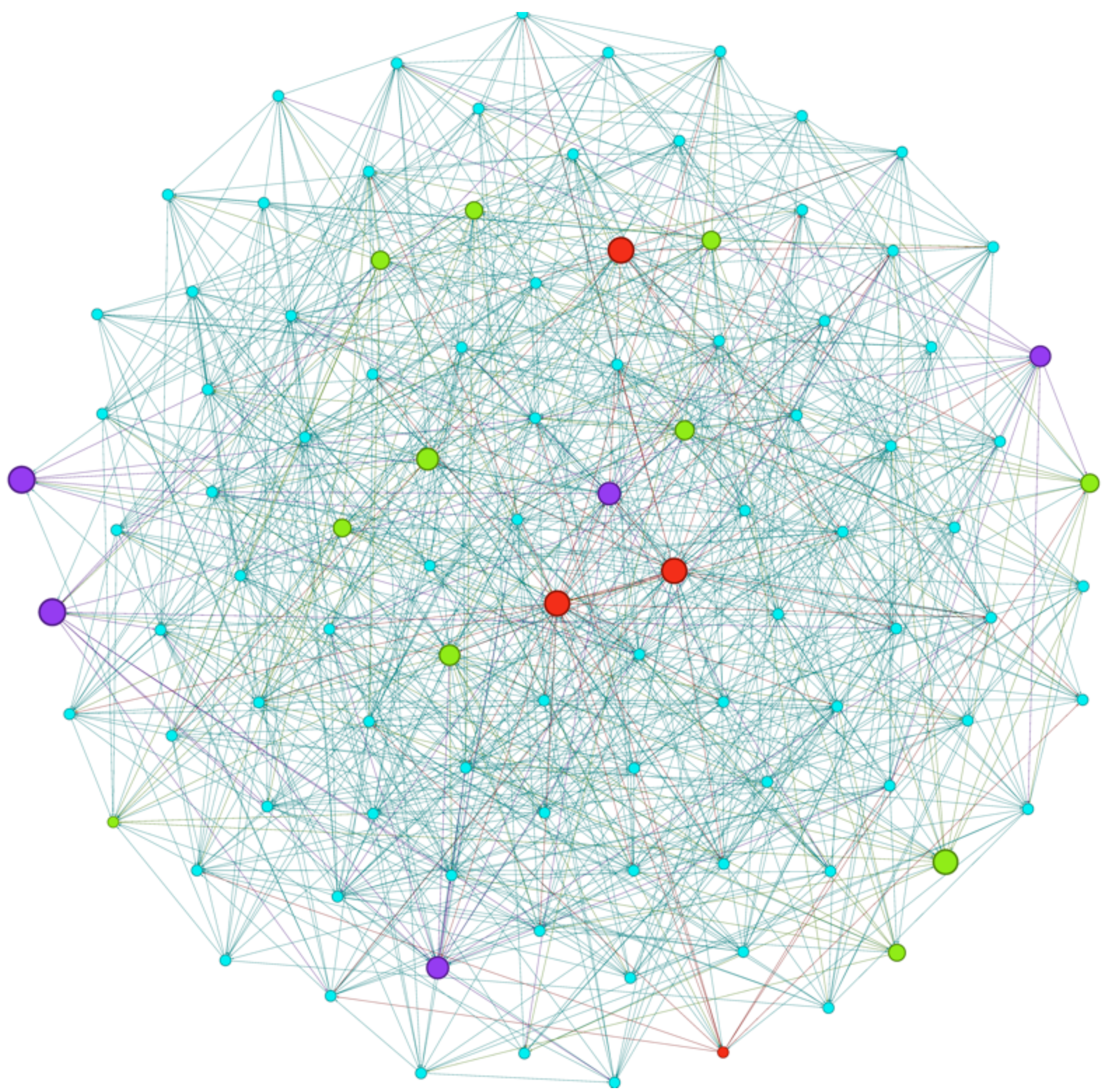
Experimental guidance



- Users on Twitter tend to follow more than being followed.
- The rates for followers and followings is roughly equal.
- The follow rate is much smaller.

K. White, private comm

Producing a random graph



- With probability p, place an edge between two nodes.

Total number of edges: $$E = \frac{N(N-1)}{2}$$

Total number of expected edges: $$E = p\left(\frac{N(N-1)}{2}\right)$$

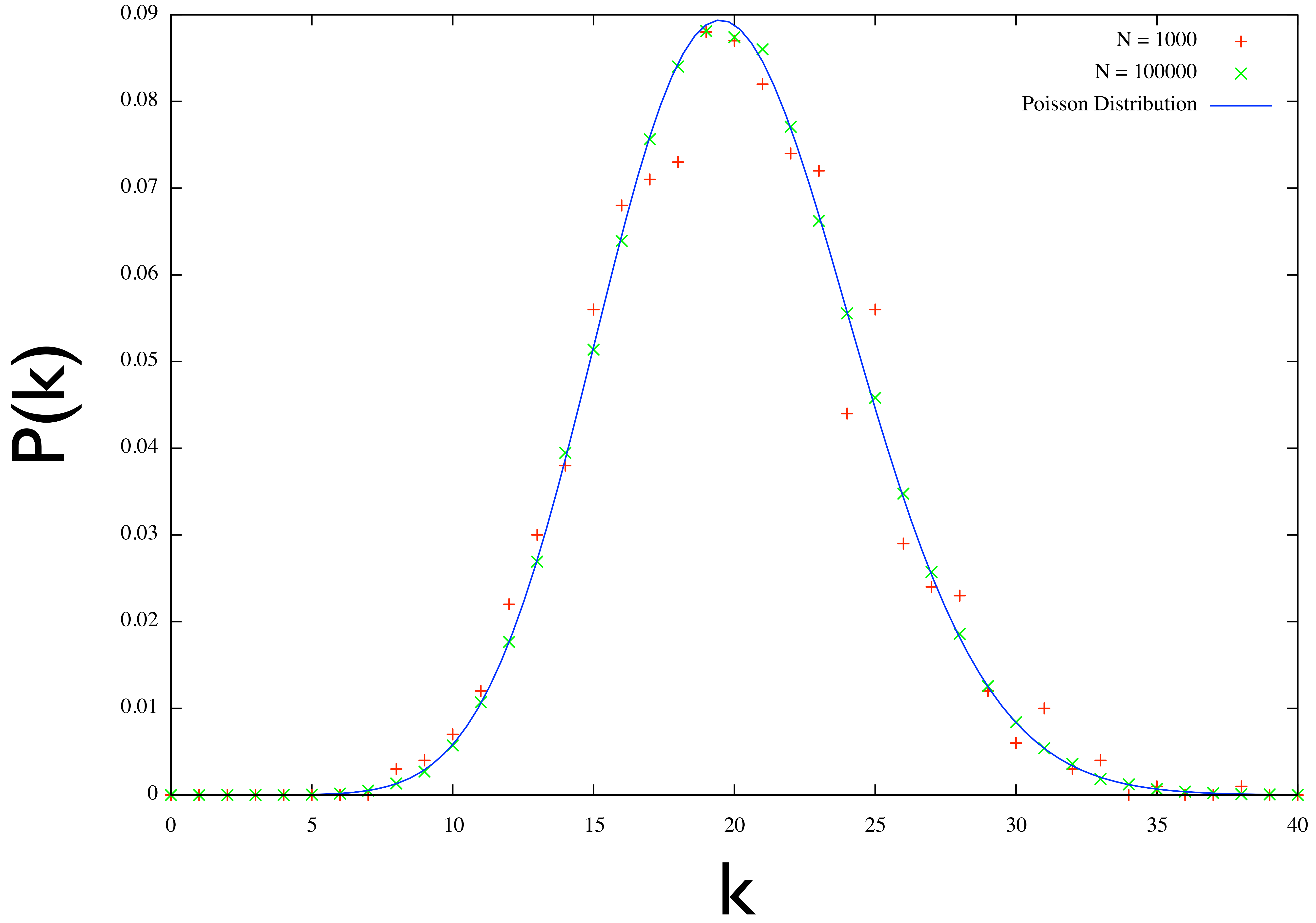
Example:

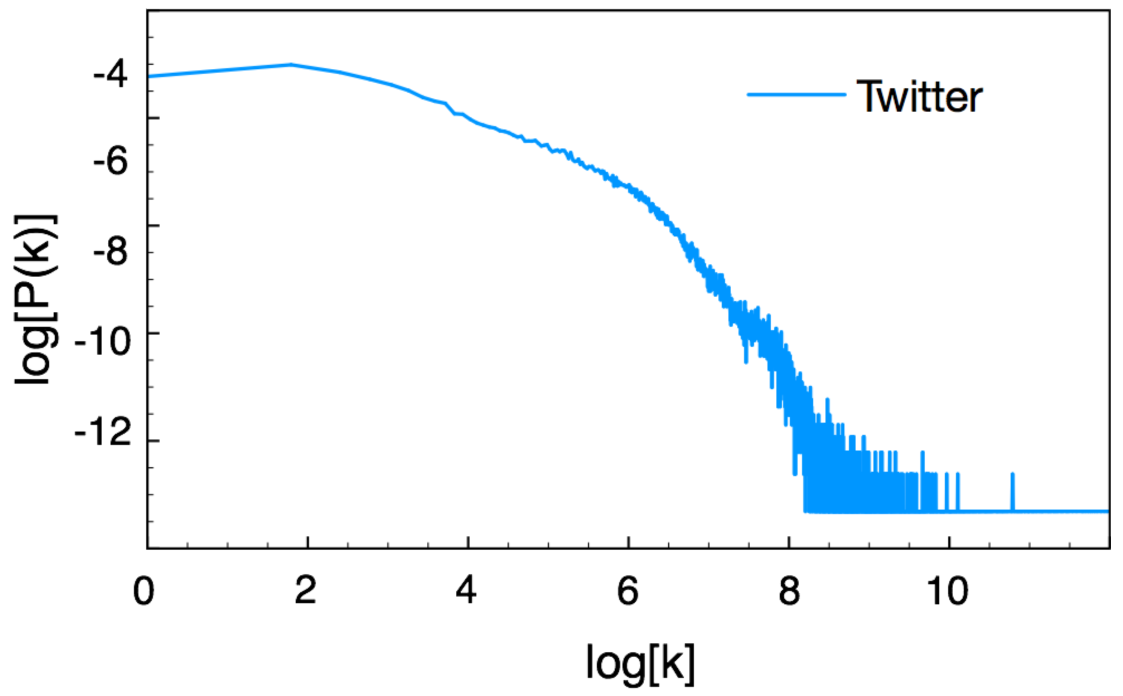
$$N = 10$$
$$E = (10 \cdot 9)/2 = 45$$
$$\text{if } p = 0.2$$
$$E(0.2) = 9$$

Degree distribution of a random graph

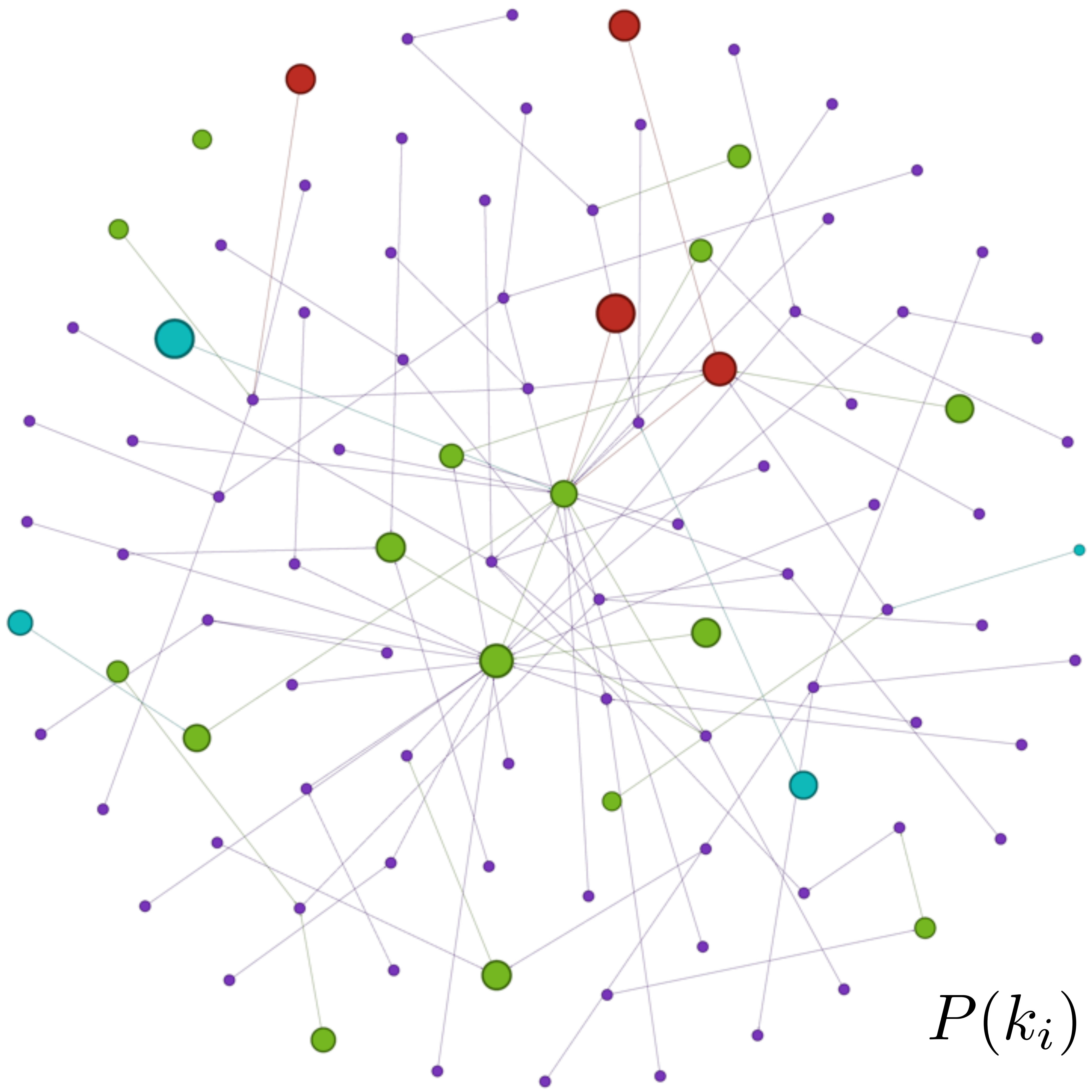
typo

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx \frac{\langle k \rangle^k e}{k!}$$





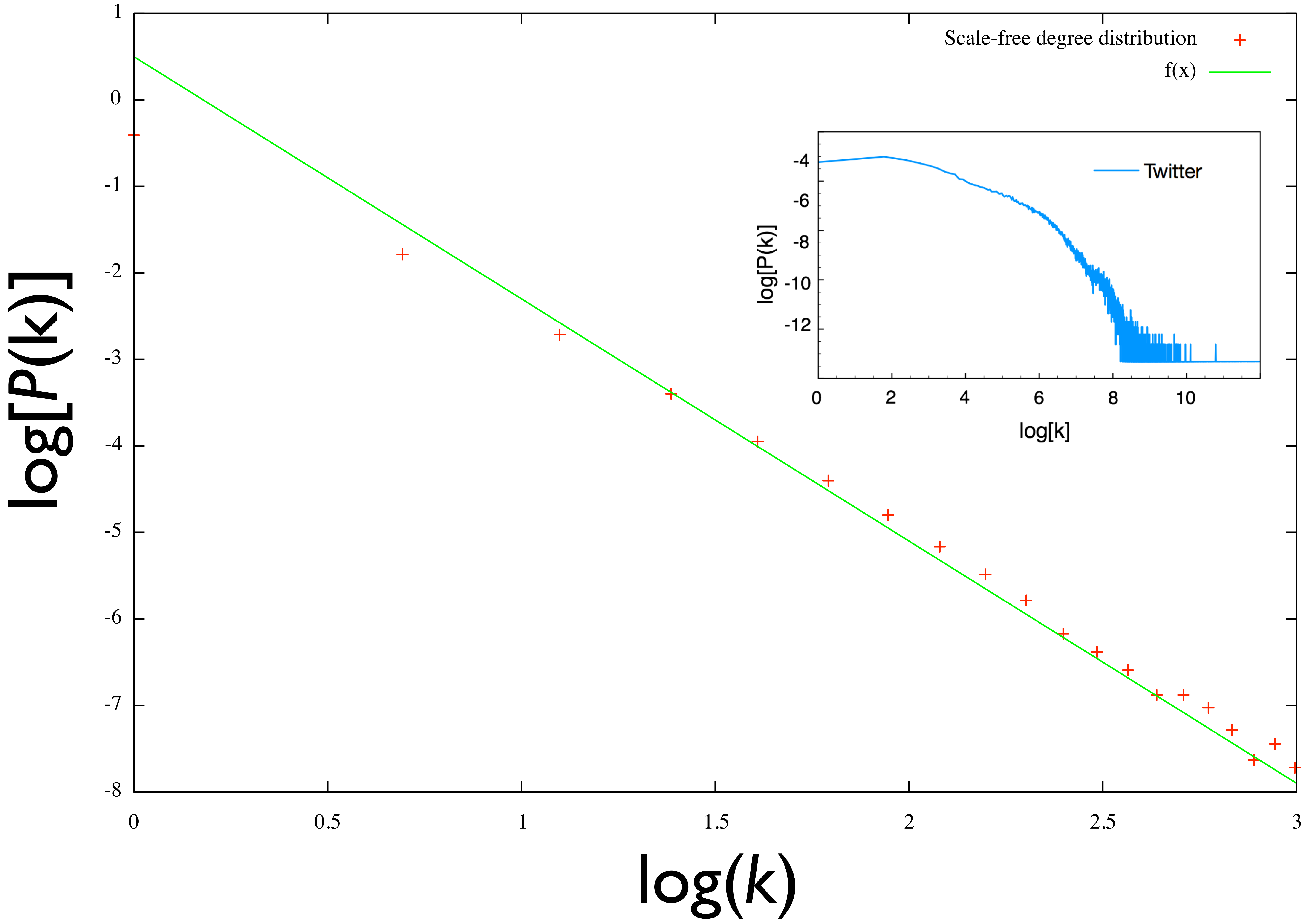
Following via preferential attachment



$$P(k_i) = \frac{k_i}{\sum_{j=1}^{N-1} k_j}$$

Degree distribution of a scale-free network

$$P(k) \propto k^{-\gamma} \text{ where } \gamma \approx 3$$

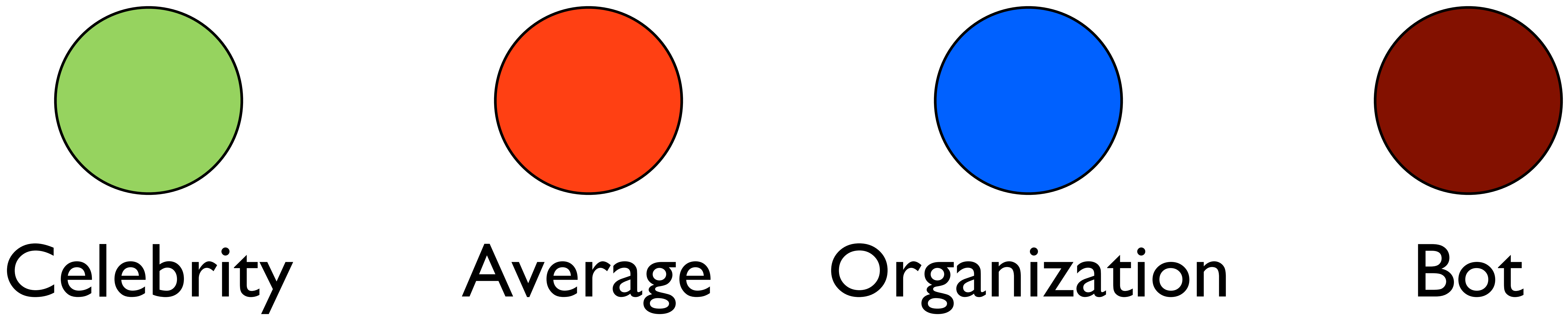


Following another entity

1. Random - follow randomly.
2. Preferential - follow based on the in-degrees of the entities.
3. Entity - follow based on their <u>exterior</u> titles.

3-Follow by entity type

- Follow an entity based on the exterior title.
- On Twitter, there are different types of users



@Pontifex

Celebrity

76 tweets,
+ 5 million
followers

@Canada

Organization

40,000 followers on 1st day

Average

Popularity and influence
comes from activity *within*
the network

Bot

Following another entity

1. Random - follow randomly.
2. Preferential - follow based on the degrees of the entities.
3. Entity - follow based on their exterior titles.
4. Preferential-entity - look to the exterior title first, and then follow based on the degrees of the specific entity types.
5. Message content - based on messages tweeted by entities

Network =

Agent 0	Agent 1	Agent 2	Agent 3	...	Agent N

Array of structs

- id = 1
- Agent type
- Number of tweets, retweets
- Region
- Creation time
- Language
- Ideology
- Humour preference
- Music preference
- List of chatty people
- Following and follower set

Network = Agent 0 | Agent 1 | Agent 2 | Agent 3 | ... | Agent N

Array of structs

- id = 1
- Agent type
- Number of tweets, retweets
- *Region*
- Creation time
- *Language*
- *Ideology*
- *Humour preference*
- *Music preference*
- List of chatty people
- Following and follower set

Network =

Agent 0	Agent 1	Agent 2	Agent 3	...	Agent N

Array of structs

- id = 1
- Average Joe
-
- *Canada*
-
- *English*
- *Politics* = ***Blue***
- *Funny*
- *Likes Jay-Z*
-
-

Network =

Agent 0	Agent 1	Agent 2	Agent 3	...	Agent N

Array of structs

- id = 3
- Average Joe
-
- *Canada*
-
- *English*
- *Politics* = ***Red***
- *Funny*
- *Likes ABBA*
-
-

Agent 1 Agent 3

Agent 1 and 2 share the same geographical location. They both appreciate funny things. They differ politically and musically, however.

Agent 1 can generate funny, *apolitical* tweets

Agent 3 may appreciate those tweets. She may follow Agent 1 and rebroadcast such content

Agent 1 Agent 3

Agent 1 and 2 share the same geographical location. They both appreciate funny things. They differ politically and musically, however.

Agent 1 can also generate politically charged tweets

Agent 3 will not rebroadcast these, as they are inconsistent with her political views

Agent 1 Agent 3

Agent 1 and 2 share the same geographical location. They both appreciate funny things. They differ politically and musically, however.

Agent 1

Agent 3

these agents can communicate because they *share the same language*

Agent 1 can also label tweets with #tags

A political tweet related to Canadian politics	#poli #cdn
…or music	#music
…or a joke	#funny

Network =

Agent 4	Agent 5	Agent 6	Agent 7	...	Agent N

Array of structs

- id = 7
- Average Joe
-
- Australia
-
- *English*
- *Politics* = ***Blue***
- *Funny*
- *Likes Jay-Z*
-
-

Agent 1 can also label tweets with #tags

#poli #cdn

…or #music

…or #funny

Agent 7 lives in a different region, thus she is interested only in #poli #aus even though both are Blue

Agent 1

can also label tweets with #tags

#poli #cdn

…or #music

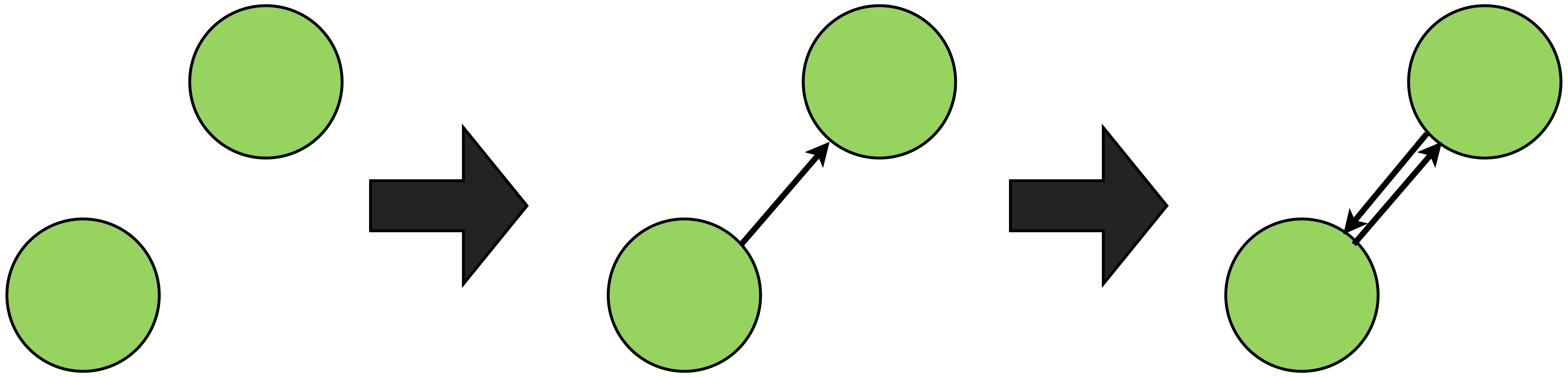
…or #funny

Agent 7

these agents agree on musical tastes and humour

Followback

- In a 15 million user Twitter data set, found that 44% of the links are reciprocal.



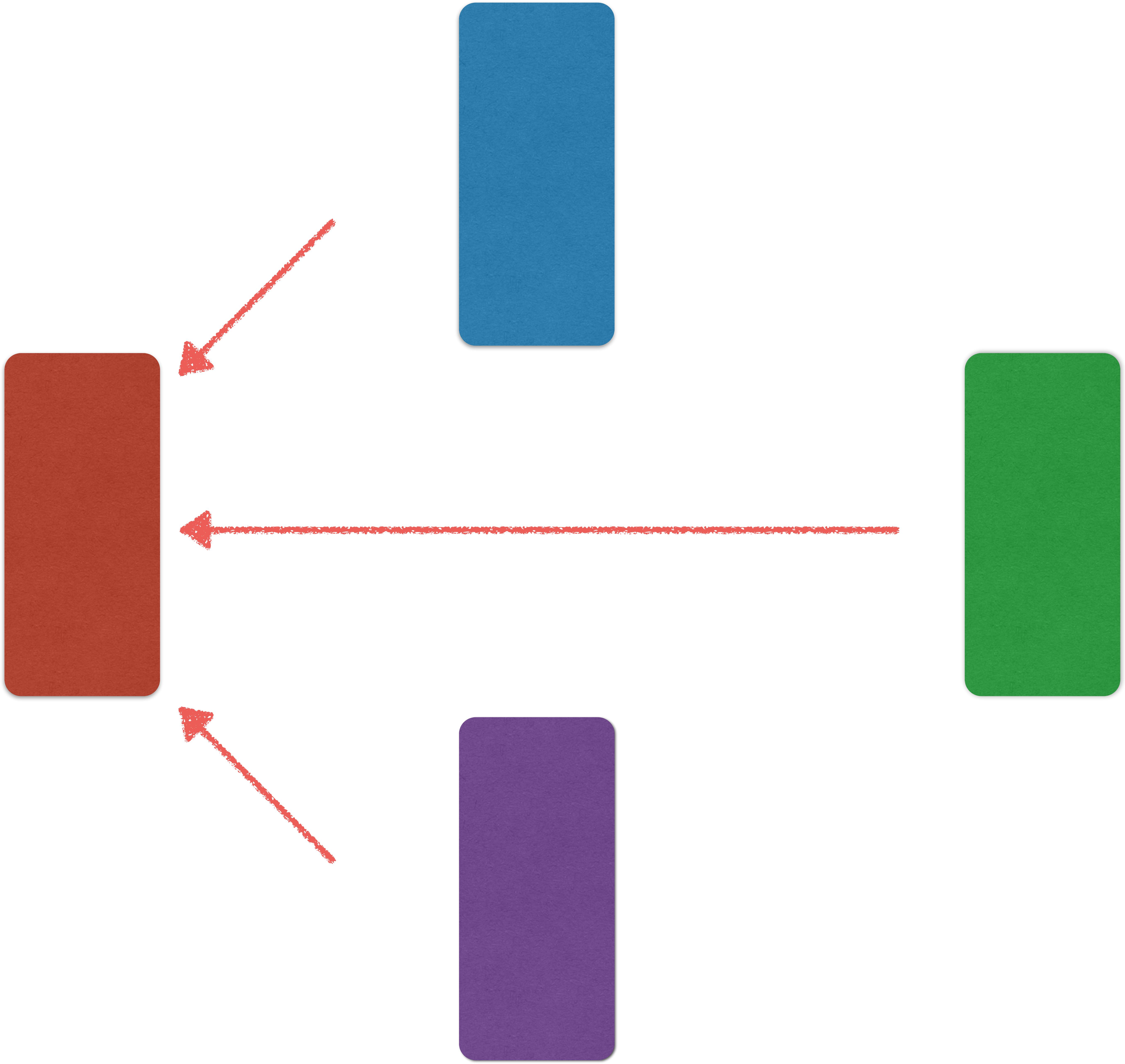
PRL **112,** 048701 (2014) PHYSICAL REVIEW LETTERS week ending 31 JANUARY 2014

Competition-Induced Criticality in a Model of Meme Popularity

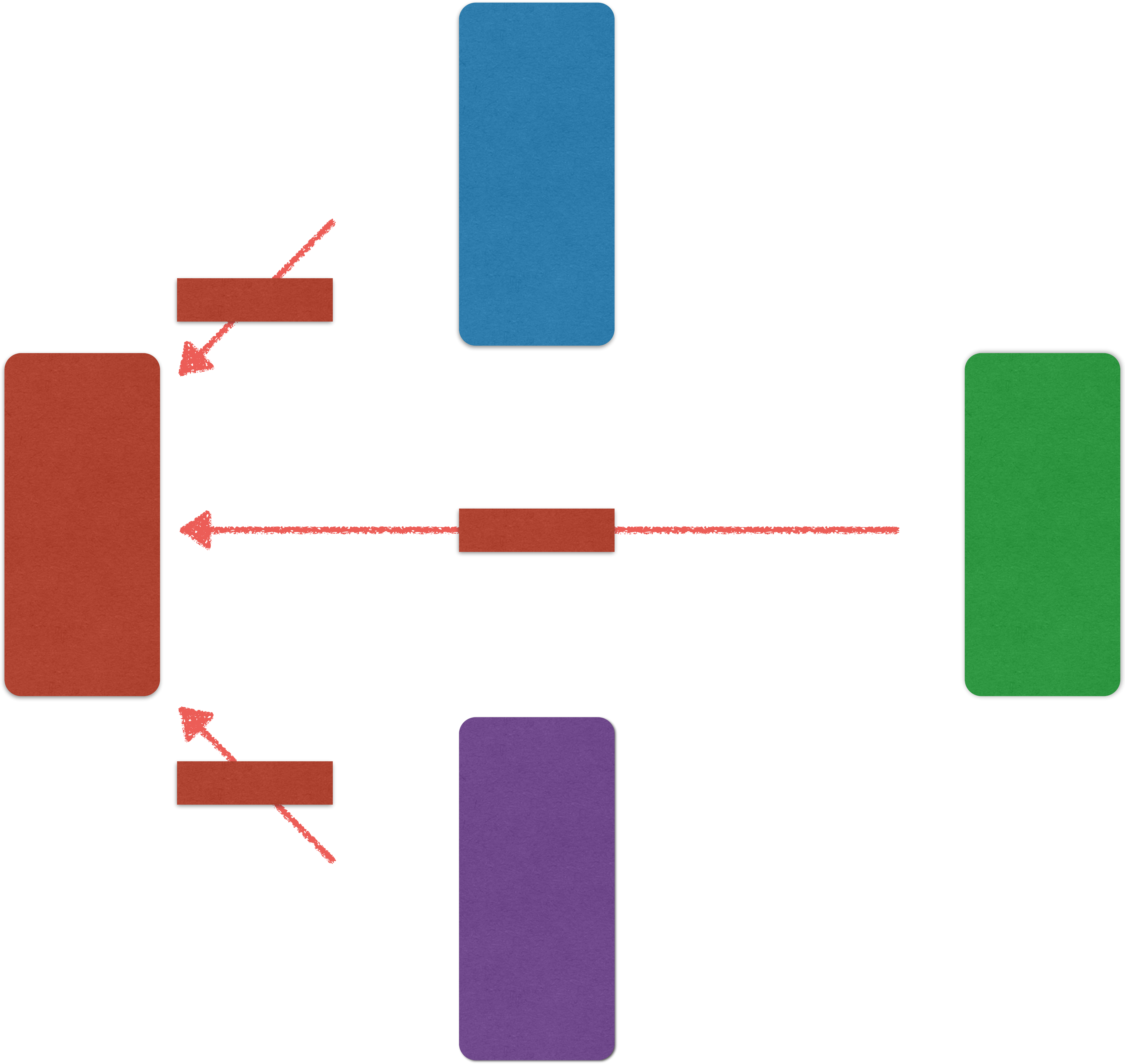
James P. Gleeson,[1*] Jonathan A. Ward,[2] Kevin P. O'Sullivan,[1] and William T. Lee[1]
[1]*MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland*
[2]*Centre for the Mathematics of Human Behaviour, Department of Mathematics and Statistics, University of Reading, Whiteknights RG6 6AH, United Kingdom*
(Received 31 May 2013; revised manuscript received 17 November 2013; published 30 January 2014)

Don’t hog the conversation

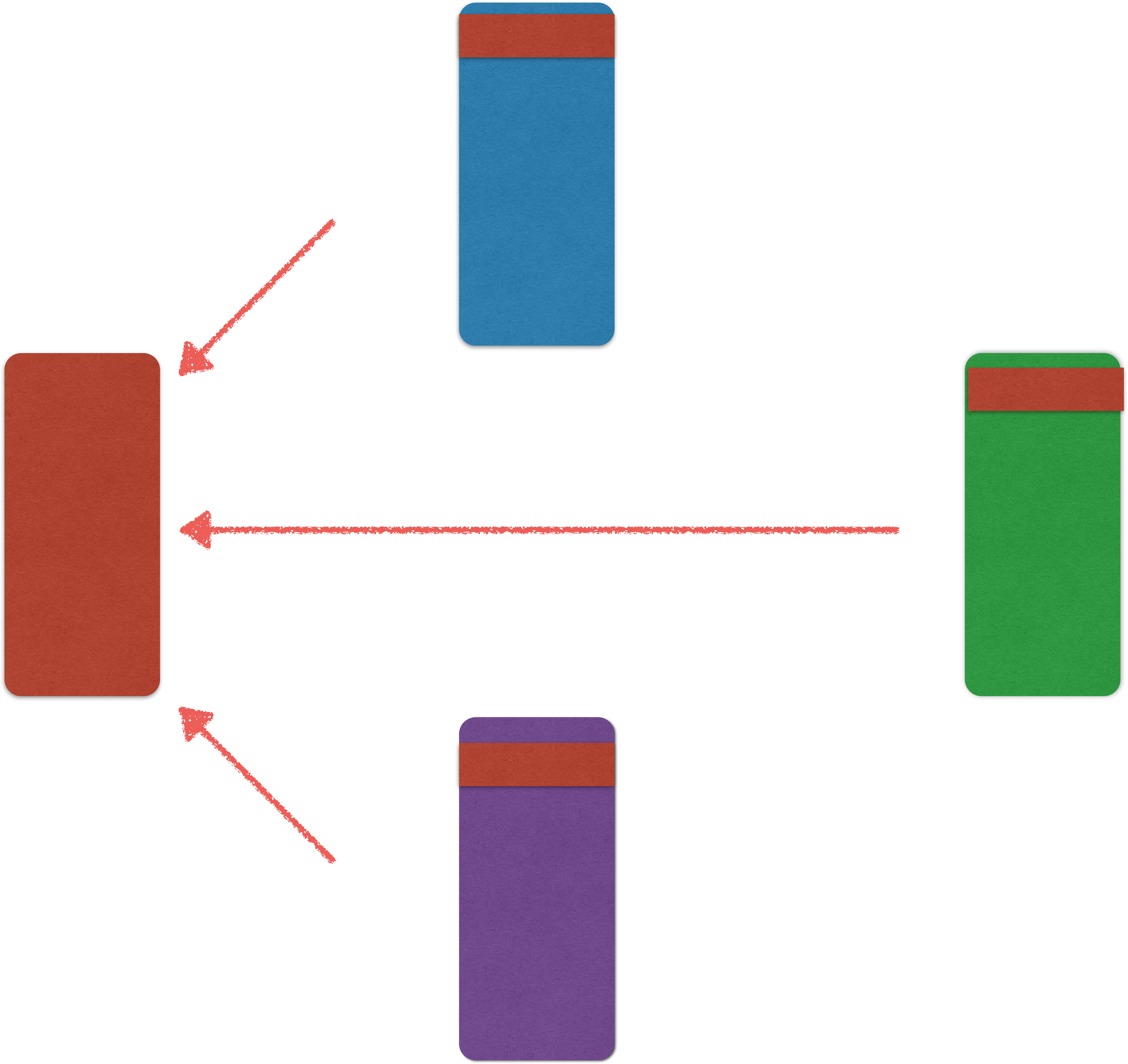
A dinner party



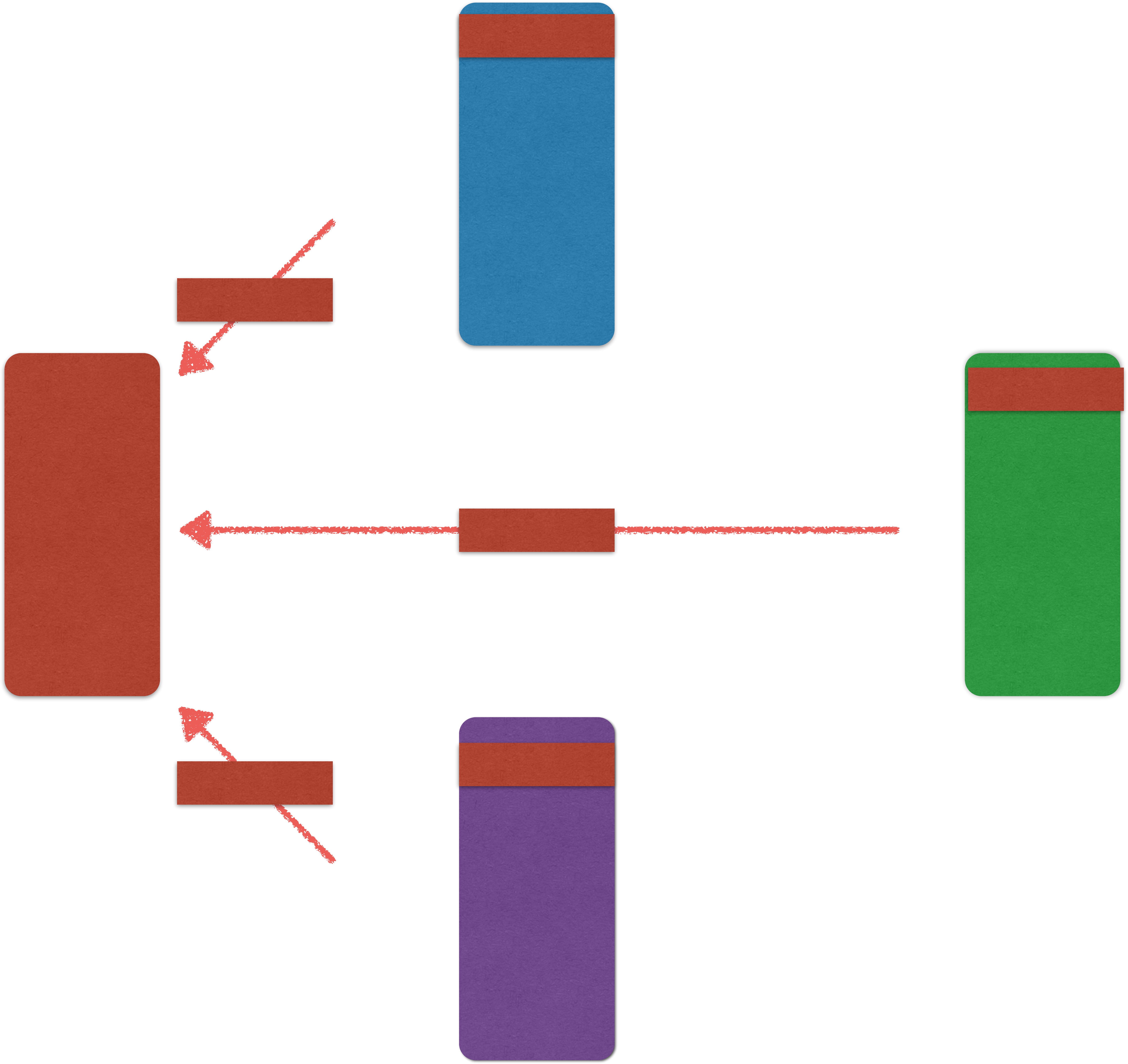
A dinner party



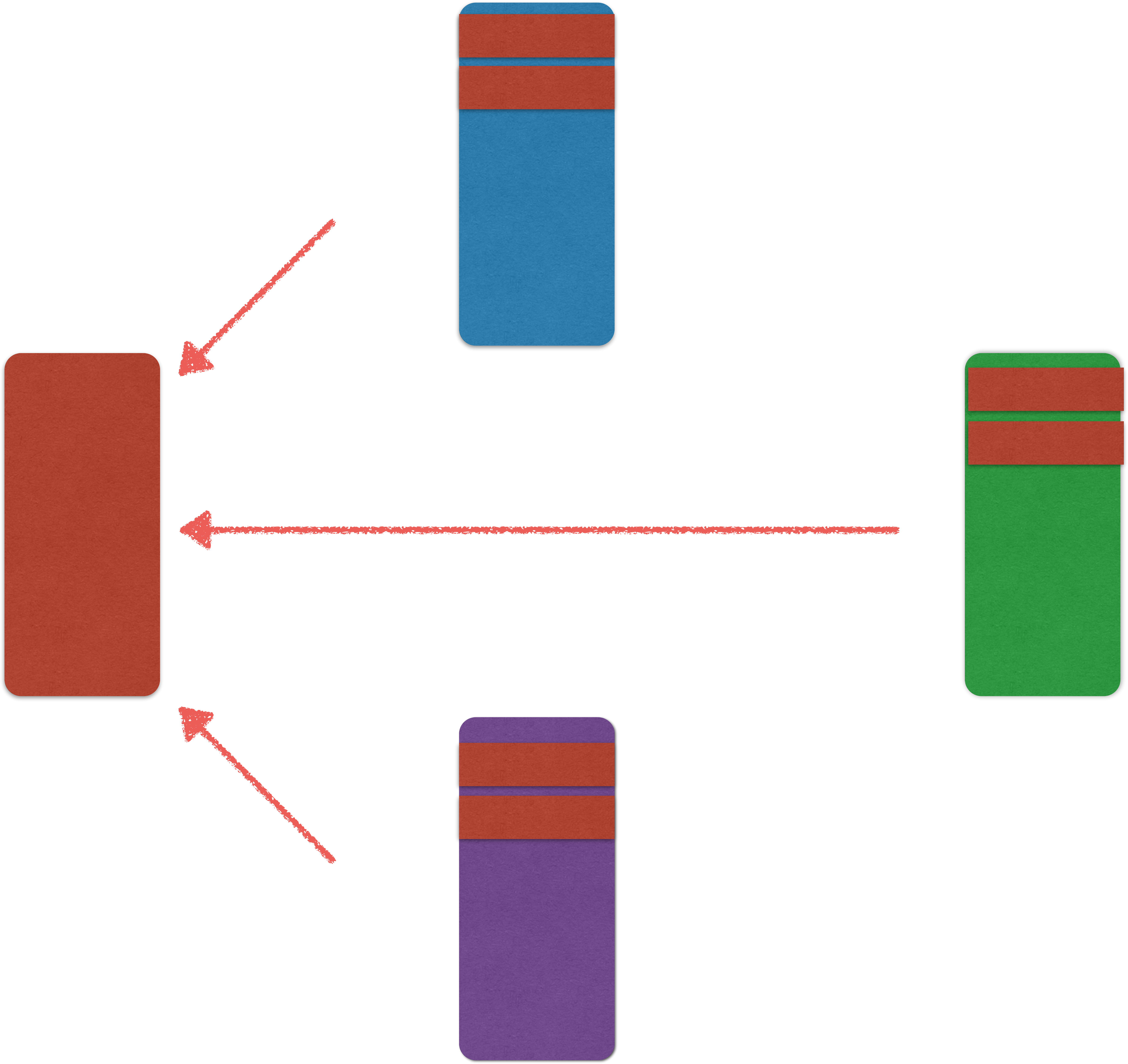
A dinner party



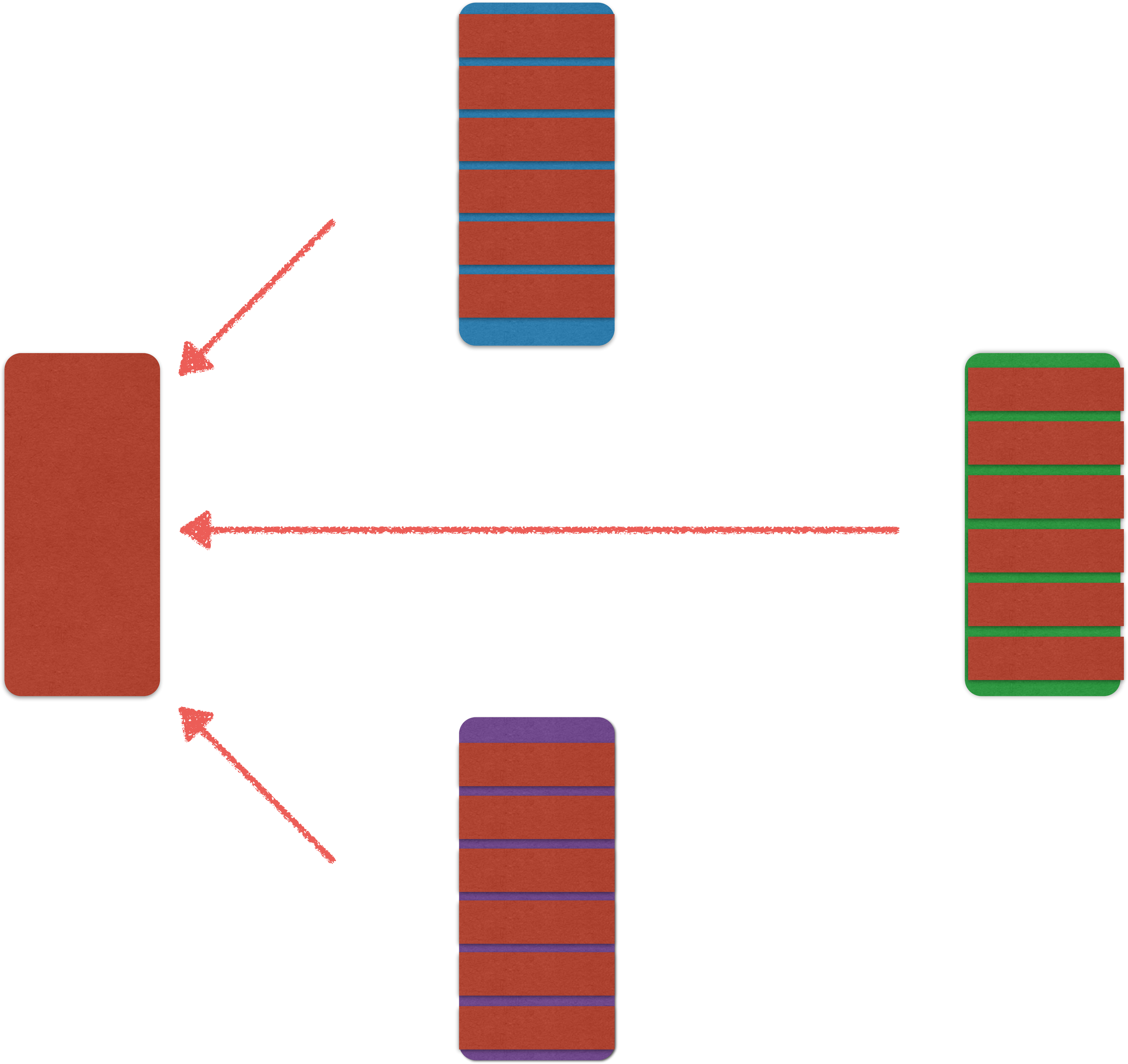
A dinner party



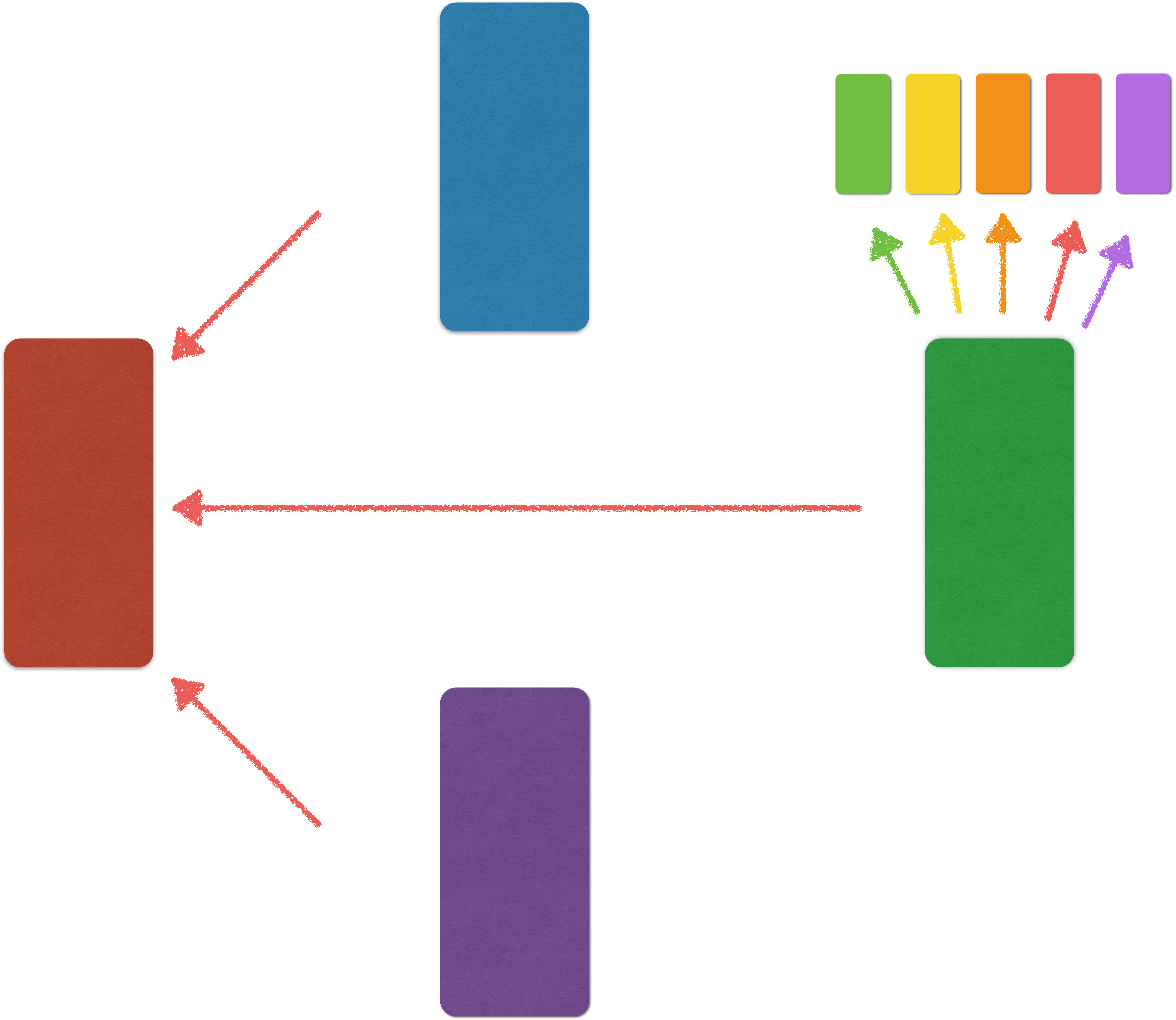
A dinner party



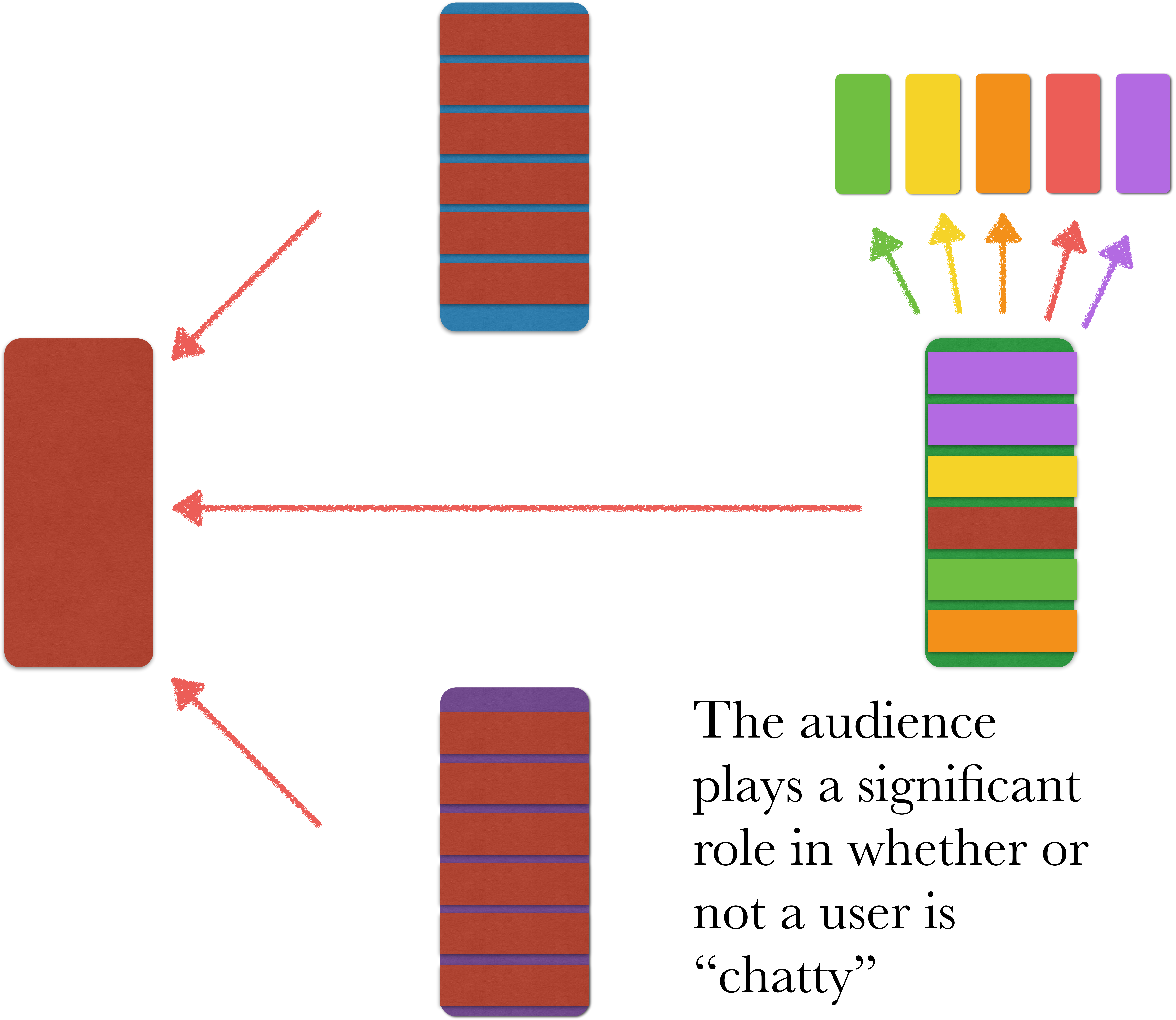
A dinner party



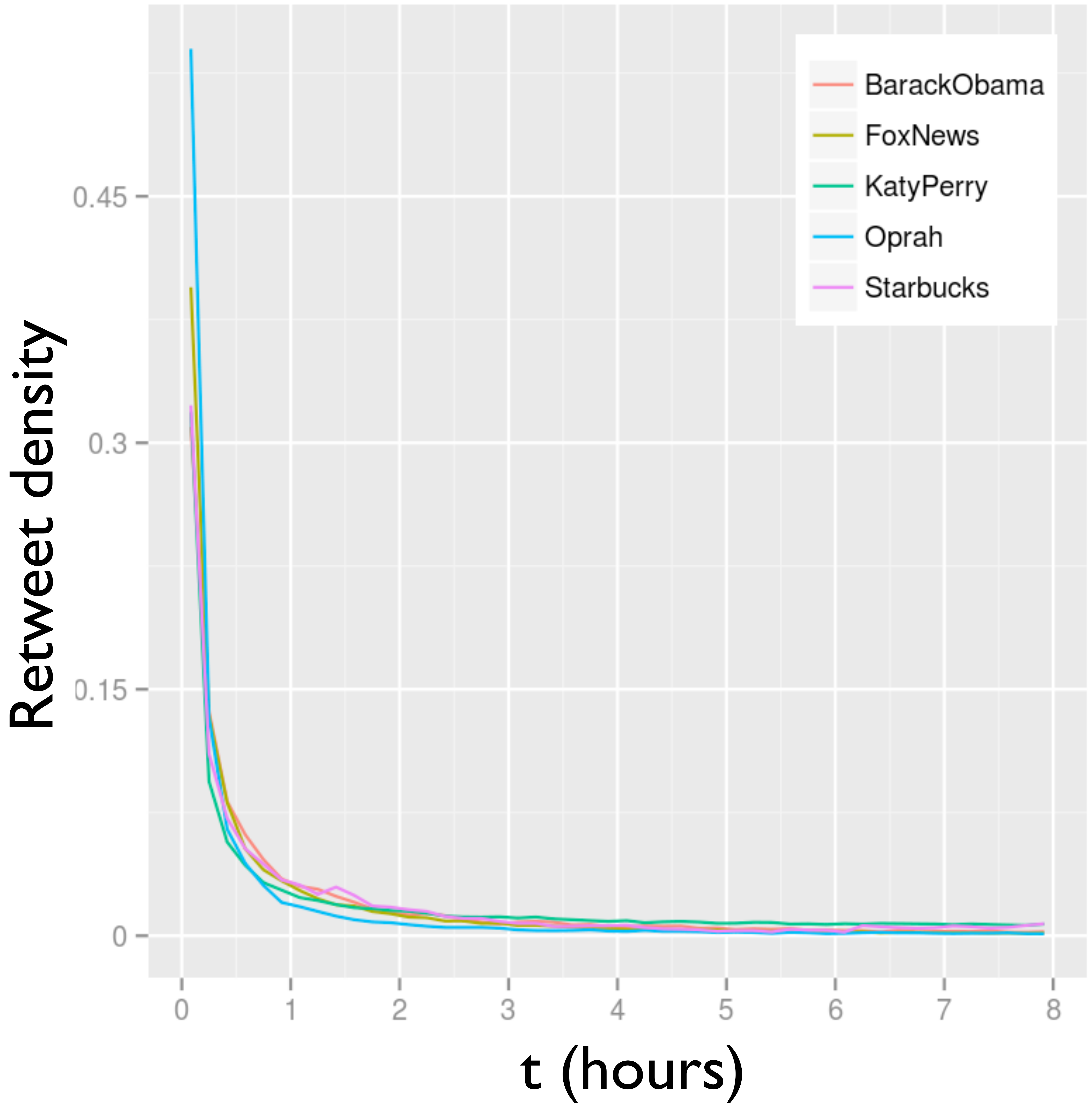
A different kind of party



A different kind of party

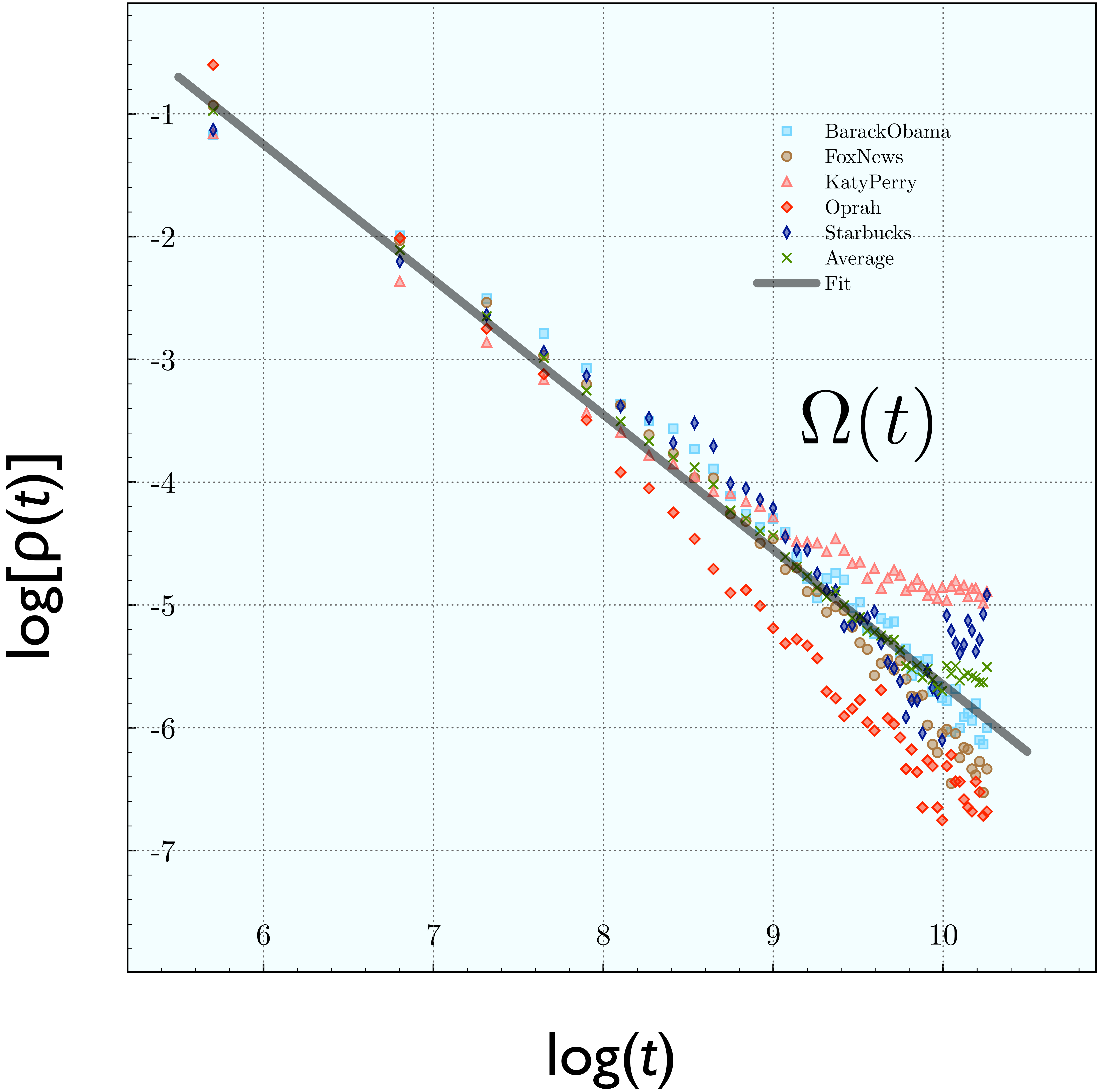


The audience plays a significant role in whether or not a user is "chatty"



BarackObama
FoxNews
KatyPerry
Oprah
Starbucks
0.45
0.3
0.15
0
0
1
2
3
4
5
6
7
8
Retweet density
t (hours)

K. White



BarackObama
FoxNews
KatyPerry
Oprah
Starbucks
Average
Fit
$\Omega(t)$
-1
-2
-3
-4
-5
-6
-7
6
7
8
9
10
$\log[\rho(t)]$
$\log(t)$

K. White

Where does $\Omega(t)$ come from?

Where does $\Omega(t)$ come from?

Where does $\Omega(t)$ come from?

Where does $\Omega(t)$ come from?

Where does $\Omega(t)$ come from?

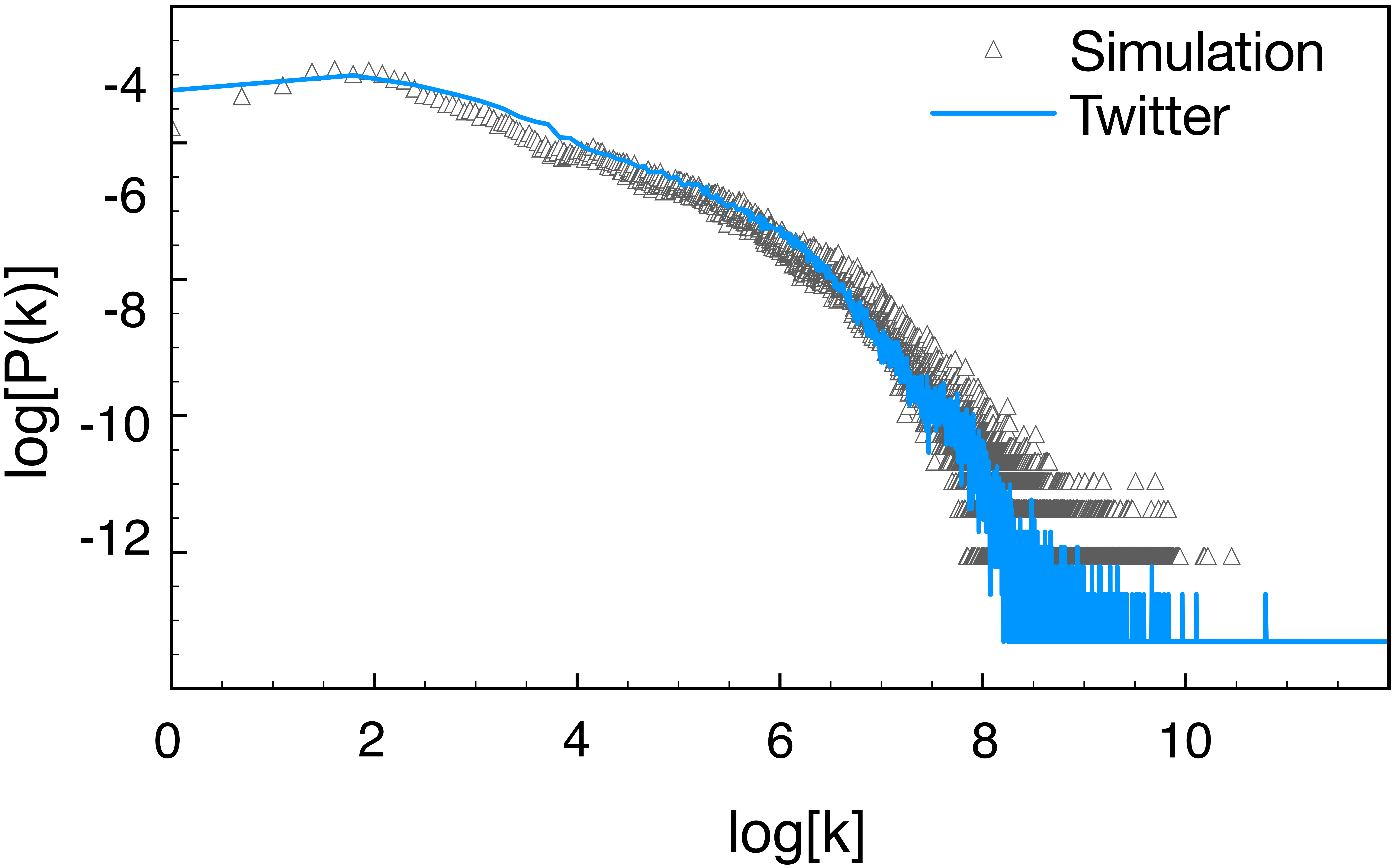
Where does $\Omega(t)$ come from?

Where does $\Omega(t)$ come from?

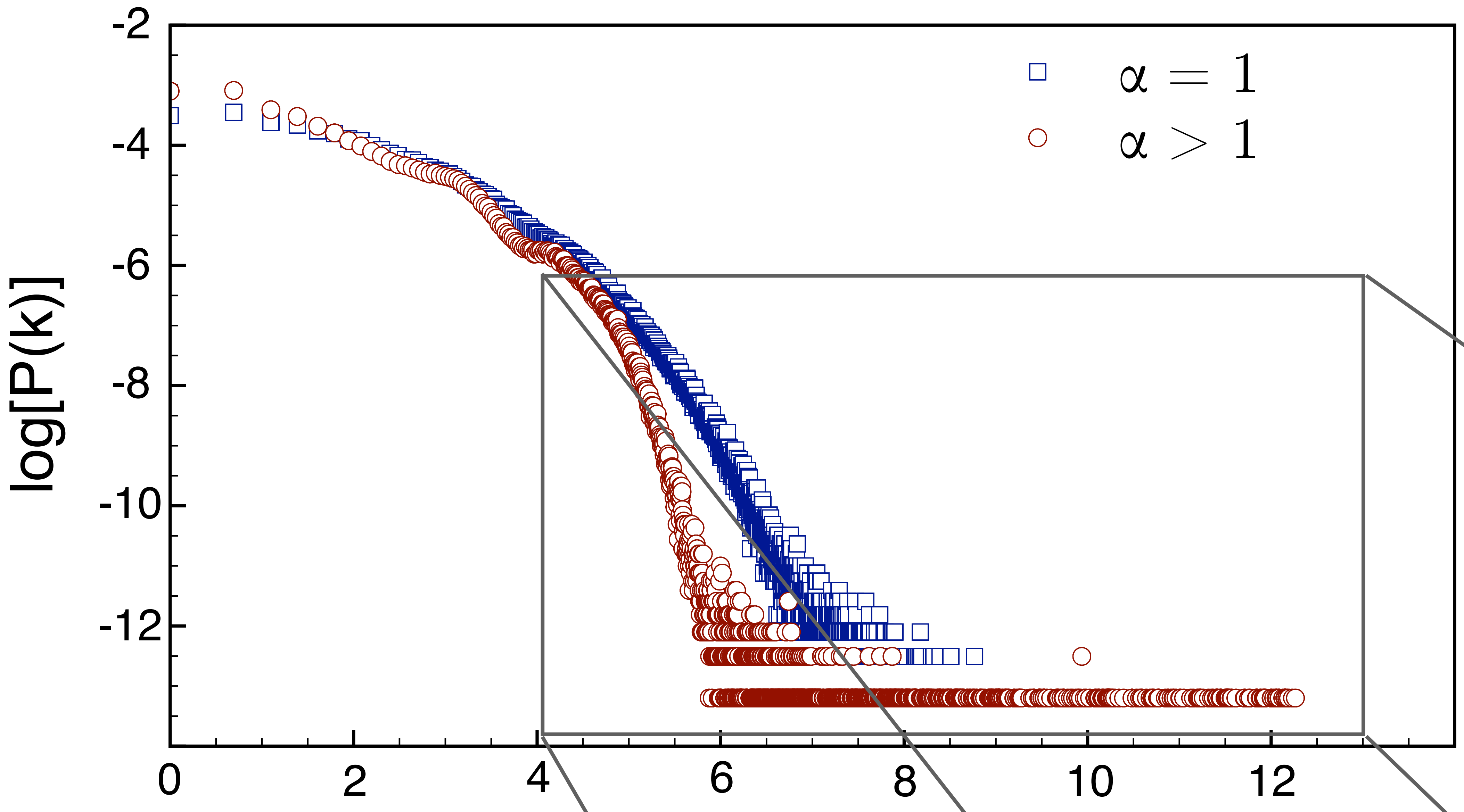
- Users can only see previous events when they appear on screen

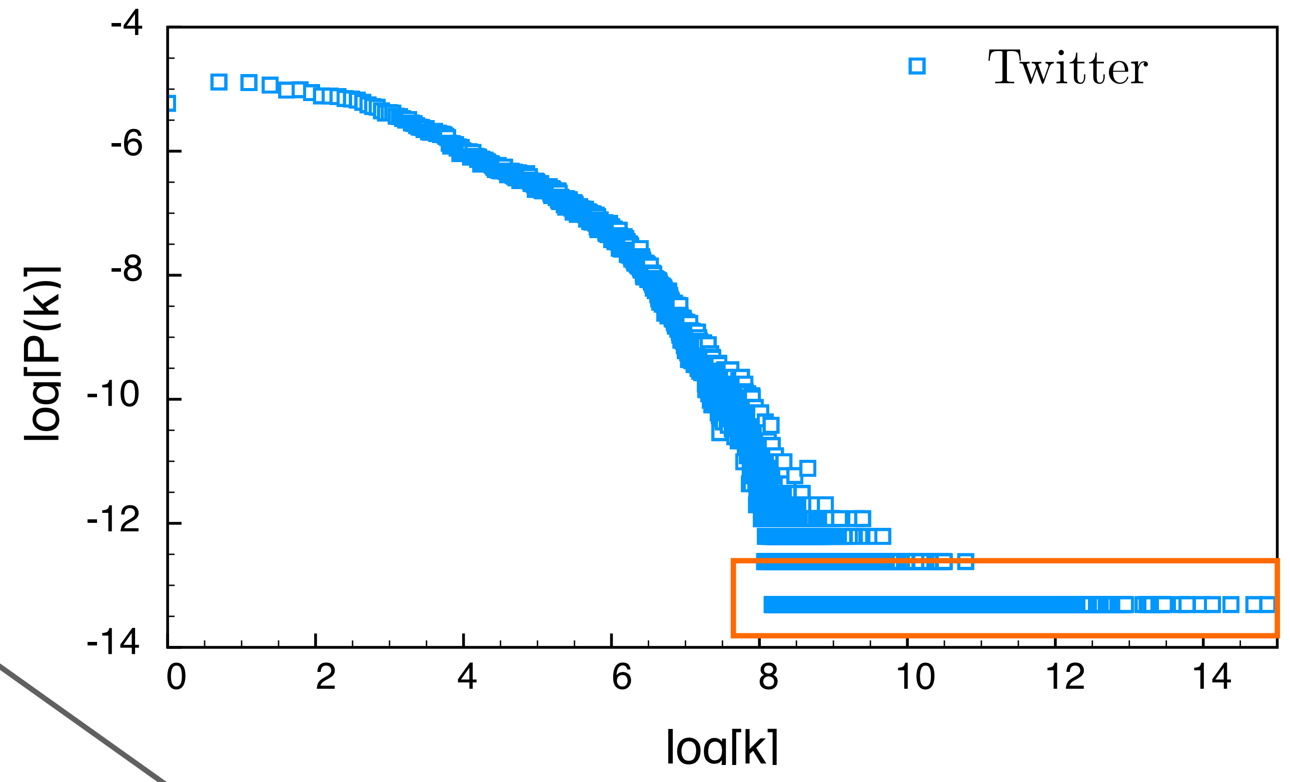
- Users who follow a large number of talkative accounts have "short *time*" vision

- We use $\Omega(t)$ rather than simulate every screen

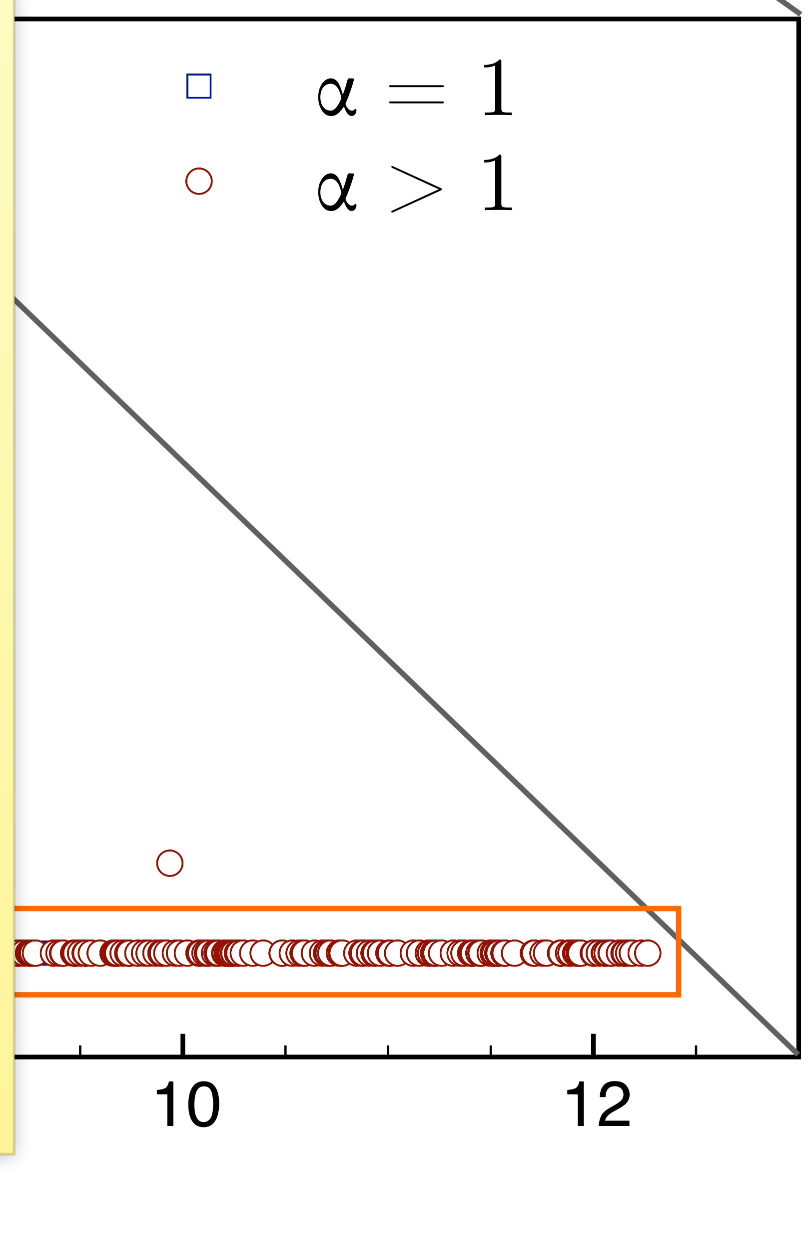


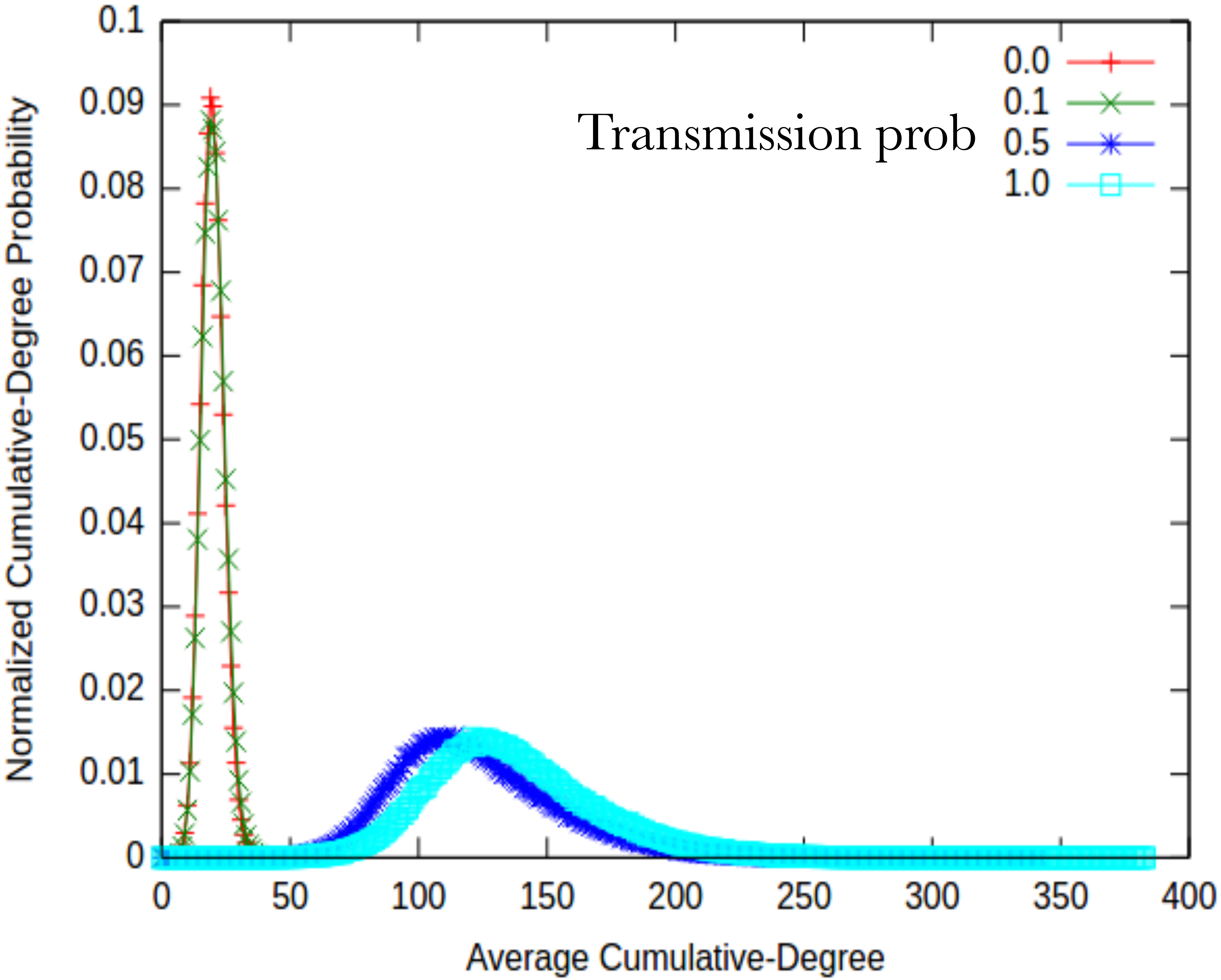
Simulation
Twitter
-4
-6
-8
-10
-12
0
2
4
6
8
10
log[P(k)]
log[k]

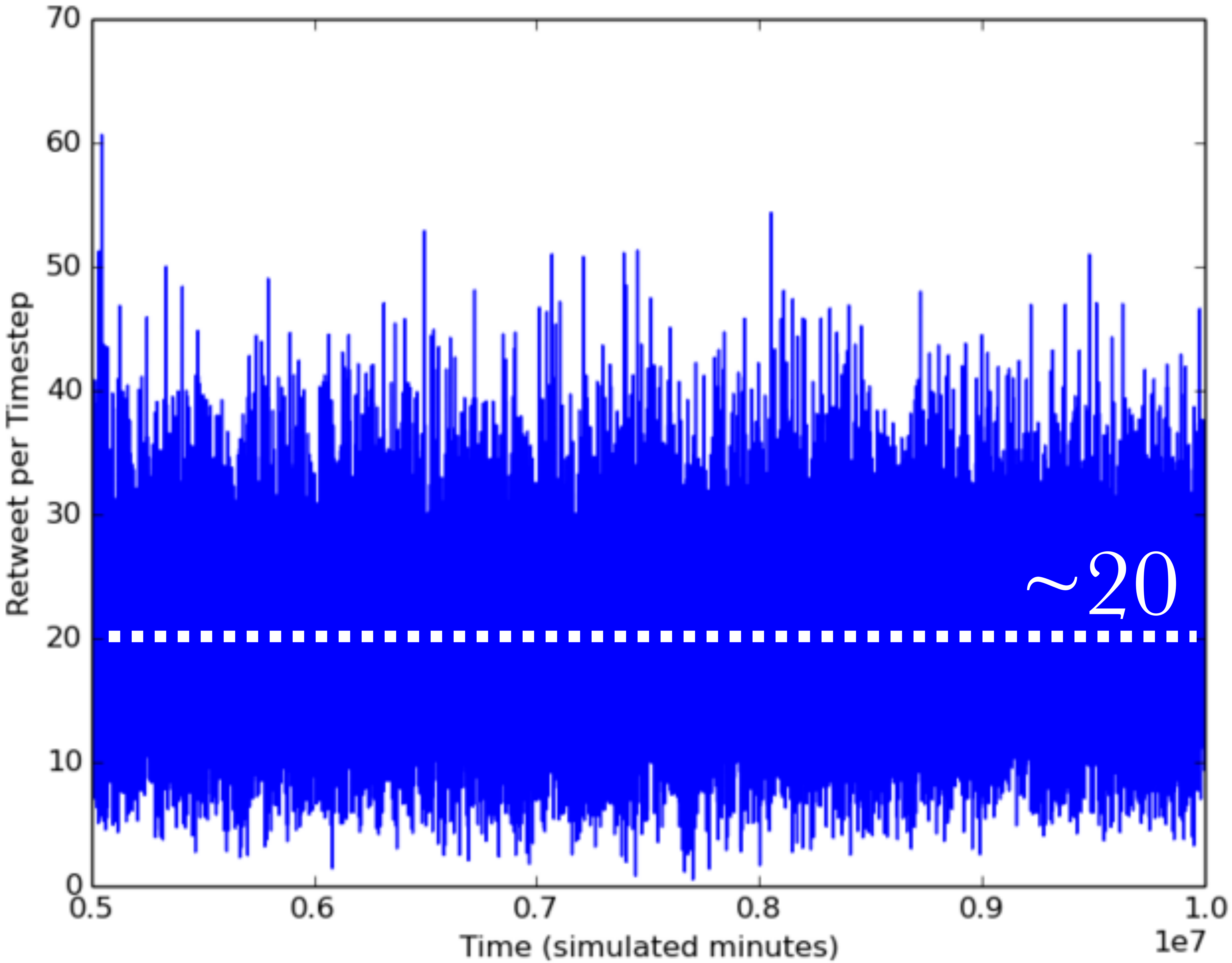




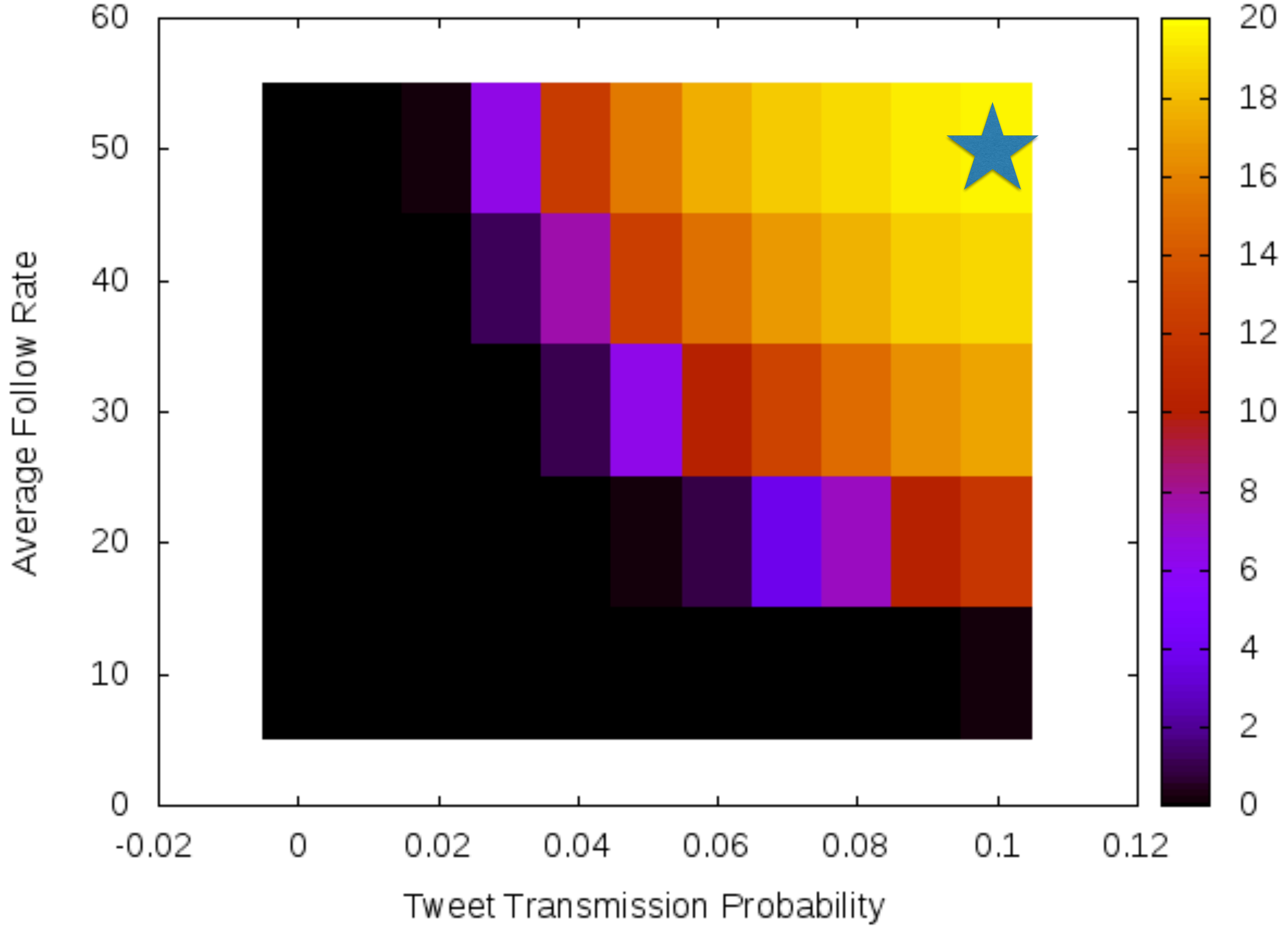
1) Gamma is the probability to follow node i with degree k, it is the mathematical form for preferential attachment.
2) I am zooming into the tail, because when the exponent is > 1 the tail is the part that is most changing and where the highly connected celebrities would lie. It suggests that the exponent alpha is greater than one for some celebrities.



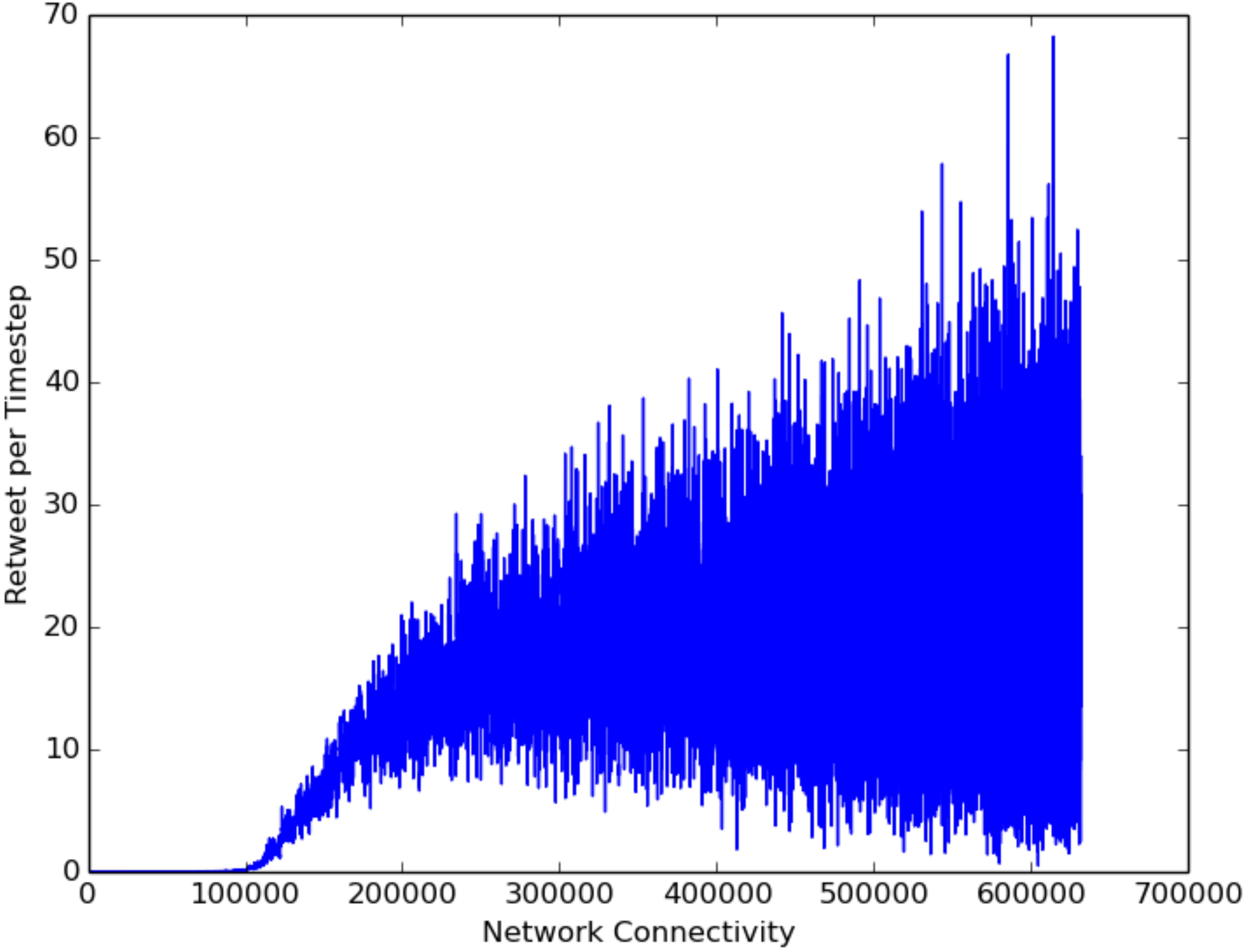

Transmission prob
0.0
0.1
0.5
1.0
Normalized Cumulative-Degree Probability
Average Cumulative-Degree
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
0
50
100
150
200
250
300
350
400



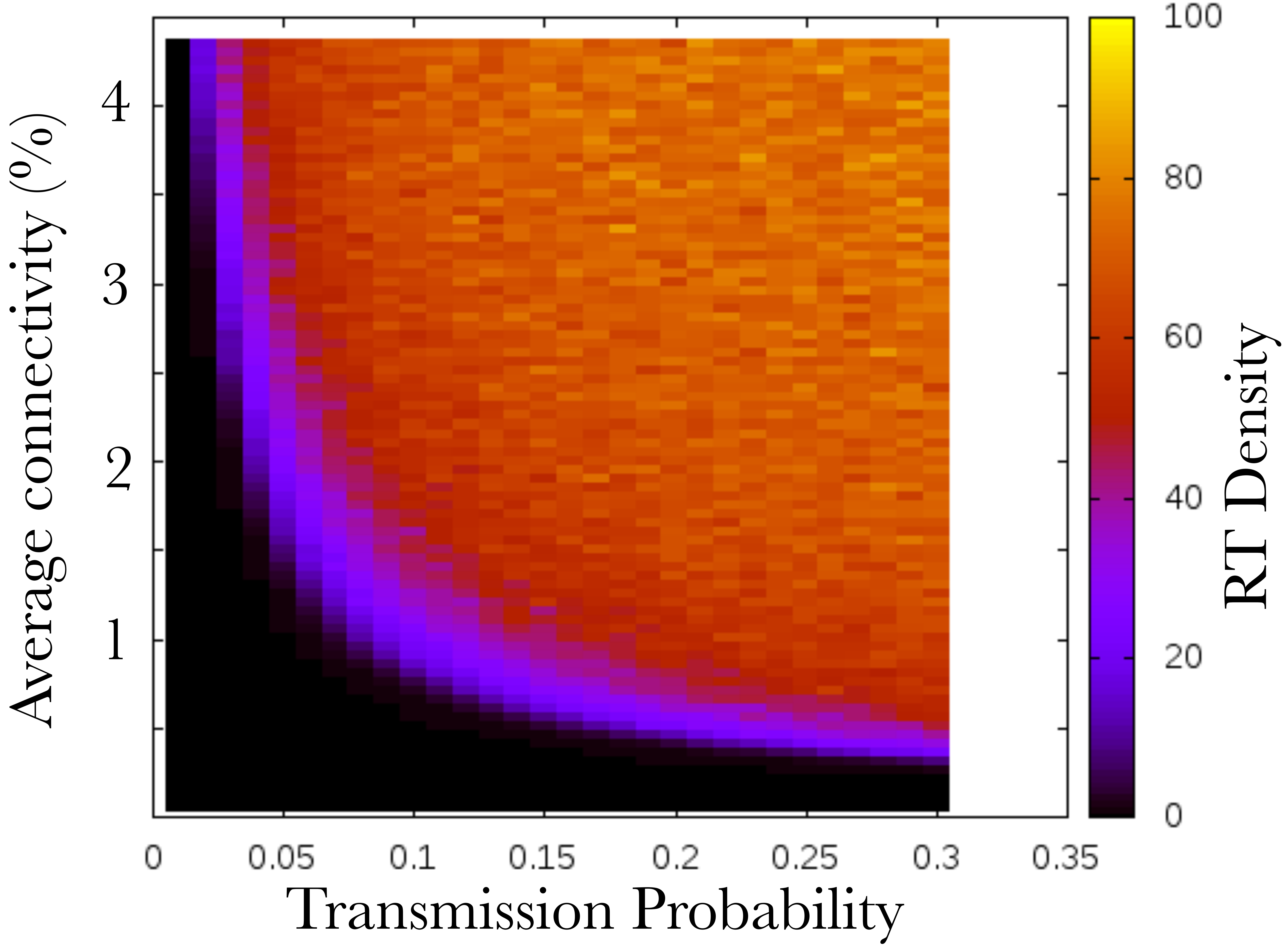
Retweet per Timestep
Time (simulated minutes)
1e7
~ 20

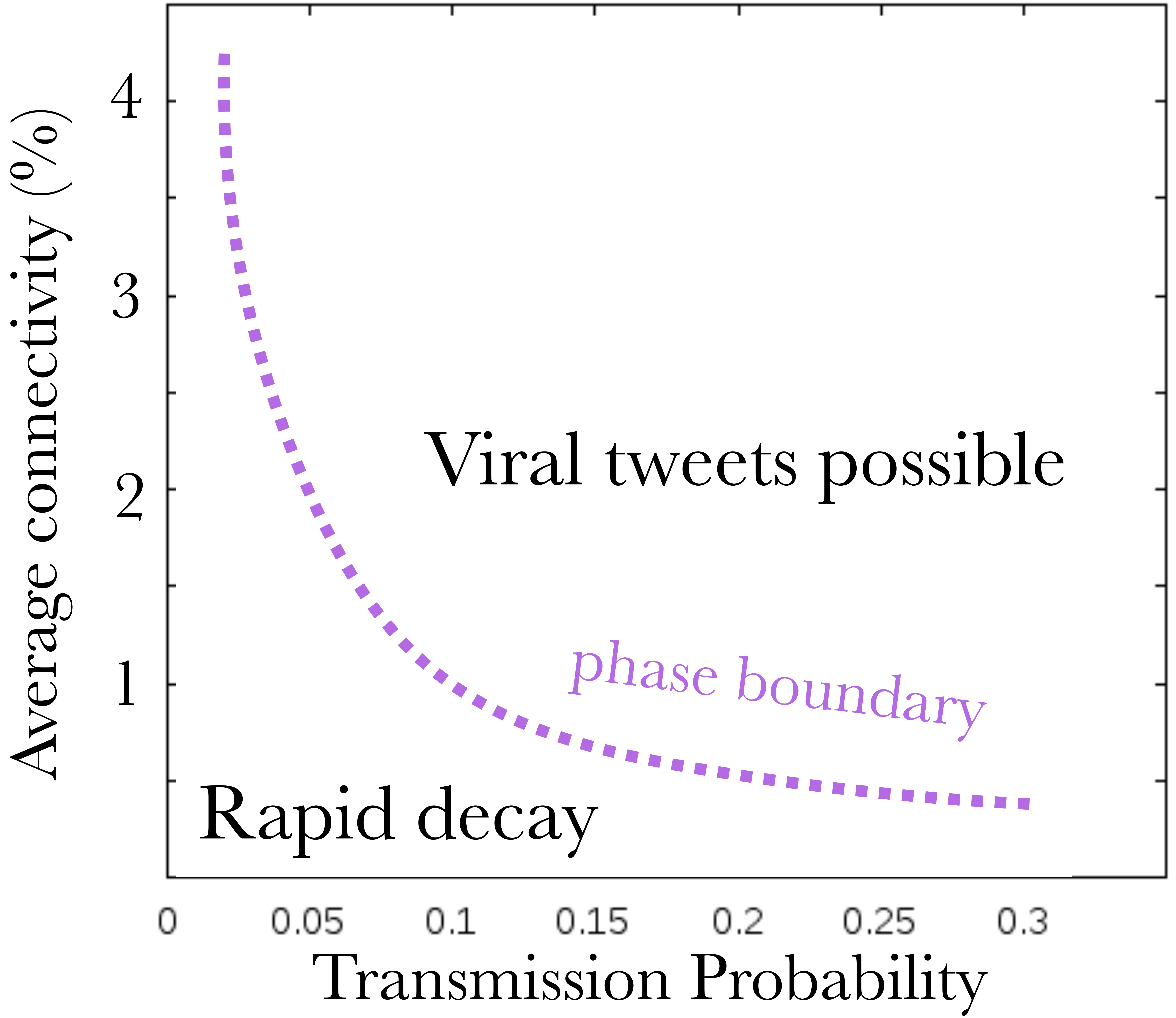


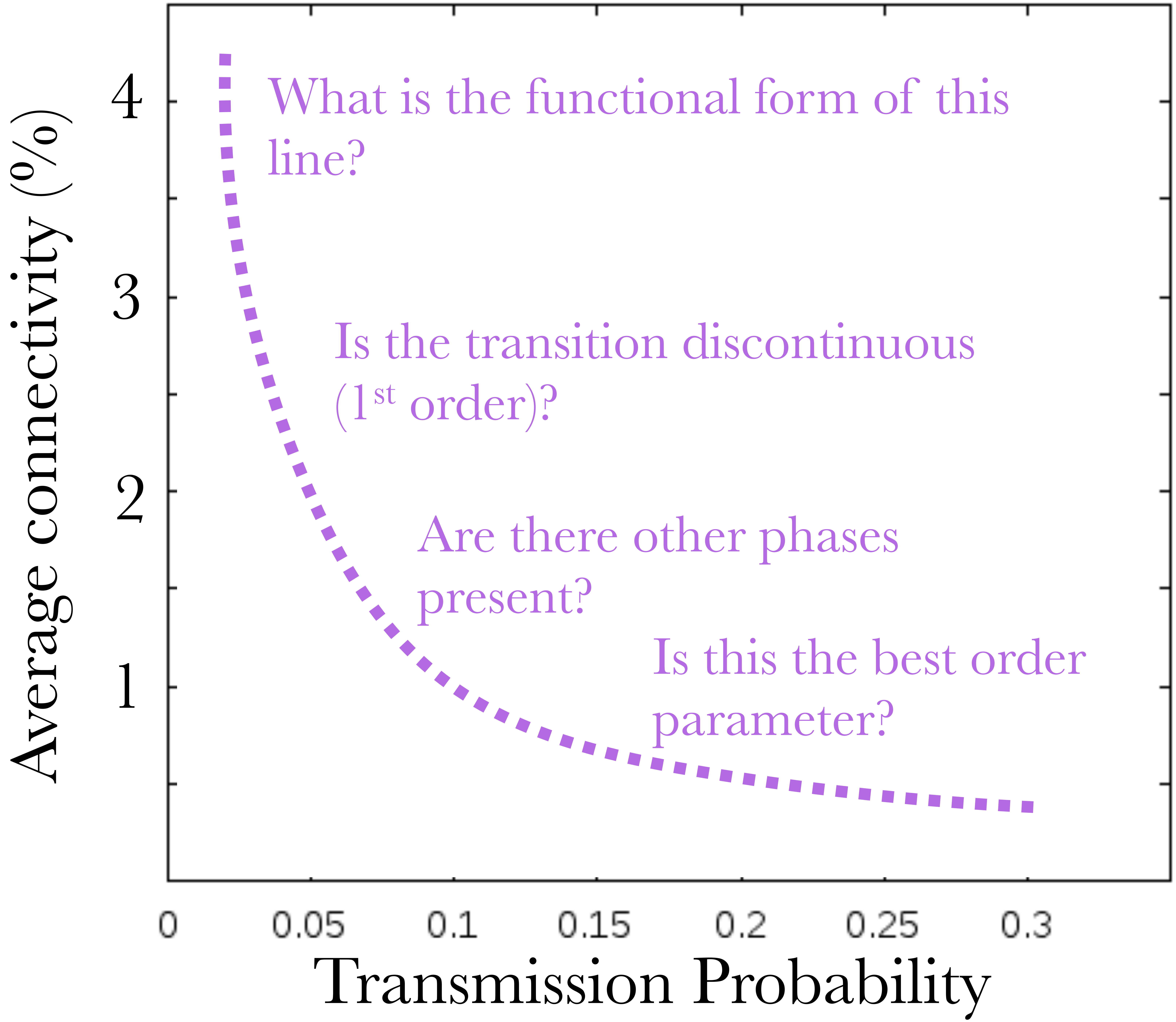
Average Follow Rate
Tweet Transmission Probability
60
50
40
30
20
10
0
-0.02
0
0.02
0.04
0.06
0.08
0.1
0.12
20
18
16
14
12
10
8
6
4
2
0

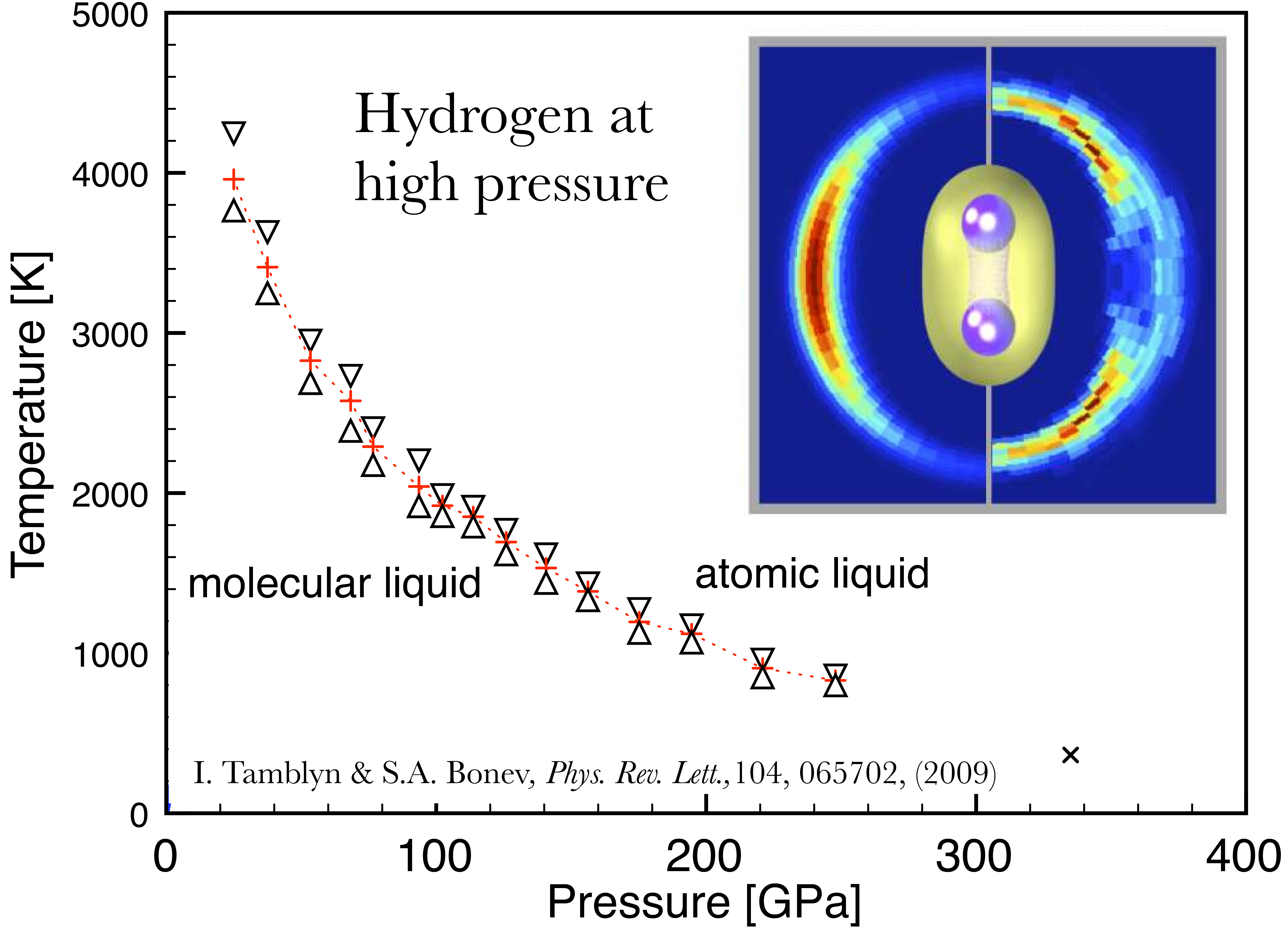

Retweet per Timestep
70
60
50
40
30
20
10
0
0
100000
200000
300000
400000
500000
600000
700000
Network Connectivity


Random Attachment
Preferential Attachment
Total Number of Follows
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
Tweet Transmission Probability
Tweet Transmission Probability
Retweets per Timestep
100
80
60
40
20
0


Average connectivity (%)
4
3
2
1
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
Transmission Probability
100
80
60
40
20
0
RT Density


Average connectivity (%)
4
3
2
1
Viral tweets possible
phase boundary
Rapid decay
0
0.05
0.1
0.15
0.2
0.25
0.3
Transmission Probability

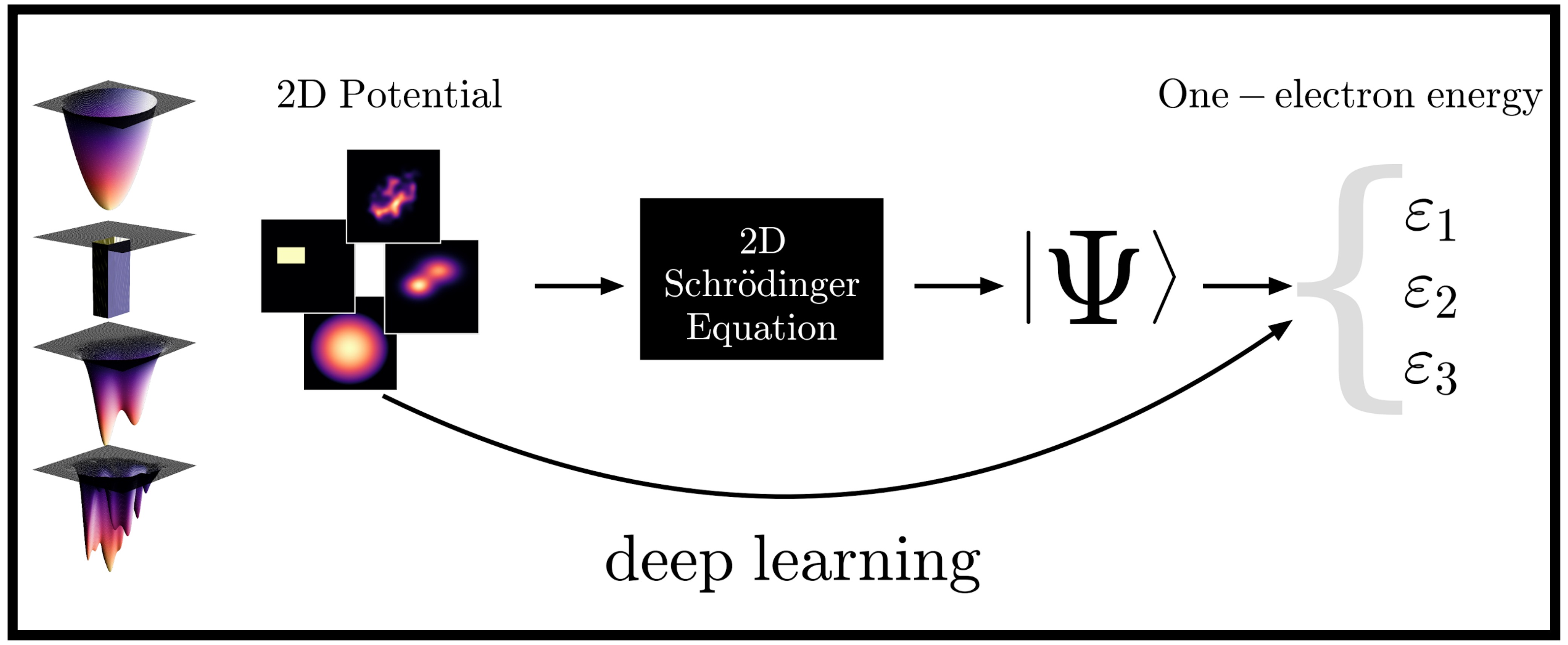

What is the functional form of this line?
Is the transition discontinuous (1st order)?
Are there other phases present?
Is this the best order parameter?
Average connectivity (%)
4
3
2
1
0
0.05
0.1
0.15
0.2
0.25
0.3
Transmission Probability


Hydrogen at high pressure
molecular liquid
atomic liquid
Temperature [K]
Pressure [GPa]
5000
4000
3000
2000
1000
0
0
100
200
300
400
I. Tamblyn & S.A. Bonev, *Phys. Rev. Lett.*, 104, 065702, (2009)

Conclusions

- Simulation accurately describes growth of Twitter
- Chattiness is in the eye of the beholder
- Models are very general: panspermia, religions, yawning, laughter, artistic styles, disease, political ideology - anything where information can be rebroadcast
- http://hashkat.org

(We also do deep learning :)



K. Ryczko, A.Domurad, N. Buhagiar, I. Tamblyn, *Social Network Analysis and Mining*, 7: 4, (2017)

NSERC CRSNG

CLEAN Computational Laboratory for Energy And Nanoscience