

# Compare graphs and activity between datasets

Antoine Vendeville

UPMC

March 18, 2019

# Overview

## 1 General stats

## 2 Degrees stats

- Degrees
- In-Degrees
- Out-Degrees

## 3 Connected components stats

- Weakly connected components
- Strongly connected components

## 4 Activity stats

- Tweets
- Retweets
- Retweeted

## 5 $\lambda, \mu, \nu$

# General stats

	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)
Nb users	2 297 179	93 892		1 340 816	90 908	100 854	65 391
Nb relations	9 653 115	517 421		26 294 030	443 399	120 577	76 869
Mean nb relations	8.40	11.02		39.22	9.75	2.39	2.35
Nb events	27 916 100	1 945 365	X	33 540 167	X	3 999 128	1 952 832
Mean nb events/u	12.15	10.71	X	25.01	X	39.65	29.86
% retweets	55	65	X	99	X	23	29
Mean nb tweets/u	3.15	3.71	X	0.17	X	20.55	14.64
Mean nb retweets/u	3.79	6.98	X	24.84	X	6.06	5.97
Mean nb retweeted/u	2.81						

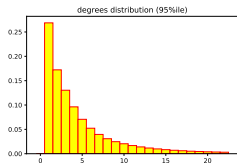
# Overview

- 1 General stats
- 2 Degrees stats
  - Degrees
  - In-Degrees
  - Out-Degrees
- 3 Connected components stats
  - Weakly connected components
  - Strongly connected components
- 4 Activity stats
  - Tweets
  - Retweets
  - Retweeted
- 5  $\lambda, \mu, \nu$

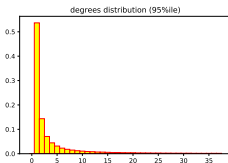
# Degrees stats

	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)	WCano(c)	Russian(c)	Weibo(c)
Mean	8.40	12.57		39.22	9.75	2.39	2.35	9.68	19.00	48.41
Mid	3	1		15	6	1	1	6	2	34
Min	1	1		1	1	1	1	1	1	1
Max	15 184	10 480		455 506	230	8 141	7 922	167	2 687	3 626
25-percentile	1	1		5	2	1	1	3	2	12
75-percentile	7	4		29	12	1	1	13	8	70
90-percentile	14	17		46	22	2	2	22	34	112
95-percentile	23	38		60	32	3	3	29	81	142
99-percentile	86	155		108	66	22	22	48	316	216

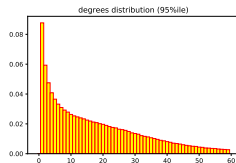
# Degrees distribution



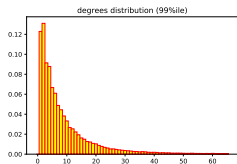
(a) WCanO



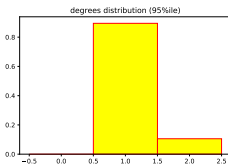
(b) Russian



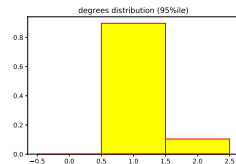
(c) Weibo(i)



(d) TDN(T)

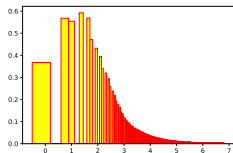


(e) TDN(i10)

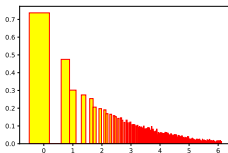


(f) TDN(i11)

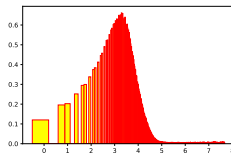
# Degrees distribution (log)



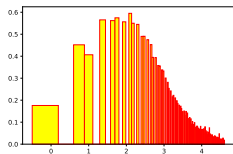
(a) WCanO



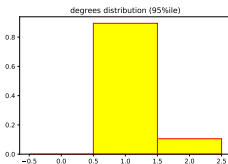
(b) Russian



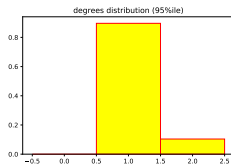
(c) Weibo(i)



(d) TDN(T)



(e) TDN(i10)



(f) TDN(i11)

# Overview

## 1 General stats

## 2 Degrees stats

- Degrees
- In-Degrees
- Out-Degrees

## 3 Connected components stats

- Weakly connected components
- Strongly connected components

## 4 Activity stats

- Tweets
- Retweets
- Retweeted

## 5 $\lambda, \mu, \nu$

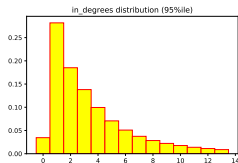


# In-Degrees stats

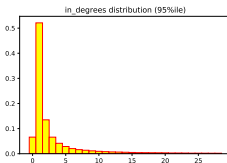
	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)
Mean	4.20	6.29		19.61	4.88	1.20	1.18
Mid	3	1		14	3	0	0
Min	0	0		0	0	0	0
Max	112	554		291	101	2 736	395
25-percentile	1	1		5	1	0	0
75-percentile	5	4		29	6	0	1
90-percentile	9	13		44	11	1	1
95-percentile	13	28		56	17	2	2
99-percentile	26	91		83	33	17	6

	WCano(c)	Russian(c)	Weibo(c)
	4.84	9.50	24.21
	3	1	17
	0	0	0
	77	1 292	453
	1	1	6
	6	4	35
	11	17	56
	14	40	71
	24	159	108

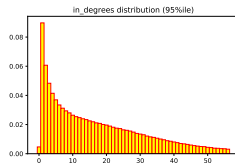
# In-Degrees distribution



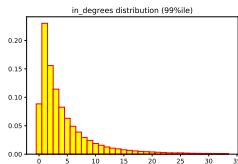
(a) WCanO



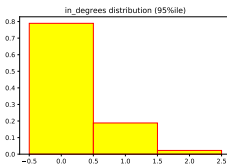
(b) Russian



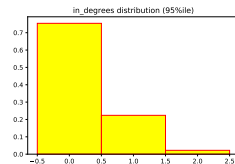
(c) Weibo(i)



(d) TDN(T)

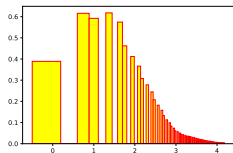


(e) TDN(i10)

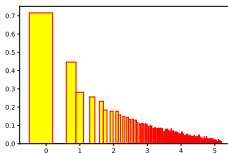


(f) TDN(i11)

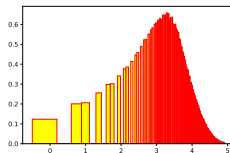
# In-Degrees distribution (log)



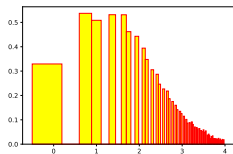
(a) WCano



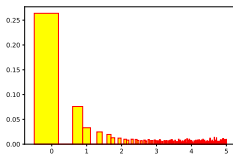
(b) Russian



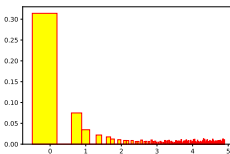
(c) Weibo(i)



(d) TDN(T)



(e) TDN(i10)



(f) TDN(i11)

# Overview

## 1 General stats

## 2 Degrees stats

- Degrees
- In-Degrees
- Out-Degrees

## 3 Connected components stats

- Weakly connected components
- Strongly connected components

## 4 Activity stats

- Tweets
- Retweets
- Retweeted

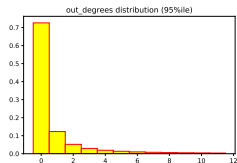
## 5 $\lambda, \mu, \nu$

# Out-Degrees stats

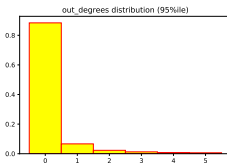
	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)
Mean	4.20	6.29		19.61	4.88	1.20	1.18
Mid	0	0		0	3	1	1
Min	0	0		0	0	0	0
Max	15 174	10 472		455 504	224	8 141	7 922
25-percentile	0	0		0	1	1	1
75-percentile	1	0		0	6	1	1
90-percentile	4	1		0	12	1	1
95-percentile	11	5		0	17	2	2
99-percentile	71	73		46	36	4	4

WCano(c)	Russian(c)	Weibo(c)
4.84	9.50	24.21
3	1	17
0	0	0
101	1 395	3 626
1	1	6
6	4	35
11	17	56
15	41	71
24	158	108

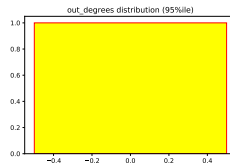
# Out-Degrees distribution



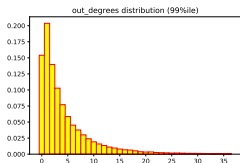
(a) WCanO



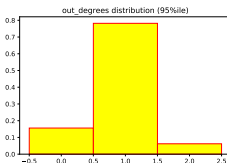
(b) Russian



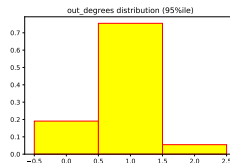
(c) Weibo(i)



(d) TDN(T)

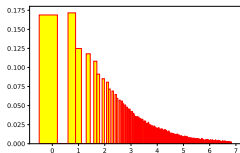


(e) TDN(i10)

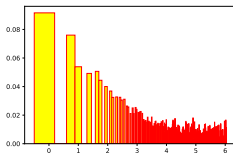


(f) TDN(i11)

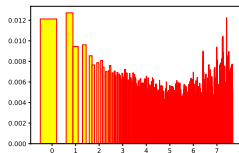
# Out-Degrees distribution (log)



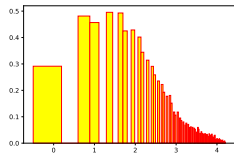
(a) WCano



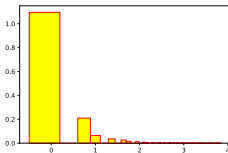
(b) Russian



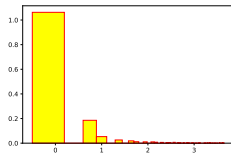
(c) Weibo(i)



(d) TDN(T)



(e) TDN(i10)



(f) TDN(i11)

# Overview

## 1 General stats

## 2 Degrees stats

- Degrees
- In-Degrees
- Out-Degrees

## 3 Connected components stats

- Weakly connected components
- Strongly connected components

## 4 Activity stats

- Tweets
- Retweets
- Retweeted

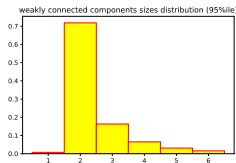
## 5 $\lambda, \mu, \nu$



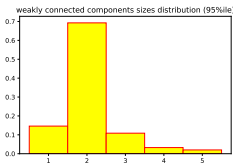
# Weakly connected components: stats

	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)	WCano(c)	Russian(c)	Weibo(c)
Number	31 706	1 659		38	1 370	2 973	1 823	31 631	1 135	38
Mean size	72.45	74.29		35 285	66.36	33.92	35.87	76.28	105.10	35 284.63
Mid size	2	2		1	2	2	2	2	2	1
Min size	1	1		1	1	1	1	1	1	1
Max size	2 191 998	119 251		1 340 768	87 349	94 094	61 076	2 309 181	116 371	1 340 768
25-percentile	2	2		1	2	2	2	2	2	1
75-percentile	3	2		2	3	2	2	3	2	2
90-percentile	4	3		2	4	3	3	5	4	2
95-percentile	6	5		2.15	5	3.4	4	7	5	2.15
99-percentile	19	11		844 685	9.31	7	9	20	10.70	844 685

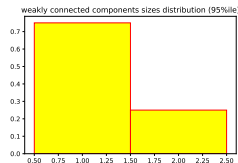
# Weakly CC: sizes distribution



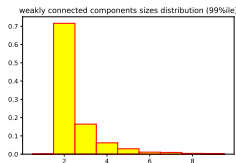
(a) WCan



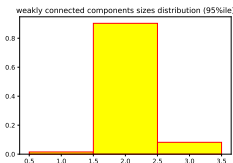
(b) Russian



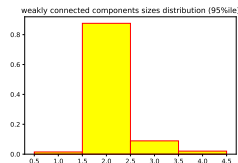
(c) Weibo(i)



(d) TDN(T)



(e) TDN(i10)



(f) TDN(i11)

# Overview

## 1 General stats

## 2 Degrees stats

- Degrees
- In-Degrees
- Out-Degrees

## 3 Connected components stats

- Weakly connected components
- Strongly connected components

## 4 Activity stats

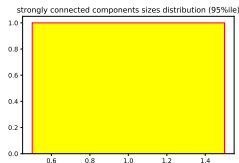
- Tweets
- Retweets
- Retweeted

## 5 $\lambda, \mu, \nu$

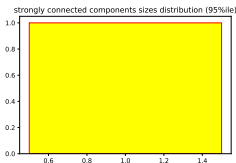
# Strongly connected components: stats

	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)	WCano(c)	Russian(c)	Weibo(c)
Number	2 297 179	116 038		1 325 415	26 822	99 945	64 692	2 412 920	8.48	8 067
Mean size	1	1.06		1.01	3.39	1.01	1.01	1	1	166.21
Mid size	1	1		1	1	1	1		1	1
Min size	1	1		1	1	1	1		105 001	1
Max size	1	7 081		14 785	59 440	43	77		1	1 332 742
25-percentile	1	1		1	1	1	1		1	1
75-percentile	1	1		1	1	1	1		1	1
90-percentile	1	1		1	2	1	1		1	1
95-percentile	1	1		1	2	1	1		1	1
99-percentile	1	1		1	4	1	1		1	1

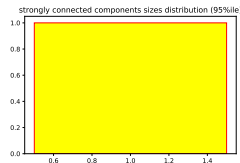
# Strongly CC: sizes distribution



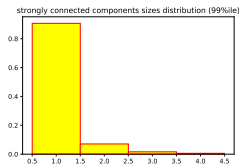
(a) WCano



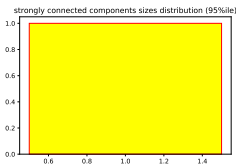
(b) Russian



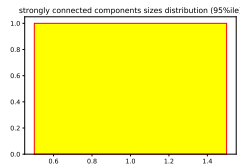
(c) Weibo(i)



(d) TDN(T)



(e) TDN(i10)



(f) TDN(i11)

# Overview

## 1 General stats

## 2 Degrees stats

- Degrees
- In-Degrees
- Out-Degrees

## 3 Connected components stats

- Weakly connected components
- Strongly connected components

## 4 Activity stats

- Tweets
- Retweets
- Retweeted

## 5 $\lambda, \mu, \nu$

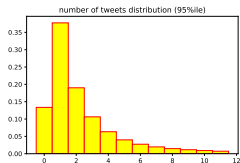
# Tweets stats

	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)
Number	12 666 338	674 292	X	232 978	X	3 088 408	1 386 888
% of events	45	35		01		77	71
Mean nb/u	3.15	3.70		0.17		20.55	14.64
Mid nb/u	2	1		0		0	0
Min nb	0	0		0		0	0
Max nb	534	4 834		3 718		3 200	3 200
25-percentile	1	0		0		0	0
75-percentile	3	2		0		1	1
90-percentile	7	6		0		3	3
95-percentile	11	12		0		8	7
99-percentile	26	55		2		557	358.8
Nb users with 1+	3 511 668	98 901		47 555		71 327	42 398
Nb users with 0	514 790	83 123		1 293 261		78 976	52 323

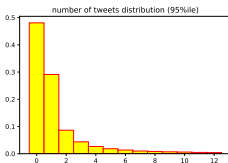
WCano(c) Russian(c) Weibo(c)

idem

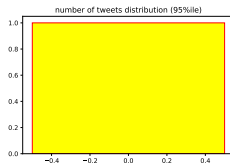
# Nb tweets/user distribution



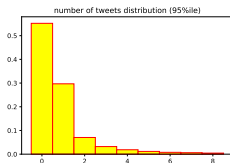
(a) WCanO



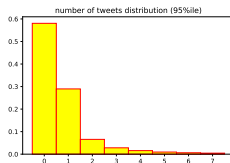
(b) Russian



(c) Weibo(i)



(d) TDN(i10)



(e) TDN(i11)



# Overview

## 1 General stats

## 2 Degrees stats

- Degrees
- In-Degrees
- Out-Degrees

## 3 Connected components stats

- Weakly connected components
- Strongly connected components

## 4 Activity stats

- Tweets
- Retweets
- Retweeted

## 5 $\lambda, \mu, \nu$

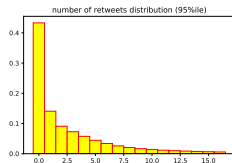
# Retweets stats

	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)
Number	15 246 762	1 271 073	X	33 307 189	X	910 720	565 944
% of events	55	65		99		23	29
Mean nb/u	3.79	6.99		24.84		6.06	5.97
Mid nb/u	1	1		17		0	0
Min nb	0	0		0		0	0
Max nb	220	2 811		1 032		3 201	3 156
25-percentile	0	0		6		0	0
75-percentile	4	2		36		0	0
90-percentile	10	10		57		1	1
95-percentile	16	25		73		2	2
99-percentile	36	133		114		46	77
Nb users with 1+	2 369 059	115 517		1 334 887		23 638	17 516
Nb users with 0	1 657 399	66 507		5 929		126 665	77 205

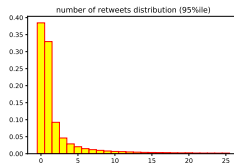
WCano(c) Russian(c) Weibo(c)

idem

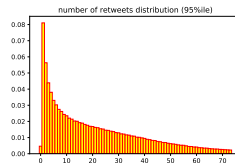
# Nb retweets/user distribution



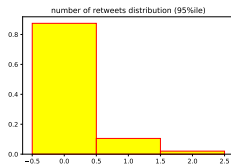
(a) WCanoe



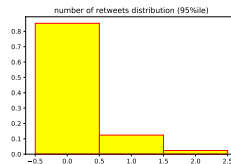
(b) Russian



(c) Weibo(i)



(d) TDN(i10)



(e) TDN(i11)

# Overview

## 1 General stats

## 2 Degrees stats

- Degrees
- In-Degrees
- Out-Degrees

## 3 Connected components stats

- Weakly connected components
- Strongly connected components

## 4 Activity stats

- Tweets
- Retweets
- Retweeted

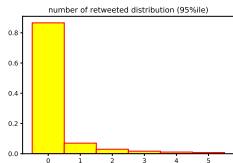
## 5 $\lambda, \mu, \nu$

# Retweeted stats

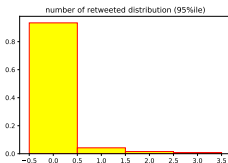
	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)
Number	11 339 937	1 271 073	X	33 307 189	X	910 720	565 944
% of events	40	65		99		23	29
Mean nb/u	2.82	6.99		24.84		6.06	5.97
Mid nb/u	0	0		0		1	1
Min nb	0	0		0		0	0
Max nb	31 303	35 539		1 107 734		274 153	225 458
25-percentile	0	0		0		0	0
75-percentile	0	0		0		2	2
90-percentile	2	1		0		6	5
95-percentile	5	3		0		12	11
99-percentile	41	48		55		42	38
Nb users with 1+	710 730	19 435		47 554		85 571	53 222
Nb users with 0	3 315 728	162 589		1 293 262		64 732	41 499

WCano(c)	Russian(c)	Weibo(c)
14 563 462	1 227 277	33 307 189
52	63	99
3.62	6.76	24.84
1	1	17
0	0	0
250	2 205	3 718
0	0	6
4	2	36
10	10	57
16	25	73
34	128	114
2 341 726	114 879	1 339 190
1 684 732	66 742	1 626

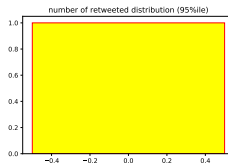
# Nb retweeted/user distribution



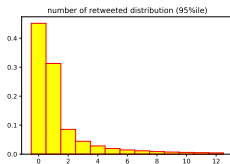
(a) WCano



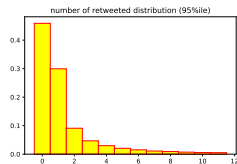
(b) Russian



(c) Weibo(i)



(d) TDN(i10)



(e) TDN(i11)

# Overview

- 1 General stats
- 2 Degrees stats
  - Degrees
  - In-Degrees
  - Out-Degrees
- 3 Connected components stats
  - Weakly connected components
  - Strongly connected components
- 4 Activity stats
  - Tweets
  - Retweets
  - Retweeted
- 5  $\lambda, \mu, \nu$

	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)
Mean	4.95 e-07	7.57 e-07	X	1.65 e-09	X	8.13 e-07	1.65 e-06
Mid	3.14 e-07	2.04 e-07		0		0	0
Min	0	0		0		0	0
Max	8.39 e-05	9.88 e-04		3.54 e-05		1.27 e-04	3.60 e-04
25-percentile	1.57 e-07	0		0		0	0
75-percentile	4.72 e-07	4.09 e-07		0		3.96 e-08	1.13 e-07
90-percentile	1.10 e-06	1.23 e-06		0		1.19 e-07	3.38 e-07
95-percentile	1.73 e-06	2.45 e-06		0		3.17 e-07	7.89 e-07
99-percentile	4.09 e-06	1.12 e-05		1.90 e-08		2.20 e-05	4.04 e-05
Nb users with >0	3 511 668	98 901		47 555		71 327	42 398
Nb users with 0	514 790	83 123		1 293 261		78 976	52 323

WCano(c) Russian(c) Weibo(c)

idem



	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)
Mean	5.95 e-07	1.43 e-06	X	2.36 e-07	X	2.40 e-07	6.73 e-07
Mid	1.57 e-07	2.04 e-07		1.62 e-07		0	0
Min	0	0		0		0	0
Max	3.46 e-05	5.74 e-04		9.82 e-06		1.27 e-04	3.56 e-04
25-percentile	0	0		5.71 e-08		0	0
75-percentile	6.29 e-07	4.09 e-07		3.43 e-07		0	0
90-percentile	1.57 e-06	2.04 e-06		5.43 e-07		3.96 e-08	1.13 e-07
95-percentile	2.52 e-06	5.11 e-06		6.95 e-07		7.92 e-08	2.25 e-07
99-percentile	5.66 e-06	2.72 e-05		1.09 e-06		1.82 e-06	8.66 e-06
Nb users with >0	2 369 059	115 517		1 334 887		23 638	17 516
Nb users with 0	1 657 399	66 507		5 929		126 665	77 205

WCano(c) Russian(c) Weibo(c)

idem

	WCano	Russian	Weibo(T)	Weibo(i)	TDN(T)	TDN(i10)	TDN(i11)
Mean	4.43 e-07	1.43 e-06	X	2.36 e-07	X	2.40 e-07	6.73 e-07
Mid	0	0		0		3.96 e-08	1.13 e-07
Min	0	0		0		0	0
Max	4.92 e-03	7.26 e-03		1.05 e-02		1.09 e-02	2.54 e-02
25-percentile	0	0		0		0	0
75-percentile	0	0		0		7.92 e-08	2.25 e-07
90-percentile	3.14 e-07	2.04 e-07		0		2.38 e-07	5.64 e-07
95-percentile	7.86 e-06	6.13 e-07		0		4.75 e-07	1.24 e-06
99-percentile	6.44 e-06	9.81 e-06		5.24 e-07		1.66 e-06	4.28 e-06
Nb users with >0	710 730	19 435		47 554		85 571	53 222
Nb users with 0	3 315 728	162 589		1 293 262		64 732	41 499

WCano(c)	Russian(c)	Weibo(c)
5.69 e-07	1.38 e-06	2.37 e-07
1.57 e-07	2.04 e-07	1.62 e-07
0	0	0
3.93 e-05	4.50 e-04	3.54 e-05
0	0	5.71 e-08
6.29 e-07	4.09 e-07	3.43 e-07
1.57 e-06	2.04 e-06	5.43 e-07
2.52 e-06	5.11 e-06	6.85 e-07
5.34 e-06	2.62 e-05	1.09 e-06
2 341 726	114 879	1 339 190
1 684 732	66 742	1 626