



isi International
Statistical
Institute







WORLD STATISTICS
CONGRESS
2025
THE HAGUE





Network Science for Official Statistics
Anonymity and Disclosure Control for Network Data

Rachel de Jong, Mark van der Loo, Frank Take
Leiden university, CBS
October 5th 2025

Towards Anonymity in Network Data

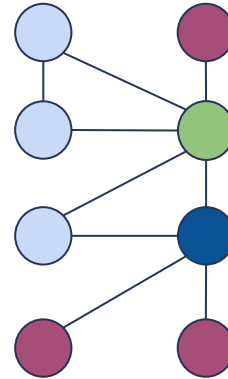
Microdata

ID	Attribute
1	
2	
3	
4	





ID	Attribute
5	
6	
7	
8	



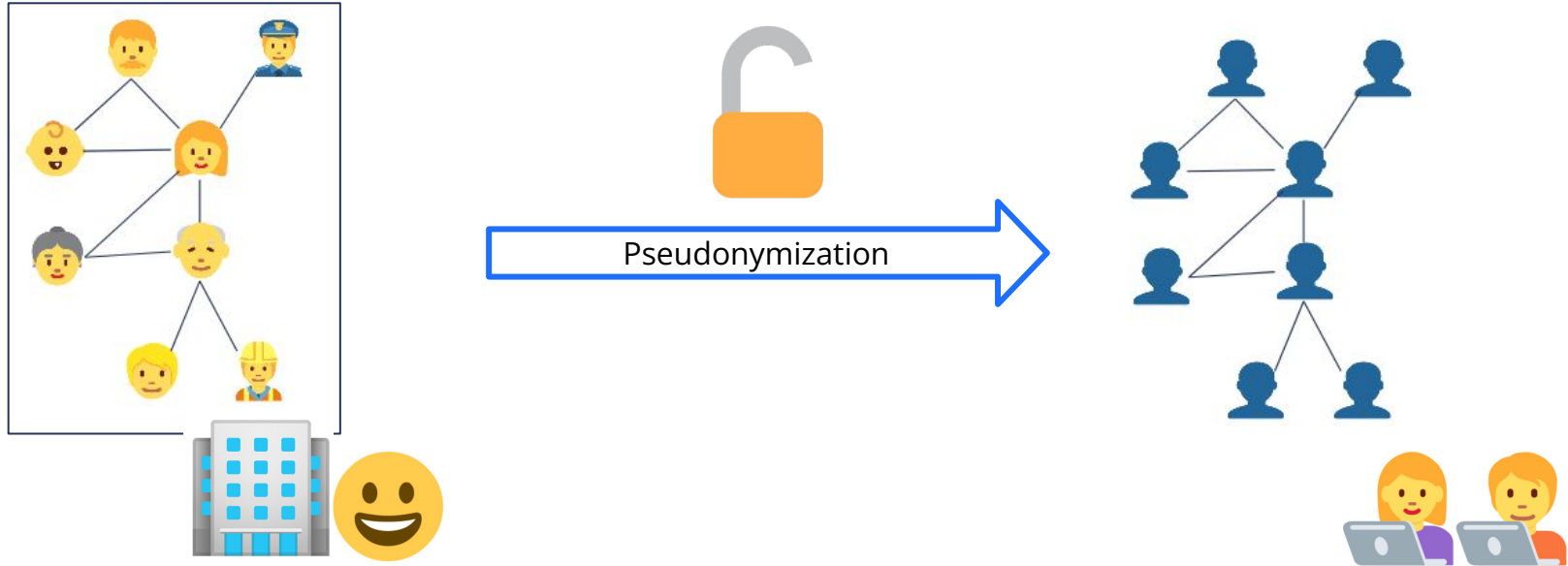
Network data



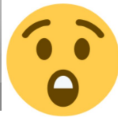
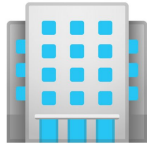
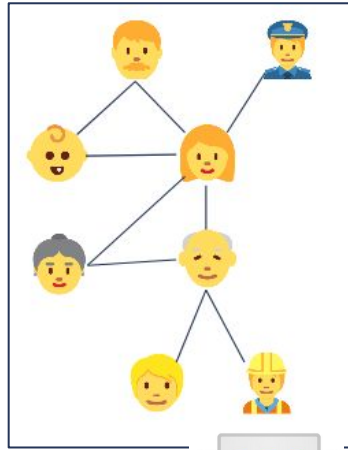
Network consists of

- Nodes 
 -  Entities: people, companies, cities, ...
- Edges 
 -  **Connections:** social ties, trade, roads, ...

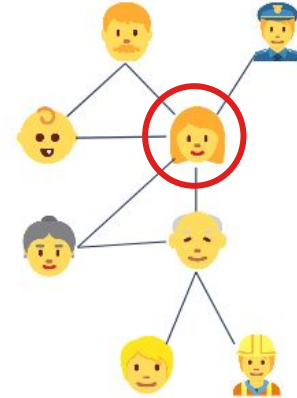
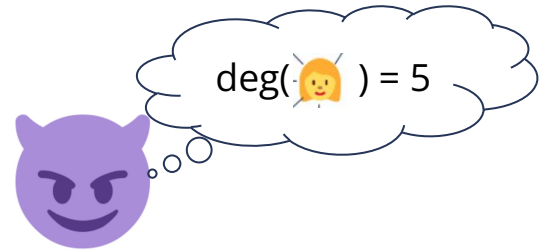
Safely publishing networks?



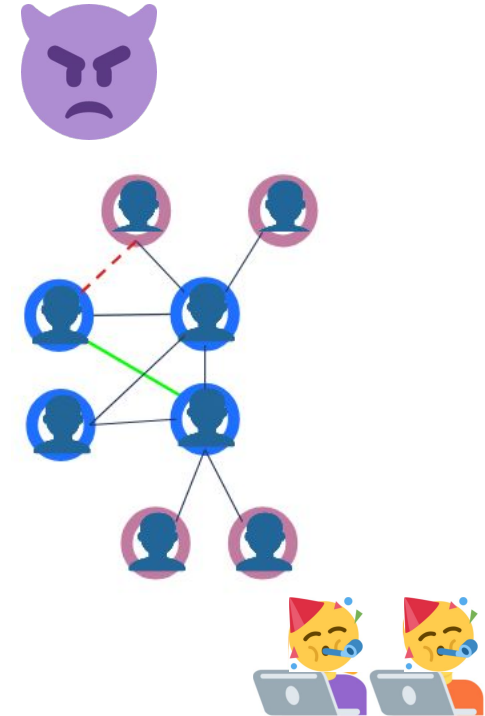
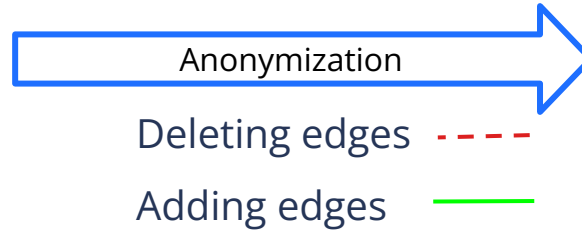
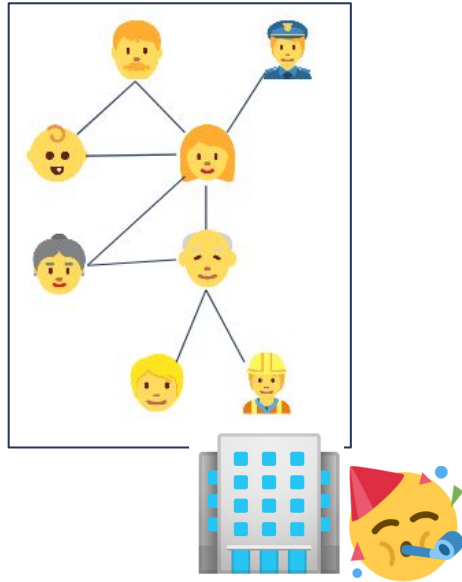
Safely publishing networks?



Pseudonymization



Safely publishing networks?



Q1: How to **measure** anonymity?

- k -Anonymity
- Measure = attacker scenario
- Uncertain information?

Q2: How to **anonymize** a network?

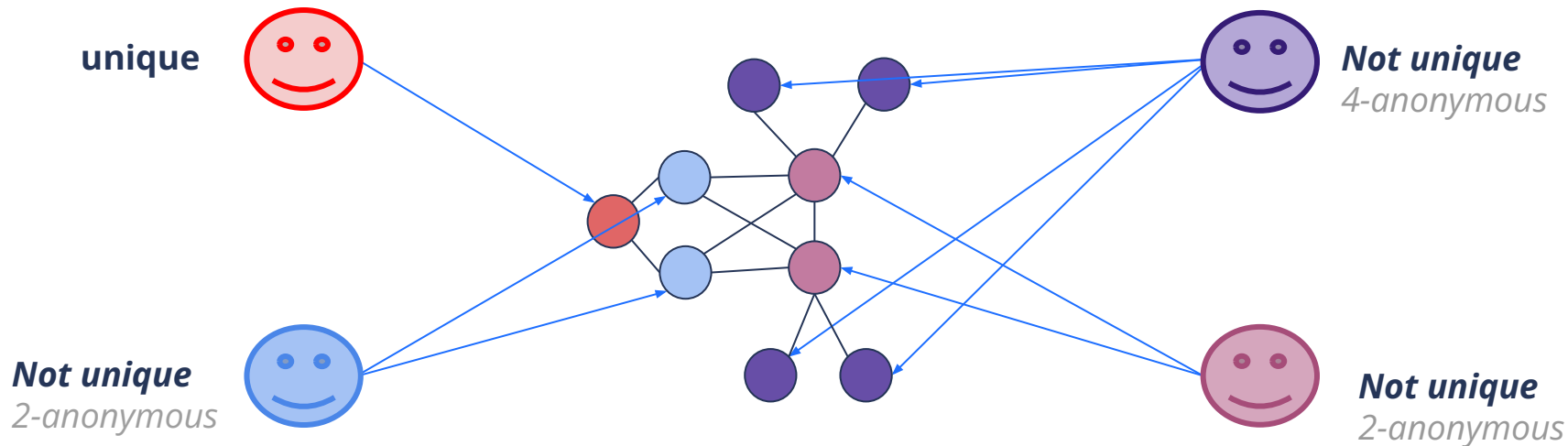


k-Anonymity

How many candidates are there for each individual?

anonymity = uniqueness: fraction of unique nodes

For now: **Not unique** → **Anonymous**



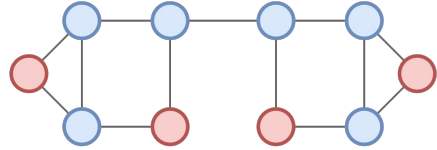
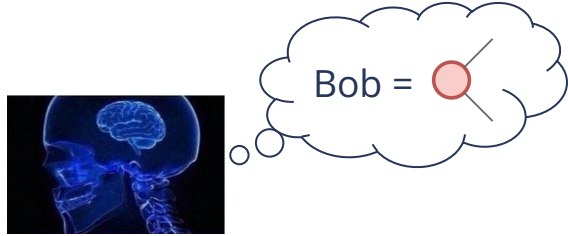
K-anonymity
Not unique



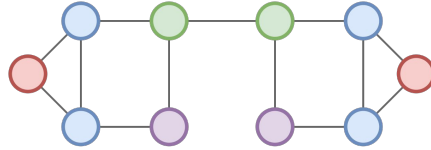
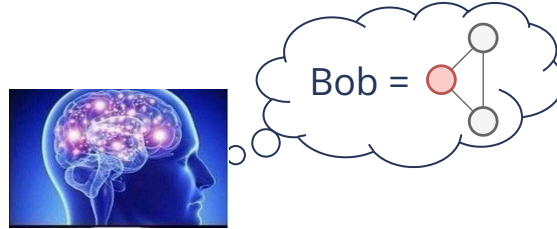
Anonymous*

*For now

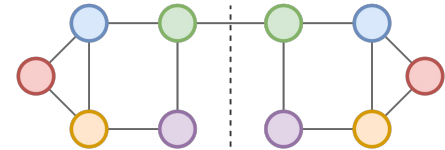
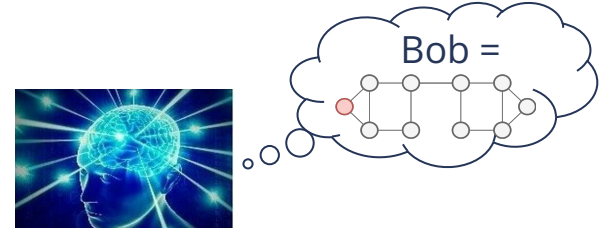
Measure = attacker scenario



Degree



1-Neighborhood

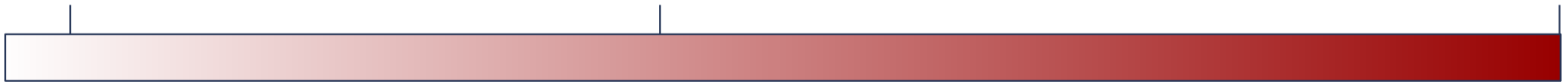


Automorphism

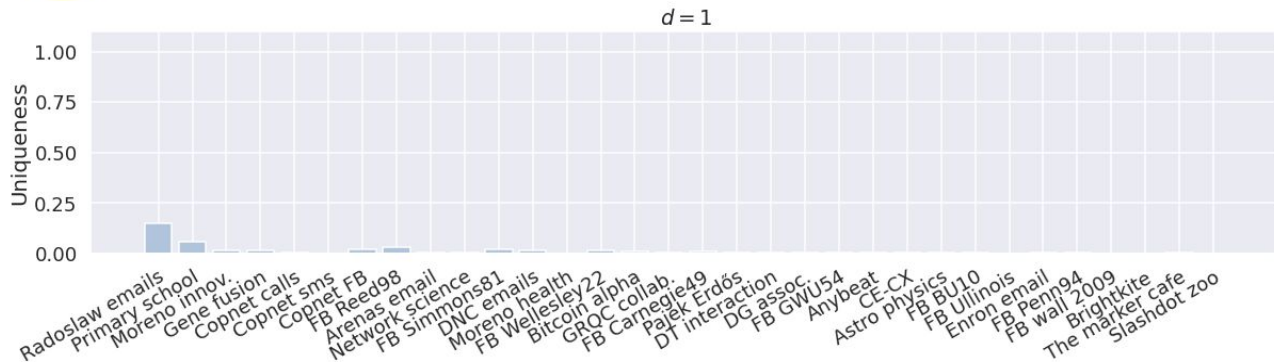
Degree

$d=1$

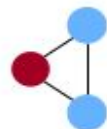
Automorphism



Measure = attacker scenario

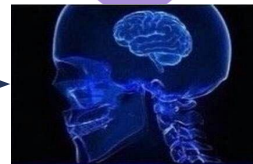


$|V|: \sim 100 - 3M$ $|E|: \sim 100 - 18M$

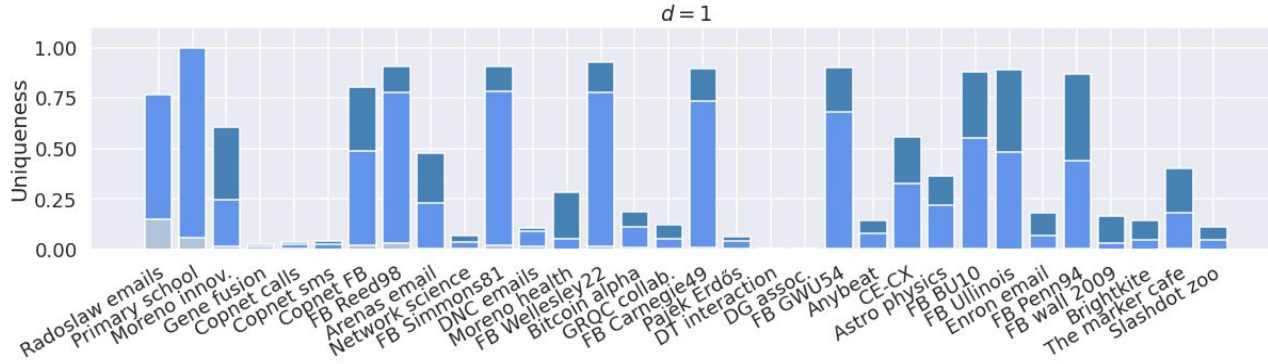


Degree

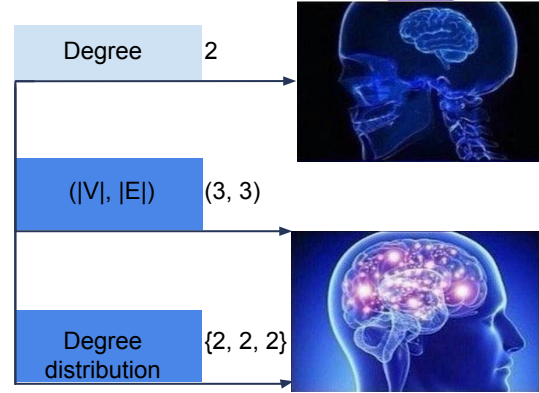
2



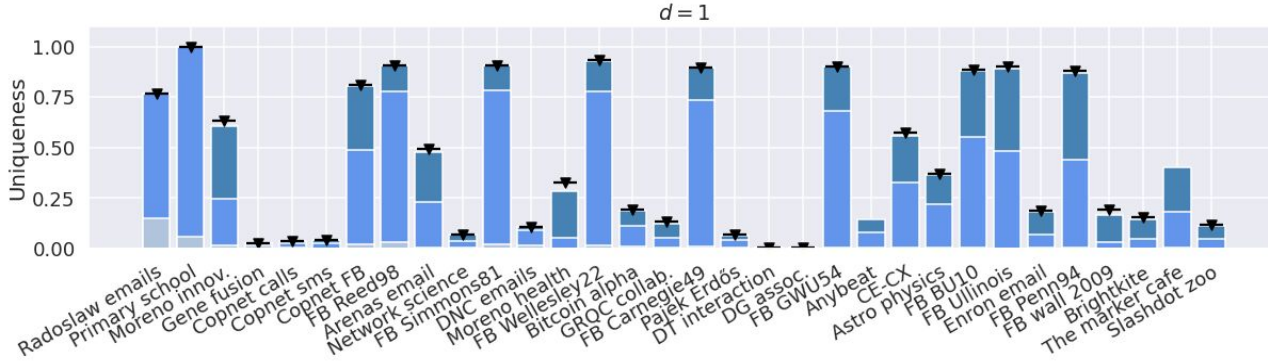
Measure = attacker scenario



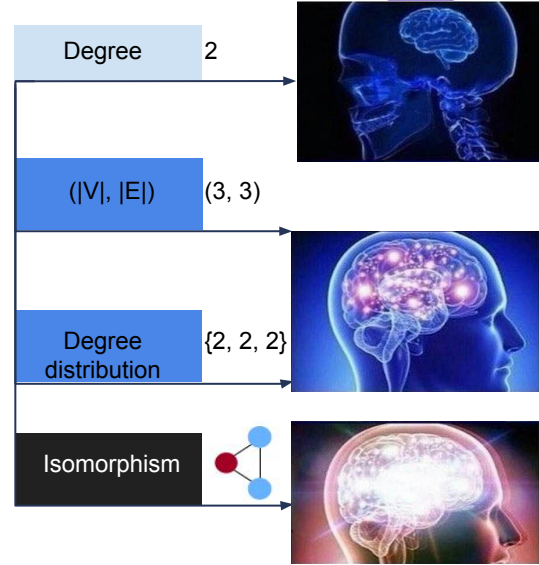
$|V|$: ~100 - 3M $|E|$: ~100 - 18M



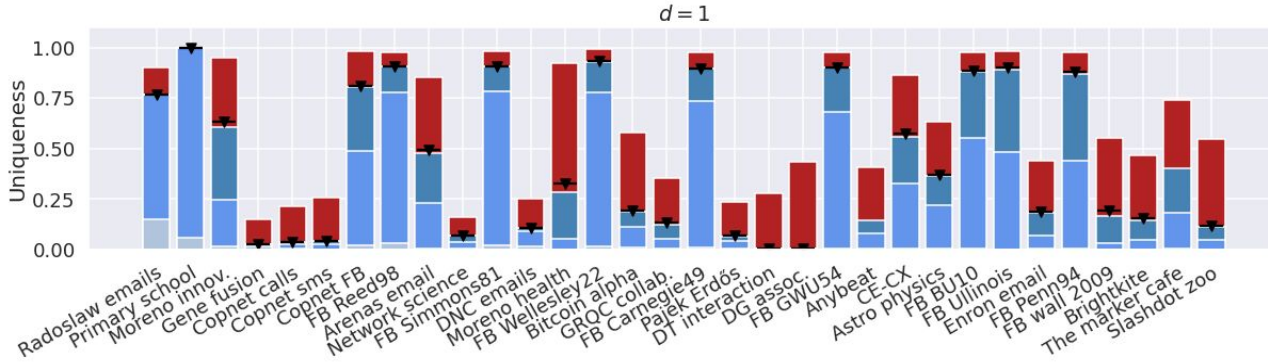
Measure = attacker scenario



$|V|$: ~100 - 3M $|E|$: ~100 - 18M

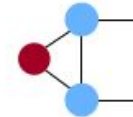


Measure = attacker scenario

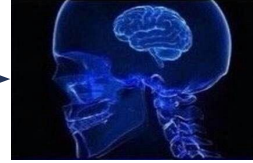


$|V|$: ~100 - 3M $|E|$: ~100 - 18M

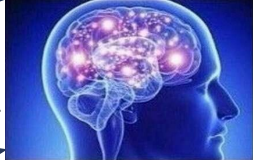
Imprecise information but looking further
is more effective!



Degree 2



$(|V|, |E|)$ (3, 3)



Degree distribution {2, 2, 2}

Isomorphism

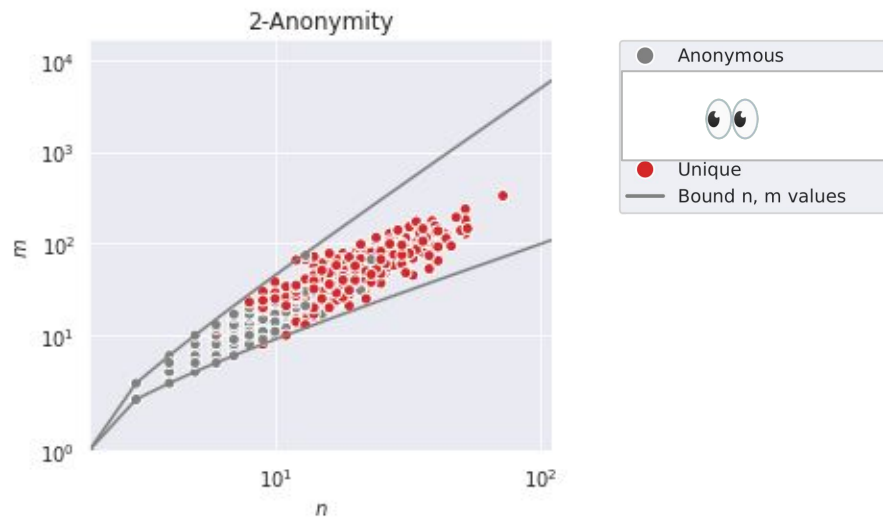


Degrees of neighbors {2, 3, 3}



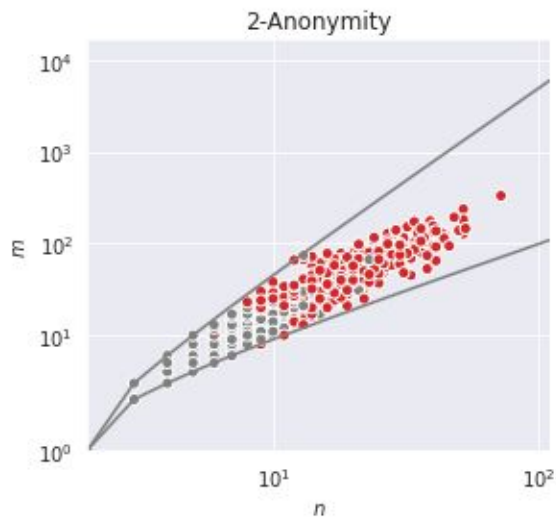
**What if
the attacker is
not so sure?**

Fuzzy k-anonymity for networks



Arenas email network

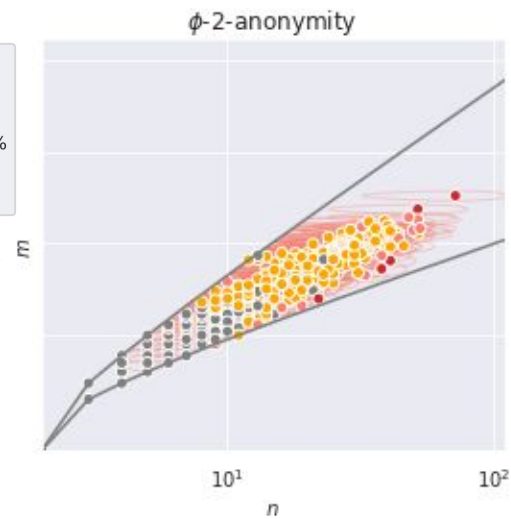
Fuzzy k-anonymity for networks



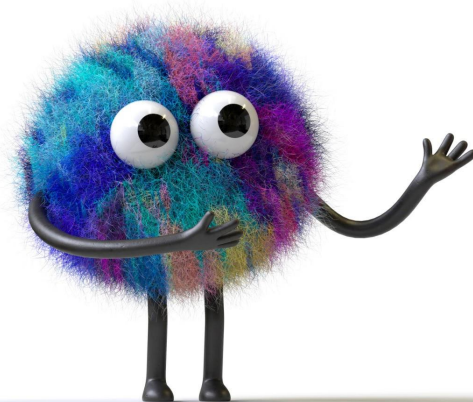
ϕ -similar if:

$$\Delta_n + \Delta_m \leq \phi(n + m)$$

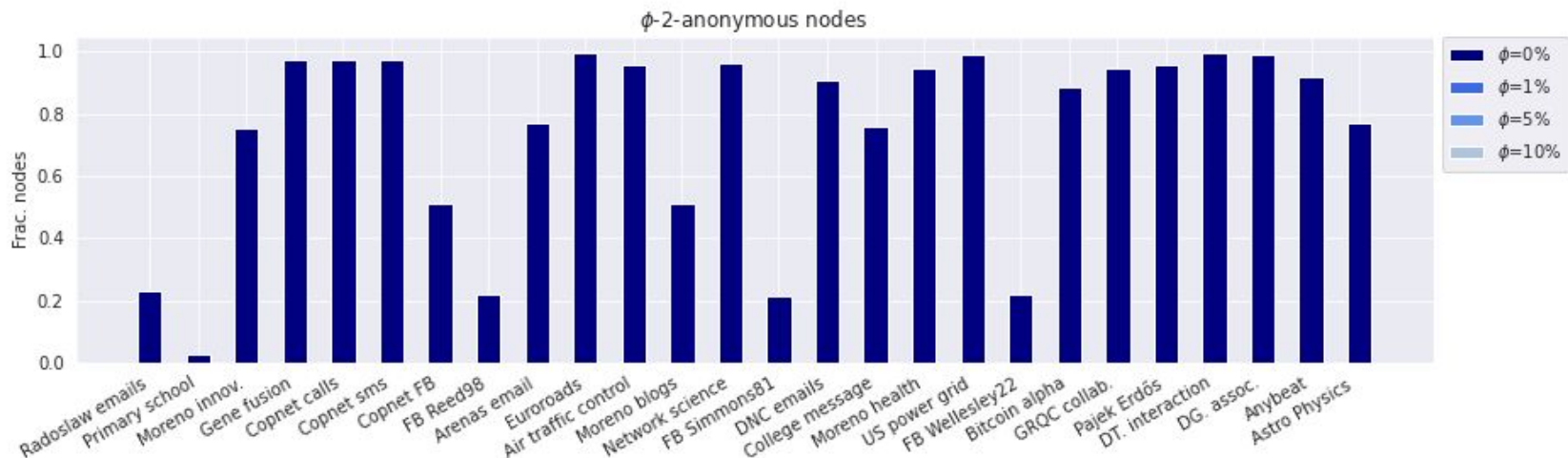
- Anonymous
- Anonymous at $\phi=1\%$
- Anonymous at $\phi=5\%$
- Anonymous at $\phi=10\%$
- Unique
- Bound n, m values



Arenas email network

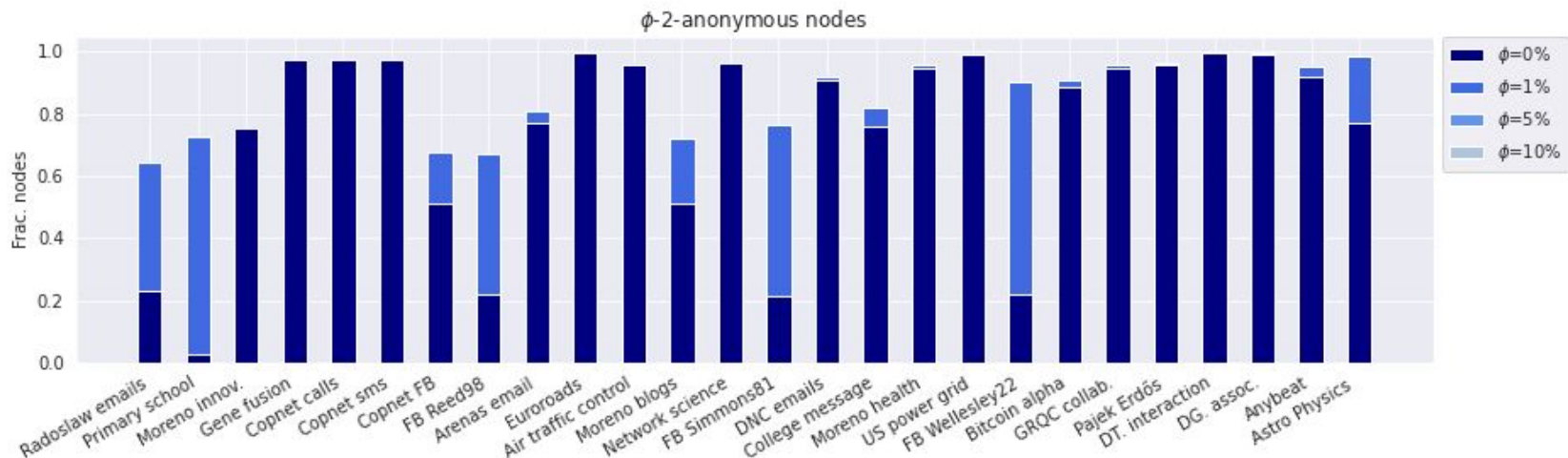


Fuzzy k-anonymity in networks



No uncertainty: **>2%** anonymous for all networks

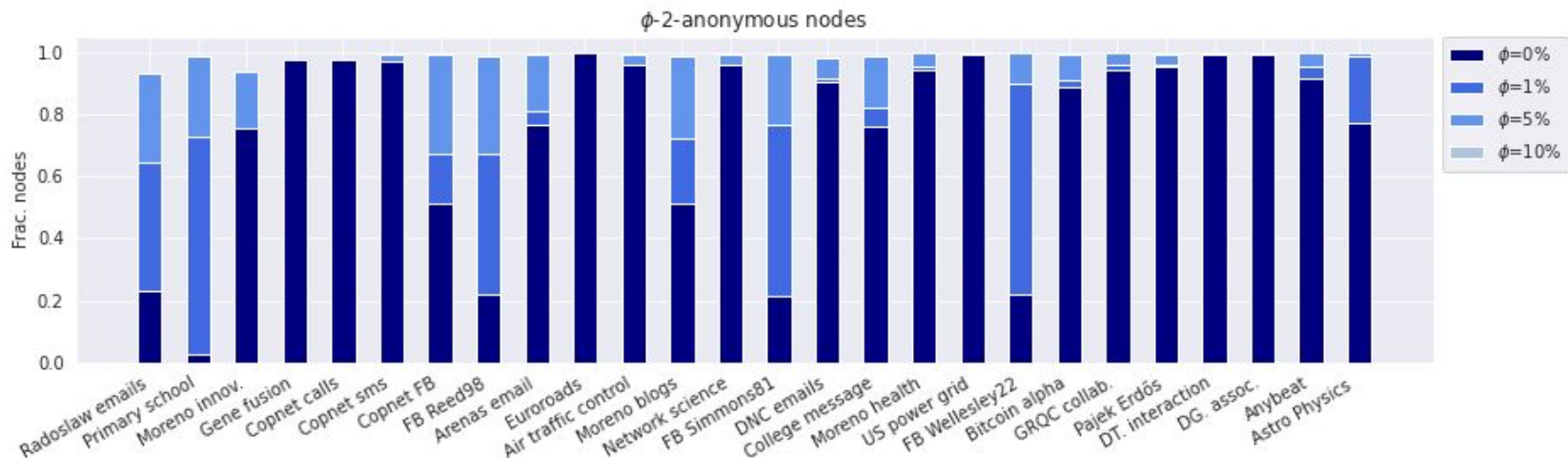
Fuzzy k-anonymity in networks



No uncertainty: **>2%** anonymous for all networks

1% uncertainty: **>60%** anonymous

Fuzzy k-anonymity in networks

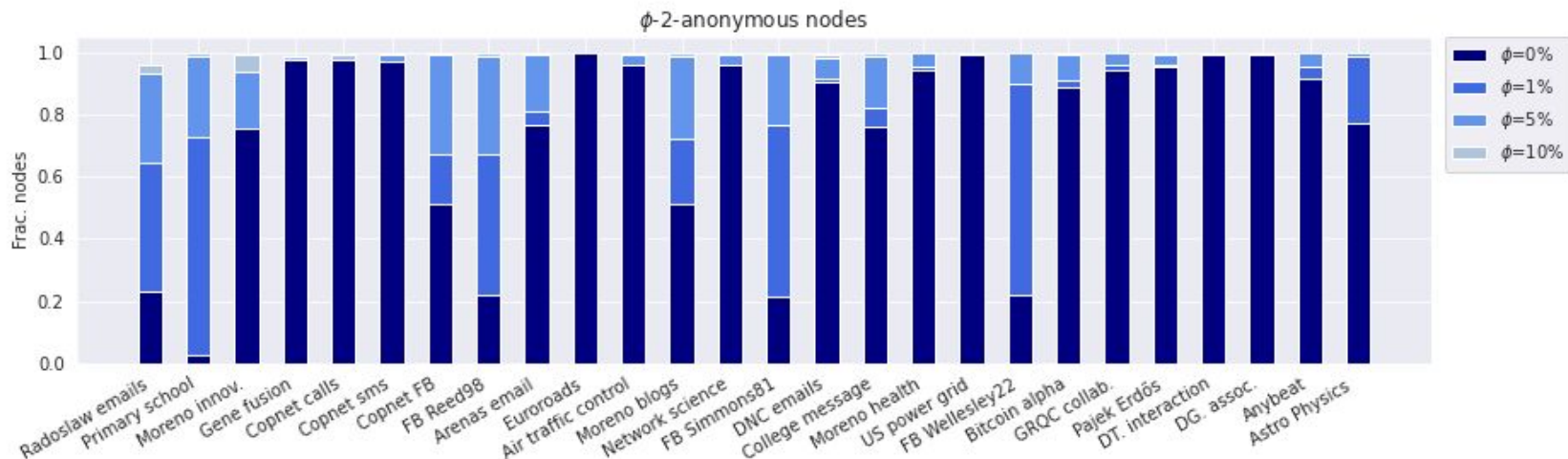


No uncertainty: **>2%** anonymous for all networks

1% uncertainty: **>60%** anonymous

5% uncertainty: **>90%** anonymous

Fuzzy k-anonymity in networks



No uncertainty: **>2%** anonymous for all networks

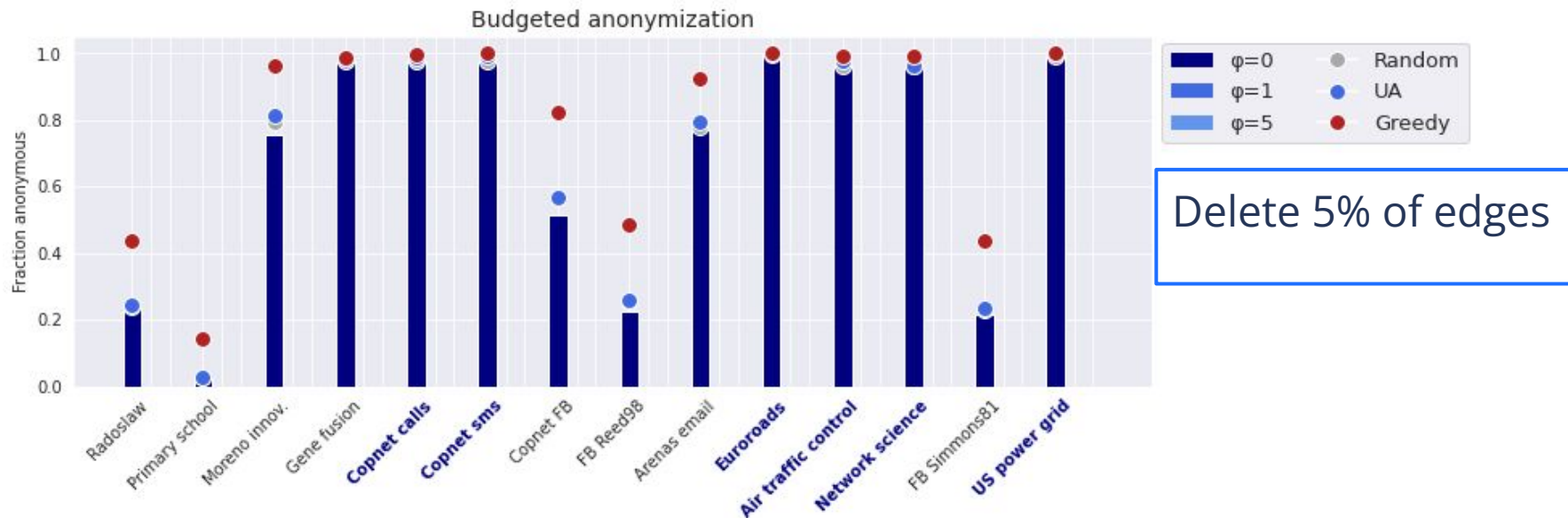
1% uncertainty: **>60%** anonymous

5% uncertainty: **>90%** anonymous

10% uncertainty: **>95%** anonymous

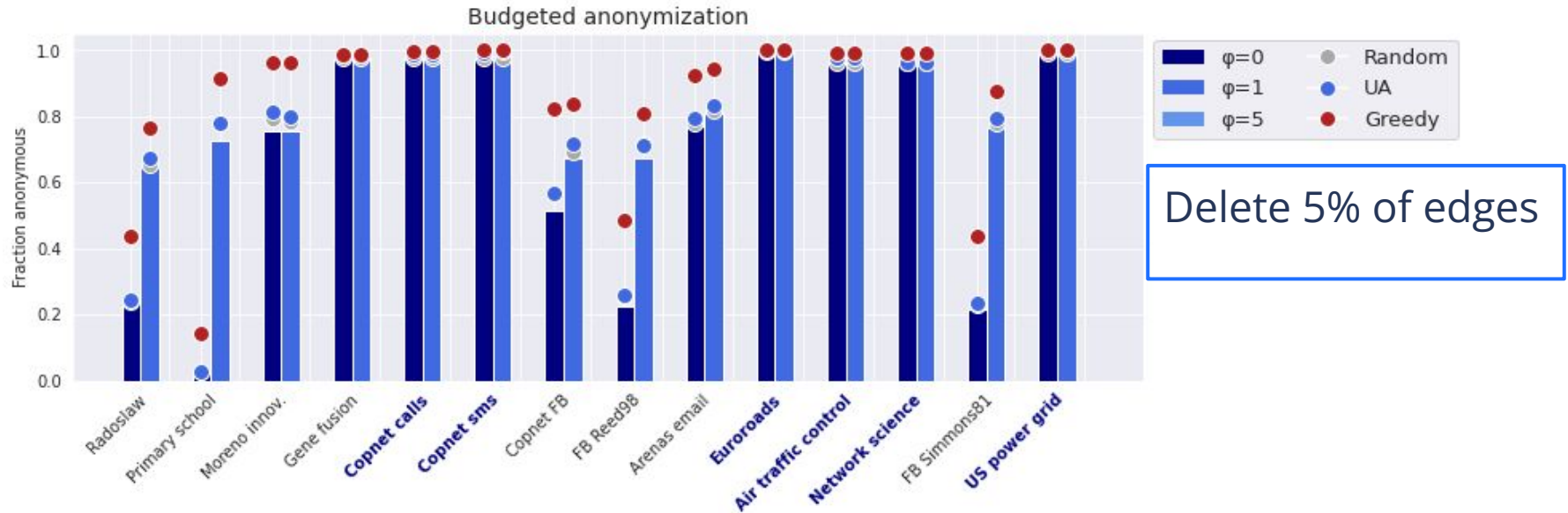
Anonymization

Anonymization



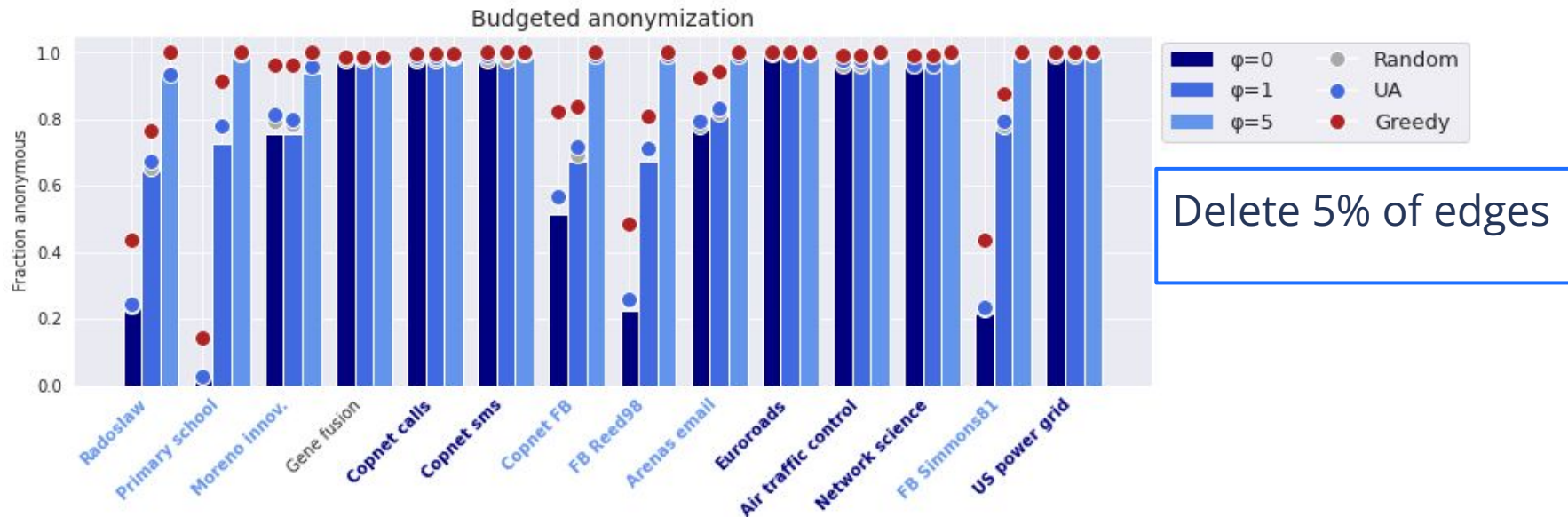
0%: anonymization limited effect

Anonymization



0%: anonymization limited effect

Anonymization



0%: anonymization limited effect
5%: most networks anonymous

Conclusion

- Measure anonymity with k -anonymity
- Attacker knowledge = anonymity measure
- Imprecise attacker knowledge
 - large increase in anonymity
- Anonymization + 5% uncertainty
 - Many networks fully anonymous

Future work

- Discussion: *which level of uncertainty is likely?*
 - Guidelines ϕ values
- Fuzzyness for other measures
- Include more network properties
 - Timestamps, node / edge labels

Rachel de Jong

Email

r.g.de.jong@liacs.leidenuniv.nl

 [linkedin.com/in/rachelgdejong/](https://www.linkedin.com/in/rachelgdejong/)

ANONET package

github.com/RacheldeJong/ANONET



- [1] **The effect of distant connections on node anonymity in complex networks.** Scientific Reports 14(1), 1156 (2024)
 - [2] **Algorithms for efficiently computing structural anonymity in complex networks.** ACM Journal of Experimental Algorithmics 28 (2023)
 - [3] **A systematic comparison of measures for k -anonymity in networks** (2024) arXiv:2407.02290
 - [4] **The anonymization problem in social networks,** preprint arXiv:2409.16163 (2024).
- Authors: R.G. de Jong, M. P. J. van der Loo, F. W. Takes

Related research

How to **measure** anonymity?*

- **Algorithms for efficiently computing structural anonymity in complex networks.** *ACM Journal of Experimental Algorithmics*, 28, 1-22. (2023).
- **The effect of distant connections on node anonymity in complex networks.** *Scientific Reports*, 14(1), 1156. (2024)
- **A systematic comparison of measures for k-anonymity in networks.** *arXiv preprint arXiv:2407.02290*. (2024)
- Paper on fuzzy k-anonymity in progress...

How to **anonymize** a network?

The anonymization problem:

- de Jong, R. G., van der Loo, M. P., & Takes, F. W. **The anonymization problem in social networks.** *arXiv preprint arXiv:2409.16163*. (2024)

Genetic algorithm for anonymization:

- Bonello, S., de Jong, R. G., Bäck, T. H., & Takes, F. W. **Utility-aware Social Network Anonymization using Genetic Algorithms.** In *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (pp. 775-778). (2025)

Simulated Annealing for network anonymization:

- E.D. Arsene, R.G. de Jong, F.W. Takes, and A.L.D. Latour, **A Simulated Annealing Approach to Social Network Anonymization** (under review)

*Authors: Rachel G. de Jong, Mark P.J. van der Loo, Frank W. Takes

Appendix

Related work: Hackathon*

Which information can be obtained?

May 3rd 2022 In 4 hours:

- 22 students found > 5,000 links about 26 volunteers
- Different sources: Facebook, linkedin, instagram (images)
- Different categories: Family, work, school
- 11 large pizzas

Co-organisers: Mark van der Loo, Jayshri Murli, Marieke de Vries, Peter-Paul de Wolf and, Frank Takes

*de Vries, M. M., et al. "The risk of identity disclosure through network structure: anecdotal evidence from a hackathon." arXiv preprint arXiv:2211.15325 (2022).

HOW ANONYMOUS ARE YOU?!

How anonymous are you in openly available data? Is it possible for hackers to extract sensitive information from these data sets, by combining them with other data sources?

In this hackathon you, as a 'potential attacker', will search for network relations of a given list of people. If you are able to find the most relations within 4 hours, you receive a prize!

MAY 3, 15:00 - 19:00



SCAN TO REGISTER

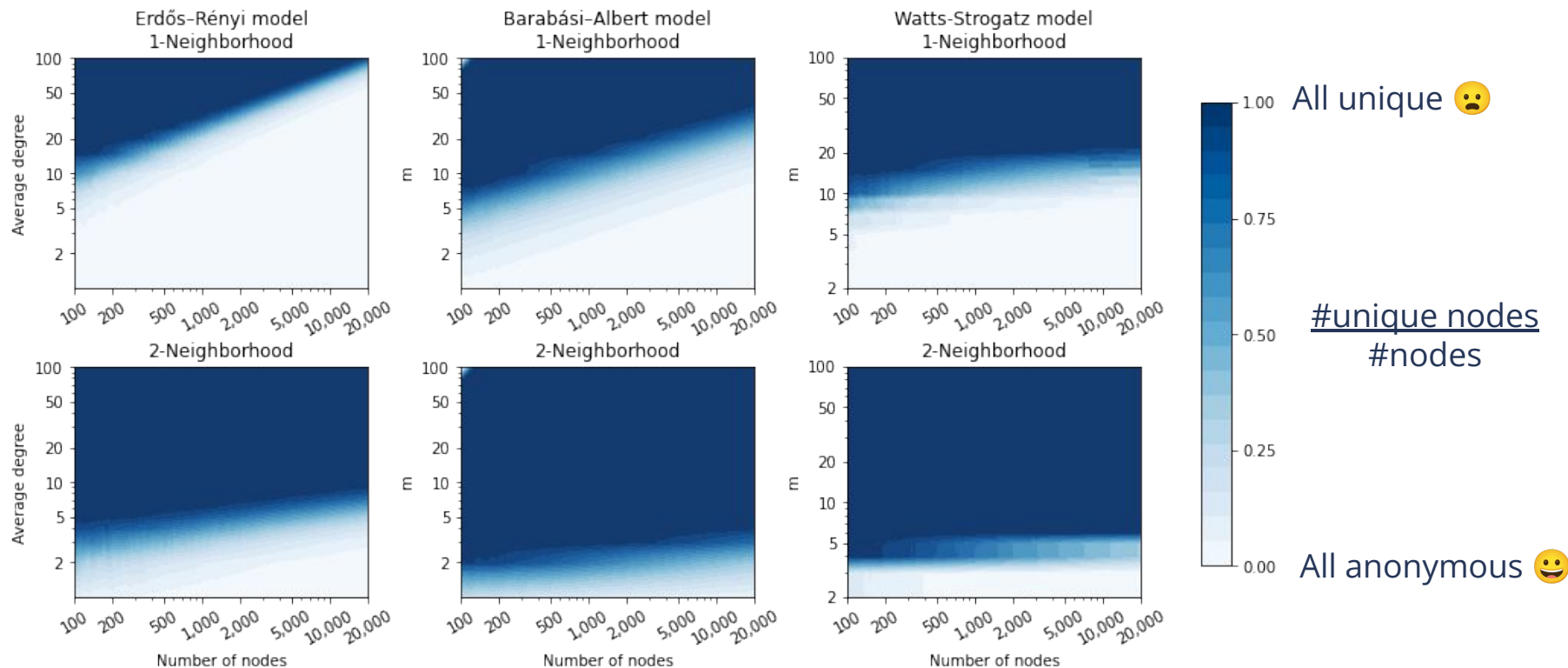
Free pizza and beverages afterwards!

Sign up before 27 April.



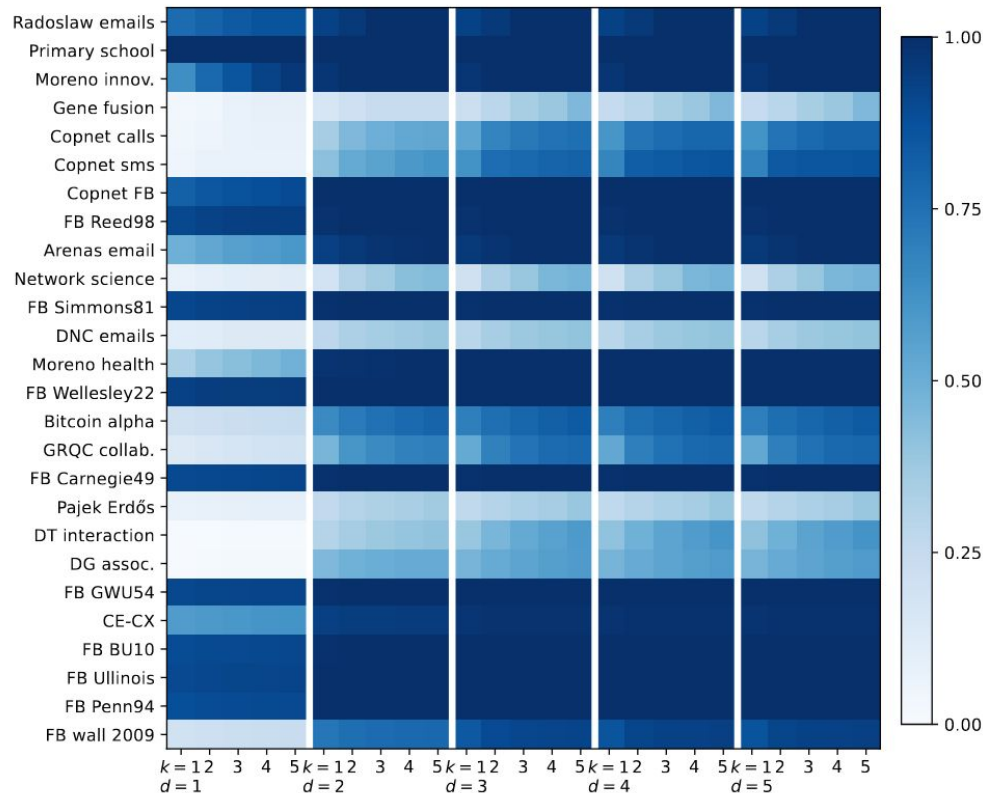
HACKATHON:
FROM PERSON TO
OPEN DATA

d - k -Anonymity on graph models



Almost no anonymity for $d=2$. Size has small / no effect

d - k -Anonymity: $d > 2, k > 1$



$d=1 \rightarrow d=5$

□ $d=1$ is not enough

□ After $d=2$ often not much change

$k=1 \rightarrow k=5$

□ Average: +0.05 up to 0.08

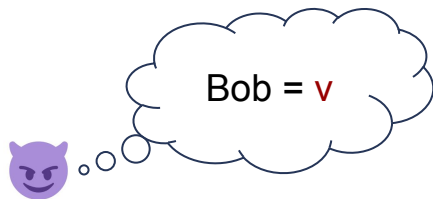
□ For some networks larger effect

Focus on:

□ 2-neighborhood

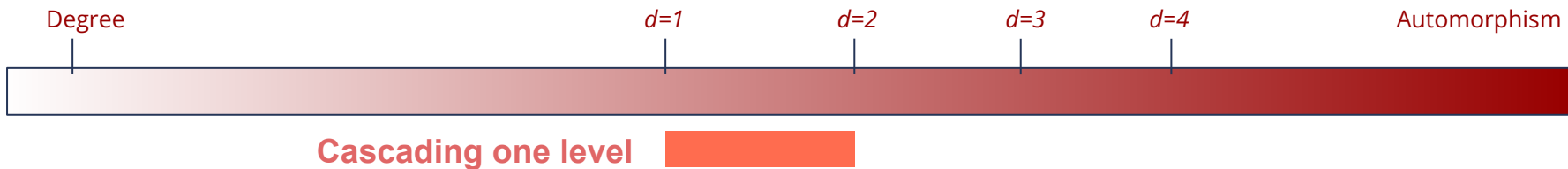
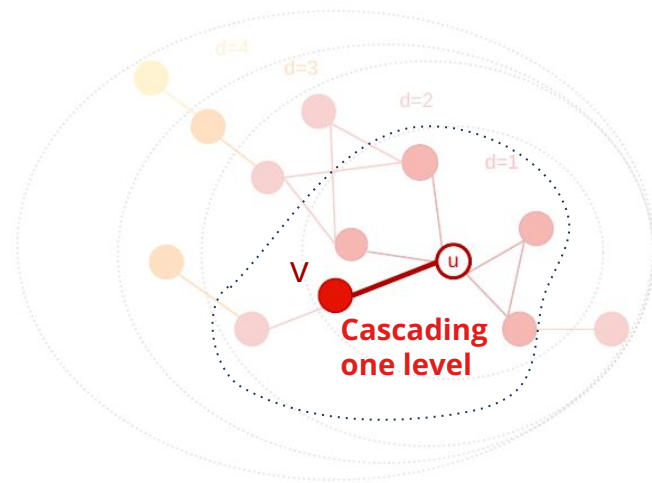
□ Uniqueness ($k=1$)

Approach 2: Anonymity-cascade

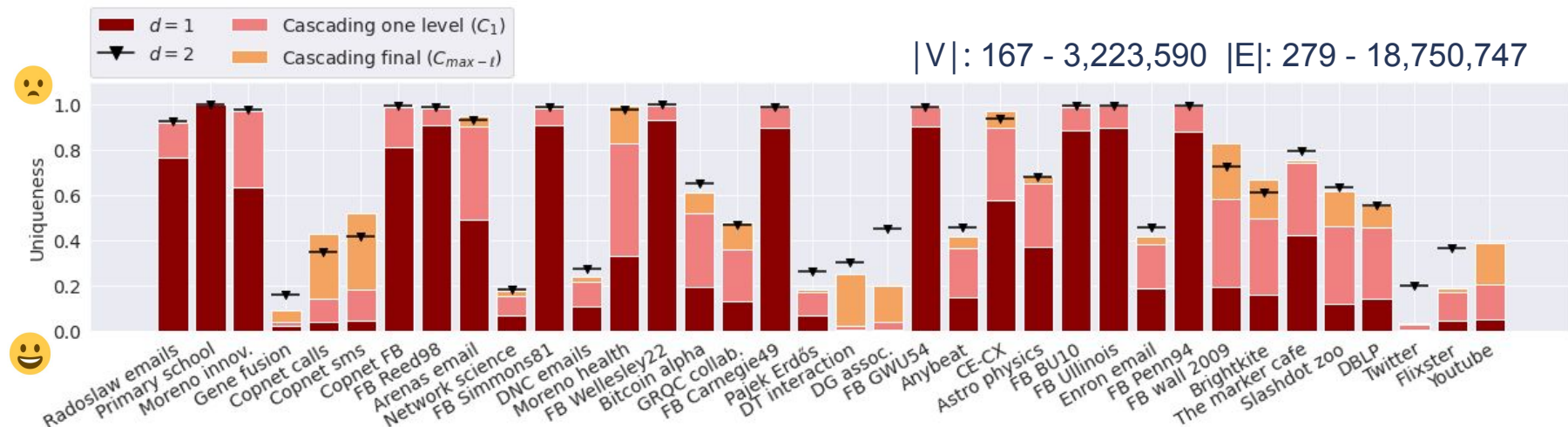


Knowledge:

- 1-neighborhood
- **Connected to unique node**
 - *"Infecting nodes with uniqueness"*



Anonymity-Cascade on real networks



*More networks due to lower runtimes $d=2$ and cascading

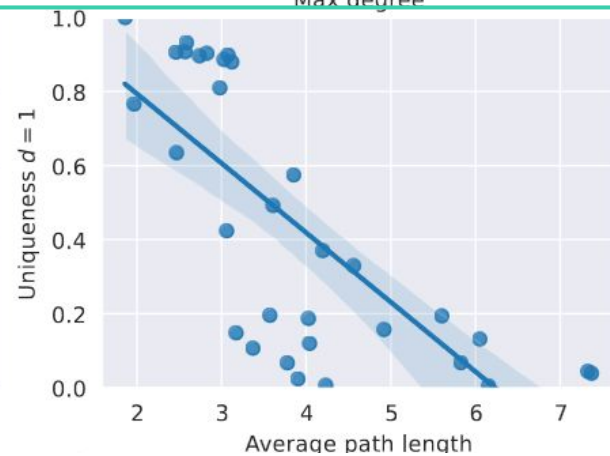
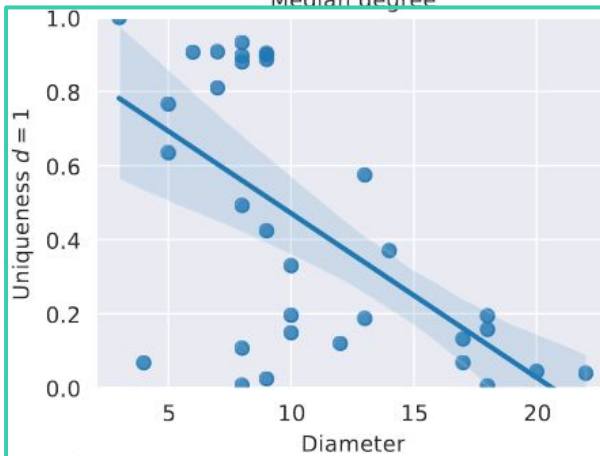
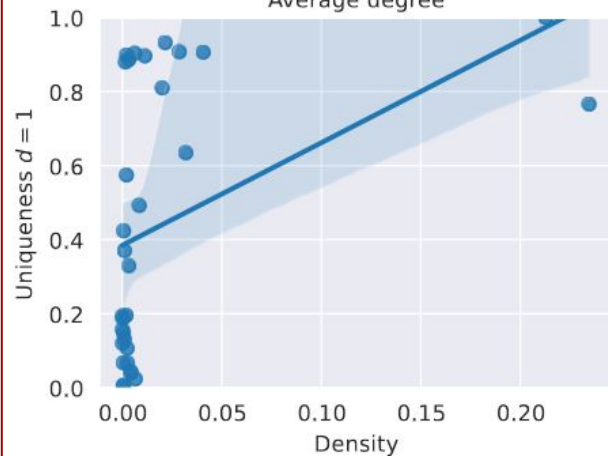
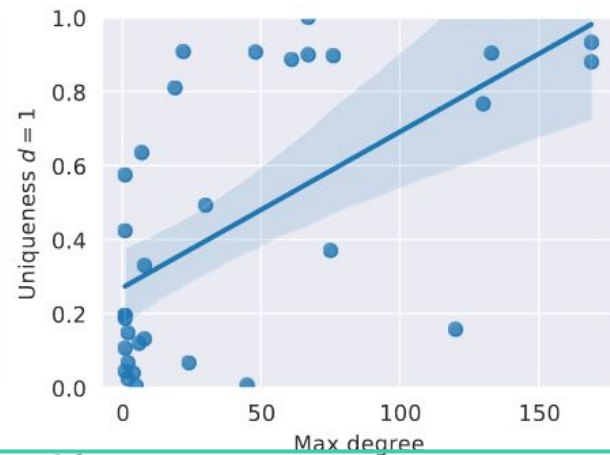
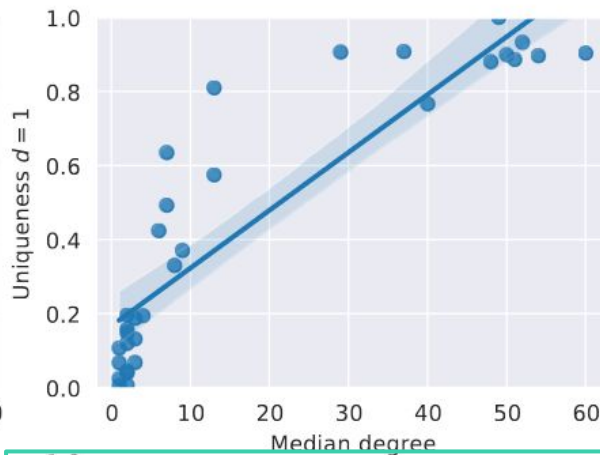
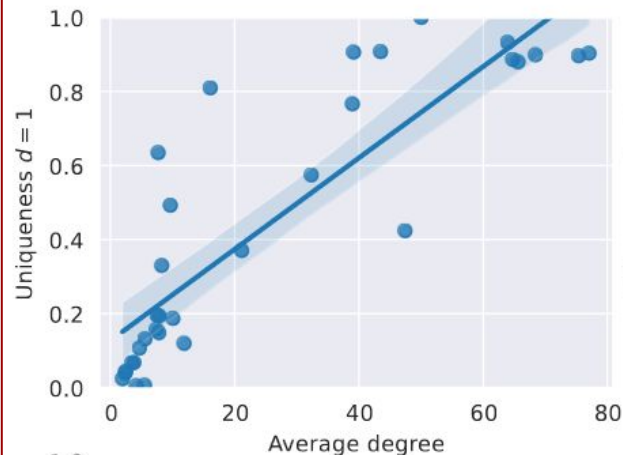
>100% increase: **19** / **5** / 36

Large increase in uniqueness with knowledge of one connection

Approximate d - k -anonymity with $d=2$

Increase for some networks with cascading final

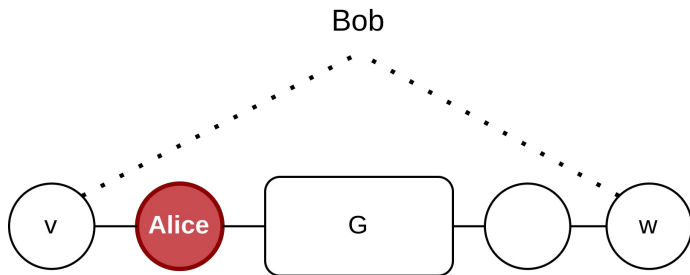
More edges \rightarrow less anonymity



More "spread" network \rightarrow more anonymity

Limitation: Twin Nodes

So far we have assumed that k candidates are distant



Bob - Alice ?

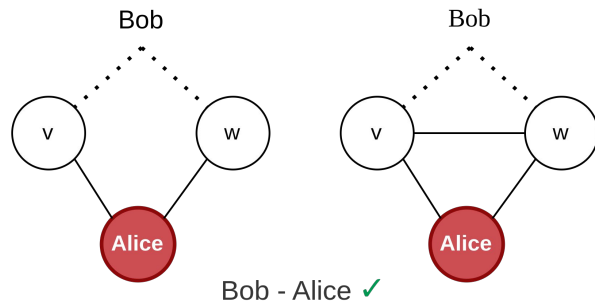
v and w structurally indistinguishable
(same orbit)



Edge anonymity



What if they are *both* connected to Alice?



Bob - Alice ✓

Open twins:

$$N_1(v) - \{v\} = N_1(w) - \{w\}$$

Closed twins:

$$N_1(v) = N_1(w)$$

→ v , w structurally indistinguishable



Edge anonymity

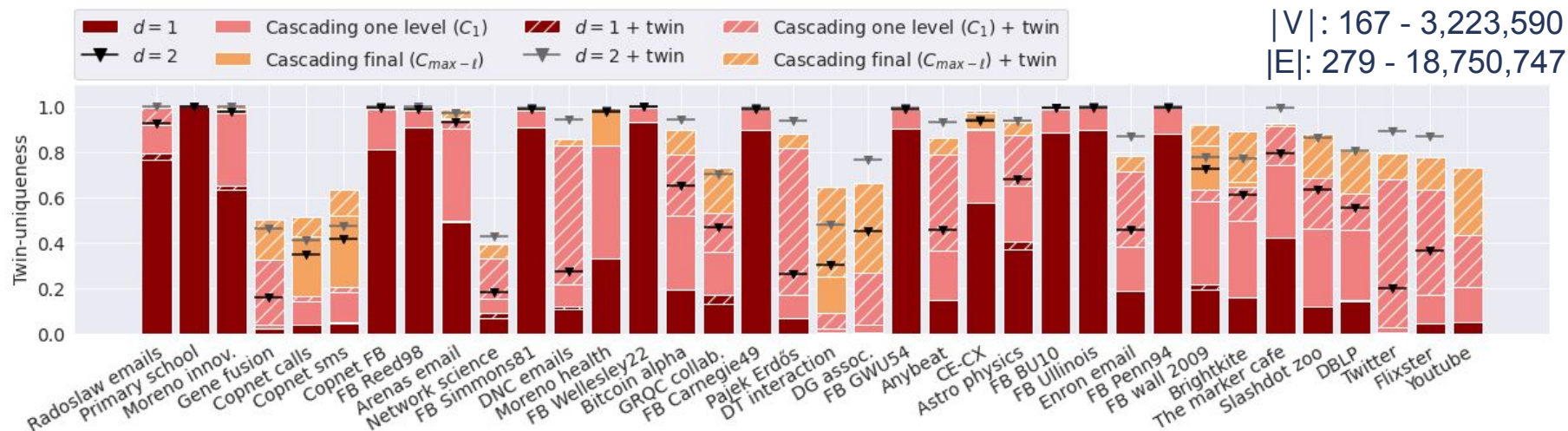


	V	E	Fraction twins	ℓ_{max}	d=1 (s)	d=2 (s)	C_1 (s)	C_f (s)	
Radoslaw emails ¹⁹	167	6,500	0.036	3	0.02 s	< 0.01 s	< 0.01 s	< 0.01 s	Communication
Primary school ²⁰	236	11,798	0.000	1	0.03 s	< 0.01 s	< 0.01 s	< 0.01 s	Human contact
Moreno innov. ¹⁹	241	1,846	0.012	3	0.00 s	0.01 s	< 0.01 s	< 0.01 s	Communication
Gene fusion ¹⁹	291	558	0.581	6	0.00 s	< 0.01 s	< 0.01 s	< 0.01 s	Bio
Copnet calls ¹	499	4,180	0.024	4	0.00 s	0.02 s	< 0.01 s	< 0.01 s	Communication
Copnet sms ¹	570	1,396	0.146	7	0.00 s	0.01 s	< 0.01 s	< 0.01 s	Communication
Copnet FB ¹	800	12,836	0.003	4	0.02 s	0.01 s	< 0.01 s	< 0.01 s	Facebook
FB Reed98 ²¹	962	37,624	0.006	3	0.10 s	0.03 s	0.01 s	0.01 s	Facebook
Arenas email ¹⁹	1,133	10,902	0.023	5	0.02 s	0.02 s	< 0.01 s	< 0.01 s	Communication
Network science ¹⁹	1,461	5,484	0.406	6	0.01 s	0.01 s	0.01 s	0.01 s	Co-autorship
FB Simmons81 ²¹	1,518	65,976	0.006	3	0.19 s	0.04 s	0.01 s	0.01 s	Facebook
DNC emails ¹⁹	1,893	8,770	0.640	4	0.02 s	0.37 s	< 0.01 s	< 0.01 s	Facebook
Moreno health ¹⁹	2,539	20,910	0.002	5	0.03 s	0.07 s	< 0.01 s	< 0.01 s	Human social
FB Wellesley22 ²¹	2,970	189,798	0.000	4	0.71 s	0.12 s	0.03 s	0.03 s	Facebook
Bitcoin alpha ²¹	3,783	28,248	0.241	5	0.07 s	0.56 s	< 0.01 s	< 0.01 s	Online social (trust)
GRQC collab. ²²	5,242	28,968	0.219	8	0.08 s	0.19 s	< 0.01 s	< 0.01 s	Co-autorship
FB Carnegie49 ²¹	6,637	499,934	0.004	4	2.47 s	0.81 s	0.09 s	0.09 s	Facebook
Pajek Erdős ¹⁹	6,927	23,700	0.658	6	0.04 s	0.32 s	< 0.01 s	< 0.01 s	Co-autorship
DT interaction ²³	7,341	30,276	0.448	12	0.09 s	5.44 s	0.01 s	0.01 s	Bio
DG assoc. ²³	7,813	42,714	0.469	8	0.21 s	5.69 s	< 0.01 s	0.01 s	Bio
FB GWU54 ²¹	12,193	939,056	0.003	4	4.62 s	1.63 s	0.14 s	0.14 s	Facebook
Anybeat ²¹	12,645	98,264	0.447	5	0.64 s	1.3 h	0.01 s	0.01 s	Online social
CE-CX ²¹	15,229	491,904	0.011	6	1.75 s	2.71 s	0.07 s	0.07 s	Bio
Astro Physics ²¹	18,771	396,100	0.089	6	1.83 s	1.013 s	0.05 s	0.05 s	Co-autorship
FB BU10 ²¹	19,700	1,275,056	0.003	4	5.45 s	2.58 s	0.19 s	0.19 s	Facebook
FB Uillinois ²¹	30,664	2,097,148	0.001	4	10.04 s	4.64 s	0.31 s	0.31 s	Facebook
Enron email ²²	36,692	367,662	0.382	6	1.51 s	2.32 m	0.05 s	0.05 s	Communication
FB Penn ²¹	41,536	2,724,440	0.001	4	15.19 s	9.57 s	0.41 s	0.42 s	Facebook
FB wall 2009 ¹⁹	46,952	366,824	0.083	8	0.69 s	2.76 s	0.05 s	0.07 s	Communication
Brightkite ²¹	58,228	428,156	0.171	8	1.13 s	22.94 s	0.07 s	0.09 s	Online social
The marker cafe ²⁴	69,413	3,289,686	0.168	5	1.12 m	19.83 m	0.58 s	0.60 s	Human contact
Slashdot zoo ¹⁹	79,116	935,462	0.207	7	5.05 s	8.86 m	0.16 s	0.20 s	Online social
Twitter ¹⁹	465,018	1,667,079	0.783	5	1.11 m	6.9 h	0.30 s	0.33 s	Online social
DBLP ¹⁹	182,702	16,689,229	0.189	10	1.08 m	1.8 h	21.80 s	22.53 s	Co-autorship
Flixster ¹⁹	2,523,387	15,837,601	0.584	8	3.54 m	9.6 h	15.59 s	16.05 s	Online social
Youtube ¹⁹	3,223,590	18,750,747	0.335	15	16.26 m	>1 week	32.94 s	33.98 s	Online social

Twin-unique:














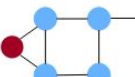

- unique or
- all candidates are twin

Incorporating Twin-uniqueness

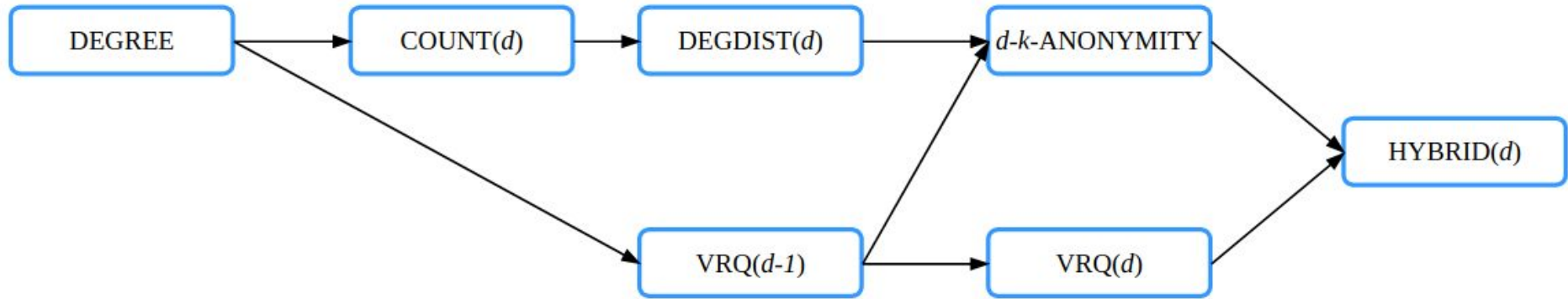


$C_1 \rightarrow$ Large increase twin-uniqueness compared to uniqueness

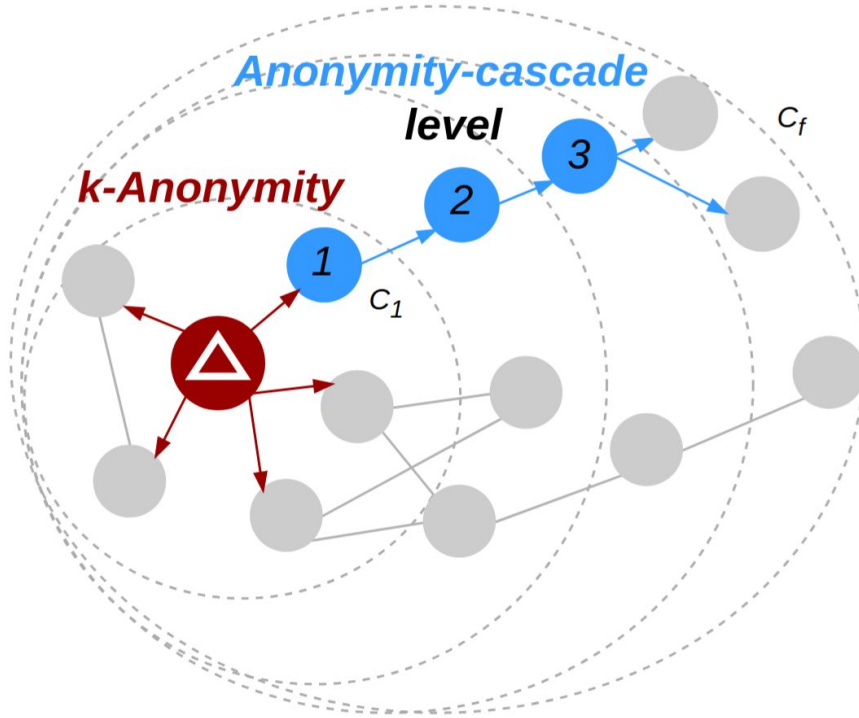
Measures

	$d = 1$		$d = 2$	
	Reach	Value	Reach	Value
Degree [26, 61–63]		2	-	-
Count [64]		(3, 3)		(5, 6)
Degree distribution [64]		{2, 2, 2}		{2, 2, 2, 3, 3}
d - k -Anonymity [10, 11, 65, 66]				
VRQ [27, 38]		{2, 3, 3}		{2, 2, 3, 3, 3}
Hybrid [67]		 {2, 3, 3}		 {2, 2, 3, 3, 3}

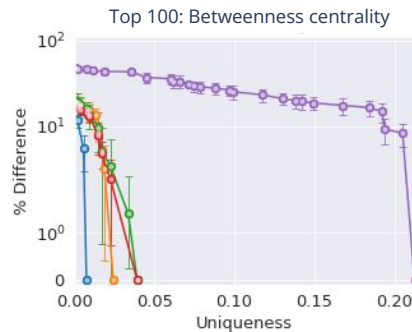
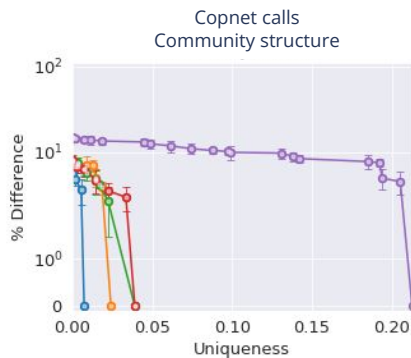
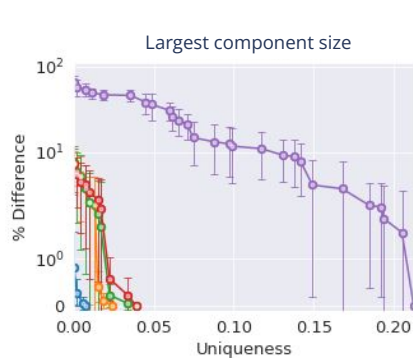
Measures: strictness



Anonymity-cascade



Anonymity vs. utility: measures



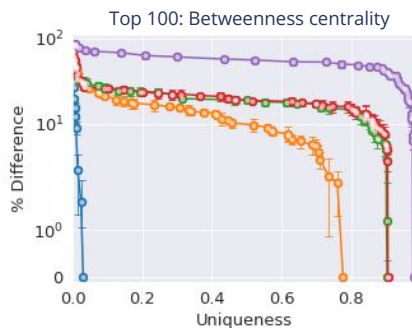
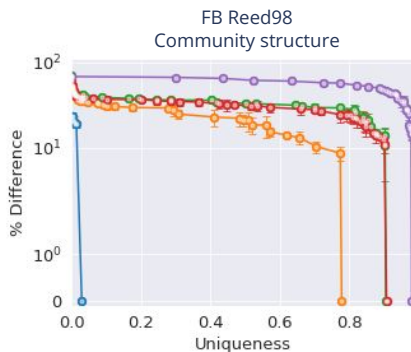
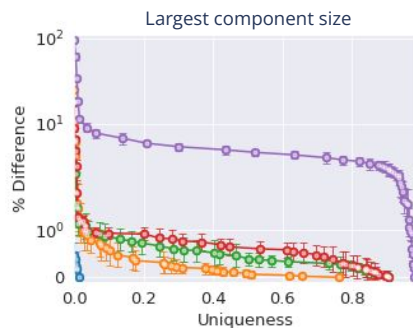
Degrees of neighbors

Isomorphism

Degree distribution

$(|V|, |E|)$

Degree

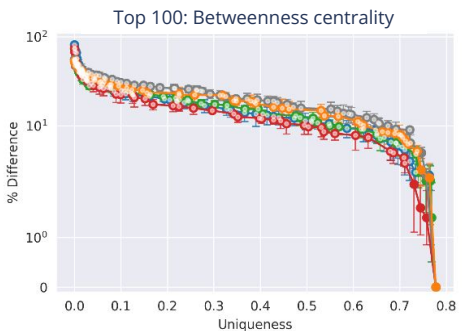
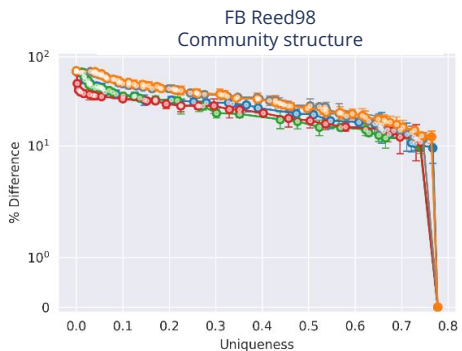
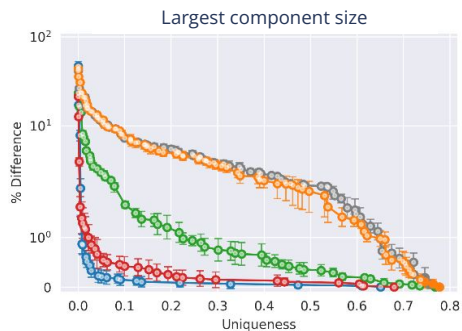
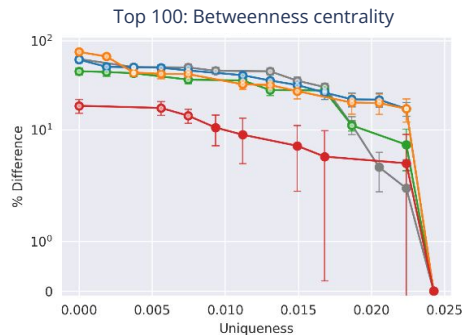
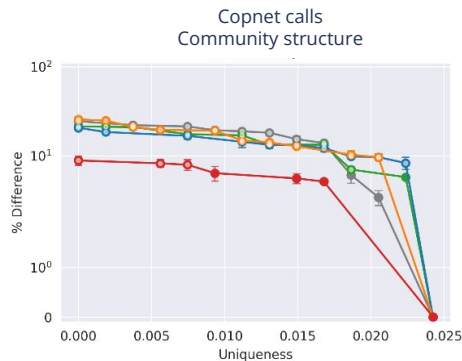
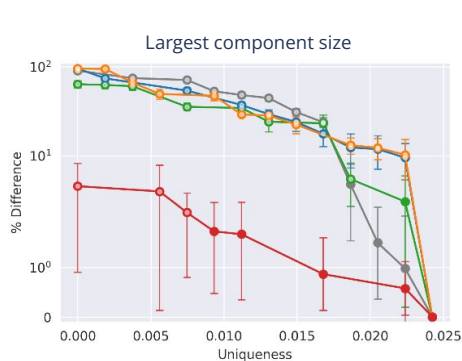


More strict measure



worse trade-off

Anonymity vs. utility: **algorithms**



Random

Structure based

Structure based

Uniqueness based

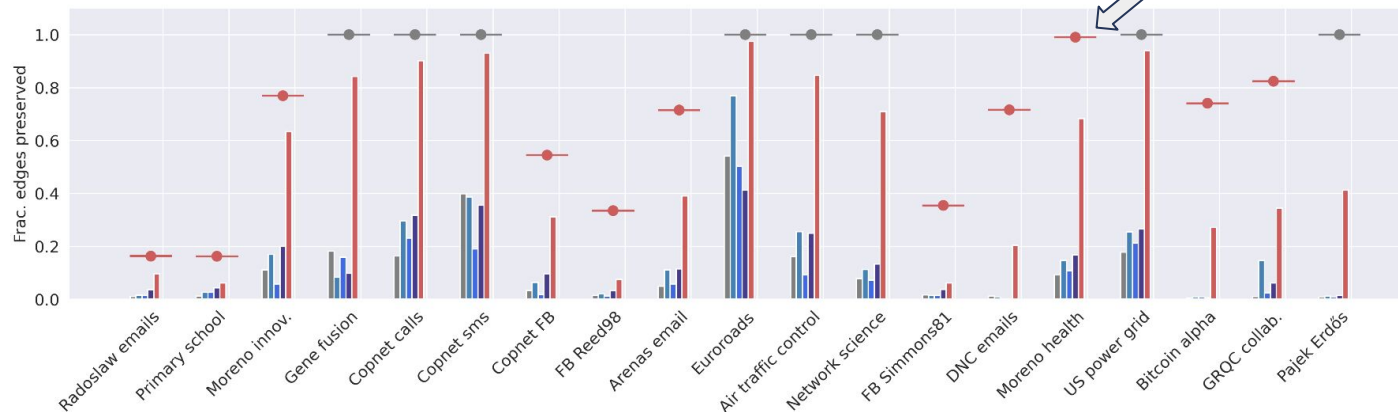
Uniqueness based

Better algorithm



better trade-off

Anonymization



95% anonymous



Random

Structure
based

Uniqueness
based

$|V|$: ~100 - 6,927 $|E|$: ~300 - 11,850

For more details
see our preprint!

