The risk of identity disclosure through network structure

Anecdotal evidence from a hackathon

M.M. de Vries, R.G. de Jong, M.P.J. van der Loo, P.-P. de Wolf, F.W. Takes

Agenda

- Problem

- Hackathon set-up

- Results

- Conclusions for SDC

Problem

Research problem

Statistics Netherlands recently developed population scale network (van der Laan et al, 2022)

- 5 types of links: family, neighbours, household members, colleagues, schoolmates
- Every person in the Netherlands

Anonymity measure developed with assumption of certain knowledge from attacker (de Jong et al, 2023a,b)

How likely is this prior knowledge?

Why a hackathon?

Online social networks (OSN) exhaustive source for finding sensitive data (Alipdrandi et al, 2014), (Koot et al, 2015)

- Open Source Intelligence (OSINT) takes advantage of online data

Research done into *what* is available, not *how much* is available

Hackathon reflects what is available, what information is harder/easier to find

Hackathon set-up

Hackathon organisation

22 students from Faculty of Science (Leiden University), split into 11 groups

Each group given 7 volunteers, asked to give as many links as possible

- 26 volunteers from CBS, Leiden University, other companies

4 hours, keeping a log

Volunteers were asked to assess validity found links

Recorded data

Edgelist_example

File Edit View Insert Format Data Tools Extensions Help

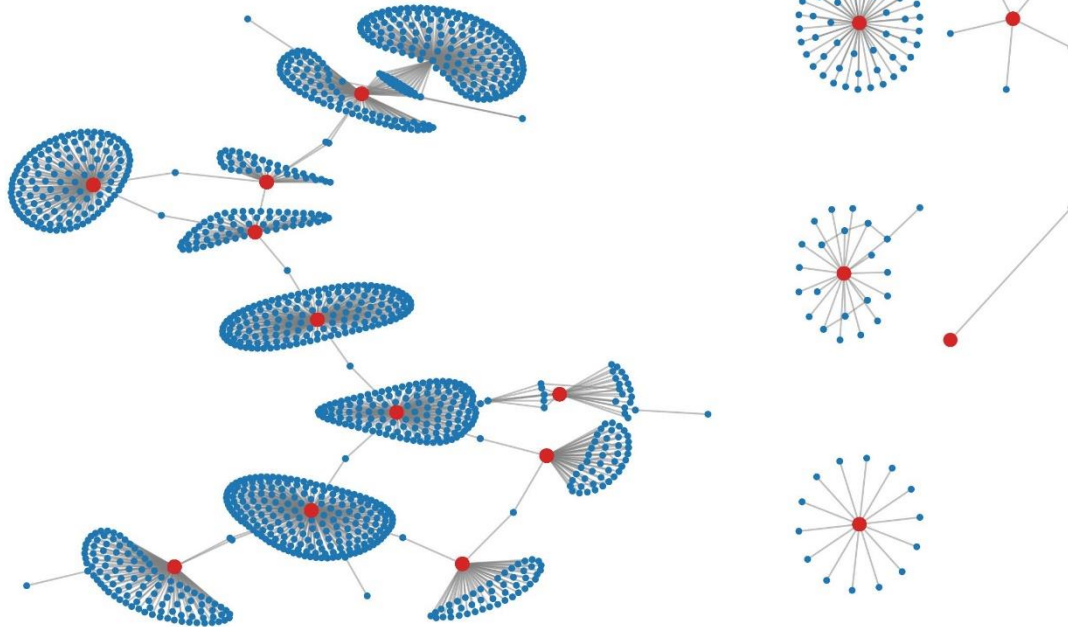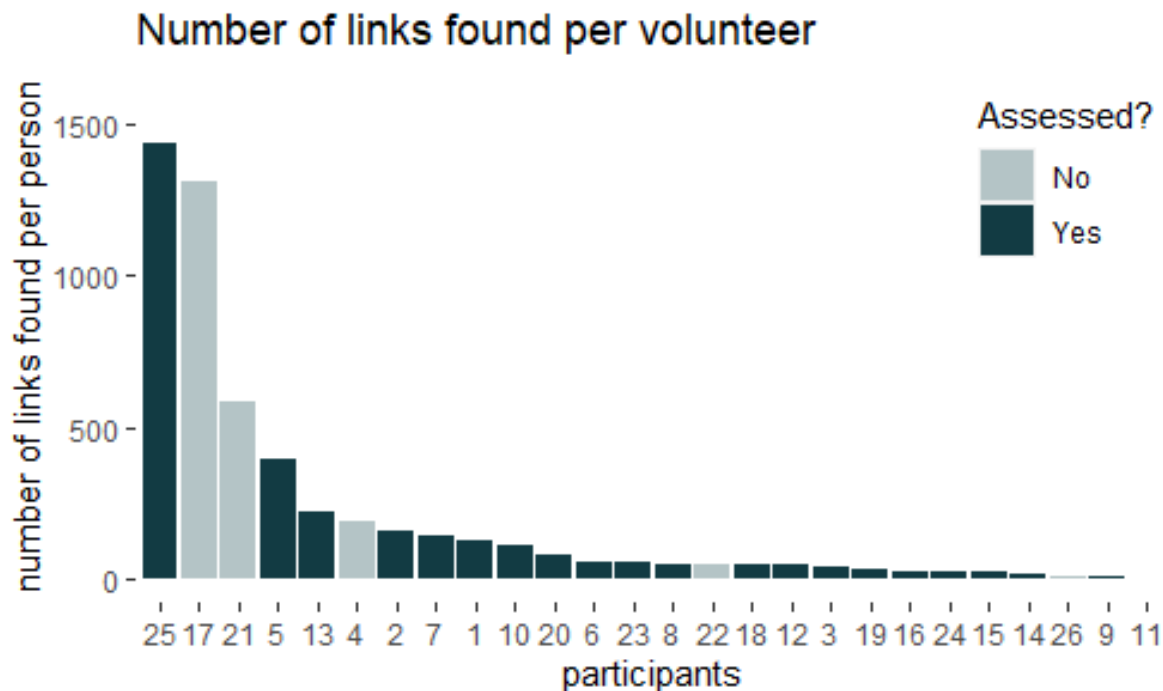|	A	B	C	D	E	F	G	H
1	Willem Alexander		Source: https://www.koninklijkhuis.nl/onderwerpen/geschiedenis/koningen-en-koninginnen/willem-alexander-koning-1967					
2	Source	Target	Type	Subtype (optional)	Distance	Reliability of link	Source	
3	Beatrix	Willem Alexande	Family	Kind	1	High	https://nl.wikipedia.org/wiki/Koninklijke_familie_van_Nederland	
4	Beatrix	Constantijn	Family	Kind	2	High	https://nl.wikipedia.org/wiki/Koninklijke_familie_van_Nederland	
5	Willem Alexande	Amalia	Family	Kind	1	High	https://nl.wikipedia.org/wiki/Koninklijke_familie_van_Nederland	
6	Amalia	Alexia	Family	Zus	2	High	https://nl.wikipedia.org/wiki/Koninklijke_familie_van_Nederland	
7	Willem Alexande	Maxima	Household	-	1	High	Story	
8	Thom de Graaf	Willem Alexande	Work	Colleague	1	High	raadvanstate.nl	
9	Bert	Ernie	Friends	-	20	High	https://nl.wikipedia.org/wiki/Bert_en_Ernie	
10								
11								
12								

8

Results

Networks found for all volunteers

Big differences between volunteers

Number of links found per volunteer

Assessed?
- No
- Yes

number of links found per person

participants

Social media deemed relatively unreliable

Types of sources and their perceived reliability
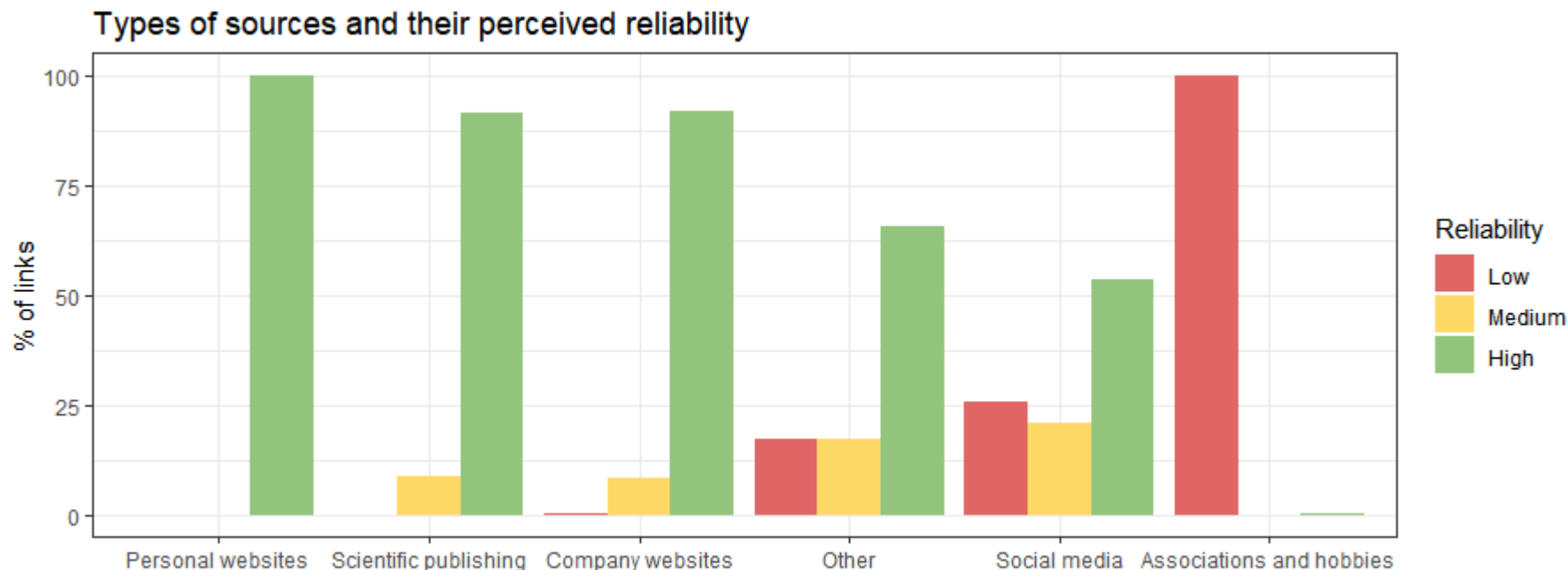
Correct assessment of validity

Perceived reliability vs accuracy
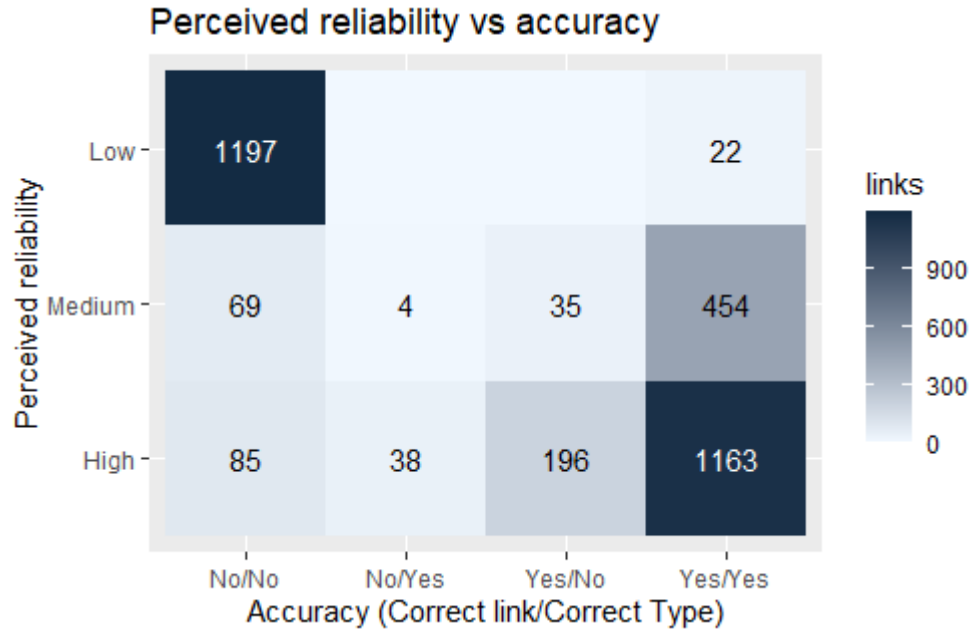
Differences between categories of links

Types of links and accuracy
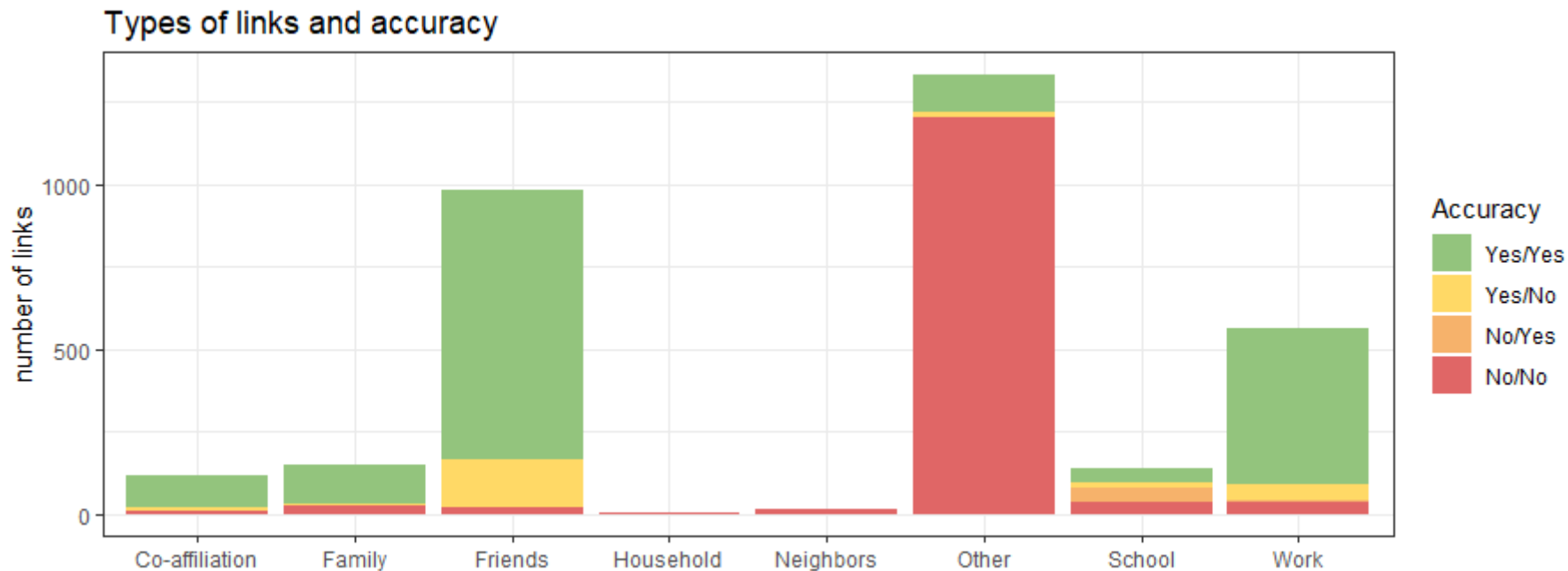
Conclusions for Statistical Disclosure Control

- Friends and colleagues easy to find and often correctly inferred

- Household members and neighbours difficult to find and often incorrect, regardless of perceived reliability

- Perceived reliability often matched accuracy

- Higher order relationships were found far less, either due to assignment or due to difficulty

Open questions

– More research needed on online availability

– Further development of anonymity measures in networks

– How to include outside sources and public information

– More generic approach for assessing risk needed:
  - Assess vulnerabilities in the attacker scenarios
  - Assess likelihood of these scenarios themselves.

Facts that matter

References

1 - van der Laan, J., E. de Jonge, M. Das, S. Te Riele, and T. Emery (2022). A whole population network and its application for the social sciences. *European Sociological Review* 39(1), 145–160.

2 - de Jong, R. G., M. P. J. van der Loo, and F. W. Takes (2023a, jun). Algorithms for efficiently computing structural anonymity in complex networks. *ACM J. Exp. Algorithmics*.

3 - de Jong, R. G., M. P. J. van der Loo, and F. W. Takes (2023b). Beyond the ego network: the effect of distant connections on node anonymity. *arXiv preprint 2306.13508*3 -

4 - Aliprandi, C., J. Irujo, M. Cuadros, S. Maier, F. Melero, and M. Raffaelli (2014, 06). Caper: Collaborative information, acquisition, processing, exploitation and reporting for the prevention of organised crime. Volume 434.

5 - Koot, G., M. Huis in 't Veld, J. Hendricksen, R. Kaptein, A. Vries, and E. van den Broek (2014, September). Foraging online social networks. In M. den Hengst, M. Israël, D. Zeng, C. Veenman, and A. Wang (Eds.), *Proceedings of the 2014 IEEE Joint Intelligence and Security Informatics Conference (JISIC2014)*, pp. 312–315. IEEE. 10.1109/JISIC.2014.62 ; Conference date: 24-09-2014 Through 26-09-2014.