

Ennetcom: e-mails zijn super graaf!

sociale-netwerk-analyse in een
PGP-dataset

Aanleiding: “Ik wil ook data!”

- data uit Canada
- Nederlands RC keurt dataverzoeken
- ... elk dataverzoek zit anders in elkaar
 - broodjes special™

NFI, doe eens?

- aanvraag uit Amsterdam met bijzonder verzoek
- groep in beeld, wie horen daar nog meer bij?
 - sociale-netwerk-analyse!
 - maar dat is nog niet triviaal...
- nauwe samenwerking met Politie

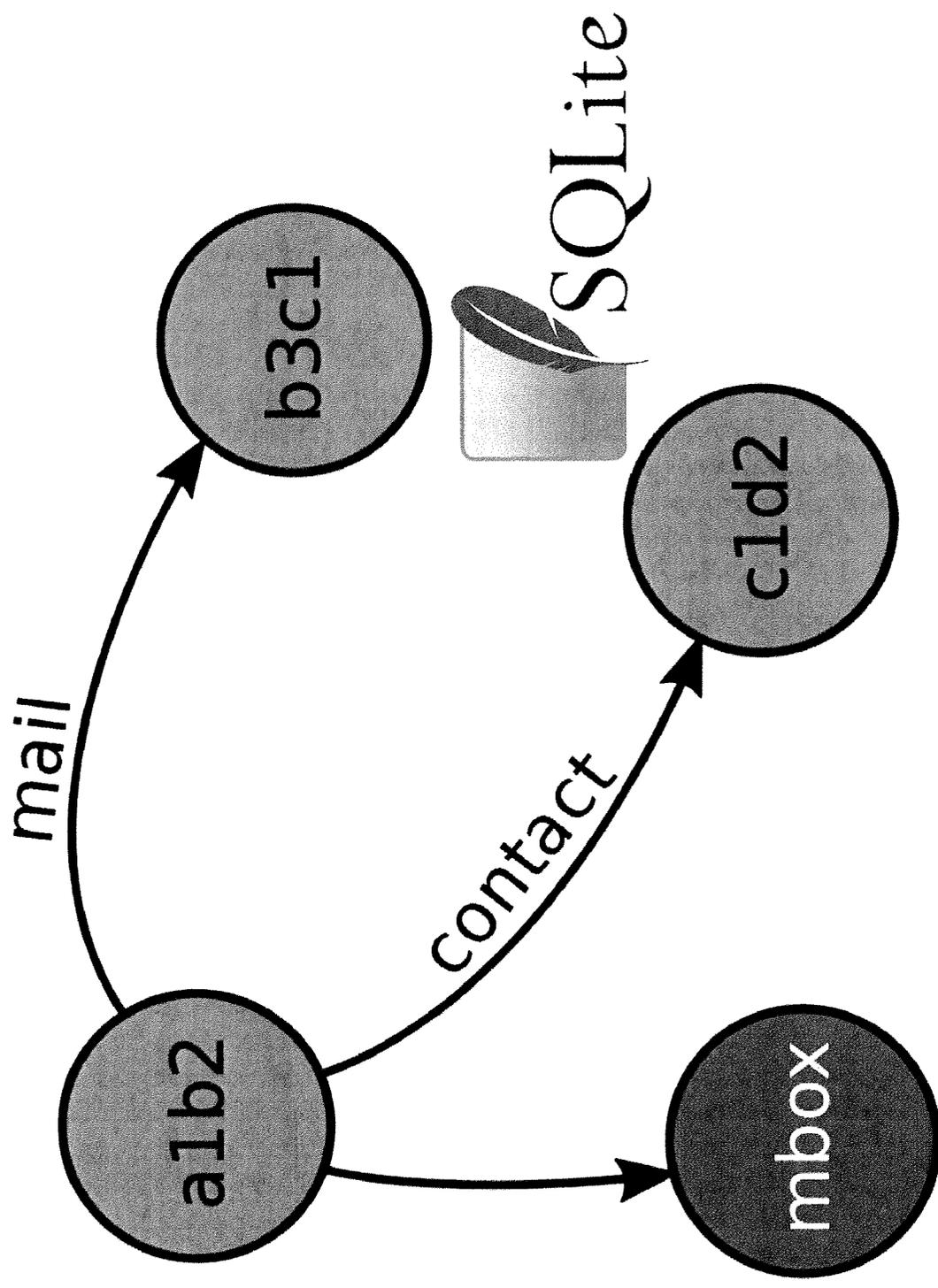
Requirements

- 'simpel' / duidelijk
- deterministisch
- nauwe selectie
- efficiënt

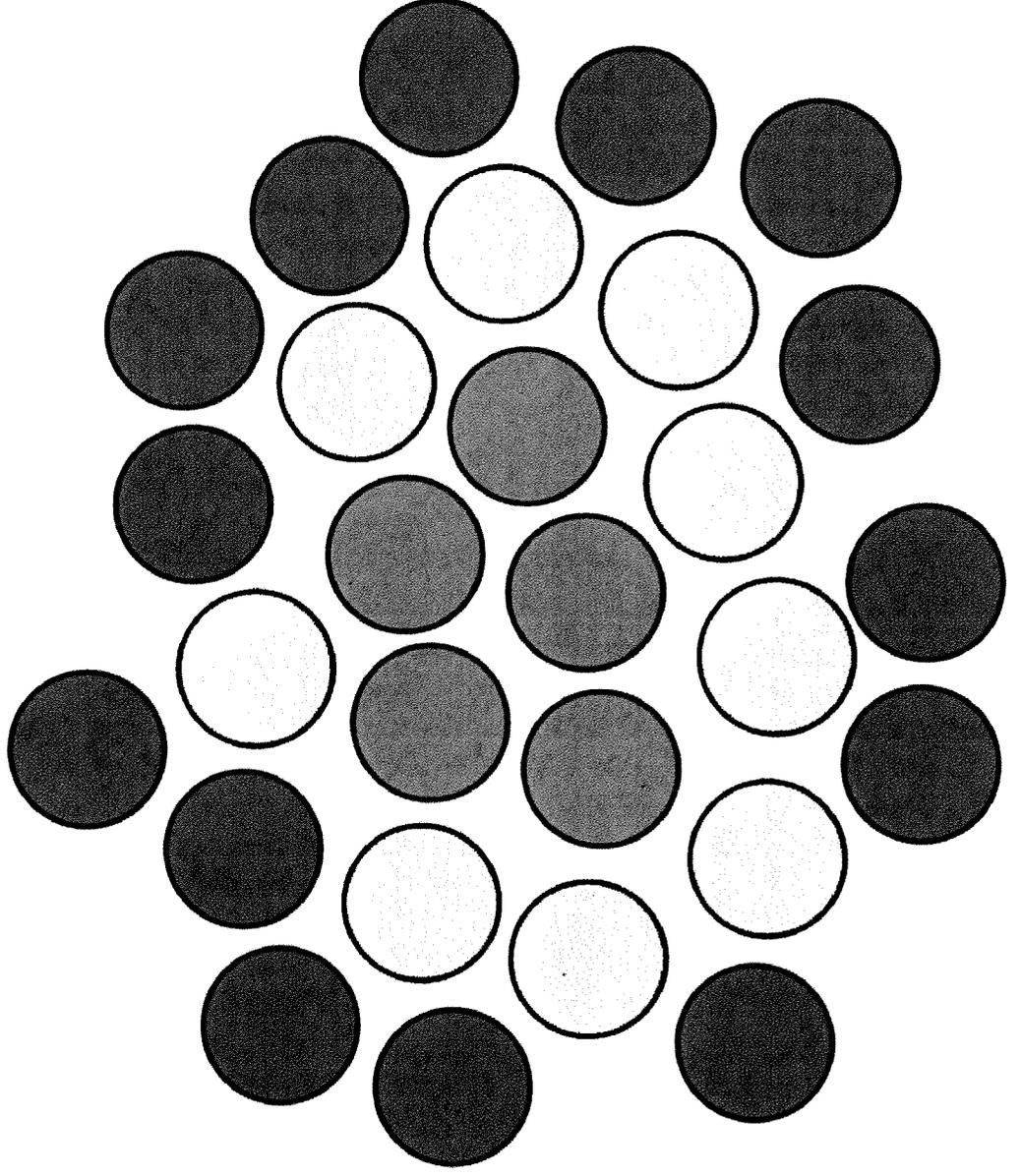
Wat hebben we eigenlijk?

- miljoenen e-mails
- verzonden van/naar een 'id'
- 'id' opgeslagen als "gekke henkie"
- allemaal onderverdeeld in mailboxen

Extractie uit Hansken



Het ABC-algoritme™



Rapportage

Kader A: 5 identiteiten (aangeleverd door aanvrager)

1 mailbox in dit kader, 0.0% van het totaal aantal mailboxen tot en met dit kader.

1234abcd	5678efgh	1357qrst	tech-support	uvwxyz2468
----------	----------	----------	--------------	------------

Kader B: 12 identiteiten

11 mailboxen in dit kader, 0.1% van het totaal aantal mailboxen tot en met dit kader.

dskjhf4389	1337-telefoon	sdk389sddd	2349sdkfnm	sdkrfjgn39
sdh520454	sdh520454			

Legenda

Kader

176

- 12345abcde Identiteit aangetroffen in de data waar waarschijnlijk geen mailbox in de data bij hoort.
- 12345abcde Identiteit waar waarschijnlijk één of meerdere mailboxen in de data bij horen die niet worden vrijgegeven.
- 12345abcde Identiteit waar waarschijnlijk één of meerdere mailboxen in de data bij horen die niet worden vrijgegeven.
- 12345abcde Identiteit van bekende medewerker.
- 12345abcde Identiteit van bekende reseller.

Resultaat

- simpel / duidelijk
 - algoritme past op 1 slide :)
- deterministisch
 - geen fuzzy links, 'harde' metadata
- nauwe selectie
 - mailboxen van kaders A en B
 - identiteiten van kader C
- efficiënt
 - graven in losstaande database
 - oplevering per mailbox

Resultaat

- nauwe samenwerking met team Canadadata en onderzoeksteam Amsterdam
- 3 standaardmogelijkheden
 - keywords
 - **ABC**
 - tijdsbestek

vragen / opmerkingen

zie ook de Hansken-stand ;)

