

Unfriend your Boss

Mapping Organizations Social Networks for
Red Team Engagments

HamburgSides 2016 - Quentin Kaiser ([@qkaiser](#))



who ?

- Quentin Kaiser
- Security Engineer / Penetration Tester
-  contact@quentinkaiser.be
-  [@qkaiser](https://twitter.com/qkaiser)
-  <https://qkaiser.github.io>

Disclaimer

This research was prepared and accomplished in my personal capacity. The opinions expressed in this talk are my own and do not reflect the view of past, current, or future employer.

Agenda

- The Origins
- Scraping Facebook 101
- Social Network Analysis
- Leveraging results during red team gigs
- Possible developments
- Mitigating Exposure

The Origins

When it comes to phishing, most pentest shops do it like this:

The Origins

When it comes to phishing, most pentest shops do it like this:

- scraping emails off the Internet

The Origins

When it comes to phishing, most pentest shops do it like this:

- scraping emails off the Internet
- finding profiles, finding email pattern, generating addresses

The Origins

When it comes to phishing, most pentest shops do it like this:

- scraping emails off the Internet
- finding profiles, finding email pattern, generating addresses
- hit "send"

The Origins

When it comes to phishing, most pentest shops do it like this:

- scraping emails off the Internet
- finding profiles, finding email pattern, generating addresses
- hit "send"

That's cool, but does it really reflect advanced attackers ?

The Origins

I want to fully take advantage of social **networks**

Research Questions

- how much data is **exposed** by an organization members through social media ?
- what information can be **derived** from such data ?
- how can we **exploit** that information in red team gigs ?

Enter Facebook Graph Search

Facebook Graph Search

- Semantic search engine
- Introduced back in March 2013
- User send query in natural language to search for entities such as pages, places, events, check-ins, status updates, and people

Facebook Graph Search

Facebook thwarted privacy concerns around Graph Search, explaining that entities are indexed based on privacy settings that Facebook users can tweak.

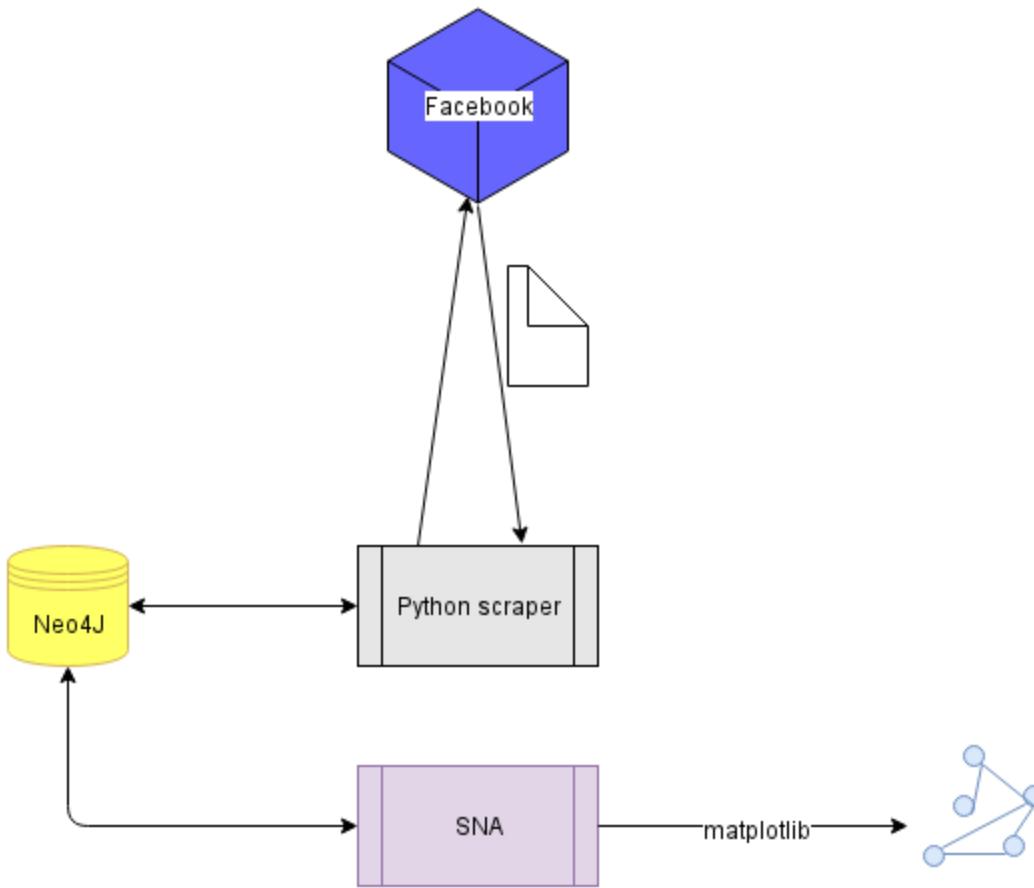
Facebook Graph Search

- Using the privacy settings as an argument for privacy protection is a fallacy.
- if *some* members of a graph under scrutiny are sharing their relationships publicly, each person linked to that overly open person will be included in the graph. Even if they don't want to.
- you decide **attributes** you share, not **relations**

Methodology

1. Scrape data off Facebook
2. Store it in graph database
3. Apply SNA
4. Exploit

The Stack



Data Acquisition

- wrote a scraper in Python based on initial work with recon-ng
- Facebook's continuous integration continuously breaks it
- but when it works 

Data Acquisition

1. Find company page

The screenshot shows a Facebook search results page for the query "google". The search bar at the top contains "google". Below the search bar, the navigation menu includes Home, Find Friends, a user icon, and a lock icon. The main content area displays three company pages:

- Google** (verified) · Website · 21,533,387 likes · Organizing the world's information and making it universally accessible and useful · [Like](#)
- Google Chrome** (verified) · Software · 18,252,382 likes · Julien Janssen likes this · Download Google Chrome: <http://www.google.com/chrome/?brand=fb> · [Learn More](#) · [Like](#)
- Google Maps** (verified) · Website · 3,228,622 likes · Mathis Charneux likes this · News, tips and candid tales of Google Maps · [Like](#)

Data Acquisition

2. Employees retrieval

The screenshot shows a Facebook search results page with a blue header bar containing the Facebook logo, the search query 'People who work at Tesla Motors', and a magnifying glass icon. Below the header are three search results, each featuring a user profile picture (a blue silhouette), a blurred name, an 'Add Friend' button, and a more options button. The first result is for a 'Senior Manufacturing Engineer at Tesla Motors' who studied at California Polytechnic State University (Cal Poly) and lives in San Jose, California. The second result is for a 'Firmware Engineer at Tesla Motors' who studies Computer Engineering at Stanford University and lives in Seattle, Washington. The third result is for a 'Production Associate at Tesla Motors' who went to Berkeley High School (Berkeley, California) '90 and lives in Oakland, California.

Role	Education	Location
Senior Manufacturing Engineer	Studied at California Polytechnic State University (Cal Poly)	Lives in San Jose, California
Firmware Engineer	Studies Computer Engineering at Stanford University	Lives in Seattle, Washington
Production Associate	Went to Berkeley High School (Berkeley, California) '90	Lives in Oakland, California

```
/search/C_ID/employees/present
```

Data Acquisition

3. Relationships retrieval

The screenshot shows a Facebook search results page with the query '...' 's friends who work at Google'. The results list three profiles:

- Security Moonshots at Google**
Friends with [REDACTED] since March 2007
Studied Master's Degree at University of Maryland, Baltimore Co...
- Software Security Engineer at Google**
Friends with [REDACTED] since August 2015
Listens to Meshuggah, Tesseract and Periphery
- Works at Google**
Friends with [REDACTED] since May 2015
Went to Seneca Valley Senior High School

```
/search/C_ID/employees/present/U_ID/friends/intersect
```

Data Acquisition Challenges

- User-Agent blacklisting

Data Acquisition Challenges

- User-Agent blacklisting
- Session token

Data Acquisition Challenges

- User-Agent blacklisting
- Session token
- Lazy Loading

Data Acquisition Challenges

- User-Agent blacklisting
- Session token
- Lazy Loading
- Parsing

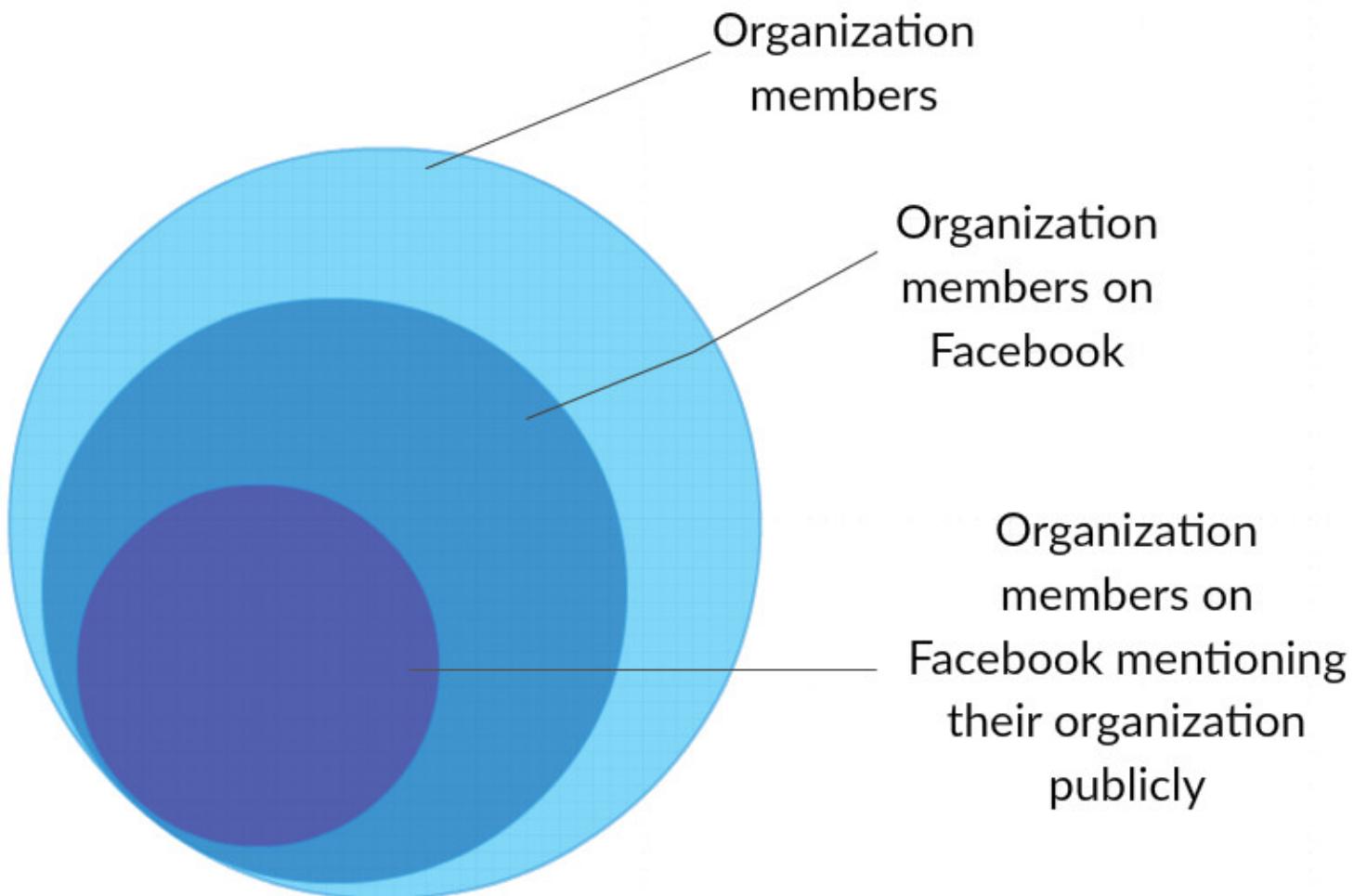
Data Acquisition Challenges

- User-Agent blacklisting
- Session token
- Lazy Loading
- Parsing
- Rate limiting

Data Acquisition Challenges

- User-Agent blacklisting
- Session token
- Lazy Loading
- Parsing
- Rate limiting
- CAPTCHA

Data Acquisition Results



Data Acquisition Results

- data set comprised 20 large companies with > 10k employees
- coverage oscillate around 20%
- 3000 nodes per company on average

Data Acquisition Results

- data set comprised 20 large companies with > 10k employees
- coverage oscillate around 20%
- 3000 nodes per company on average

Data is incomplete and I don't have the processing power of a server farm.

Data Acquisition Results

- data set comprised 20 large companies with > 10k employees
- coverage oscillate around 20%
- 3000 nodes per company on average

Data is incomplete and I don't have the processing power of a server farm.

→ It's therefore **qualitative** analysis.

Social Network Analysis

Social Network Analysis

- Degree Centrality

Social Network Analysis

- Degree Centrality
- Betweenness Centrality

Social Network Analysis

- Degree Centrality
- Betweenness Centrality
- Community Detection

Social Network Analysis

- Degree Centrality
- Betweenness Centrality
- Community Detection
- Hierarchy Reconstruction

Social Network Analysis

- Degree Centrality
- Betweenness Centrality
- Community Detection
- Hierarchy Reconstruction 😕

SNA / Degree Centrality

Degree centrality is a measure of influence within large and complex networks.

SNA / Degree Centrality

Nodes with highest degree:

- HR department workers
- union representatives
- "social beasts"

SNA / Betweenness Centrality

A measure for quantifying the control of a human on the communication between other humans in a social network.

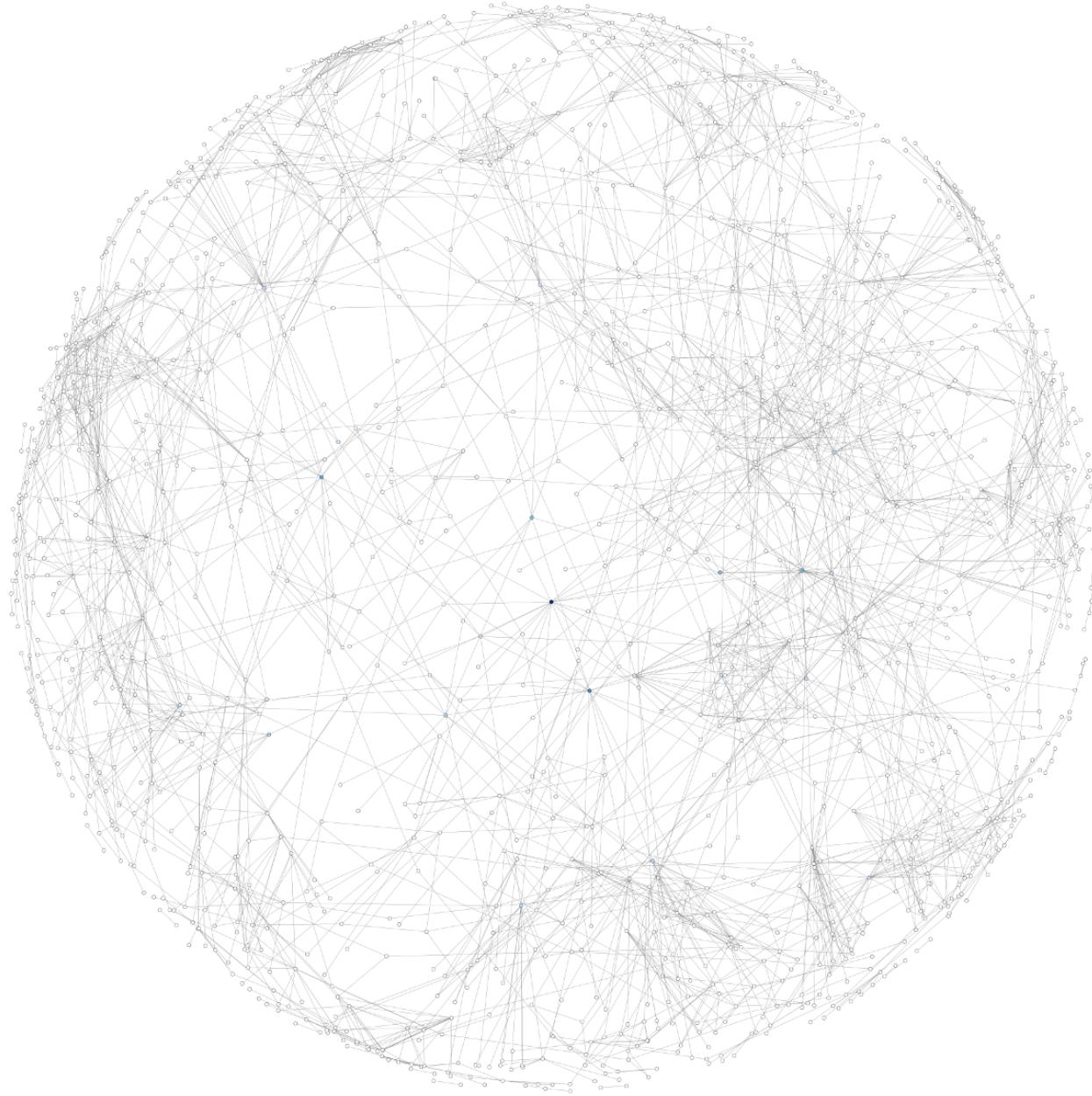
SNA / Betweenness Centrality

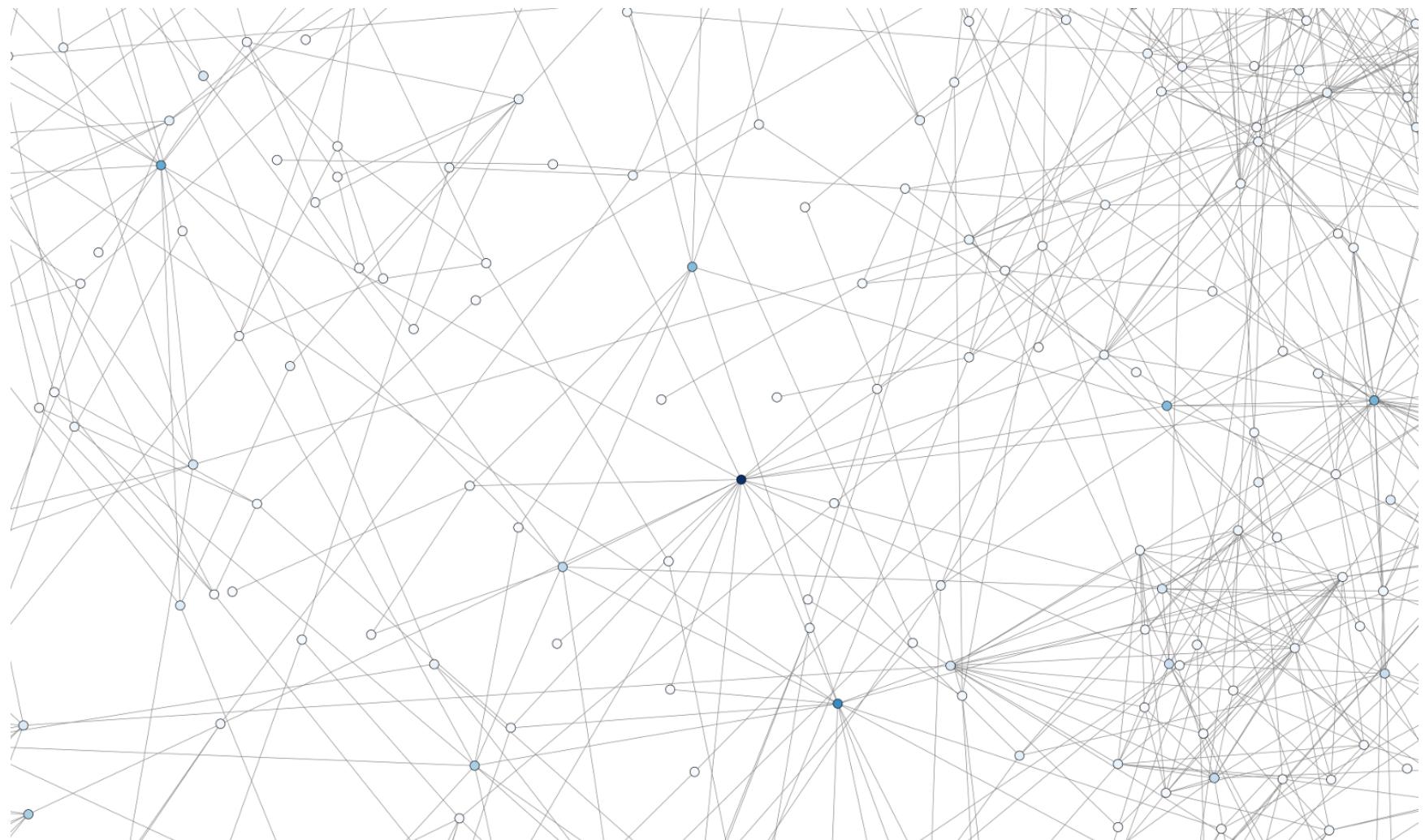
Nodes with highest degree are high value targets holding coordinators profiles (executives, directors).

SNA / Betweenness Centrality

Nodes with highest degree are high value targets holding coordinators profiles (executives, directors).

Really close to a "executives top 10".



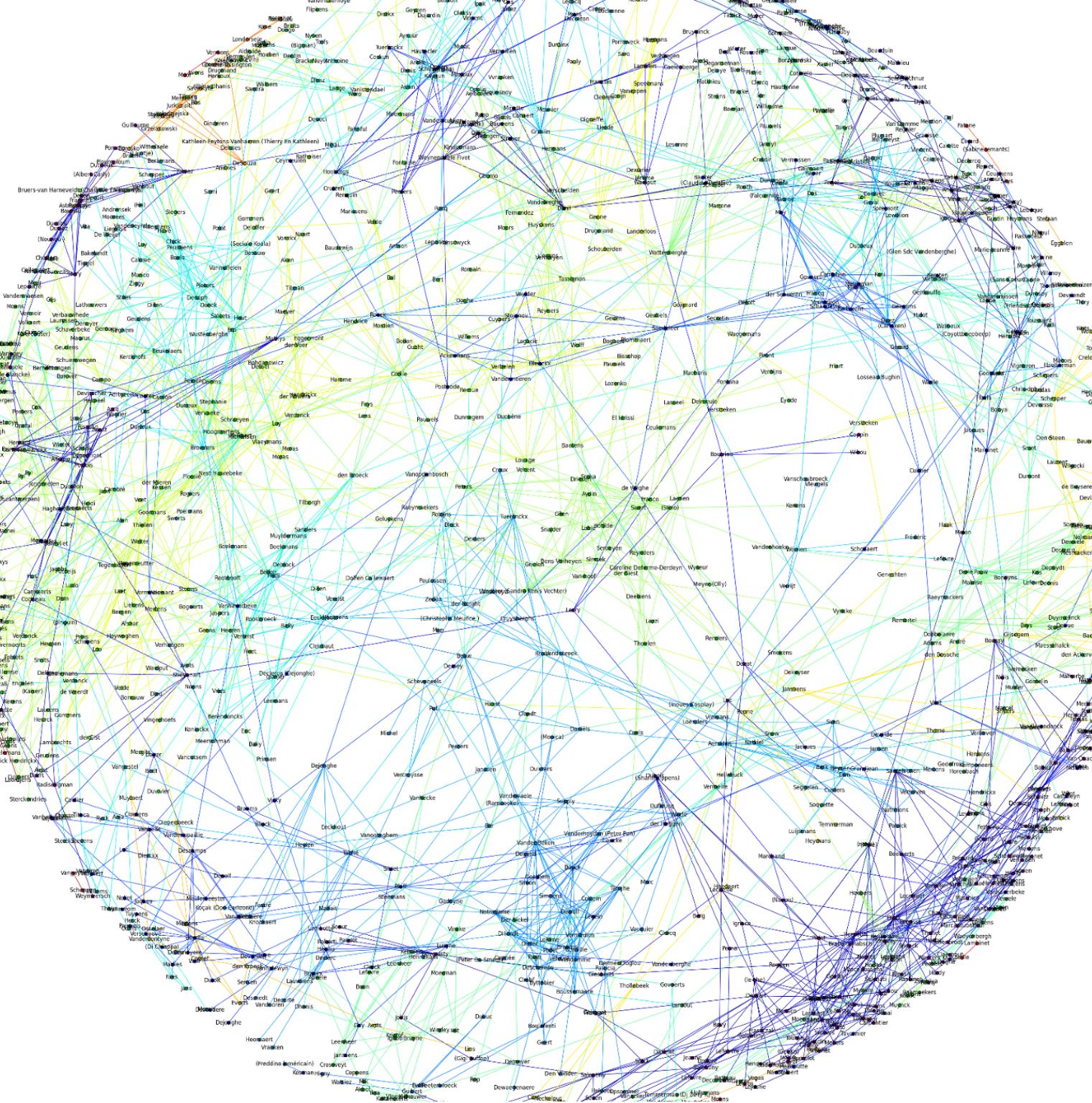


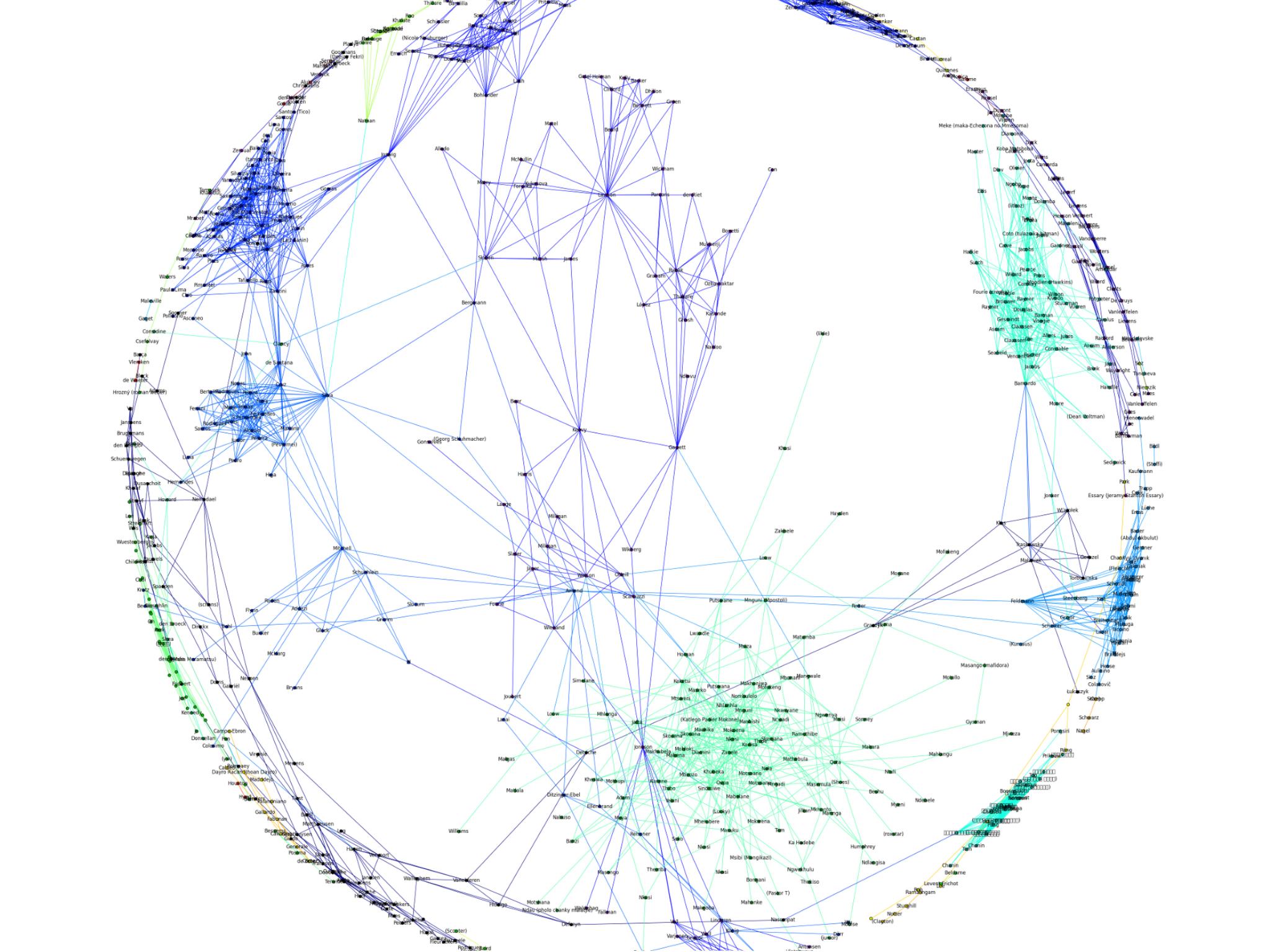
SNA / Community Detection

Humans have a tendency to homophily, mainly caused by geographical and organizational foci. Such a tendency can be visualized when applying community detection algorithms to the acquired social networks.

SNA/ Community Detection

- Relied on Louvain algorithm
- Observed communities foci:
 - geographical locations
 - internal departments
- could potentially identify hidden communities
 - internal commissions
 - working groups
 - shared common interest (sports club, political affiliation)





SNA/ Community Detection

SNA / Hierarchy Reconstruction

Hierarchy reconstruction can't be done because edges are **not weighted** and **undirected**.

SNA / Hierarchy Reconstruction

Hierarchy reconstruction can't be done because edges are **not weighted** and **undirected**. 😢

Leveraging results during red team gigs

- phishing on steroid
- social engineering tactics
- emulating corporate espionage
- organizational destabilization

Possible developments

- extension to other online social networks
- inclusion of profile attributes
 - geolocation (home town, work address, geolocated posts, checkins)
 - time (employmentship duration, friendship duration)

Extension to other online social networks

Methodology could be applied to Twitter.

1. find target organization twitter profile
2. search profiles mentioning @org in their bio
3. get relationships by intersecting each member's followers with profiles set

Extension to other online social networks

I actually did it.

```
qkaiser@localhost$ python main.py tesla
[+] Connecting to Neo4J ...
[+] Connection established (Neo4J 2.2.5)
[+] Searching for tesla ...
    [0] TeslaMotors
    [1] elonmusk
    [2] TeslaMotorsClub
Please select a profile:0
Looking for @TeslaMotors employees...
```

Extension to other online social networks

Some kind of NLP is required to differentiate:

Extension to other online social networks

Some kind of NLP is required to differentiate:

- **fan boys**

[*] ChrisLaylin - Aspiring @TeslaMotors owner

[*] dpetersson - early adopter & @TeslaMotors fan

Extension to other online social networks

Some kind of NLP is required to differentiate:

- **fan boys**

[*] ChrisLaylin - Aspiring @TeslaMotors owner
[*] dpetersson - early adopter & @TeslaMotors fan

- **ex-employees**

[*] JordanLevine - Former: @TeslaMotors
[*] clairissima - formerly @teslamotors

Extension to other online social networks

Some kind of NLP is required to differentiate:

- **fan boys**

```
[*] ChrisLaylin - Aspiring @TeslaMotors owner  
[*] dpetersson - early adopter & @TeslaMotors fan
```

- **ex-employees**

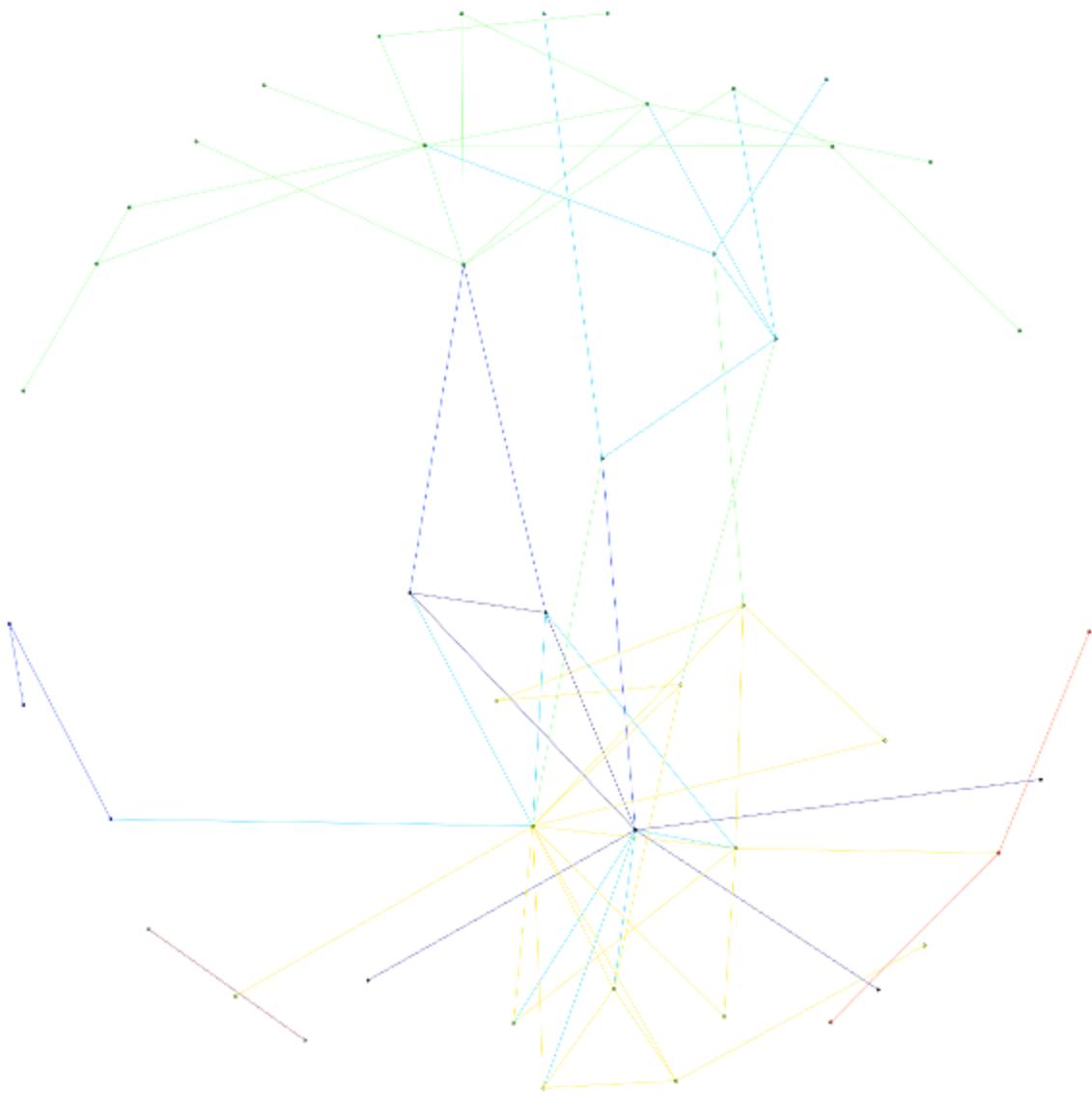
```
[*] JordanLevine - Former: @TeslaMotors  
[*] clairissima - formerly @teslamotors
```

- **actual employees**

```
[*] rcpopple - Finance Director @TeslaMotors  
[*] markmuraoka - Battery Engineering Intern @TeslaMotors
```

Extension to other online social networks

- broadcast-oriented networks 
- not enough coverage 
- still, can enrich prior analysis 



Mitigating Exposure

It is difficult to safely reveal limited information about a social network

- Joseph Bonneau in "*Eight friends are enough: Social graph approximation via public listings.*"

Mitigating Exposure

- Facebook
 - exposure limitation as function of node distance
 - anti-scrapping measures

Mitigating Exposure

- Organizations
 - find a balance between PR benefit and potential threat of exposure
 - legal department issue
 - security awareness training

Mitigating Exposure

- End users
 - remove as much information as possible

I haz find bug, pls send 500\$

I didn't get in contact with FB because of this:

This is a case where privacy can get complicated [...]. We've chosen to focus more on privacy controls around your content and personal information, since trying to maintain privacy by **limiting discoverability is often an illusion**. Since Facebook is a network designed for social participation, it's nearly impossible for it to work properly and let people stay completely hidden - there are many ways to discover a profile or friendship beyond friend lists or searches. But even if someone discovers your profile, **you have a great degree of control about what they can then access**.

Facebook bug bounties - the unofficial treasure map, June 2016.

<https://www.facebook.com/notes/phwd/facebook-bug-bounties-the-unofficial-treasure-map/1020506894706001>

Thanks for your attention! 👍

Any questions ?

Links ! Links ! Links !

- Toolkit <https://github.com/qkaiser/segraph>
(code broke again, I'll push it as soon as I fixed it.)
- Blog post <https://qkaiser.github.io/>
- White paper <https://qkaiser.github.io/publications>

Related Work

aka "cool reads"

- **The Power of Local Information in Social Networks** by *Borgs et al.*
- **Inferring the Maximum Likelihood Hierarchy in Social Networks** by *Maiya et al.*
- **Finding Hierarchy in Directed Online Social Networks** by *Gupte et al.*
- **Eight Friends Are Enough: Social Graph Approximation via Public Listings** by *Bonneau et al.* 

Food for thoughts

EDITION: UNITED STATES ▾

 **REUTERS**

Business Markets World Politics Tech Commentary Breakingviews Money Life

WORLD NEWS | Sat Dec 24, 2016 | 9:52am EST

Turkish authorities investigating 10,000 social media users: ministry



Guess how.