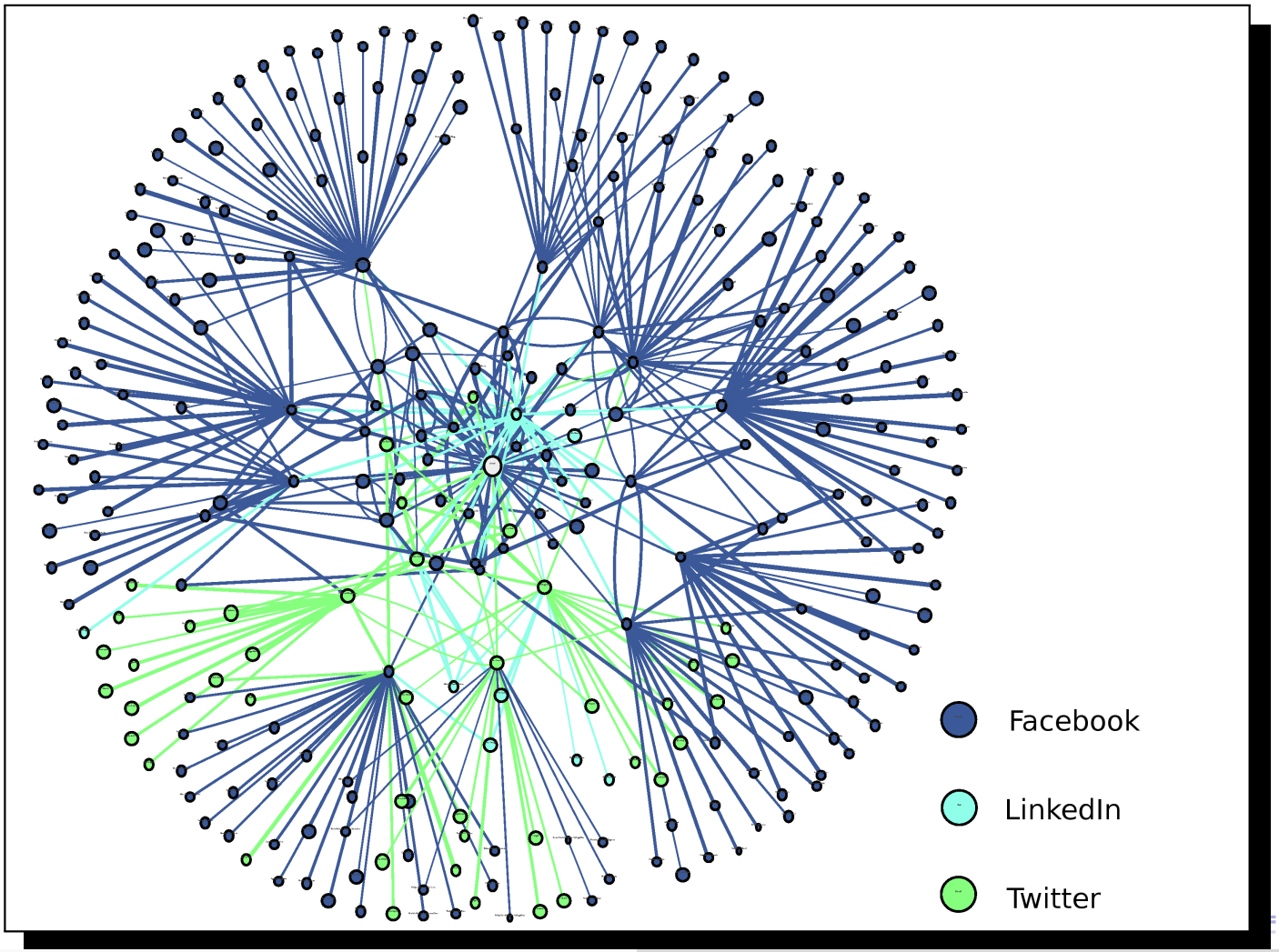
LeakedOut: the Social Networks You Get Caught In

José I. Orlicki

CoreLabs - Core Security Technologies
PhD program - ITBA
(BA-Con 2008 Buenos Aires)

October 8, 2008

Introduction - You are here!



Who am I

- I am 26 years old.
- I am from Buenos Aires, Argentina (where you are located).
- Computer scientist.
- Researcher at CoreLabs for 2 years. Core sells pen-test software and services.
- Phd student at ITBA.
- Main interests: Open Source Intelligence, Attack Simulation and Planning, Complex Network Analysis.
- Mostly harmless, interested in Social Networks after getting sucked into LinkedIn.

What is this talk about

- *Not* about WebApps Security, check previous talk at BH-USA "Black Hat 2008: Satan Is On My Friends List", OpenSocial section recommended.

What is this talk about

- *Not* about WebApps Security, check previous talk at BH-USA "Black Hat 2008: Satan Is On My Friends List", OpenSocial section recommended.
- *Not* about closing your Facebook account, you can do it sending an *email* to the appropriate people.

What is this talk about

- *Not* about WebApps Security, check previous talk at BH-USA "Black Hat 2008: Satan Is On My Friends List", OpenSocial section recommended.
- *Not* about closing your Facebook account, you can do it sending an *email* to the appropriate people.
- Open Source Intelligence (OSINT): processing public/leaked info to generate actionable intelligence. (Not GPL!)

What is this talk about

- *Not* about WebApps Security, check previous talk at BH-USA "Black Hat 2008: Satan Is On My Friends List", OpenSocial section recommended.
- *Not* about closing your Facebook account, you can do it sending an *email* to the appropriate people.
- Open Source Intelligence (OSINT): processing public/leaked info to generate actionable intelligence. (Not GPL!)
- Centered on Internet, Search Engines, Social Network Services, IMs, Web 0.2, etc.

What is this talk about

- *Not* about WebApps Security, check previous talk at BH-USA "Black Hat 2008: Satan Is On My Friends List", OpenSocial section recommended.
- *Not* about closing your Facebook account, you can do it sending an *email* to the appropriate people.
- Open Source Intelligence (OSINT): processing public/leaked info to generate actionable intelligence. (Not GPL!)
- Centered on Internet, Search Engines, Social Network Services, IMs, Web 0.2, etc.
- Discusses a small prototype called Exomind.

What's funny part?

- Navigate color graphics with dots and lines.

What's funny part?

- Navigate color graphics with dots and lines.
- Learn to control your leaked profile.

What's funny part?

- Navigate color graphics with dots and lines.
- Learn to control your leaked profile.
- Build chatbot replicants after watching *Bladerunner* (director's cut).

What's funny part?

- Navigate color graphics with dots and lines.
- Learn to control your leaked profile.
- Build chatbot replicants after watching *Bladerunner* (director's cut).
- You are allowed to talk hype after this talk, using expressions like Cloud Computing and Attention Economy.

What's funny part?

- Navigate color graphics with dots and lines.
- Learn to control your leaked profile.
- Build chatbot replicants after watching *Bladerunner* (director's cut).
- You are allowed to talk hype after this talk, using expressions like Cloud Computing and Attention Economy.
- Fortunately understand real privacy and security threats to you and your organization.

What's funny part?

- Navigate color graphics with dots and lines.
- Learn to control your leaked profile.
- Build chatbot replicants after watching *Bladerunner* (director's cut).
- You are allowed to talk hype after this talk, using expressions like Cloud Computing and Attention Economy.
- Fortunately understand real privacy and security threats to you and your organization.
- Bottom line: now with the Wikinomy and Web 0.2 there is a lot of public/leaked info available, you may need to measure the consequences !?!

Client-side is a pentest subgenre

Automatic OSINT has already been included in professional pen-testing.

- Phishing and client apps exploitation led to the development of client-side pen-tests.

Client-side is a pentest subgenre

Automatic OSINT has already been included in professional pen-testing.

- Phishing and client apps exploitation led to the development of client-side pen-tests.
- Automated client-side pentests include open crawling of emails, deceiving mail content and crafted urls or files exploting vulns.

Client-side is a pentest subgenre

Automatic OSINT has already been included in professional pen-testing.

- Phishing and client apps exploitation led to the development of client-side pen-tests.
- Automated client-side pentests include open crawling of emails, deceiving mail content and crafted urls or files exploting vulns.
- Core Security Consulting Services kindly provided these stats:
  - 2007: 23% pentests, had CS.
  - 2008: 11% pentests, had CS.

Client-side is a pentest subgenre

Automatic OSINT has already been included in professional pen-testing.

- Phishing and client apps exploitation led to the development of client-side pen-tests.
- Automated client-side pentests include open crawling of emails, deceiving mail content and crafted urls or files exploting vulns.
- Core Security Consulting Services kindly provided these stats:
  - 2007: 23% pentests, had CS.
  - 2008: 11% pentests, had CS.
- So, involves OSINT, social engineering and binary exploits.

(FriendAds: check talk "Smartphones (in)security" at Ekoparty 2008, is about client-side binary exploits)

Related work

- Paterva is a small company, leaded by Raelof Temmingh, developing OSINT software called Maltego [1][2][3].

Related work

- Paterva is a small company, leaded by Raelof Temmingh, developing OSINT software called Maltego [1][2][3].
- They are focused on Transforms (e.g.: domain2emails) and a client-server architecture specification. In our case we called Expanders to something similar to *Transforms*.

Related work

- Paterva is a small company, leaded by Raelof Temmingh, developing OSINT software called Maltego [1][2][3].
- They are focused on Transforms (e.g.: domain2emails) and a client-server architecture specification. In our case we called Expanders to something similar to *Transforms*.
- They use a rich set of entities like Person, Domain, Host, etc.

Related work

- Paterva is a small company, leaded by Raelof Temmingh, developing OSINT software called Maltego [1][2][3].
- They are focused on Transforms (e.g.: domain2emails) and a client-server architecture specification. In our case we called Expanders to something similar to *Transforms*.
- They use a rich set of entities like Person, Domain, Host, etc.
- Also PyMaltego library is being developed by the grugq, including client/server capabilities.

What's new

- Multiple social/anything network system harvester and analyzer (Exomind):
  - (Almost) pluggable modules, Bots implementing Expanders.
  - Online fusion of people/node aliases, during crawling.

What's new

- Multiple social/anything network system harvester and analyzer (Exomind):
  - (Almost) pluggable modules, Bots implementing Expanders.
  - Online fusion of people/node aliases, during crawling.
- Online weighting of nodes and edges allowing filtering and more accurate results.

What's new

- Multiple social/anything network system harvester and analyzer (Exomind):
  - (Almost) pluggable modules, Bots implementing Expanders.
  - Online fusion of people/node aliases, during crawling.
- Online weighting of nodes and edges allowing filtering and more accurate results.
- Using the information gathered:
  - Vocabulary impersonation to tune your chatbots.
  - Programmatic framework for analysis algorithms and exporting to other frameworks.

What's new

- Multiple social/anything network system harvester and analyzer (Exomind):
  - (Almost) pluggable modules, Bots implementing Expanders.
  - Online fusion of people/node aliases, during crawling.
- Online weighting of nodes and edges allowing filtering and more accurate results.
- Using the information gathered:
  - Vocabulary impersonation to tune your chatbots.
  - Programmatic framework for analysis algorithms and exporting to other frameworks.
- Extra! Search Engine Based Chatbot.

Exomind design overview

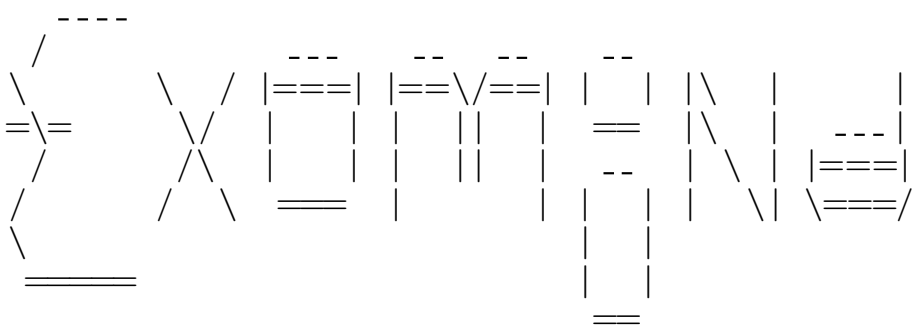
- `exomind` class: a Python command line interface.

Exomind design overview

- `exomind` class: a Python command line interface.
- `Exomind` class: a programmatic interface.

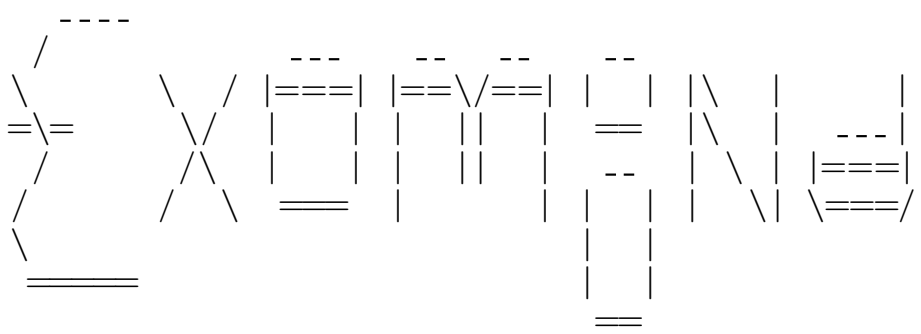
Exomind design overview

- `exomind` class: a Python command line interface.
- `Exomind` class: a programmatic interface.
- `SQLGraph`: a graph abstraction over MySQL, similar to that of the `networkx` library (but persistent).
  - Simple entity model: `node`, `edge`, `node_attr` and `edge_attr`. Also attrs. have `type` and `count`.



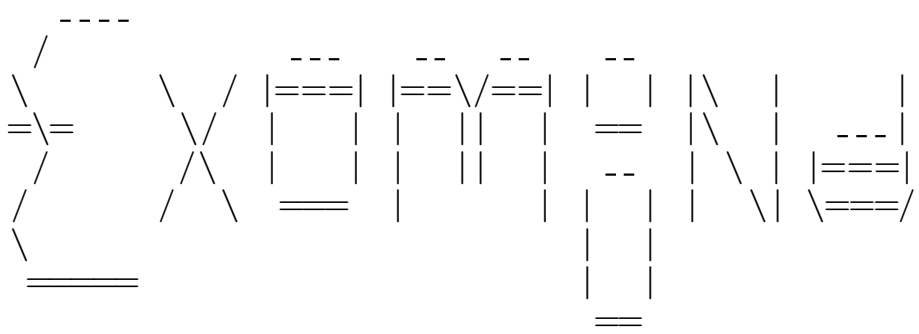
Exomind design overview (cont.)

- `BlackWidow`: the crawler, allows batch-only concurrency, BFS and DFS (using `SQLQueue`).



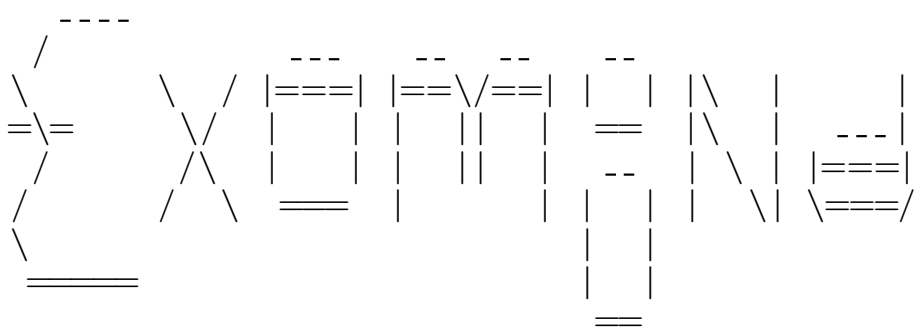
Exomind design overview (cont.)

- `BlackWidow`: the crawler, allows batch-only concurrency, BFS and DFS (using `SQLQueue`).
- `LambHerd`: initializes the bots.



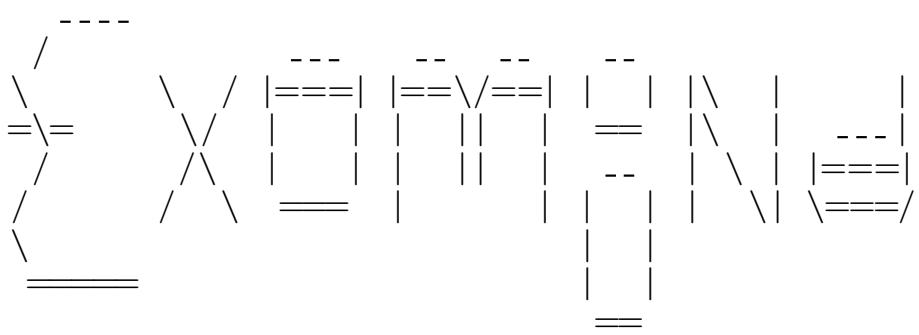
Exomind design overview (cont.)

- `BlackWidow`: the crawler, allows batch-only concurrency, BFS and DFS (using `SQLQueue`).
- `LambHerd`: initializes the bots.
- `mechanize` for web scripting [4], e.g. login .



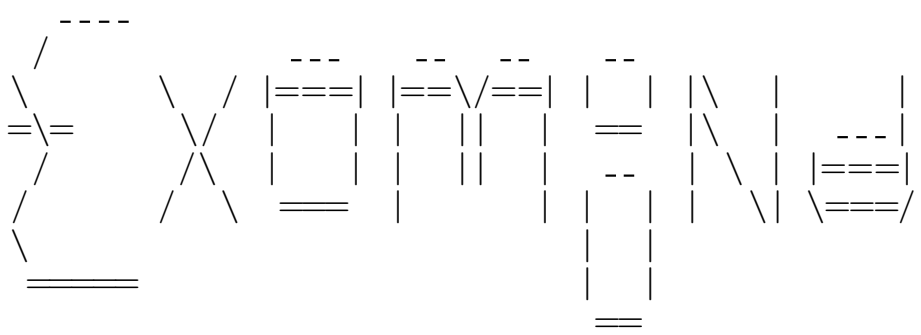
Exomind design overview (cont.)

- `BlackWidow`: the crawler, allows batch-only concurrency, BFS and DFS (using `SQLQueue`).
- `LambHerd`: initializes the bots.
- `mechanize` for web scripting [4], e.g. login .
- `pygraphviz+networkx` for visualization [5].



Exomind design overview (cont.)

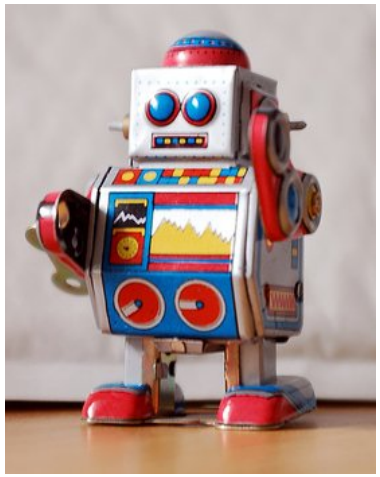
- `BlackWidow`: the crawler, allows batch-only concurrency, BFS and DFS (using `SQLQueue`).
- `LambHerd`: initializes the bots.
- `mechanize` for web scripting [4], e.g. login .
- `pygraphviz+networkx` for visualization [5].
- `msnlib` for IM chatbot example [6].



Bot Modules

Retrieve and format information but could also be interactive:

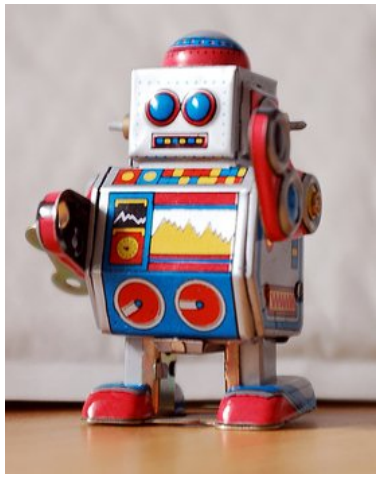
- Social/anything network services.



Bot Modules

Retrieve and format information but could also be interactive:

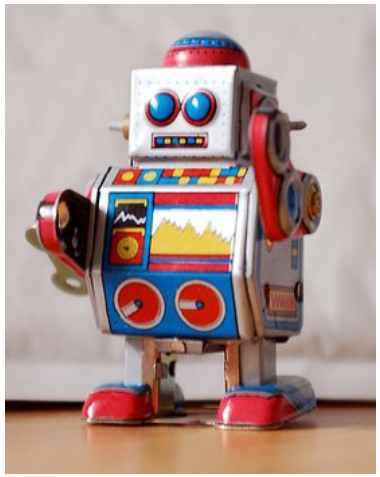
- Social/anything network services.
- Search engines general info.



Bot Modules

Retrieve and format information but could also be interactive:

- Social/anything network services.
- Search engines general info.
- Graph related algorithms to process existing data.



Bot Methods

Bots have (some) of the ability to implement:

- Expanders: return neighbors/contacts of a node (plus attributes).

Bot Methods

Bots have (some) of the ability to implement:

- Expanders: return neighbors/contacts of a node (plus attributes).
- Weigh Scale: measure some node or edge weight and filters.

Bot Methods

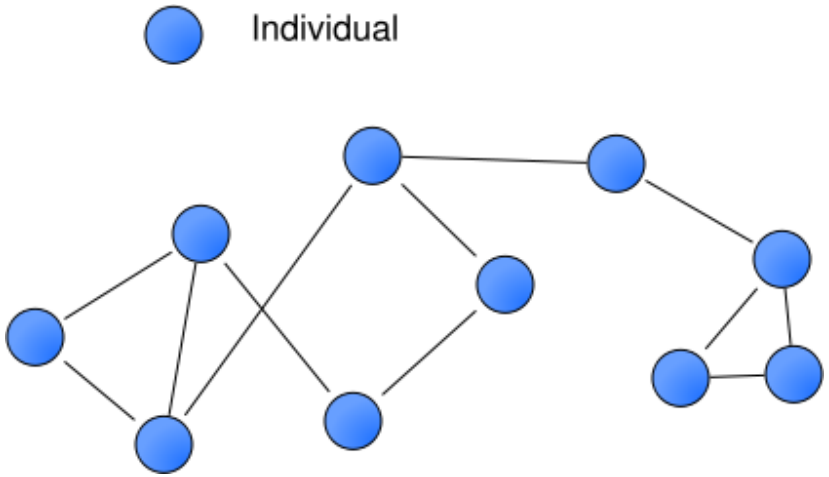
Bots have (some) of the ability to implement:

- Expanders: return neighbors/contacts of a node (plus attributes).
- Weigh Scale: measure some node or edge weight and filters.
- ChatBot commands: contact people and semi/fully automatically talk to them.

Expanders: Social Network Services

These bots include:

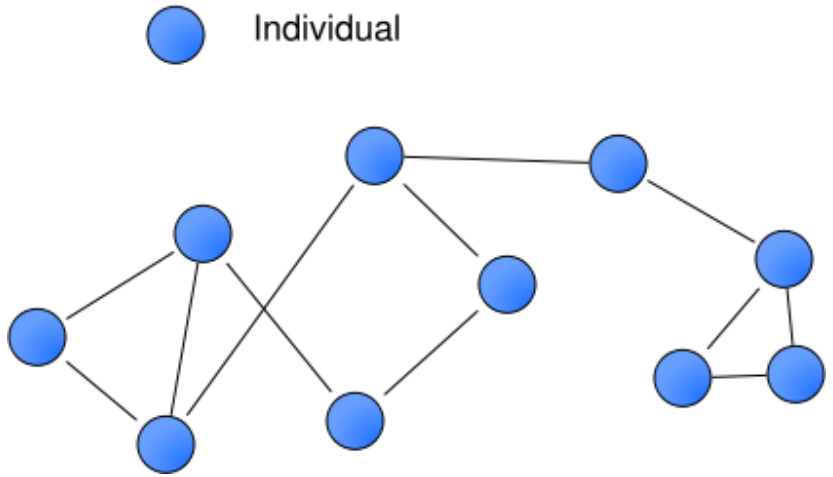
- Logic to detect the affiliation to the service.



Expanders: Social Network Services

These bots include:

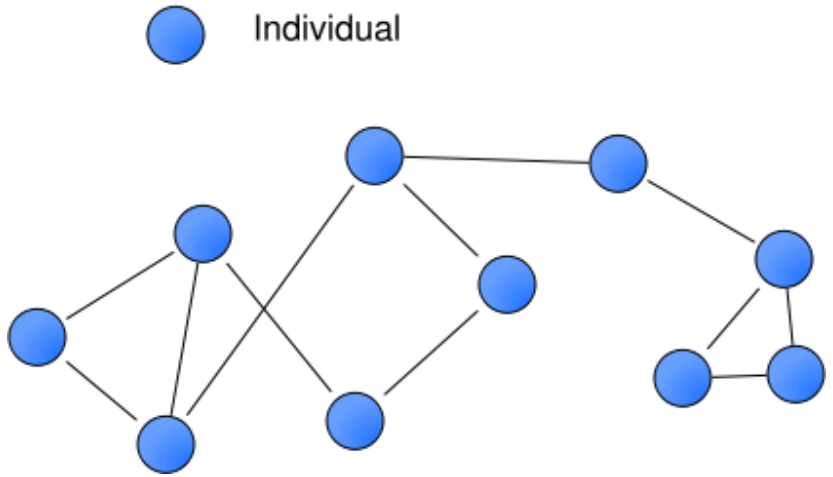
- Logic to detect the affiliation to the service.
- Logic to login using existing user/pass if necessary.



Expanders: Social Network Services

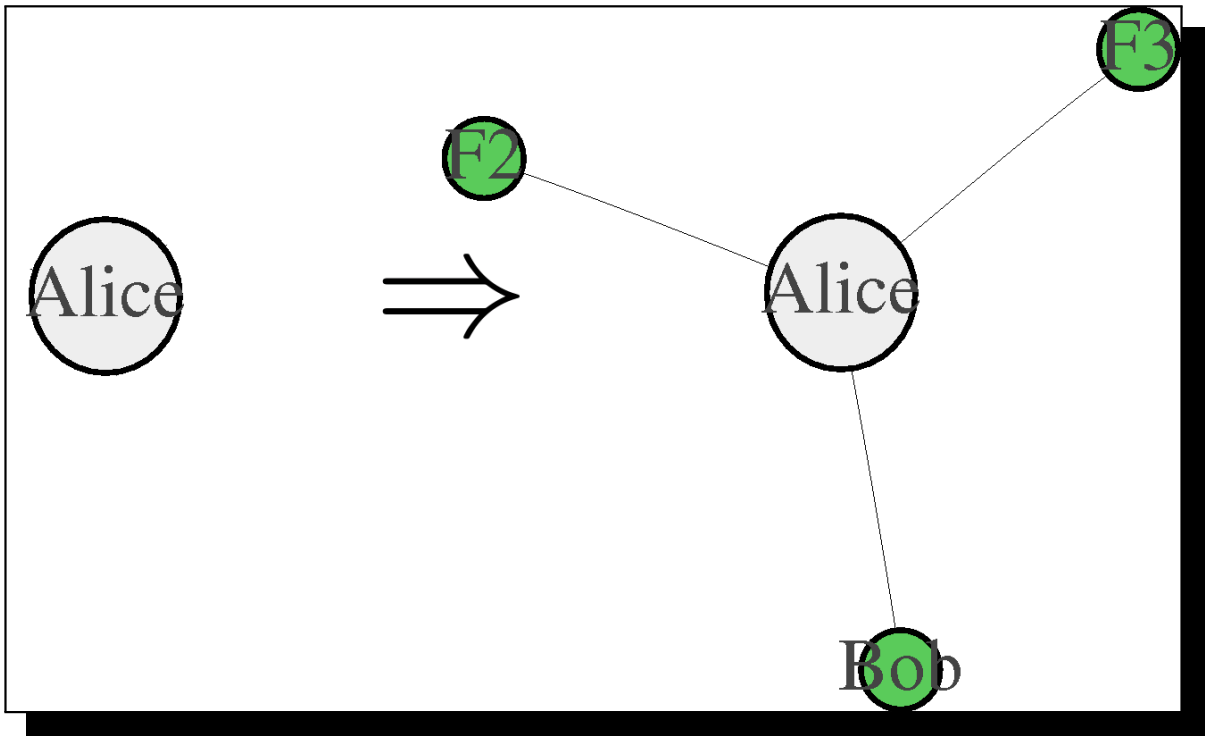
These bots include:

- Logic to detect the affiliation to the service.
- Logic to login using existing user/pass if necessary.
- Regex or html/xml parsing to locate the info.



Expander

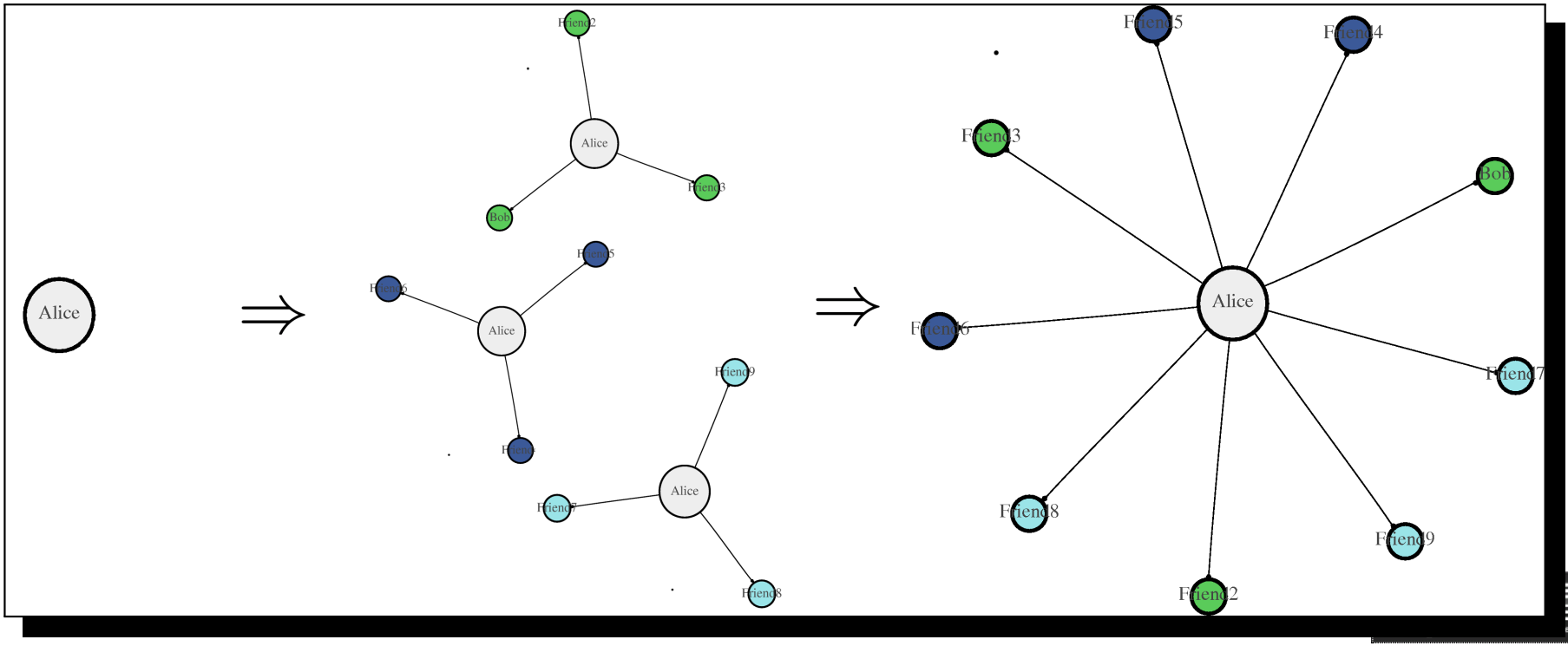
For example, for ViralBuddy service you can implement `ViralBuddyBot` that expands `Alice`:



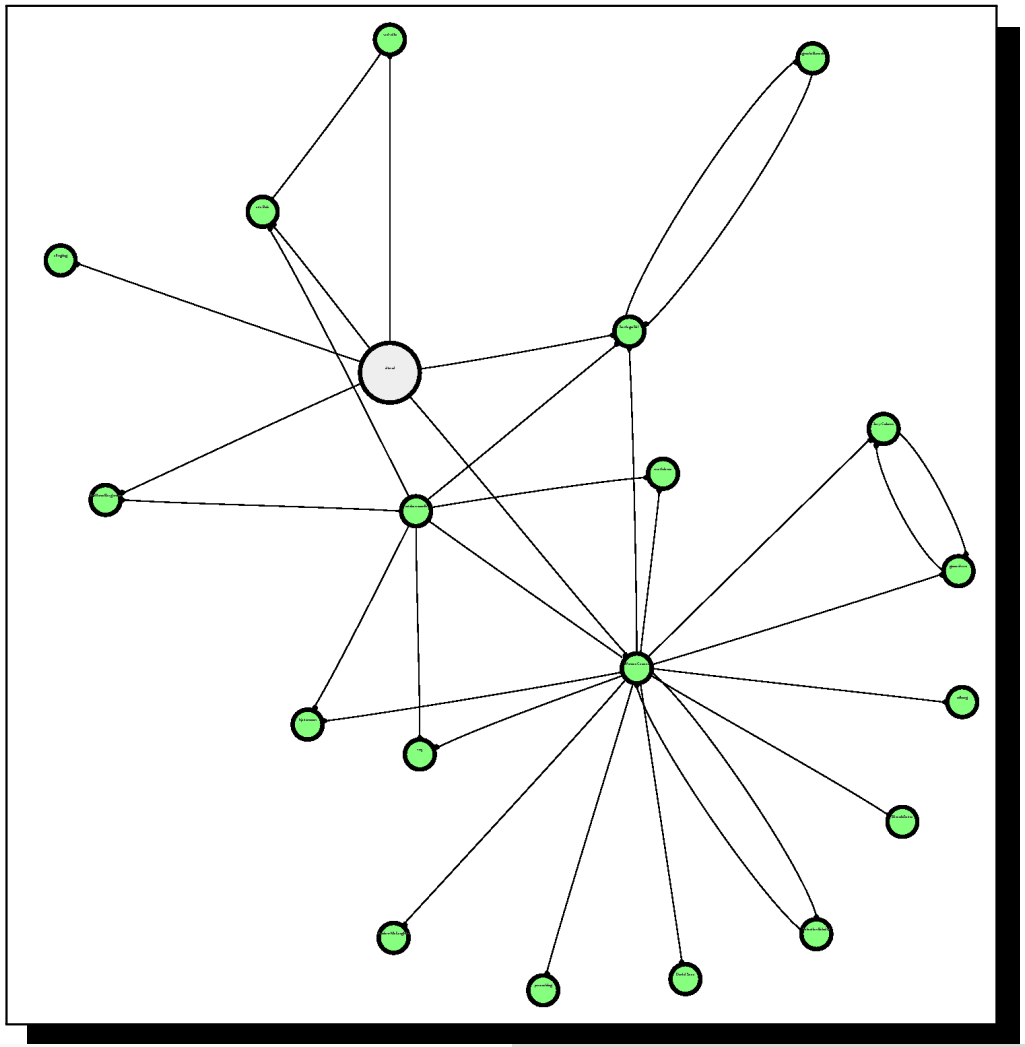
Plus extra attributes for `Alice`, each friend and each separate link.

Multiple Social Networks Fusion

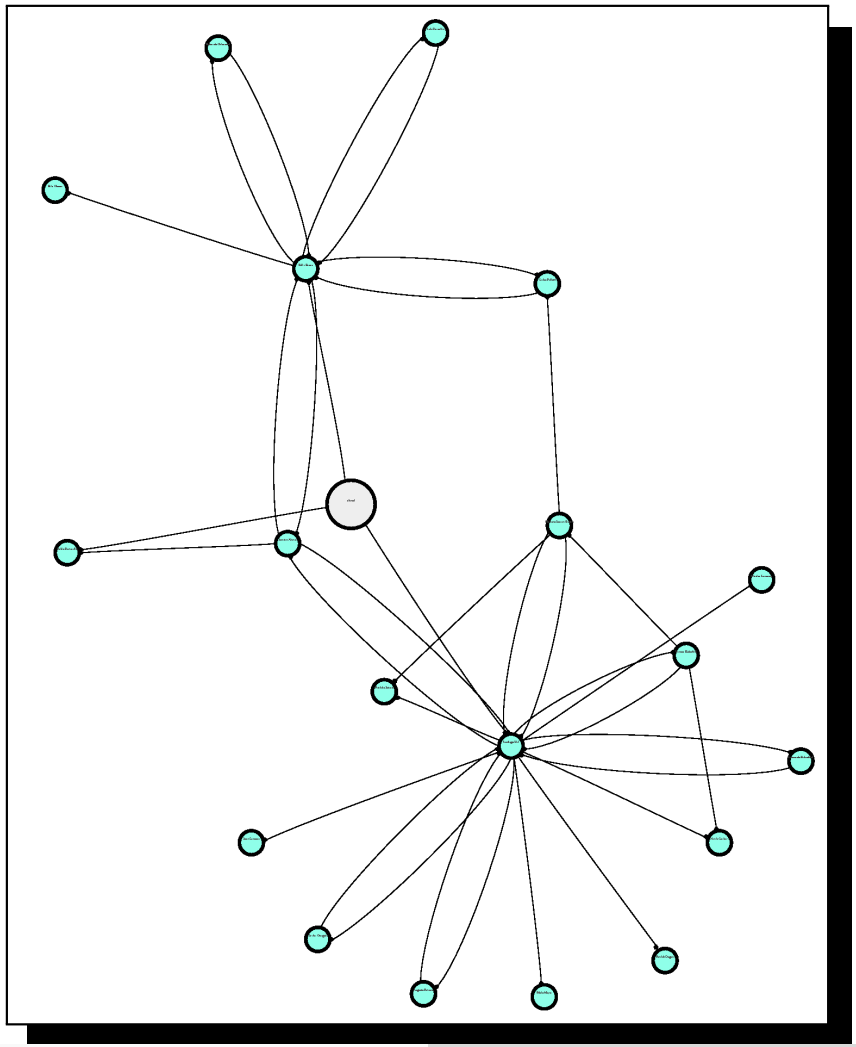
`Alice` contacts from various social networks can be merged together:



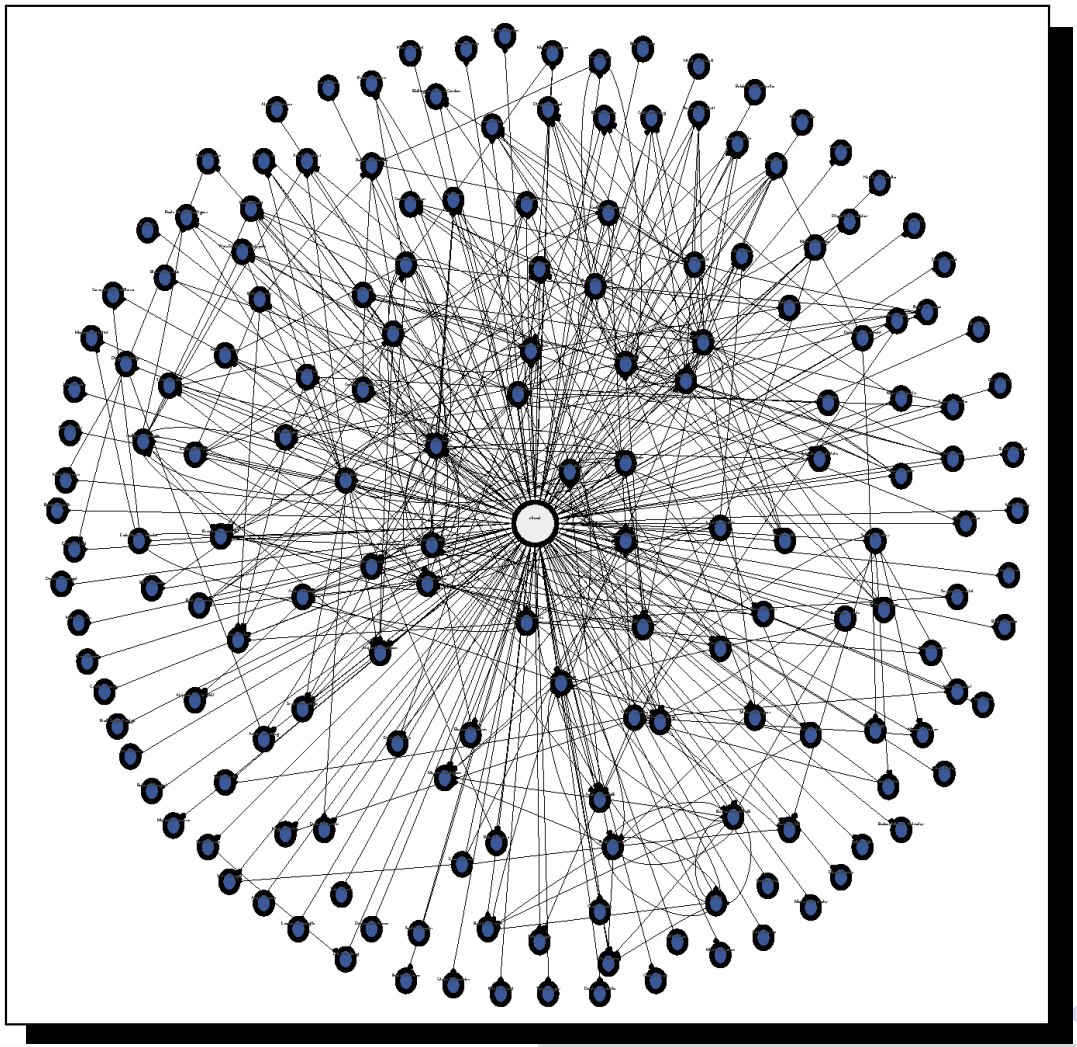
Twitter Following



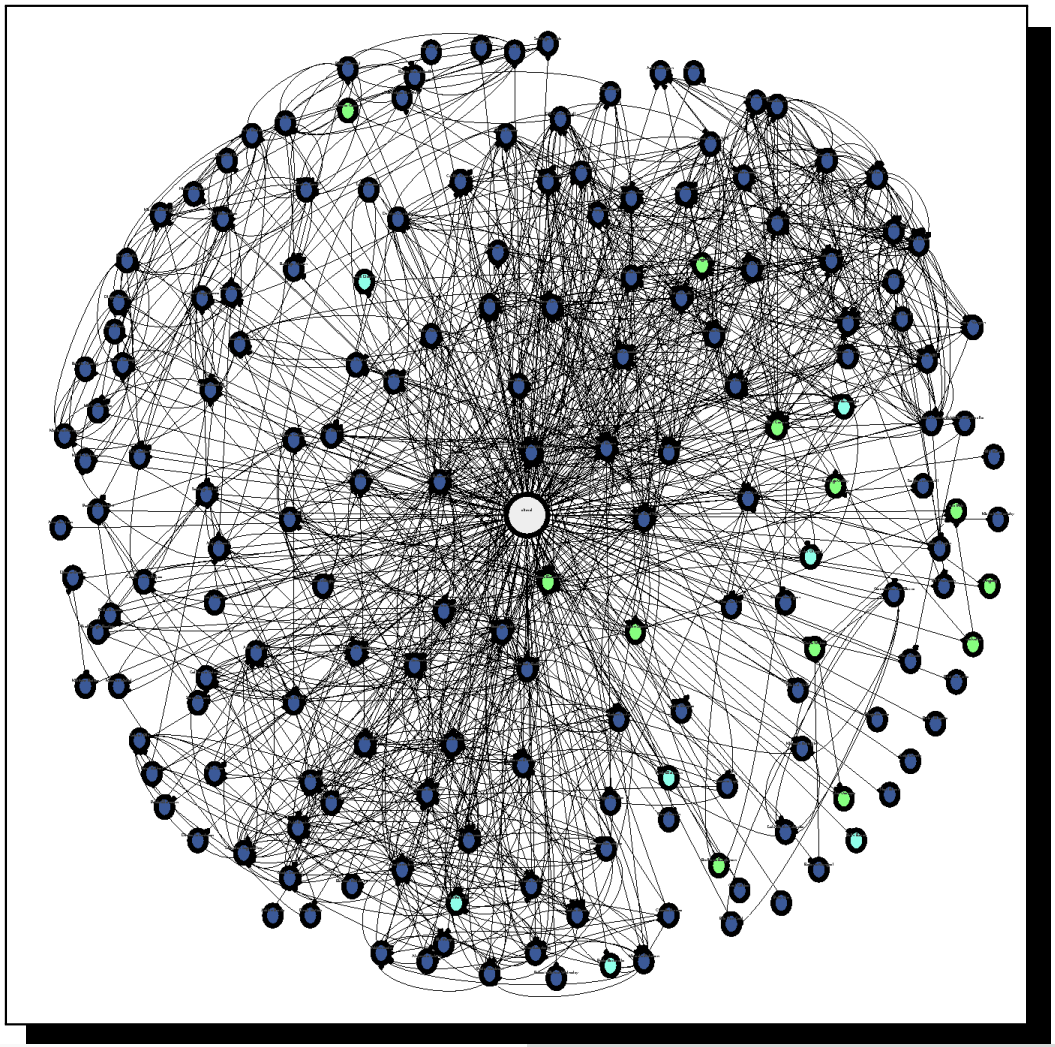
LinkedIn Recommendations



Facebook Friends



When crawled together...



Examples

```
>>> all_expanders()
LinkedInBot::recommendations
LinkedInBot::also_viewed
GraphBot::neighbors
GraphBot::with_all
SearchEngineBot::domain_to_emails
SearchEngineBot::name_to_self_emails
SearchEngineBot::name_to_emails_strong
SearchEngineBot::domain_to_emails_strong
SearchEngineBot::name_to_emails
SearchEngineBot::vocabulary
FacebookBot::friends
TwitterBot::following
TwitterBot::followers
TwitterBot::favorites
```

Recursive Regex Matching

General and unstructured information coming from search engine results (`SearchEngineBot`) can be retrieved and processed with specific purposes.

- Phones.
- Physical Addresses.
- E-mails (prototype).

```
comments. Main Content. English Wikipedia references for Coresecurity.com 1-10 of
10 ...From: CORE Security Technologies Advisories *advisories@coresecurity.com* ... Cor
Security Technologies Advisory http://www.coresecurity.com Axis Network ...Gerardo
Richarte and CoreLabs. *gera@coresecurity.com*. MD5 to be considered . ... \xa1Gracias!
gera (Gerardo Richarte) *gerardo.richarte@coresecurity.com* . ...or http:/
/oss.coresecurity.com/uhooker/release/1.2/uhooker_v1.2.zip (zip) checkout the doc pages
because Im constantly posting new stuff like sample scripts ...Advisory URL: http:/
/www.coresecurity.com/?action=item&amp;id=2035 ... http://www.coresecurity.com/files
/attachments/CORE-2007-1004-VLC-tutoria ...Browser Defender site report for coresecurity
```

Recursive Regex Matching (cont.)

To harvest specific data from search engines snippets or full pages your need powerful (*and a lot of!*) regular expression to link data.

- E-mail: firstlastname user@domain.com. E.g., if you have the user:

```
tokregex+'␣'+tokregex+'␣'+user+'␣at␣'+tokregex
```

Recursive Regex Matching (cont.)

To harvest specific data from search engines snippets or full pages your need powerful (*and a lot of!*) regular expression to link data.

- E-mail: firstlastname user@domain.com. E.g., if you have the user:

```
tokregex+'␣'+tokregex+'␣'+user+'␣at␣'+tokregex
```

- Many heuristics to post-process the regex matches.

```
Alice Smith a.....@bla.com -->
                              Alice Smith asmith@bla.com
_ _ alice.smith@bla.com -->
                              Alice Smith alice.smith@bla.com
```

Recursive Regex Matching (cont.)

To harvest specific data from search engines snippets or full pages your need powerful (*and a lot of!*) regular expression to link data.

- E-mail: firstlastname user@domain.com. E.g., if you have the user:

```
tokregex+'␣'+tokregex+'␣'+user+'␣at␣'+tokregex
```

- Many heuristics to post-process the regex matches.

```
Alice Smith a..... @bla.com -->
                              Alice Smith asmith@bla.com
_ _ alice.smith@bla.com -->
                              Alice Smith alice.smith@bla.com
```

- Iterate, if near assume related and define Expander!

Preventing Abuse and Bans

Stealthness is a main concern of professional pentesters, so to reduce the noise generated all bots include:

- `sleep_regular_secs`: sleep t seconds when desired.

Preventing Abuse and Bans

Stealthness is a main concern of professional pentesters, so to reduce the noise generated all bots include:

- `sleep_regular_secs`: sleep t seconds when desired.
- `sleep_random_bool`: sleep time between ops. is $2 * t * rand(0, 1)$ if enabled.

Preventing Abuse and Bans

Stealthness is a main concern of professional pentesters, so to reduce the noise generated all bots include:

- `sleep_regular_secs`: sleep t seconds when desired.
- `sleep_random_bool`: sleep time between ops. is $2 * t * rand(0, 1)$ if enabled.
- `sleep_module_gets`: sleep after this number of ops.

Preventing Abuse and Bans

Stealthness is a main concern of professional pentesters, so to reduce the noise generated all bots include:

- `sleep_regular_secs`: sleep t seconds when desired.
- `sleep_random_bool`: sleep time between ops. is $2 * t * rand(0, 1)$ if enabled.
- `sleep_module_gets`: sleep after this number of ops.
- Famous line `Press Enter to continue:` can be included in your bots to incorporate human interaction (and slow down!).

Demo1: targeted email crawling!

Weighting and filtering

We have a lot of info, but the information is too weak:

- Common names or nicks pollute the info from the cloud, e.g. Paul and John .

Weighting and filtering

We have a lot of info, but the information is too weak:

- Common names or nicks pollute the info from the cloud, e.g. Paul and John .
- Little/no context: the information is ambiguous.

Weighting and filtering

We have a lot of info, but the information is too weak:

- Common names or nicks pollute the info from the cloud, e.g. Paul and John .
- Little/no context: the information is ambiguous.
- Detecting aliases and duplicates is a complex problem.

Weighting and filtering

We have a lot of info, but the information is too weak:

- Common names or nicks pollute the info from the cloud, e.g. Paul and John .
- Little/no context: the information is ambiguous.
- Detecting aliases and duplicates is a complex problem.
- But at least we can filter, only precise information will be allowed.

Weighting and filtering

We have a lot of info, but the information is too weak:

- Common names or nicks pollute the info from the cloud, e.g. Paul and John .
- Little/no context: the information is ambiguous.
- Detecting aliases and duplicates is a complex problem.
- But at least we can filter, only precise information will be allowed.
- How? Use a universal reference corpus, some big search engine!

Search Engine Hits

If I am impersonating `Alice Smith`, is contact `Bob Johnson` a good target?

- If `Bob` has a very common name: NO.
- If `Bob` is not very related to `Alice`: NO.

$$h(\texttt{alice smith}) = 289,000$$

$$h(\texttt{bob johnson}) = 861,000$$

$$h(\texttt{alice smith,bob johnson}) = 825$$

Search Engine Hits (cont.)

Normalize the # of hits (h) and transform them into entropy (!):

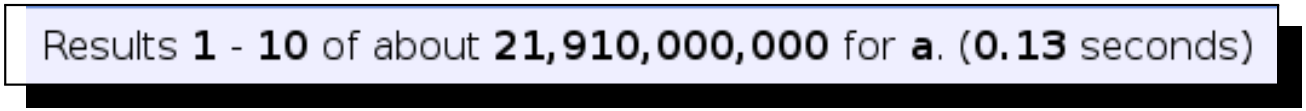
- Total pages: $M = 21,910,000,000$

Results **1** - **10** of about **21,910,000,000** for **a**. (**0.13** seconds)

Search Engine Hits (cont.)

Normalize the # of hits (h) and transform them into entropy (!):

- Total pages: $M = 21,910,000,000$

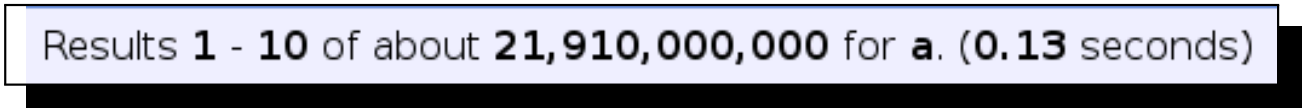


- Max. entropy (1 hit): $-\log_2 M \simeq 34.35$ bits

Search Engine Hits (cont.)

Normalize the # of hits (h) and transform them into entropy (!):

- Total pages: $M = 21,910,000,000$

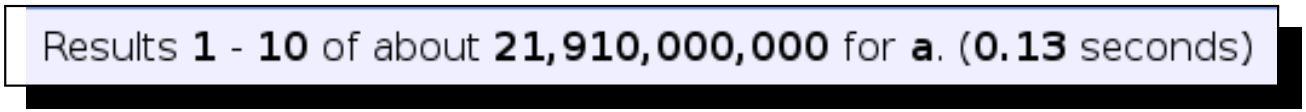


- Max. entropy (1 hit): $-\log_2 M \simeq 34.35$ bits
- Bob's fraction: $\frac{861{,}000}{21{,}910{,}000{,}000} \simeq 3.929 \cdot 10^{-5}$

Search Engine Hits (cont.)

Normalize the # of hits (h) and transform them into entropy (!):

- Total pages: $M = 21,910,000,000$

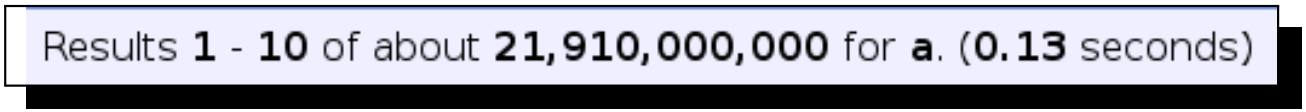


- Max. entropy (1 hit): $-\log_2 M \simeq 34.35$ bits
- Bob's fraction: $\frac{861{,}000}{21{,}910{,}000{,}000} \simeq 3.929 \cdot 10^{-5}$
- Entropy: $-\log_2(3.929 \cdot 10^{-5}) \simeq 14.635$ bits

Search Engine Hits (cont.)

Normalize the # of hits (h) and transform them into entropy (!):

- Total pages: $M = 21,910,000,000$

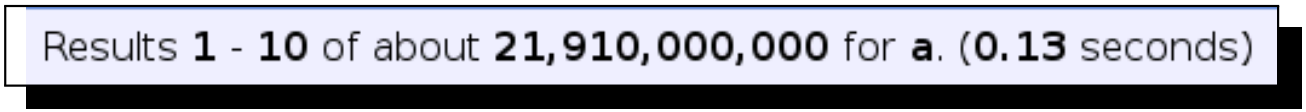


- Max. entropy (1 hit): $-\log_2 M \simeq 34.35$ bits
- Bob's fraction: $\frac{861{,}000}{21{,}910{,}000{,}000} \simeq 3.929 \cdot 10^{-5}$
- Entropy: $-\log_2(3.929 \cdot 10^{-5}) \simeq 14.635$ bits
- Normalized (betw. 0 and 1): $\frac{34.35-14.635}{34.35} \simeq 0.573$

Search Engine Hits (cont.)

Normalize the # of hits (h) and transform them into entropy (!):

- Total pages: $M = 21,910,000,000$



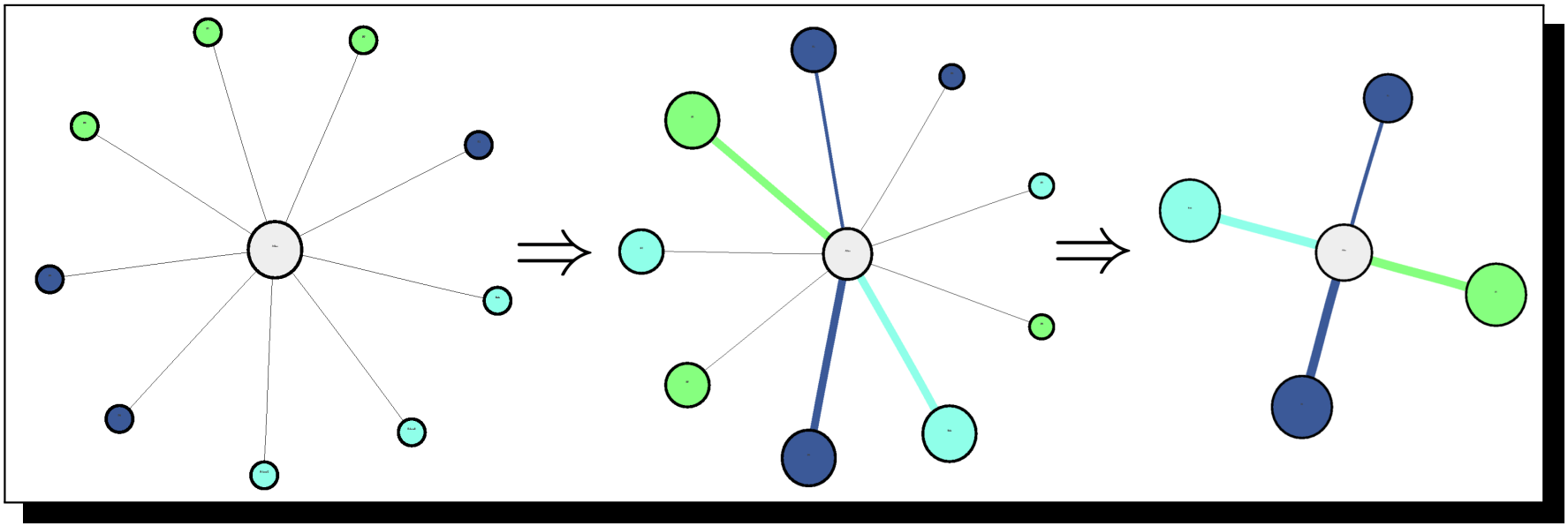
- Max. entropy (1 hit): $-\log_2 M \simeq 34.35$ bits
- Bob's fraction: $\frac{861,000}{21,910,000,000} \simeq 3.929 \cdot 10^{-5}$
- Entropy: $-\log_2(3.929 \cdot 10^{-5}) \simeq 14.635$ bits
- Normalized (betw. 0 and 1): $\frac{34.35-14.635}{34.35} \simeq 0.573$
- If for example 0.5 is a chosen threshold, Bob is dropped!

Normalized Google Distance

Same with links. Hits $h(x)$ and $h(x,y)$ are more useful if normalized (where $p(x) = \frac{h(x)}{M}$ and $p(x|y) = \frac{p(x,y)}{p(y)}$):

- Jaccard distance: $\frac{h(x,y)}{h(x)+h(y)-h(x,y)}$
- *NGD* is more robust! [7]:
$$\frac{\max\{-\log_2 p(x|y), -\log_2 p(y|x)\}}{\max\{-\log_2 p(x), -\log_2 p(y)\}} = \frac{\max\{\log_2 h(x), \log_2 h(y)\} - \log_2 h(x,y)}{\max\{-\log_2 h(x), -\log_2 h(y)\}}$$
- $NGD(\textit{Alice Smith}, \textit{Bob Johnson}) \simeq 0.545$ if, for example, 0.5 is our chosen threshold, the link to Bob is dropped, he's is not a good target!

Weighting and filtering result



Observation: A context can be set to restrict the corpus! E.g. `security`, `argentina`. etc.

Weigh Scales in Exomind

```
>>> all_weigh_scales()
SearchEngineBot::jaccard_distance
SearchEngineBot::se_hits
SearchEngineBot::normalized_se_entropy
SearchEngineBot::normalized_se_distance
SearchEngineBot::hits_distance
GraphBot::obfuscate
>>>
```

Weigh scales can apply transformations on node neighborhoods, for example `GraphBot::obfuscate` encrypts the data collected. Also to blindly detect possible leaked links...

Demo2: Network blind reconstruction! Brute-forcing each possible link!

Using the Data Collected

- Any kind of information available can be collected. Just add new `Attributes` and extract it in old or new expanders.

Using the Data Collected

- Any kind of information available can be collected. Just add new `Attributes` and extract it in old or new expanders.
- A special attribute called `TAG` was devised to build word bags defining fuzzy profiles for persons.

Using the Data Collected

- Any kind of information available can be collected. Just add new `Attributes` and extract it in old or new expanders.
- A special attribute called `TAG` was devised to build word bags defining fuzzy profiles for persons.


```
add_node_attr('Bob_Johnson', Attributes.TAG, 'best')
add_node_attr('Bob_Johnson', Attributes.TAG, 'reverse')
add_node_attr('Bob_Johnson', Attributes.TAG, 'engineer')
add_node_attr('Bob_Johnson', Attributes.TAG, 'security')
```

Vocabulary Impersonation

If the fuzzy profile is centered on written word of the target node, we are collecting his/her vocabulary. Use common patterns:

- `...Alice said...`,
- `...posted by Alice...`,
- `...Alice wrote...`.

```
list: [('co-founder', 6L), ('sets', 7L), ('technical', 8L), ('direction', 6L),
('company', 17L), ('responsible', 5L), ('overseeing', 5L), ('development', 3L),
('managed', 4L), ('catch', 4L), ('chief', 43L), ('techology', 3L), ('officer', 40L),
('talk', 4L), ('new', 8L), ('class', 3L), ('vulnerability', 15L), ('thats', 2L),
('talks', 2L), ('recent', 3L), ('update', 2L), ('interview', 6L), ('said', 98L), ('view',
5L), ('information', 4L), ('provided', 3L), ('free', 4L), ('public', 8L), ('search', 5L),
('listing', 4L), ('friends', 5L), ('photos', 6L), ('videos', 6L), ('join', 5L),
('attacker', 2L), ('remotely', 1L), ('execute', 1L), ('code', 12L), ('exploit', 9L),
('bugs', 10L), ('user', 4L), ('interaction', 3L), ('releasing', 1L), ('customers', 1L),
('week', 7L), ('technology', 39L), ('flaw', 9L), ('demonstrates', 1L), ('writes', 1L),
('discusses', 1L), ('security', 37L), ('threats', 2L), ('tools', 4L), ('importance', 1L),
('presentation', 3L), ('titled', 2L), ('esteemed', 2L), ('reply',...
```

Vocabulary Impersonation (cont.)

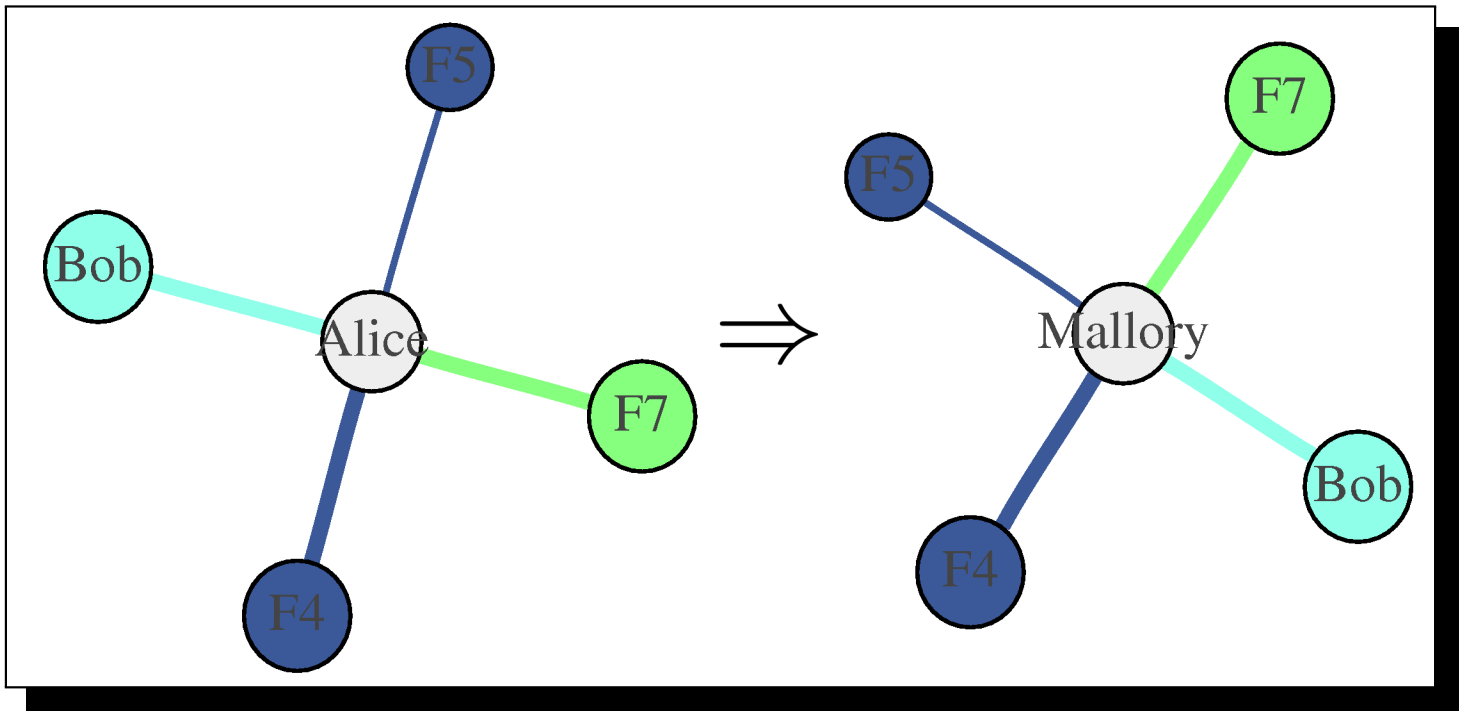
(Previous fun with languages at [8]).

- Module called *Dino*, based on a thesaurus and lacking ambiguity checks at this moment ;(, NLP hackers needed.

Bob: tell me
(Fake Alice: *tell* me that)
After vocabulary translation!
Fake Alice: state me that
Bob: tell me
(Fake Alice: *do* your friends *put* money in your *pocket*?)
After!
Fake Alice: execute your friends risk money in your collect?

Social network infiltration

`Mallory` wants to automatically impersonate `Alice` and interact with her contact `Bob`, possibly leveraging stolen credentials. Any channel available.



Social network infiltration (cont.)

In `Exomind` is included a MSN chatbot infiltration:

- `Mallory` uses `Alice` nickname.
- Possible MSN users a retrieved from `Alice` contacts collected.
- Initial messages send every N seconds.
- Manual or automatic responses using the Movie Chatbot or Eliza, and the vocabulary collected.

Search Engine Chatbot

What would Google (Search) say it it can talk? Question Answering systems exists [9] but we want an informal chat. Movie scripts: more or less structured format [10].

- Query `chatline` or `chatline[:-1]` until answers are found.

(You can use Google Translate if you want to change the language.)

Search Engine Chatbot

What would Google (Search) say it it can talk? Question Answering systems exists [9] but we want an informal chat. Movie scripts: more or less structured format [10].

- Query `chatline` or `chatline[:-1]` until answers are found.
- Extract all possible answers from the snippets.

(You can use Google Translate if you want to change the language.)

Search Engine Chatbot

What would Google (Search) say it it can talk? Question Answering systems exists [9] but we want an informal chat. Movie scripts: more or less structured format [10].

- Query `chatline` or `chatline[:-1]` until answers are found.
- Extract all possible answers from the snippets.
- Choose a random subset, say 20.

(You can use Google Translate if you want to change the language.)

Search Engine Chatbot

What would Google (Search) say it it can talk? Question Answering systems exists [9] but we want an informal chat. Movie scripts: more or less structured format [10].

- Query `chatline` or `chatline[:-1]` until answers are found.
- Extract all possible answers from the snippets.
- Choose a random subset, say 20.
- (optional!) Measure the Google Distance and choose the best 5 answers.

(You can use Google Translate if you want to change the language.)

Search Engine Chatbot

What would Google (Search) say it it can talk? Question Answering systems exists [9] but we want an informal chat. Movie scripts: more or less structured format [10].

- Query `chatline` or `chatline[:-1]` until answers are found.
- Extract all possible answers from the snippets.
- Choose a random subset, say 20.
- (optional!) Measure the Google Distance and choose the best 5 answers.
- Choose the most used in the subset.

(You can use Google Translate if you want to change the language.)

Search Engine Chatbot

What would Google (Search) say it it can talk? Question Answering systems exists [9] but we want an informal chat. Movie scripts: more or less structured format [10].

- Query `chatline` or `chatline[:-1]` until answers are found.
- Extract all possible answers from the snippets.
- Choose a random subset, say 20.
- (optional!) Measure the Google Distance and choose the best 5 answers.
- Choose the most used in the subset.
- Voilà!

(You can use Google Translate if you want to change the language.)

Demo3: Search Engine Movie Chatbot!

Conclusion/Countermeasures

- It's a small world.
- Don't use your real name on the web (in the attention economy information is currency, don't give it away!).
- Don't accept strangers on your social network (in the social economy contacts are currency, don't give away your friends or yourself!).
- Use internal messaging for private issues of your organization.
- Analyze your social networks with `Exomind`! [11]

Final Questions?

Thanks!

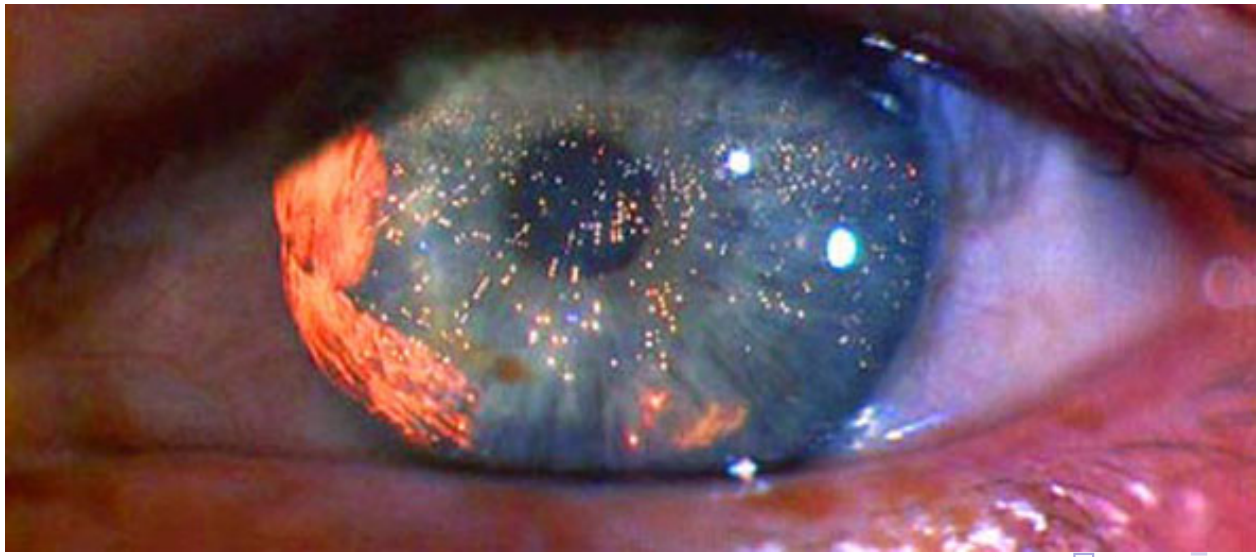
Aure: webapps and python discussions.

Alfred: LaTeX template.

Beto: Consulting stats.

ITBA team: scientific environment.

Core team: food, shelter and creative hacking environment.



References

1 www.paterva.com/maltego

2 www.first.org/conference/2007/papers/temmingh-roelof-slides.pdf

3 www.blackhat.com/presentations/bh-europe-08/Temmingh-Bohme/Presentation/bh-eu-08-temmingh-bohme.pdf

4 http://wwwsearch.sourceforge.net/mechanize/

5 https://networkx.lanl.gov/wiki/pygraphviz/

6 http://blitiri.com.ar/p/msnlib/

7 R. Cilibrasi, P.M.B. Vitanyi, Automatic meaning discovery using Google. http://xxx.lanl.gov/abs/cs.CV/0312044 (2004)

8 Look Who's Translating: Impersonations, Chinese Whispers and Fun with Machine Translation on the Internet www.mt-archive.info/EAMT-2006-Gaspari.pdf

9 Quarteroni et. al, A Chatbot-based Interactive Question Answering System

10 www.imsdb.com

11 Exomind tool, downloads soon at corelabs.coresecurity.com