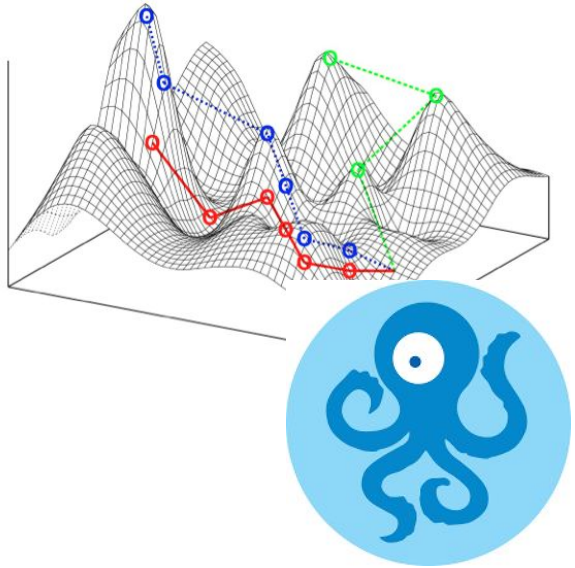


VPN Censorship, Probes and Resilience.

Building feedback loops with Open Data.

Ain Ghazal & OONI Team, July 2022.

你好!

I am Ain Ghazal.

I have questions about censorship.

github: **@ainghazal**

ICFP Fellow @OONI.

All mistakes are my own!

1. **Let’s measure VPNs.**
2. **Tools and roadmap.**
3. **Difficult questions.**

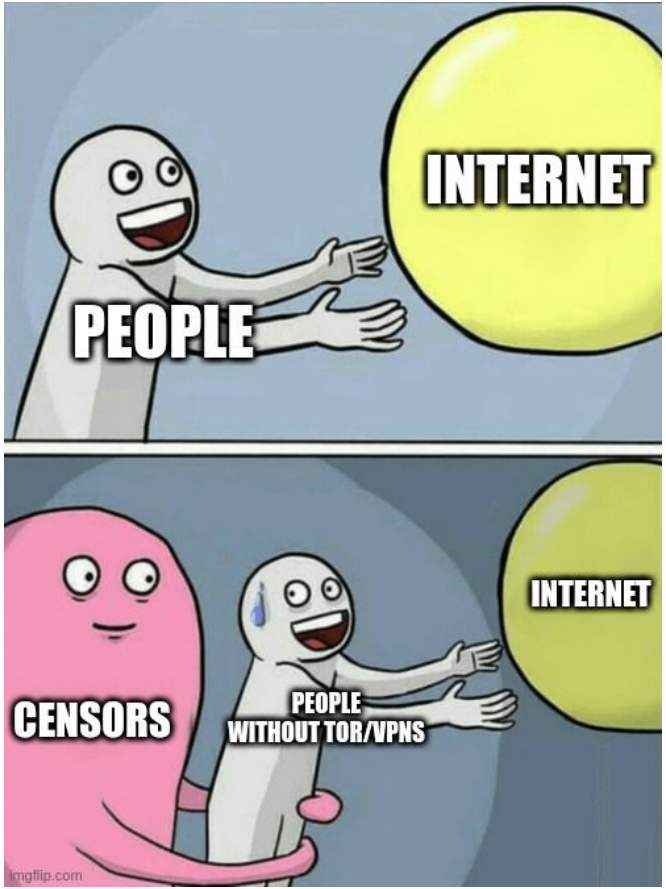
1.

Let’s measure VPN traffic!

What to measure, and why.

A (biased) history of the filternet

- 1990s: let's connect computers!
- 2000s: let's sell `$stuff` online!
- 2010s: let's shut it down!



Geofencing and IP reputation.

Connectivity is a human right, but evasion is not so simple anymore, since the “filternet mainland” changes under our feet.


Select all squares with
traffic lights
If there are none, click skip

Timnit Gebru 
@timnitGebru

Makes sense that ML community's techbroness+white saviorism along with "model said so with superhuman accuracy" with zero reflection & data analysis helps make this ideology rampant & influential in "AI."

Reading this piece reminded me how much they waste our time with this shit

11:15 PM · Jul 11, 2022 · Twitter Web App

3 Retweets **43** Likes

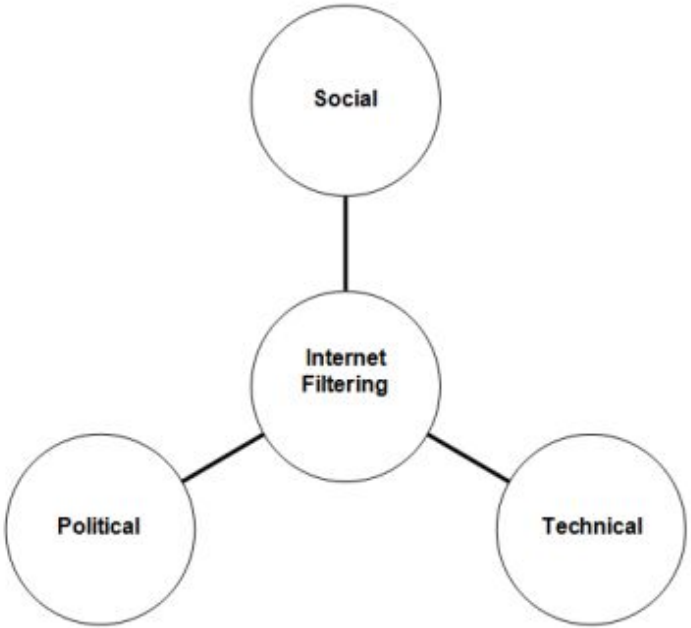
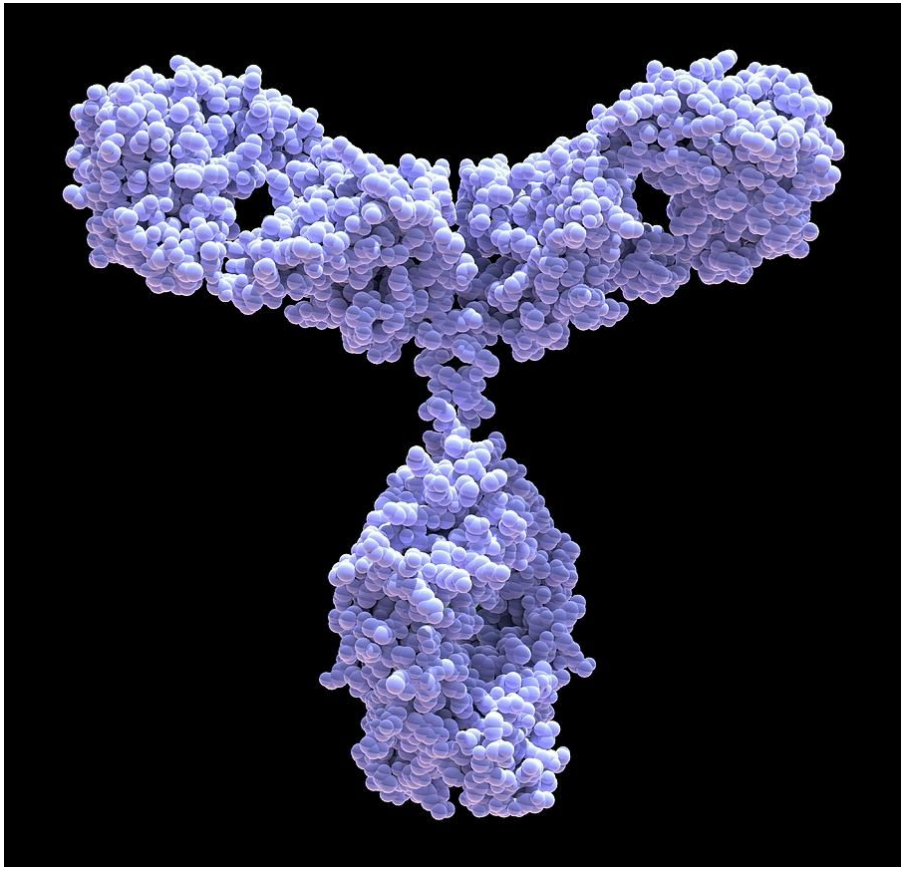


Figure 2. Three Aspects of Internet Censorship

Connectivity is a human right, ~~but evasion is not so simple anymore, since the "filternet mainland" changes under our feet.~~

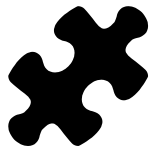
Does the internet have an immune system?

Meta question for later, but perhaps something more than a simple metaphor...



Research questions:

What do we know about VPNs in the wild?



Being banned.

$governments are increasingly pushing for disconnecting regional portions of the internet. They target VPNs / Tor.

Easy to pick with DPI.

There's a trend in sophisticated classifiers. But are they being used in the wild? (real rules are likely quite simple).

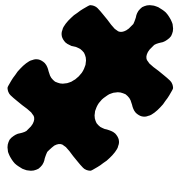
Are they throttled?

Difficult to say. We need a significant baseline before start answering that.

```
14
15          "github.com/google/gopacket"
16          "github.com/google/gopacket/layers"
17          "github.com/google/gopacket/pcap"
18
19          dissector "github.com/ainghazal/tls-dissector"
20  )
21
22  var (
23          magic        = []byte{0x16, 0x03, 0x01}
24          clientHellos int
25          doText       bool
26  )
27
```

Let’s find out!

Can we do better?



MEASURE.

We need empirical evidence, as precise and real-time as possible.

UNDERSTAND.

Make sense of the data.

Targeted experiments to validate hypotheses.

ANTICIPATE.

Keep running ahead, or better, “dancing with the censors.”

(better read to the tune by nouvelle vague.)

How OONI can help.

PROBES.

World-wide network of volunteers, **probing** real-life networks.

COLLABORATION.

Try to **understand** how censor capabilities evolve with time.

DATA ENDPOINTS.

The less noisy you are, the more likely you go under the radar?

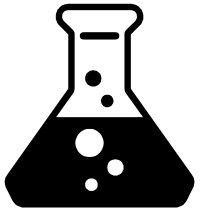
2.

Nel mezzo del cammin...

Tools, partnerships, and a map of the road ahead.

I am in the ~middle of my fellowship.

Two main protocols of interest.



OpenVPN.

Has been there for a while.

2.6 + dco might renew interest.

https://github.com/ooni/minivpn

Wireguard.

The new cool thing. (And it’s really cool!).

https://github.com/WireGuard/wireguard-go

minivpn

A minimalistic implementation of the OpenVPN protocol in Go (client only).

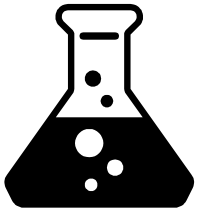
This implementation is intended for research purposes only. It has serious flaws, so please do **not** use it for any real-life situation where you need to trust it with user data.

This is not a working implementation with all the properties that you need from software that can effectively protect your privacy. If you arrived here looking for such a thing, please use misteriumnetwork/go-openvpn instead.

License

```
SPDX-License-Identifier: GPL-3.0-or-later
```

What's in a probe?



(pre) Is Provider API reachable?

VPN Handshake

- (tls) handshake (*****)
- {openvpn,wg}-handshake

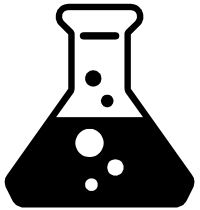
Data exchange

- ICMP ping.
- Webpage fetch

ndt7 run:

Several metrics about maximum network performance.

stretch goal: multiple streams.



Vanilla and beyond.

Obfs4.

Was broken. Not so much?

Also comes in **flavors (*).**

Other obfuscation protocols:

- Userspace wg obfuscation.
- *Shadowsocks?*
- *Hysteria?*

*** This is a quite interesting topic.**

Obfuscation

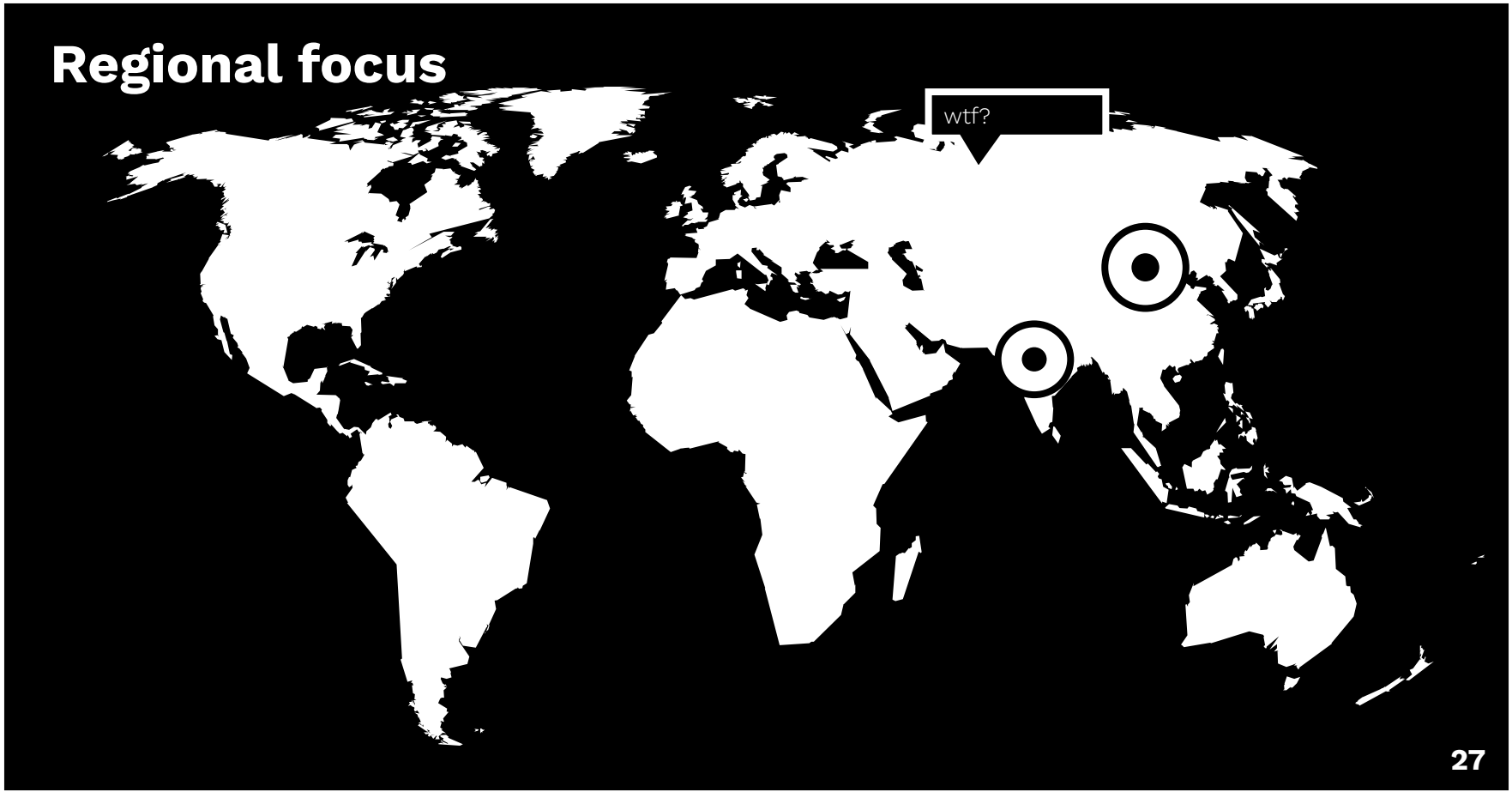
`obfs4` is supported. Add an additional entry in the config file, in this format:

```
proxy-obfs4 obfs4://RHOST:RPORT?cert=BASE64ENCODED_CERT&iat-mode=0
```

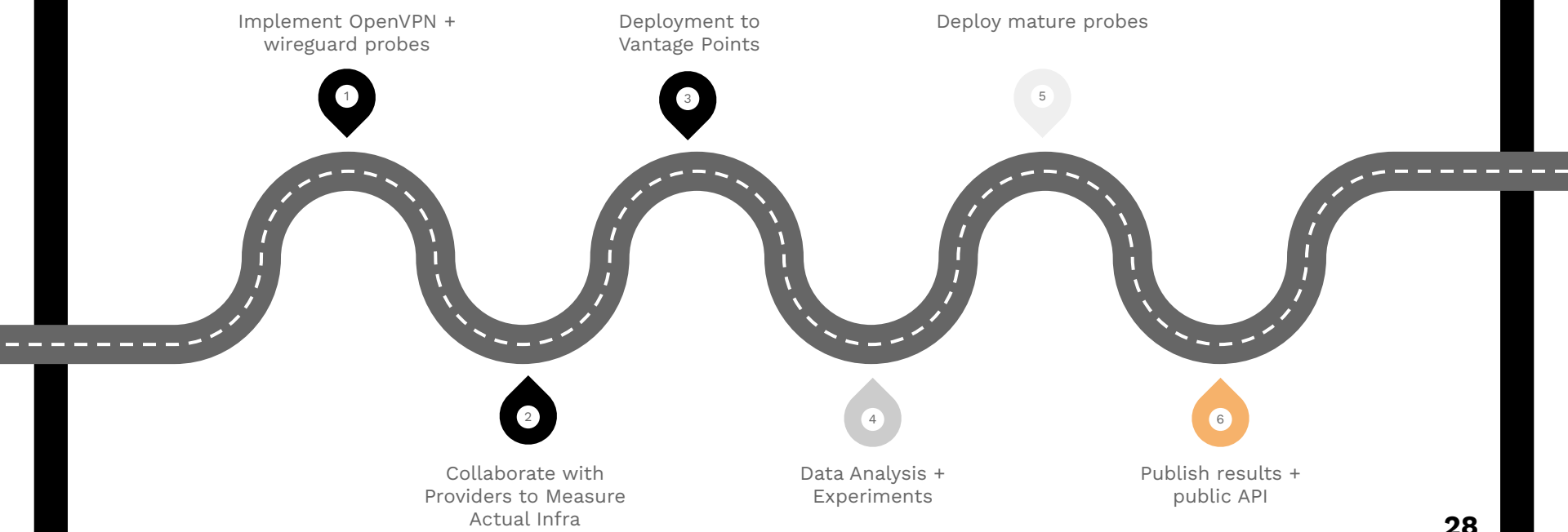
PLUGGABLE TRANSPORTS

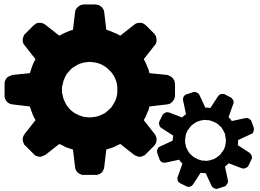
SOMEONE SAID FLAVORS?

Regional focus



Roadmap


Implement OpenVPN + wireguard probes
1
Collaborate with Providers to Measure Actual Infra
2
Deployment to Vantage Points
3
Data Analysis + Experiments
4
Deploy mature probes
5
Publish results + public API
6



Expected flow

Ongoing partnerships.

3.

Difficult questions.

In case the “easy” ones were not hard enough... :)

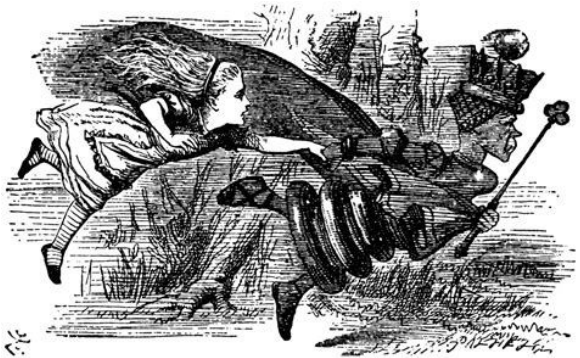
Are we measuring what we think we're measuring?

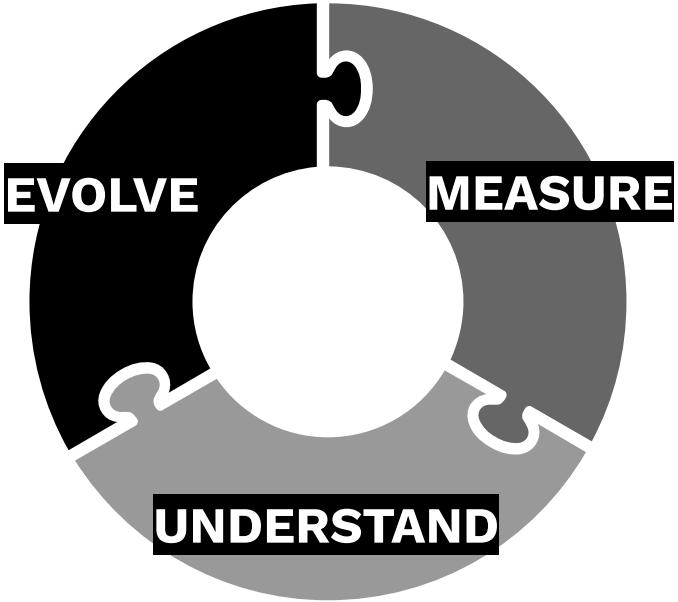
Apologies to any epistemologists (or ichthyologists) in the room.

Closed feedback loops. But how?

Observability vs. exposure.

Open Data and dynamic collaboration as the only antidote against powerful adversaries.

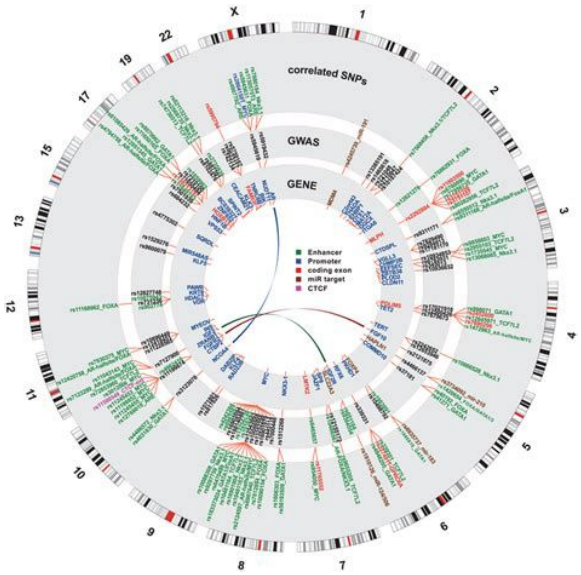



EVOLVE
MEASURE
UNDERSTAND

How to improve annotations?

Look at genomic pipelines.

How to add semantic layers to raw data from heterogeneous sources?



What can we learn from theoretical frameworks?

Optional, probably boring, epistemic detour.

Are the intertubes in an arms race?

DE GRUYTER OPEN

Proceedings on Privacy Enhancing Technologies ; 2016 (4):83–101

Tariq Elahi*, John A. Doucette, Hadi Hosseini, Steven J. Murdoch, and Ian Goldberg

A Framework for the Game-theoretic Analysis of Censorship Resistance

If the game-theoretic approach is a accurate model of the reality...

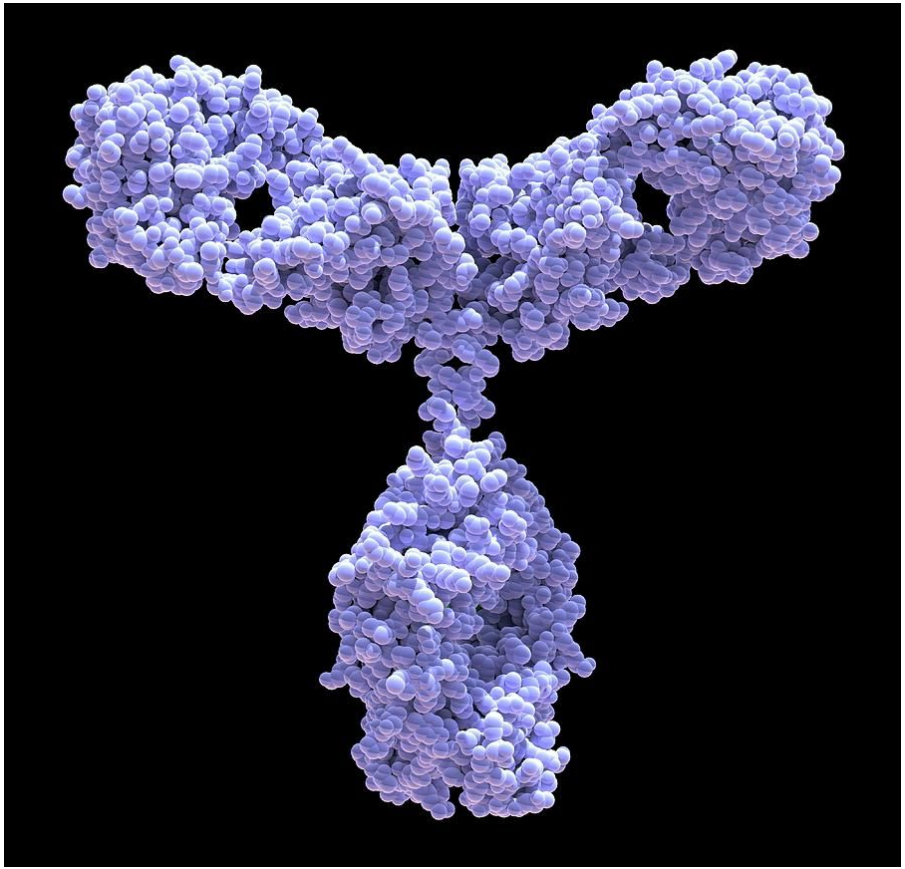
- No single winning strategy. (careful with *all eggs in one basket.*)
- Expect throttling (ESS).
- Traffic thresholds (*).
- An ever-increasing cost for all players.

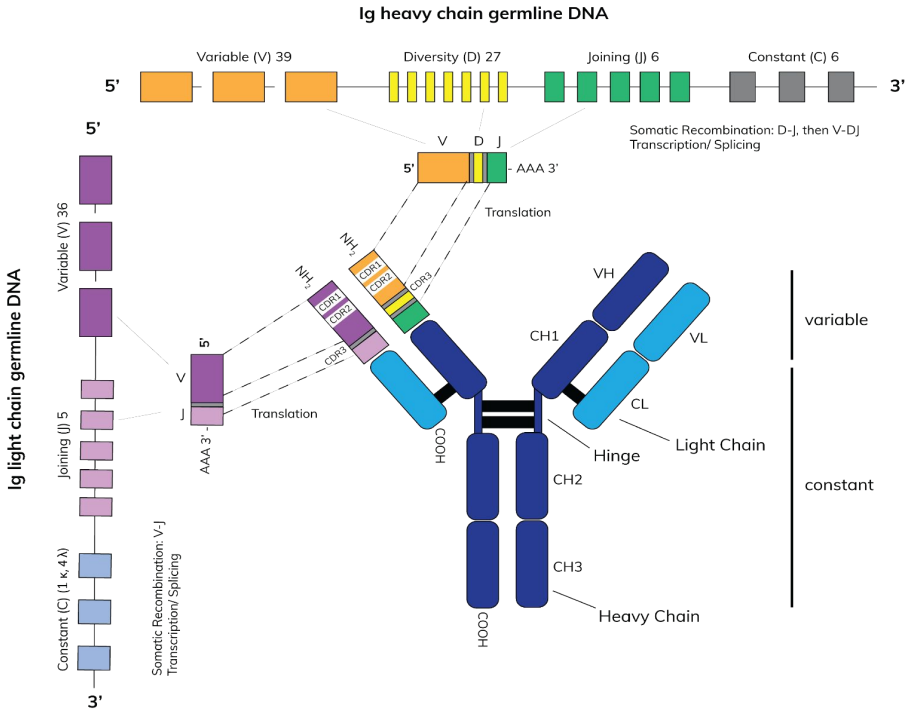
The size issue.

Global constraints impose a theoretical boundary on certain parameters in the design space.

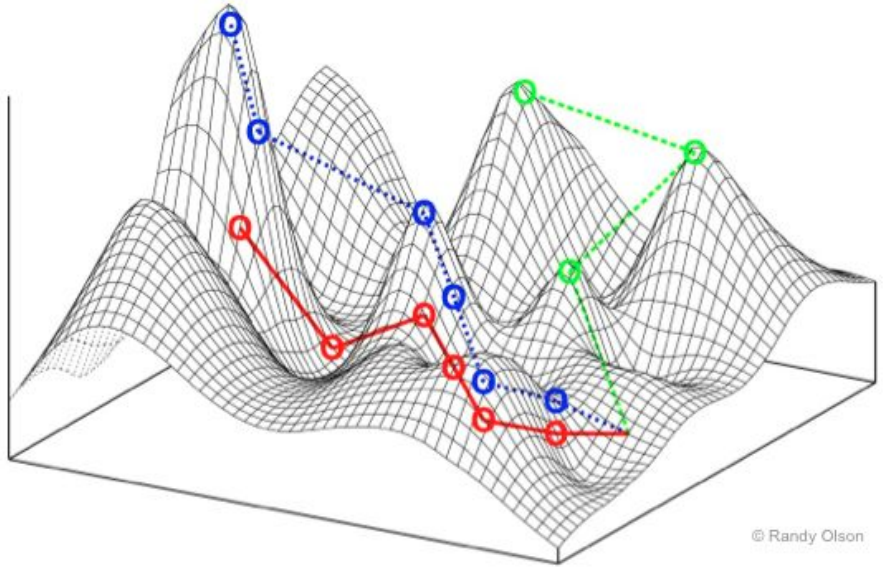
Insights from immunology.

It is co-evolutionary landscapes *all the way down*!




Ig heavy chain germline DNA
Variable (V) 39
Diversity (D) 27
Joining (J) 6
Constant (C) 6
5'
3'
Somatic Recombination: D-J, then V-DJ
Transcription/ Splicing
V
D
J
AAA 3'
Translation
Ig light chain germline DNA
Variable (V) 36
Joining (J) 5
Constant (C) (1 κ, 4 λ)
Somatic Recombination: V-J
Transcription/ Splicing
NH2
CDR1
CDR2
CDR3
VH
VL
CH1
CL
Light Chain
Hinge
CH2
CH3
Heavy Chain
COOH
variable
constant

https://en.wikipedia.org/wiki/M%C3%BCllerian_mimicry



https://en.wikipedia.org/wiki/Evolutionary_landscape

A resource arrangement that works in practice can work in theory.

Ostrom's Law

https://en.wikipedia.org/wiki/Elinor_Ostrom#Design_principles_for_Common_Pool_Resource_(CPR)_institution

Coming back to earth…

Which experiments can be designed to “dance with the censors”?

🤔🧐

Thanks!

Any questions?

You can find me at

- Github: @ainghazal
- ain@openobservatory.org

Credits

- Presentation template by SlidesCarnival