Machine Learning for Social Change

Fernand Pajot
Data Scientist
@fernandp

change.org

the world’s largest petition platform

change.org

It works

300M signatures

4000 declared victories

Victories in 121 countries

20M users experiencing victory

Smallest winning petition has 2 signatures

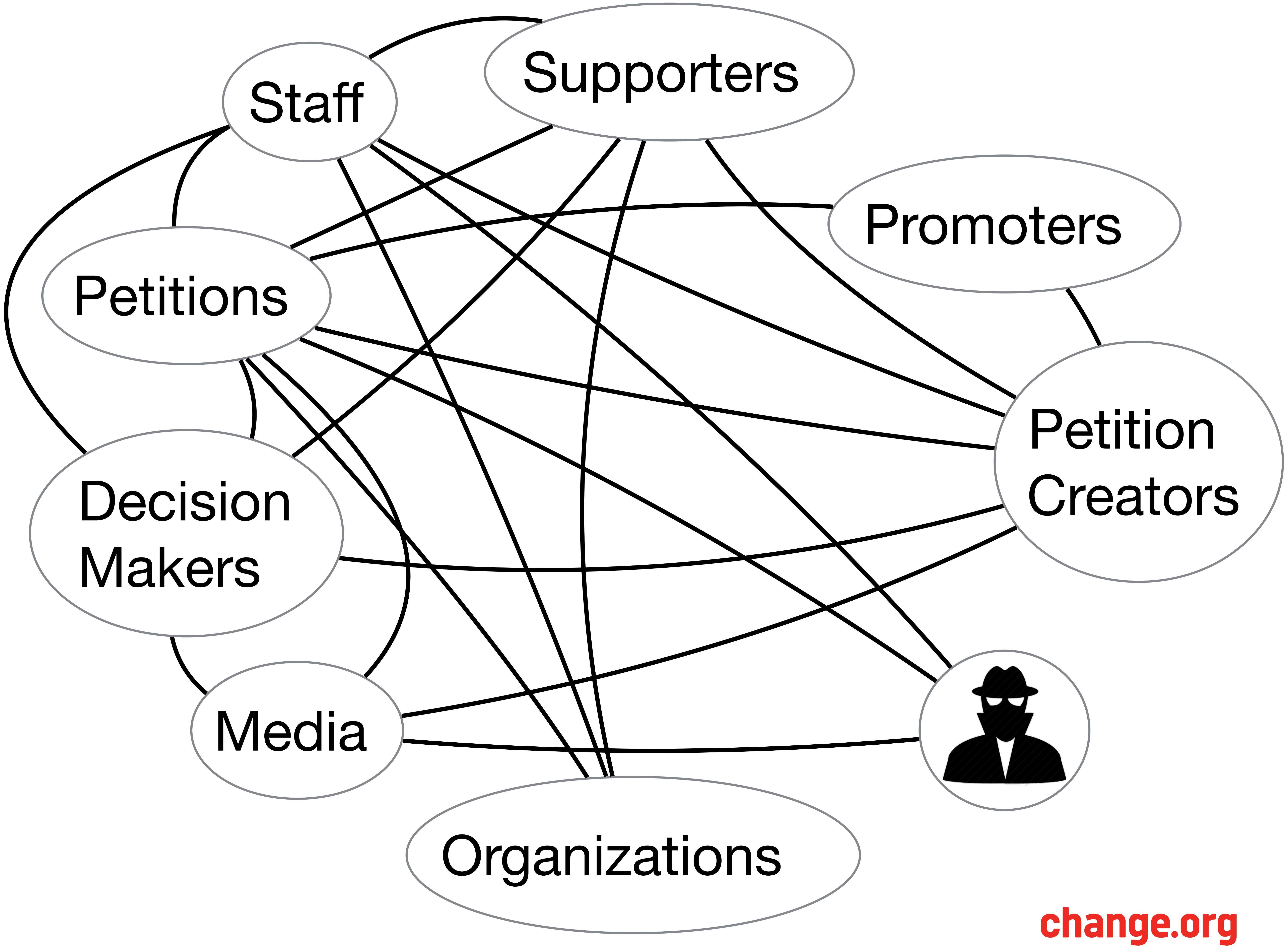
Biggest winning petition has 1.3M signatures

change.org

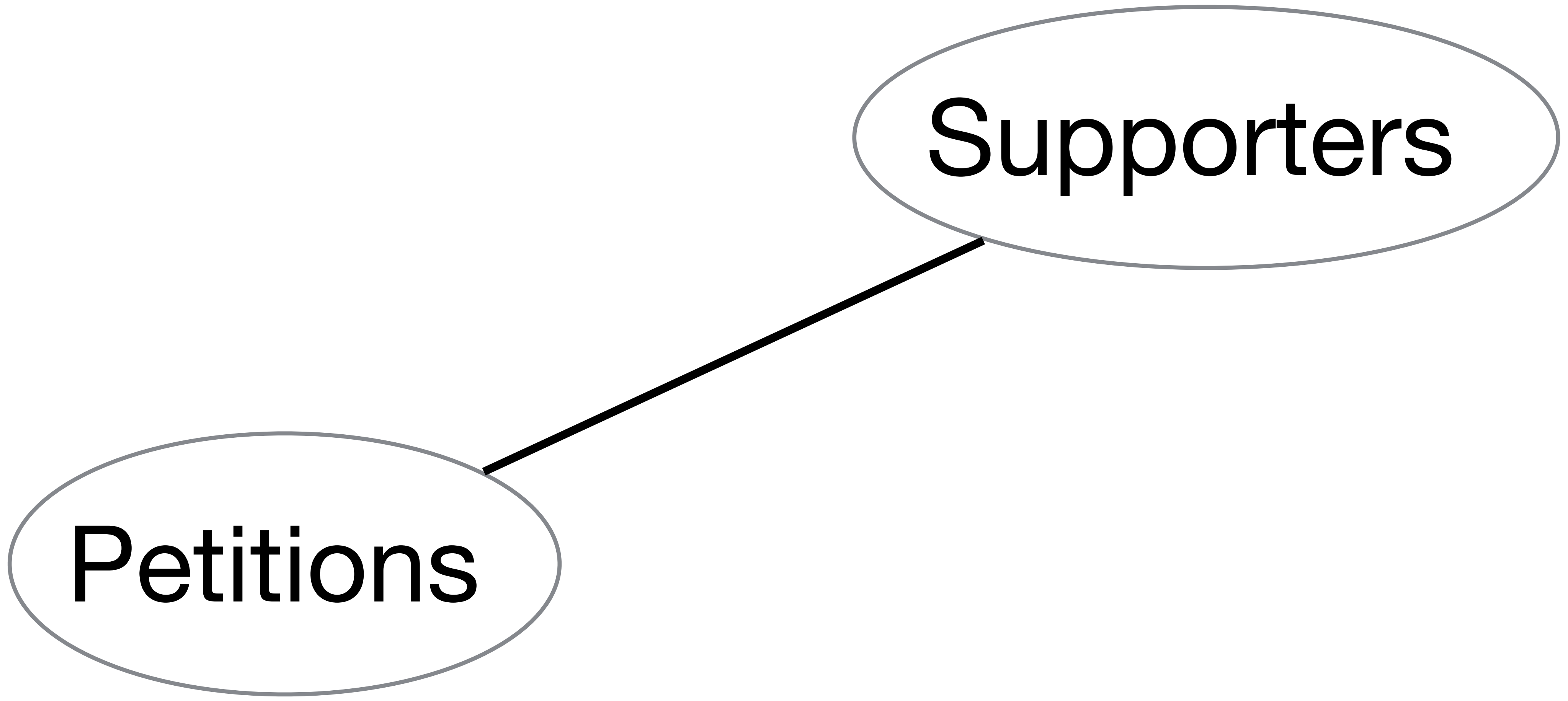

300,000
People Say
Reinstate Gay
Scout Leader
www.change.org/scouts
300,000
People Say
Reinstate Gay
Scout Leader
300,000
People Say
Reinstate Gay
Scout Leader
www.change.org/scouts

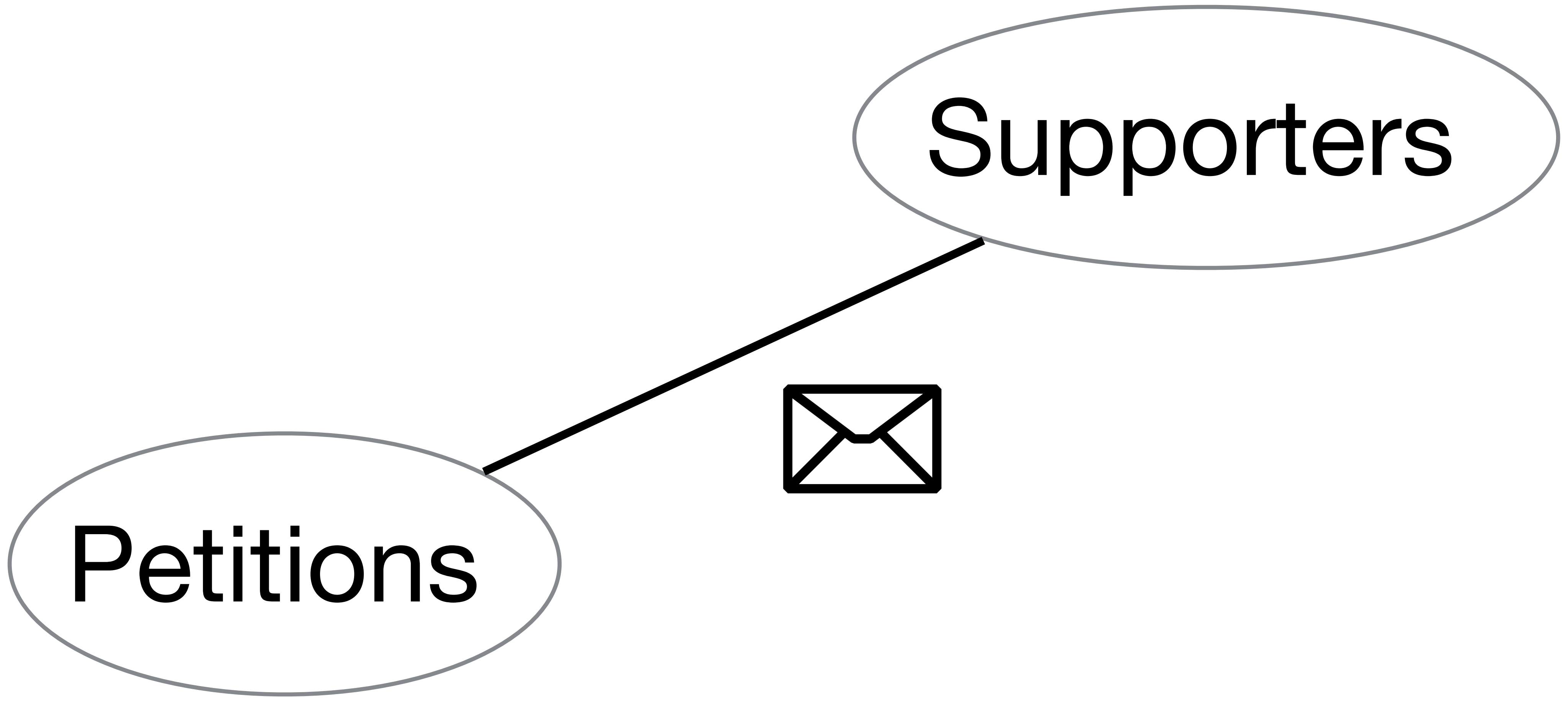
$$\max_{f}\left[f\left(\sum_{x \in \text{🌐}} \Delta_{x} \right) \right]$$

change.org

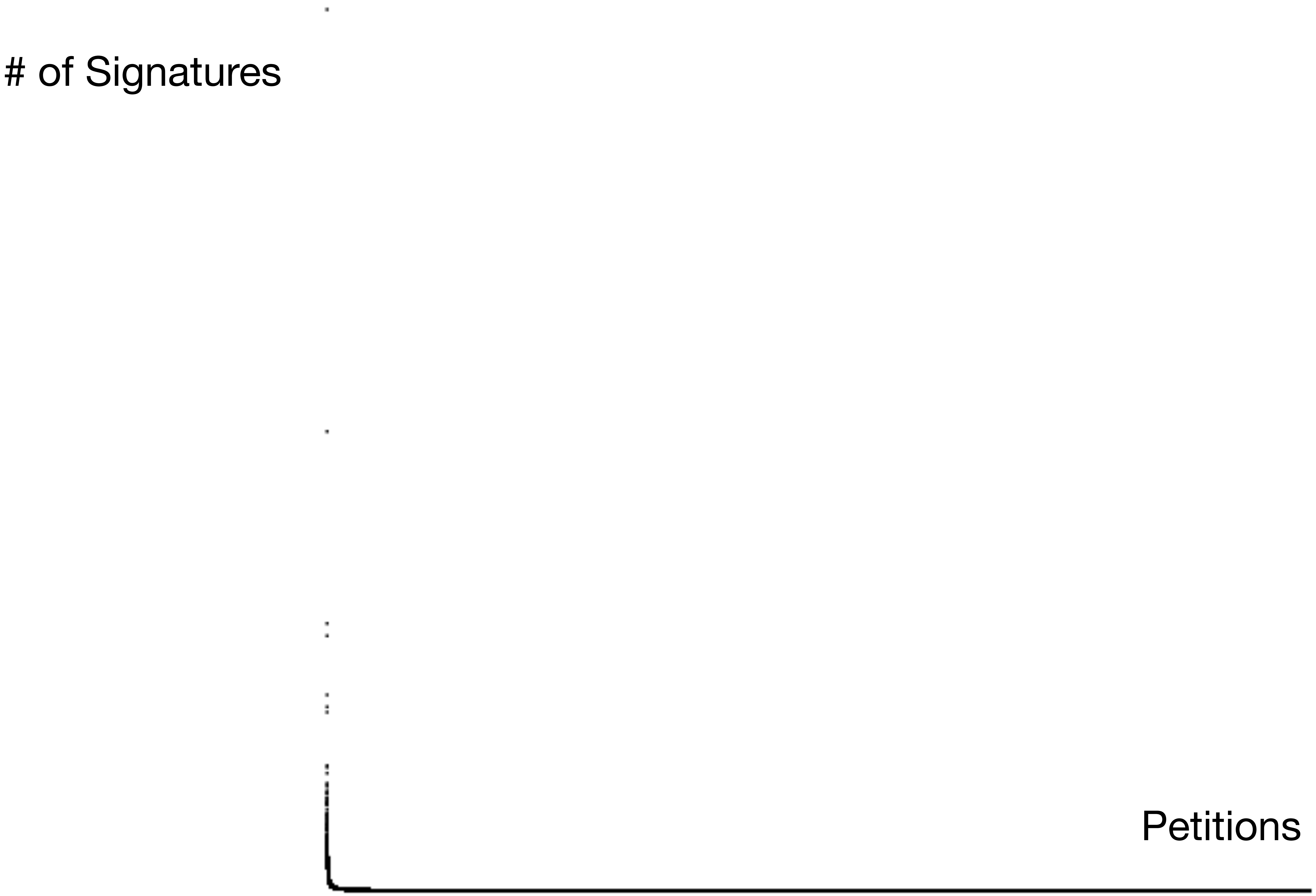

Staff
Supporters
Promoters
Petitions
Petition Creators
Decision Makers
Media
Organizations

change.org

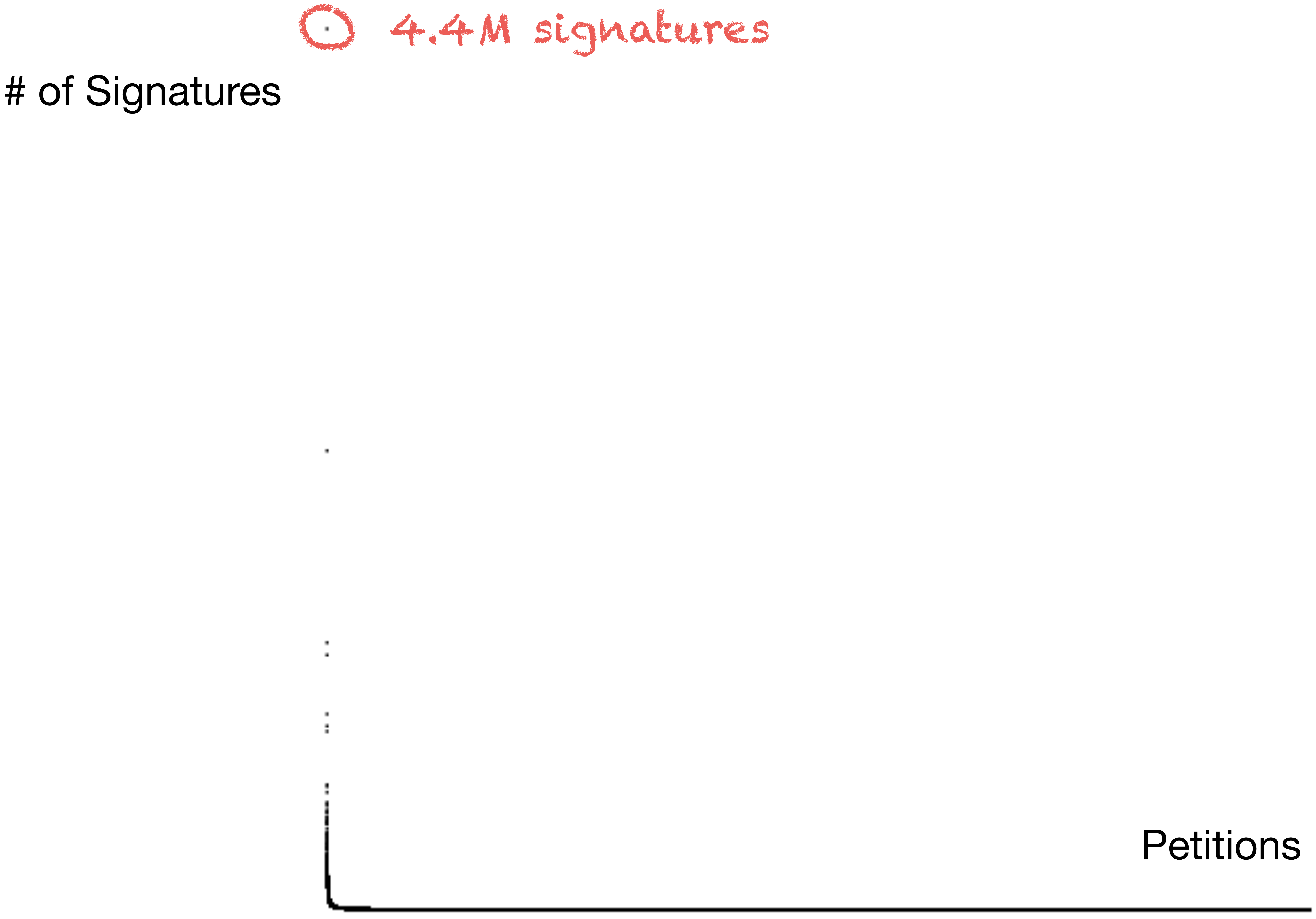

Supporters
Petitions



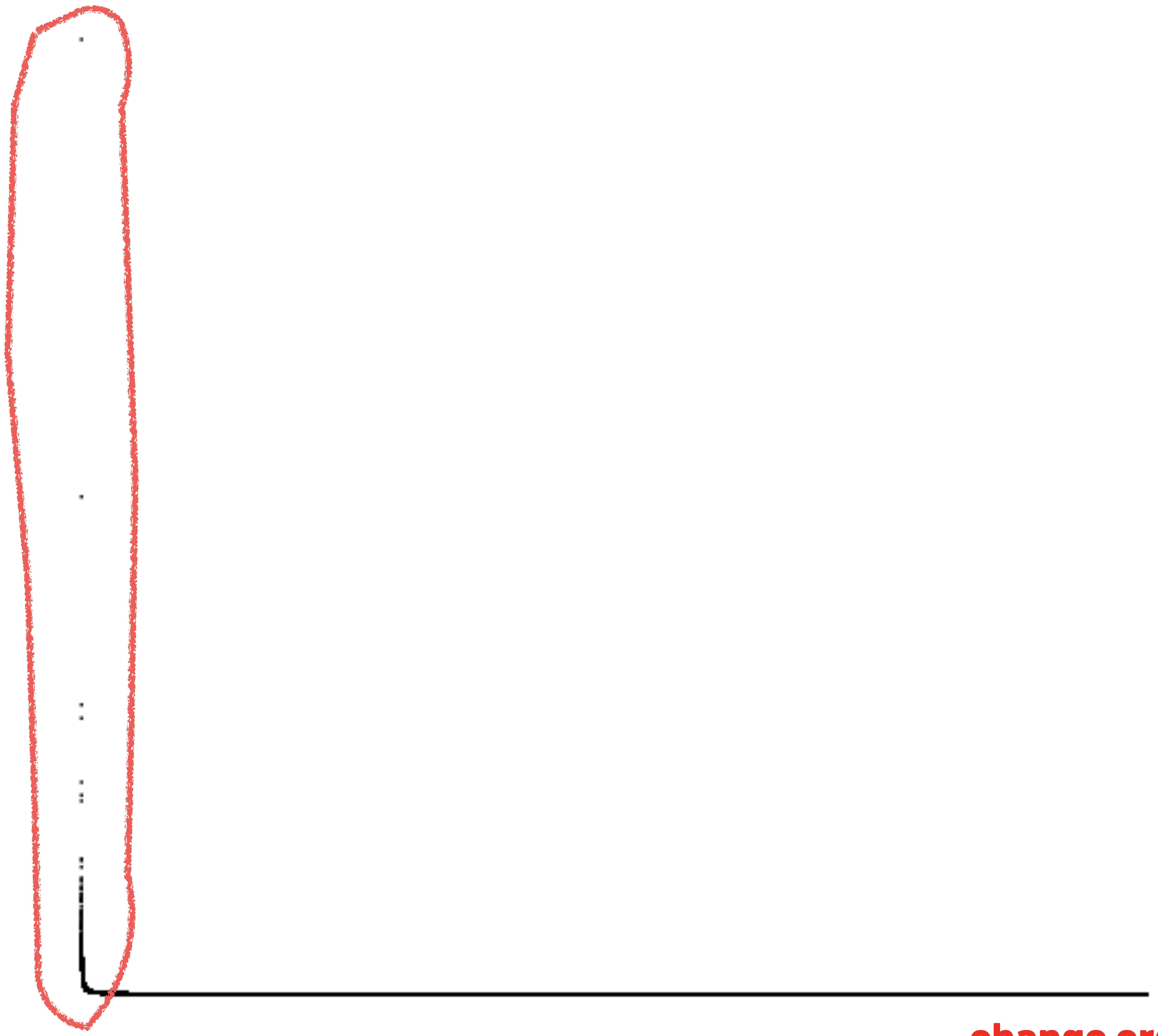
change.org



change.org



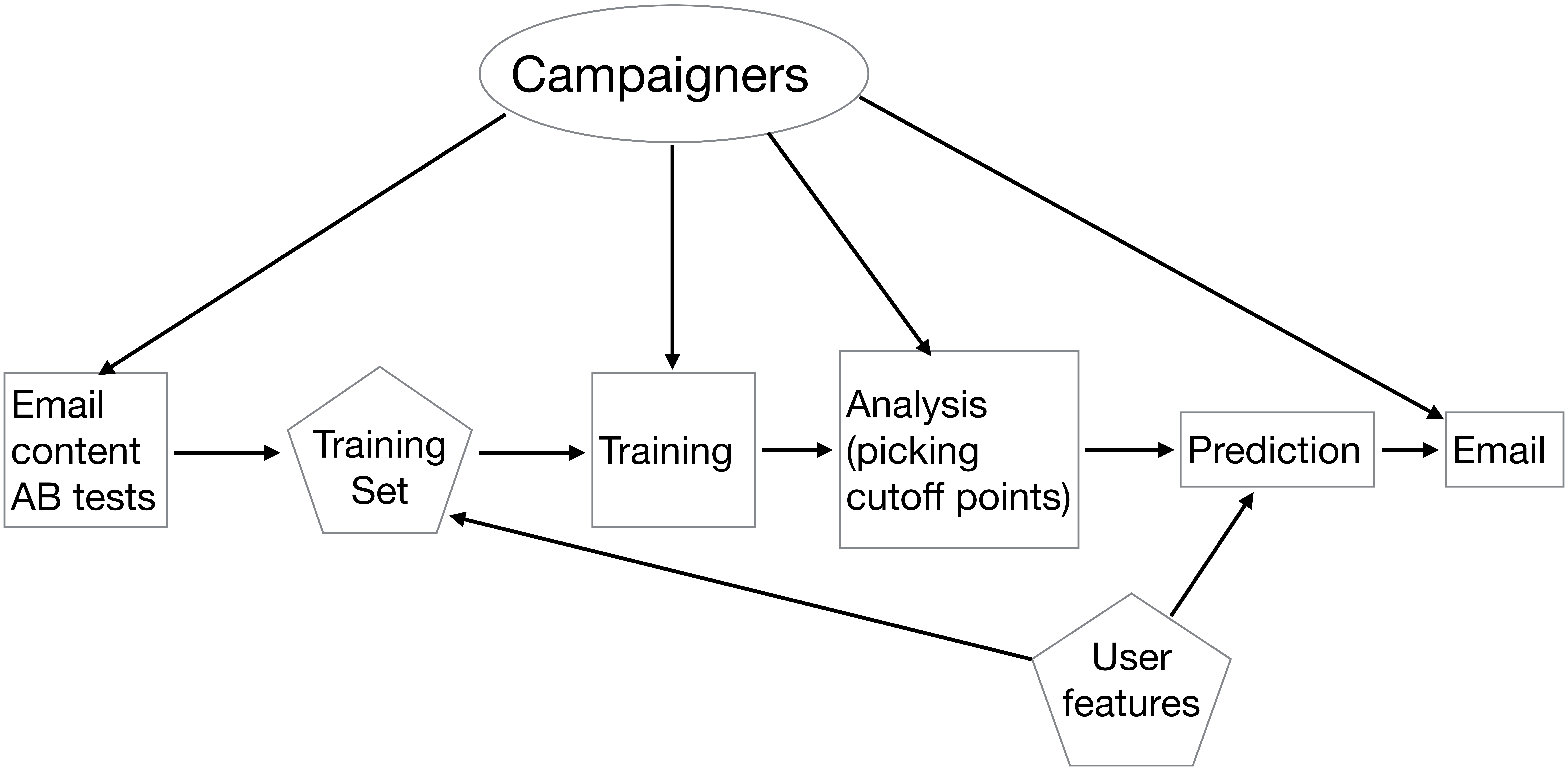
change.org

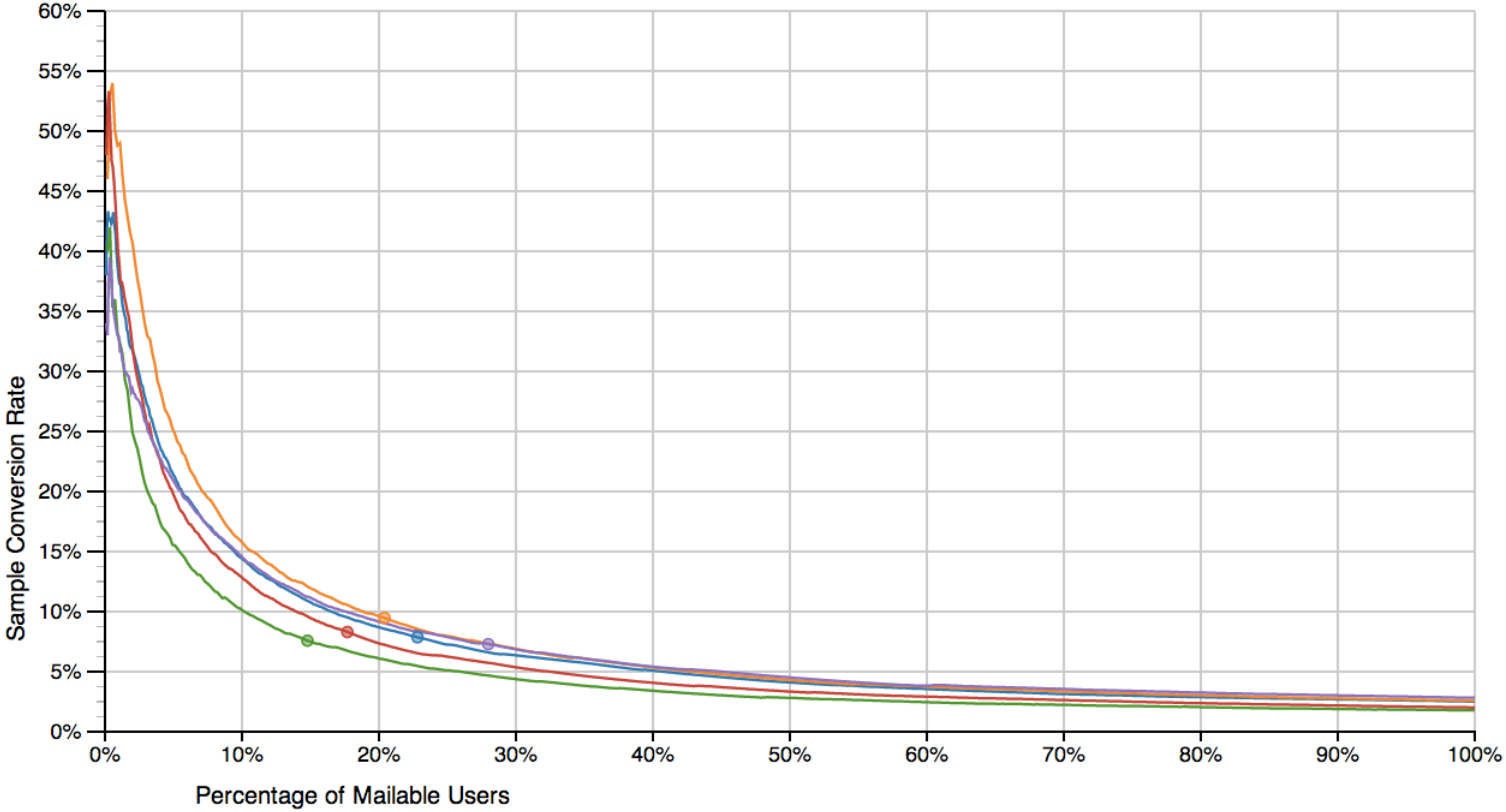


change.org

10s of petitions → 10Ms of users

~daily


Campaigners
Email content AB tests
Training Set
Training
Analysis (picking cutoff points)
Prediction
Email
User features



Sample Conversion Rate
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Percentage of Mailable Users

change.org

Learnings

Behavioral data trumps demographics & third party data sources.

Make all your features binary (via bucketing) to have flexibility on which algorithm to pick.

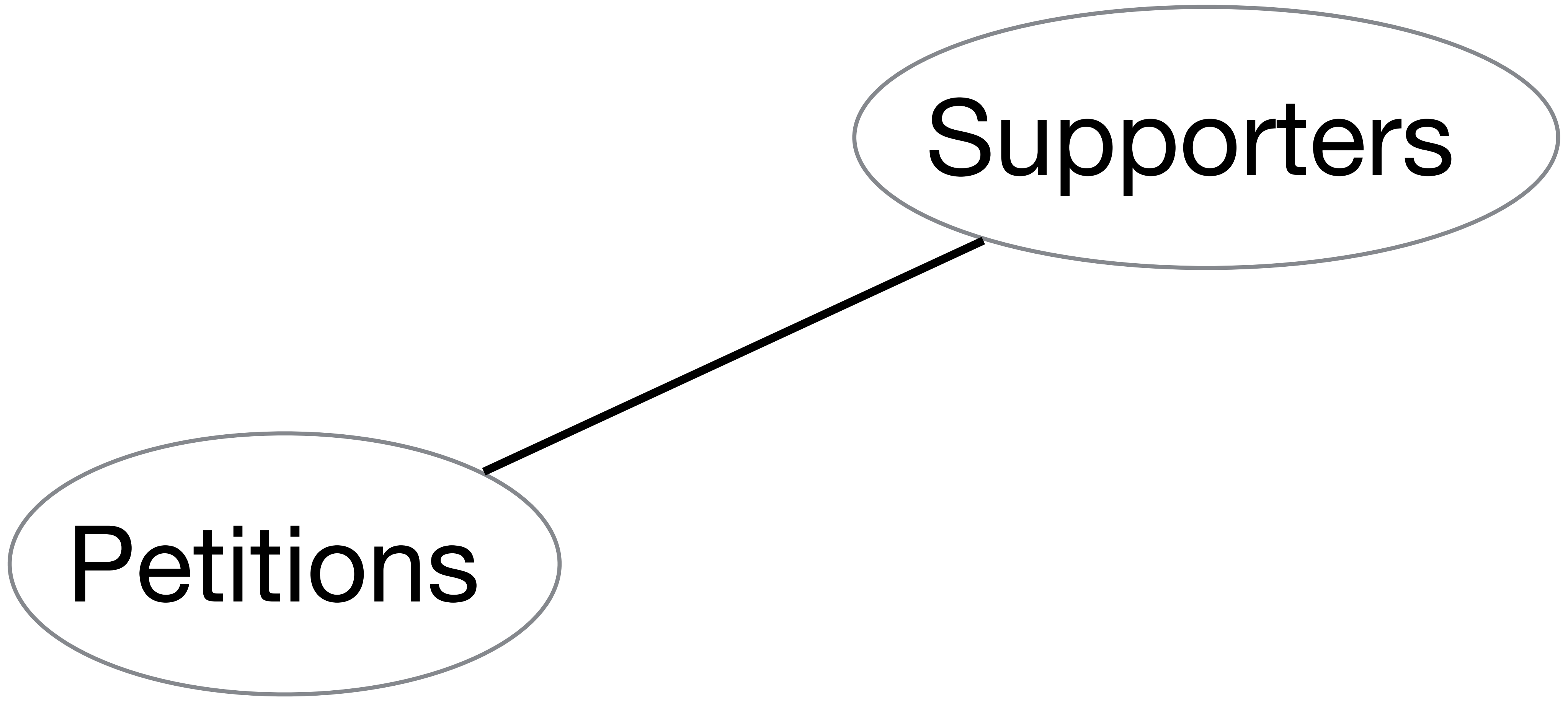
Random Forests work well for imbalanced (~2% positives) data.

Performance was not the main criteria when choosing the learning model, use case and flexibility were way more important.

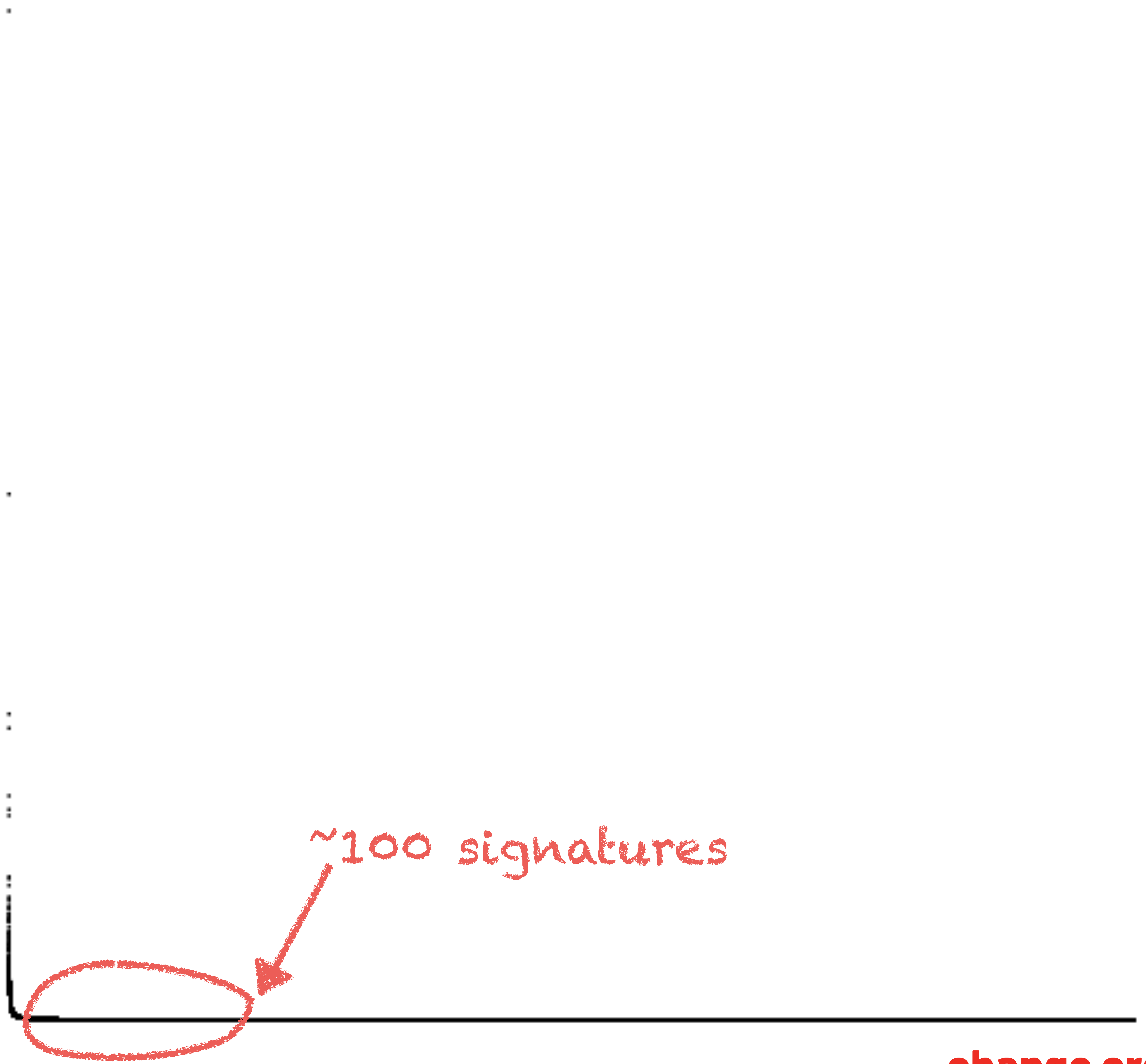
Learnings

Big RAM/CPU instances are cheap, sparse binary datasets are small. If you can do in memory training, do it.

Use statistical distributions and basic time series analysis to verify data in your ETL pipeline.



change.org



change.org

Recommendation & Discovery as a coherent Product

The data landscape

Featured Petitions
Social graph
Geography
Similar petitions
Similar users
Followed topics
Followed users

Delivery Channels

Online feed
Online contextual (like after signing)
Email push
Email digest
Mobile Push
Facebook

Recommendation & Discovery as a Product

Tons of input sources to try.

Tons of possible UXs.

Exploration vs Exploitation.

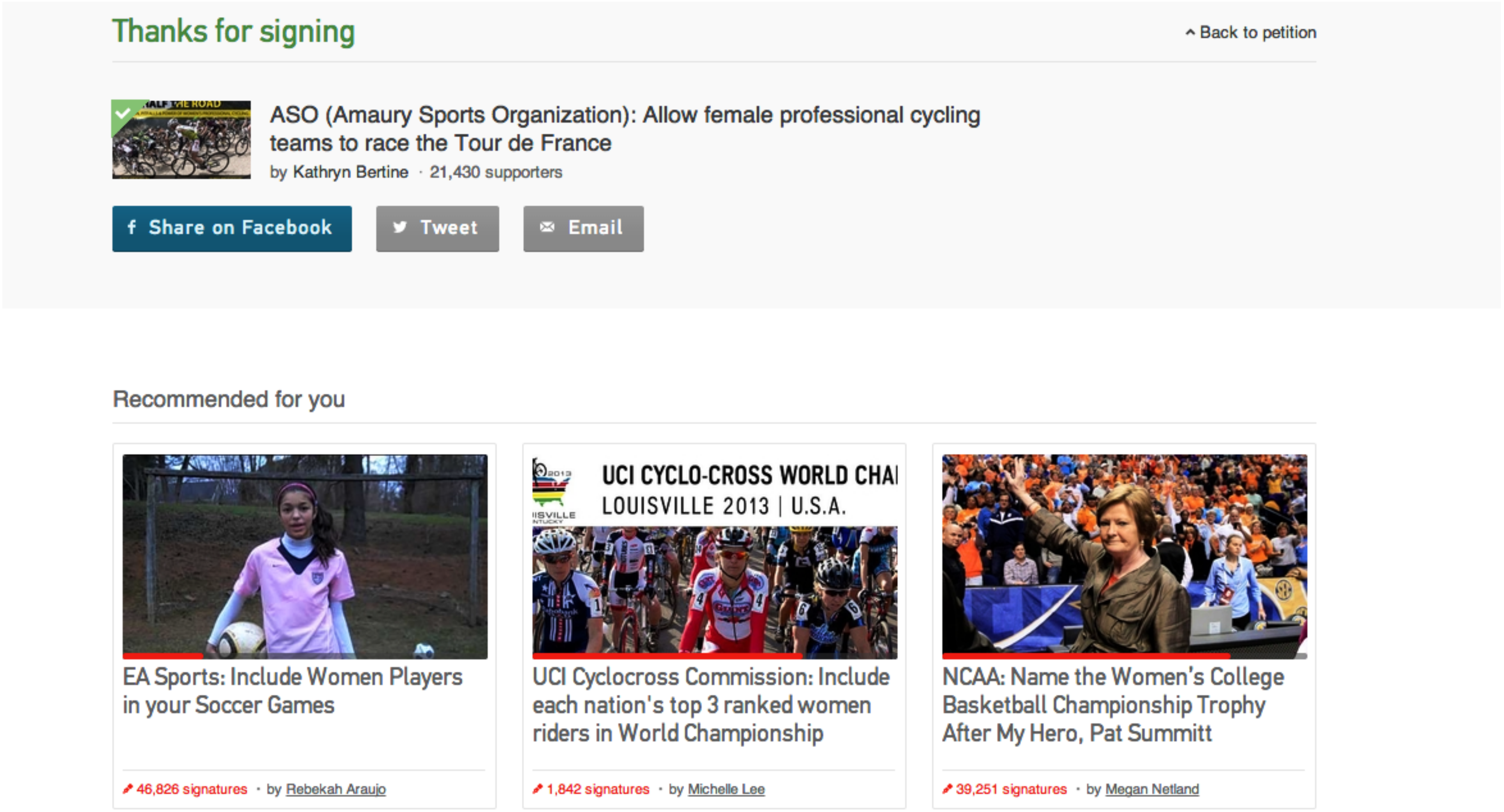
AB testing must as cheap as possible. For algorithms and features.

Which metric do you want to optimize?

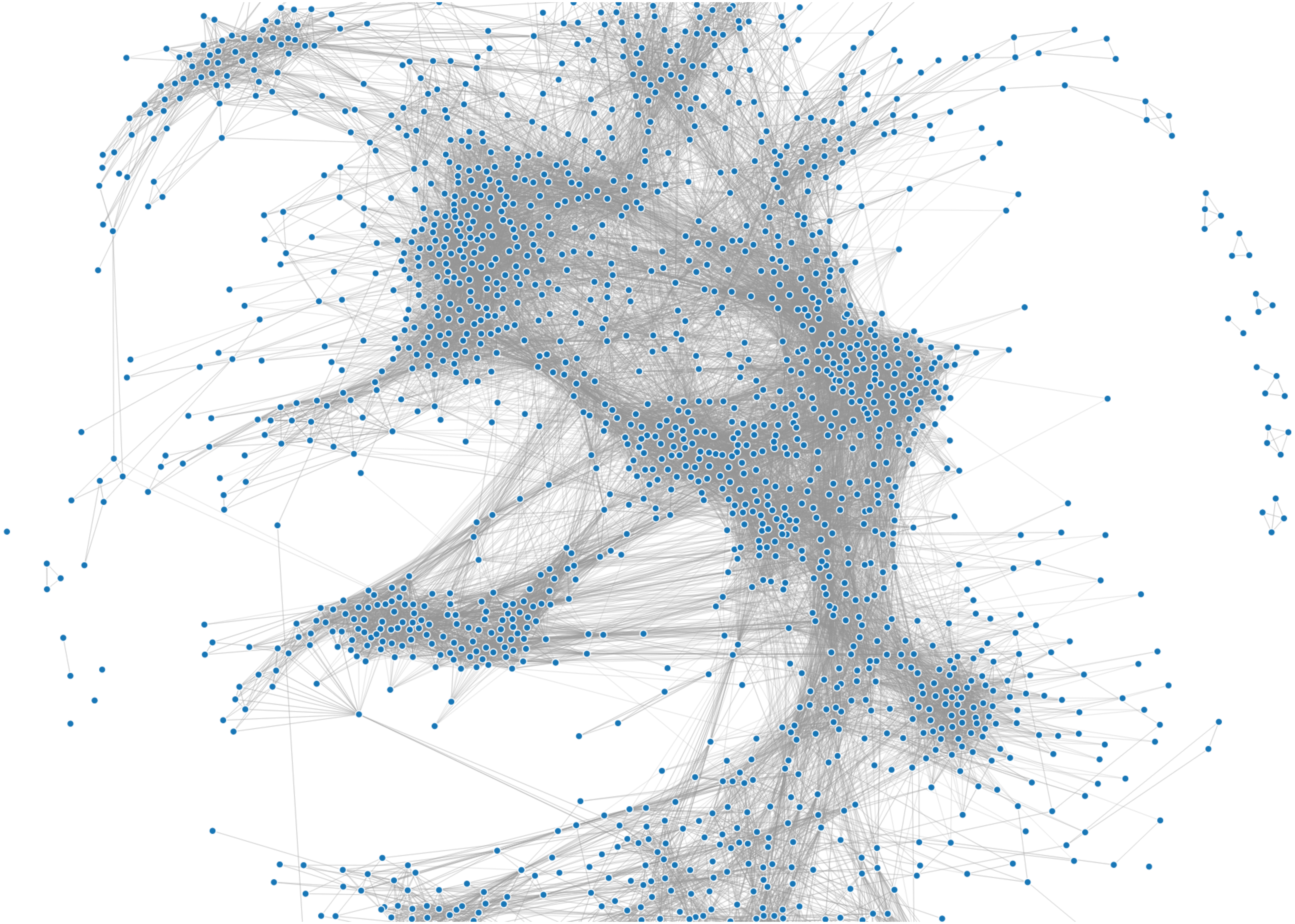
Picking the right metric is the hardest part.

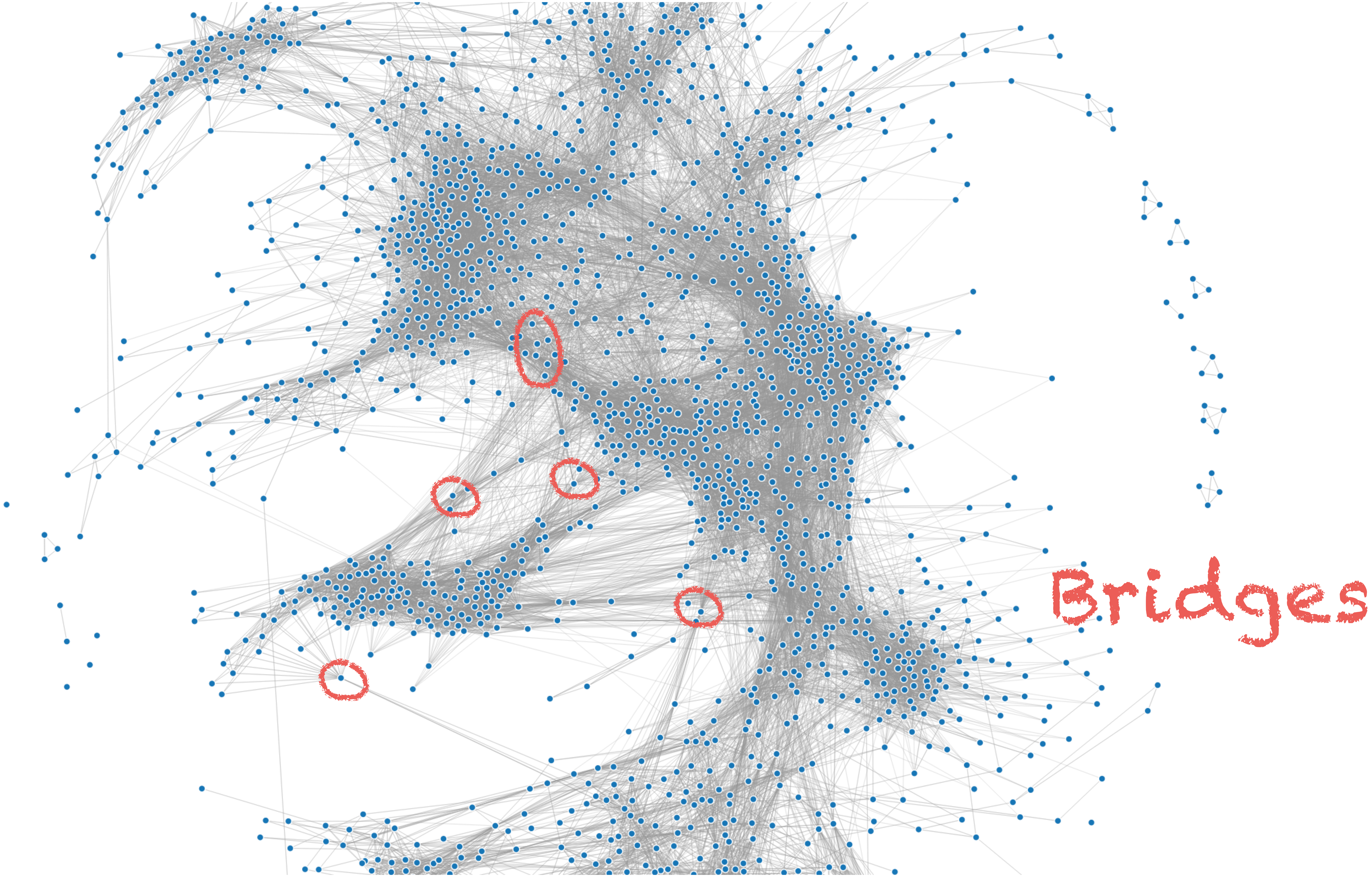
change.org

Similar petitions via Collaborative Filtering



change.org





change.org

Similar petitions via Collaborative Filtering

3x increase over baseline in additional signatures.

30% increase in overall signatures.

Always start with the simplest models such as basic similarity metrics.

Computing n^2 pairs is a lot, use company data insights and Locality Sensitive Hashing to drastically reduce that number.

Recommendation & Discovery as a Product

A more complicated model with a meager 5% improvement is usually not worth it.

Finding the right UX flow is the hardest part.

Better recommendations means less focus on the rest of the page (such as sharing).

change.org

Learnings

Recommendations' performance is highly variable depending on *context*.

A user's sensitivity to different data inputs changes drastically over time.

New users are not interested in similar petitions but highly sensitive to what they friends are signing.

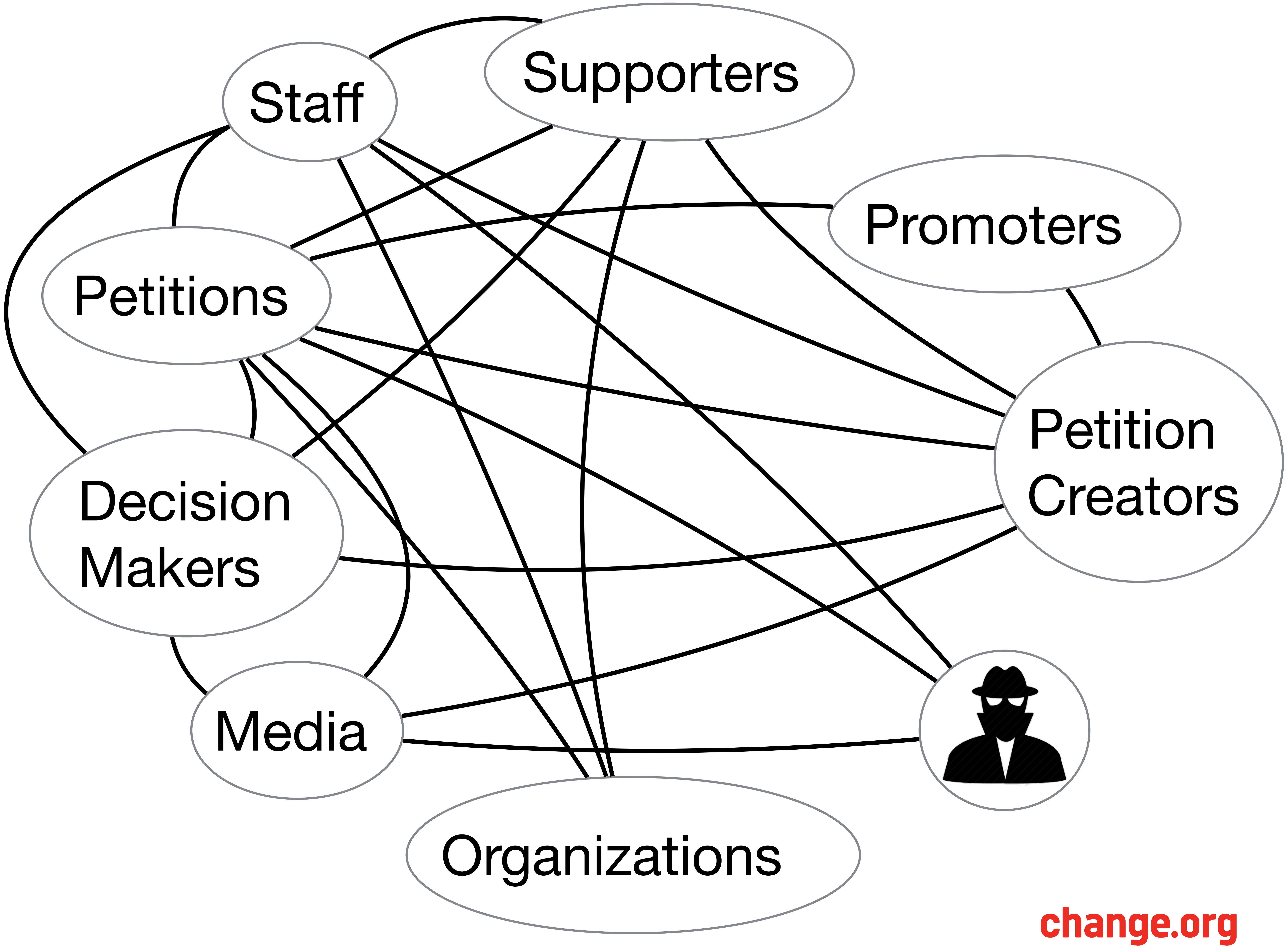
Learnings

Cold starting users is not the biggest problem.

Getting data about a user is not enough. You must first show the user the value of engaging with the platform.

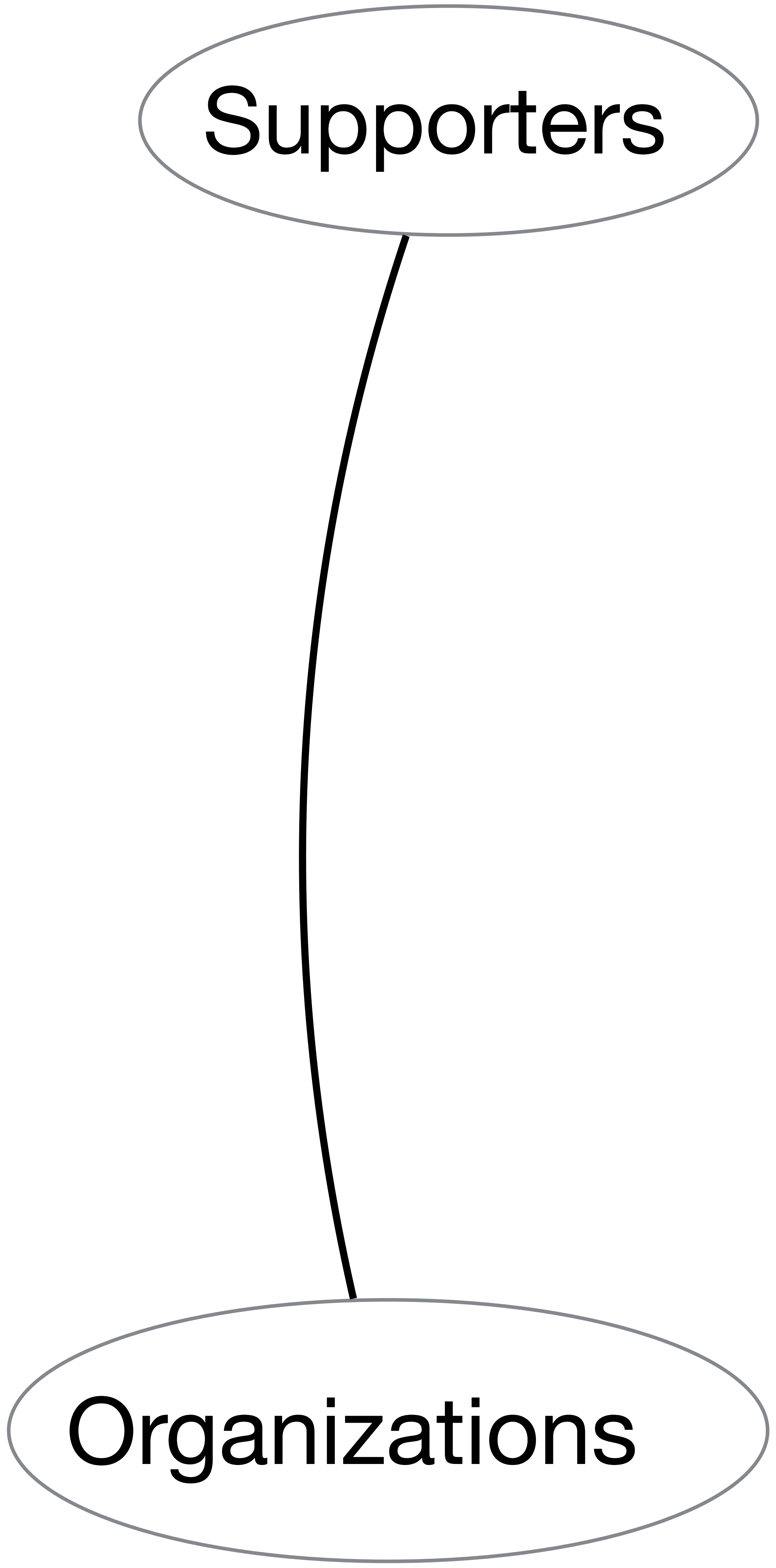
Social graph partitioning is hard and incompatible with cohort analyses.

change.org



Staff
Supporters
Promoters
Petitions
Petition Creators
Decision Makers
Media
Organizations

change.org



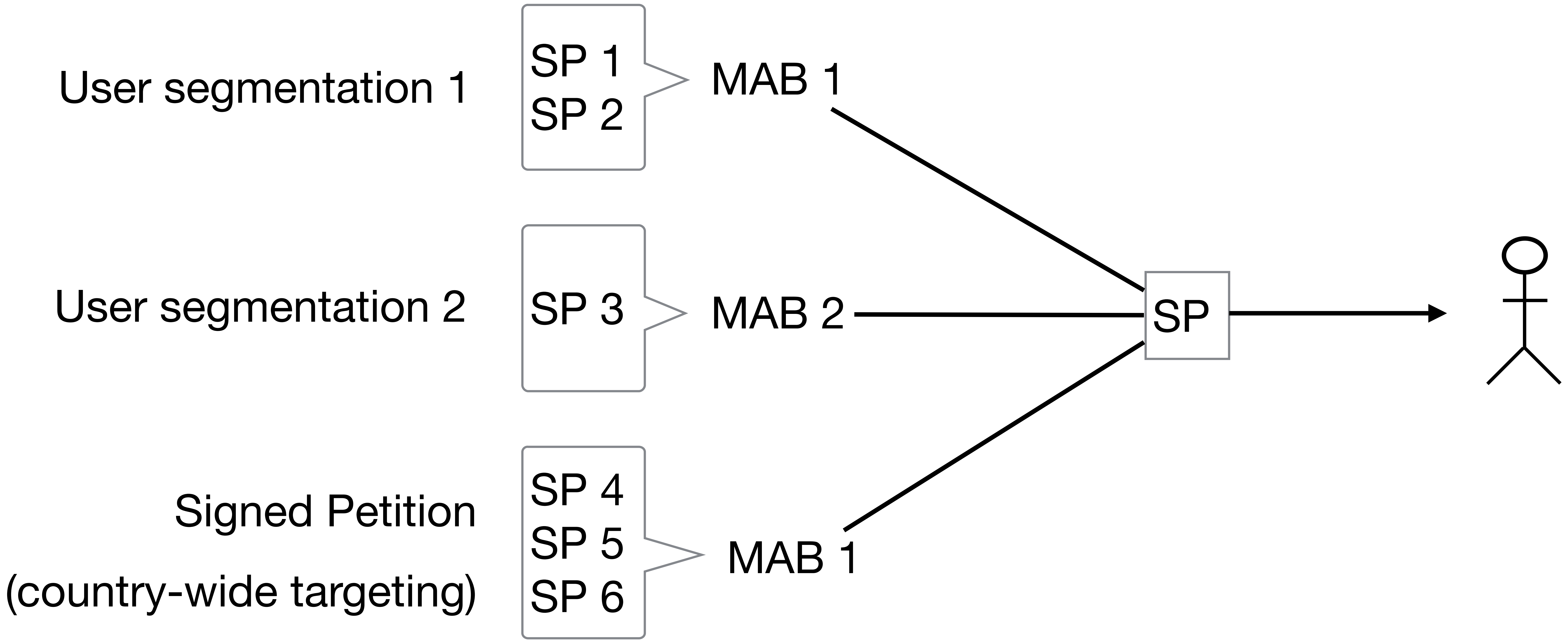
change.org

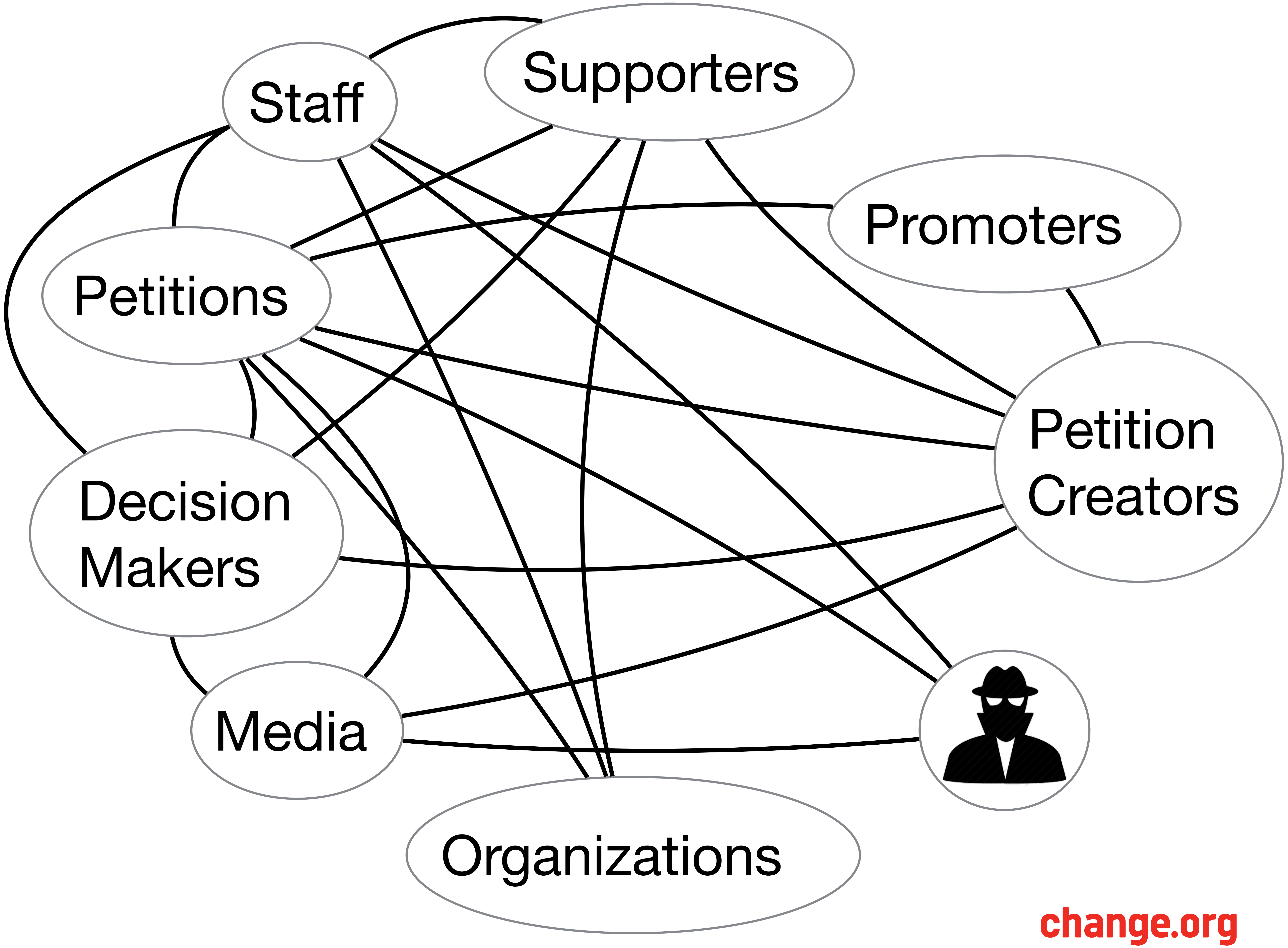
Sponsored petition (SP) targeting

Contextualized Multi-Armed Bandits (MAB) $\max_i \left(price_i \cdot CTR_i \right)$

Sponsored petition (SP) targeting

Contextualized Multi-Armed Bandits (MAB) $\max_i \left(price_i \cdot CTR_i \right)$




Staff
Supporters
Promoters
Petitions
Petition Creators
Decision Makers
Media
Organizations

change.org


300,000
People Say
Reinstate Gay
Scout Leader
www.change.org/scouts
300,000
People Say
Reinstate Gay
Scout Leader
300,000
People Say
Reinstate Gay
Scout Leader
www.change.org/scouts

We’re hiring!

change.org/careers

change.org