Modeling Online User Interactions and their Offline effects on Socio-Technical Platforms

Hitkul

PhD Thesis Defense - June 18, 2024

INDRAPRASTHA INSTITUTE of INFORMATION TECHNOLOGY DELHI

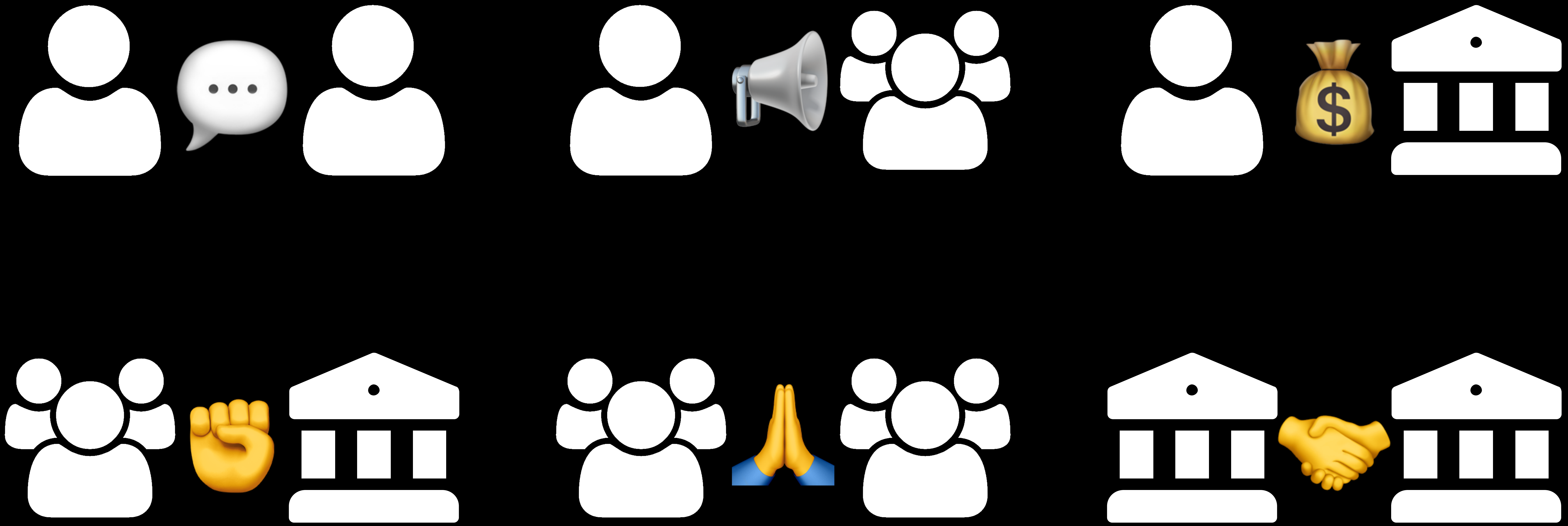




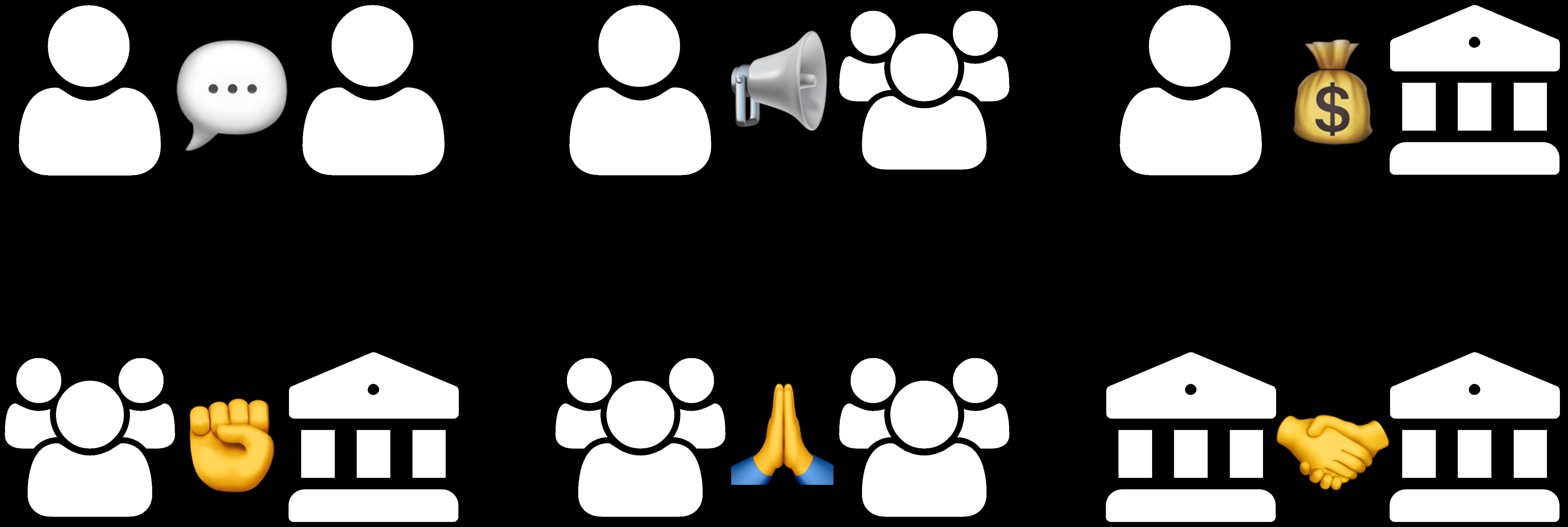
Committee Members

Interactions

Interactions



Interactions



Interactions

Interactions

Interactions

Individual

Community

Organizations

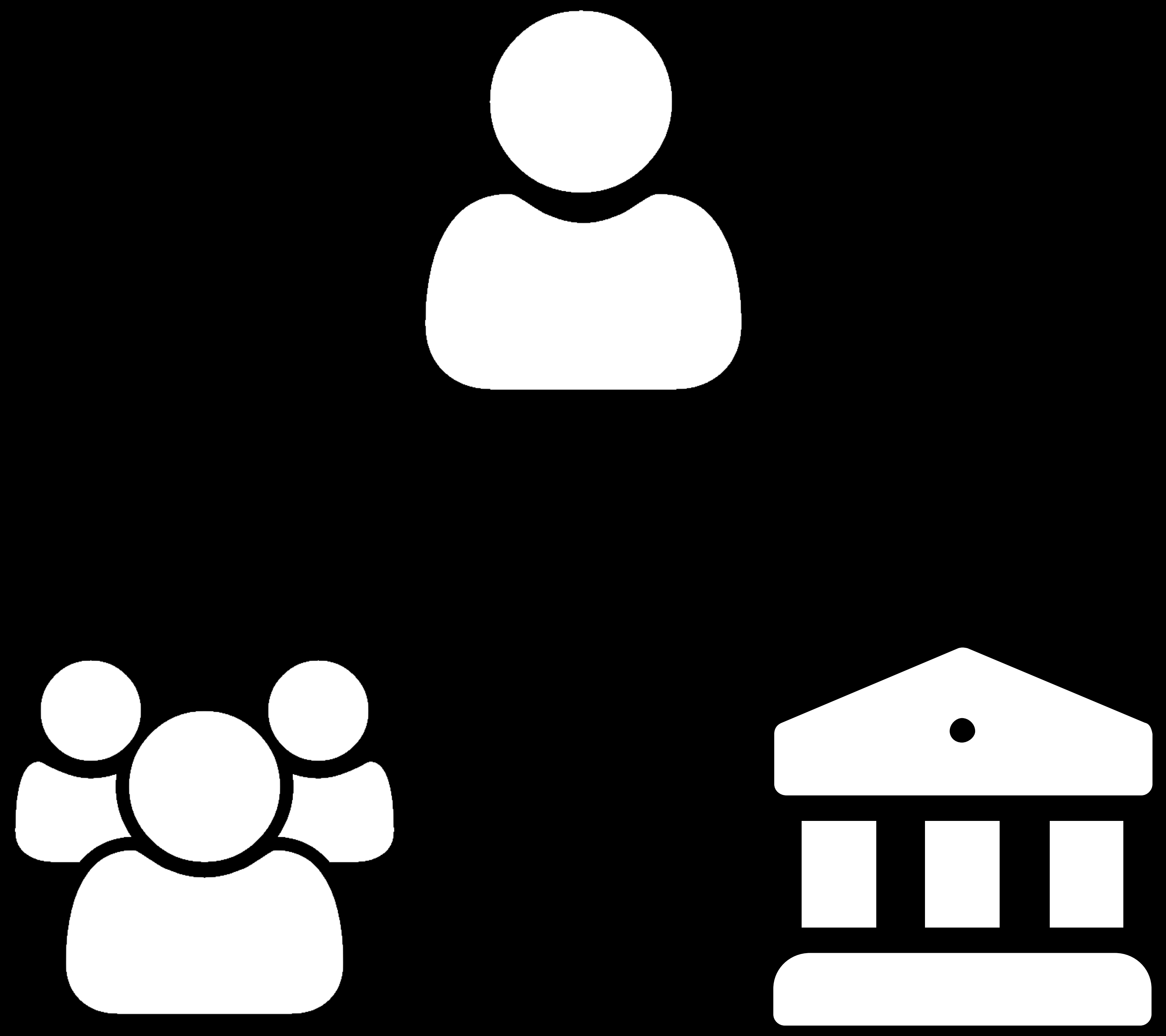
Interactions

Individual

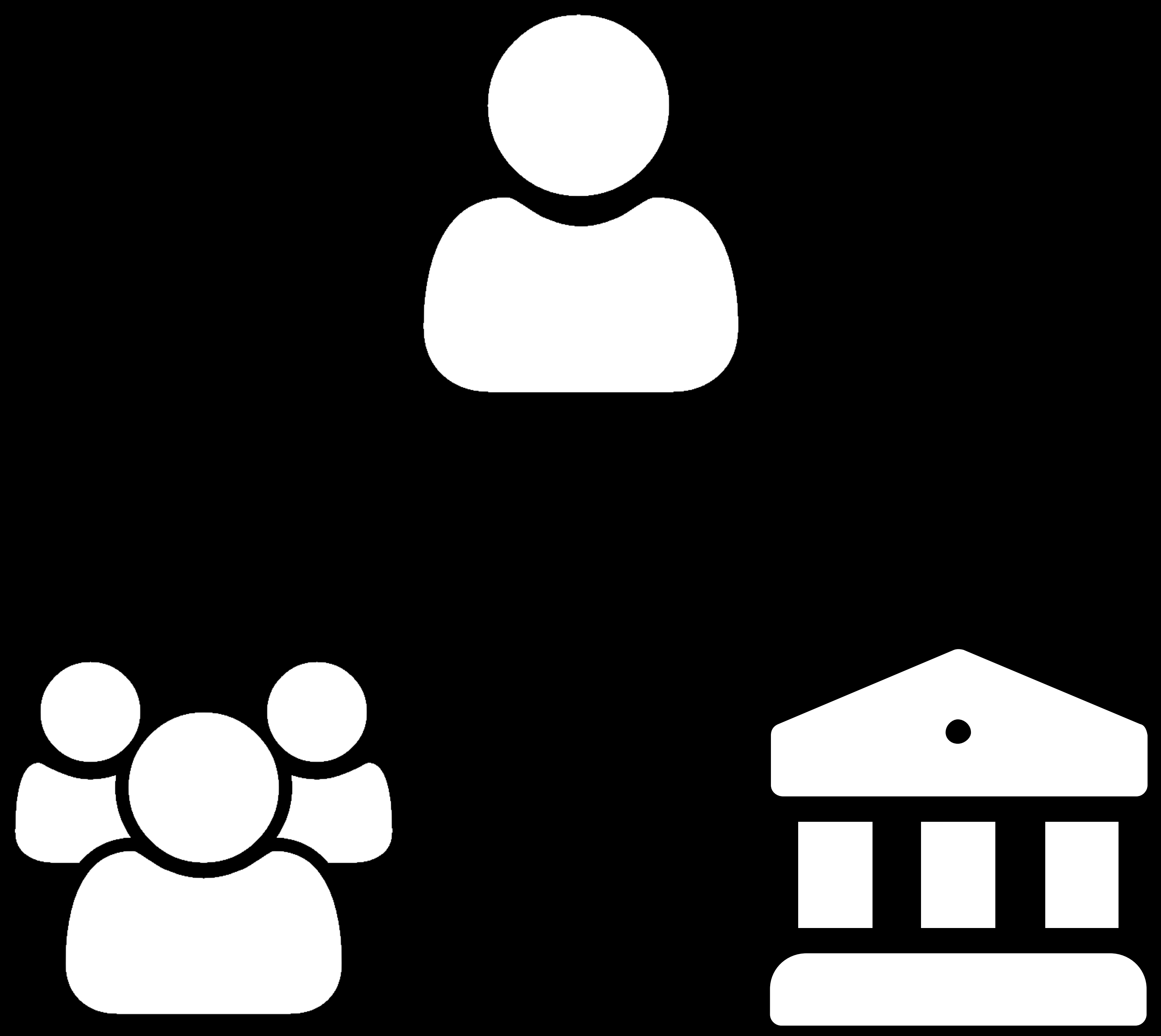
Community

Organizations

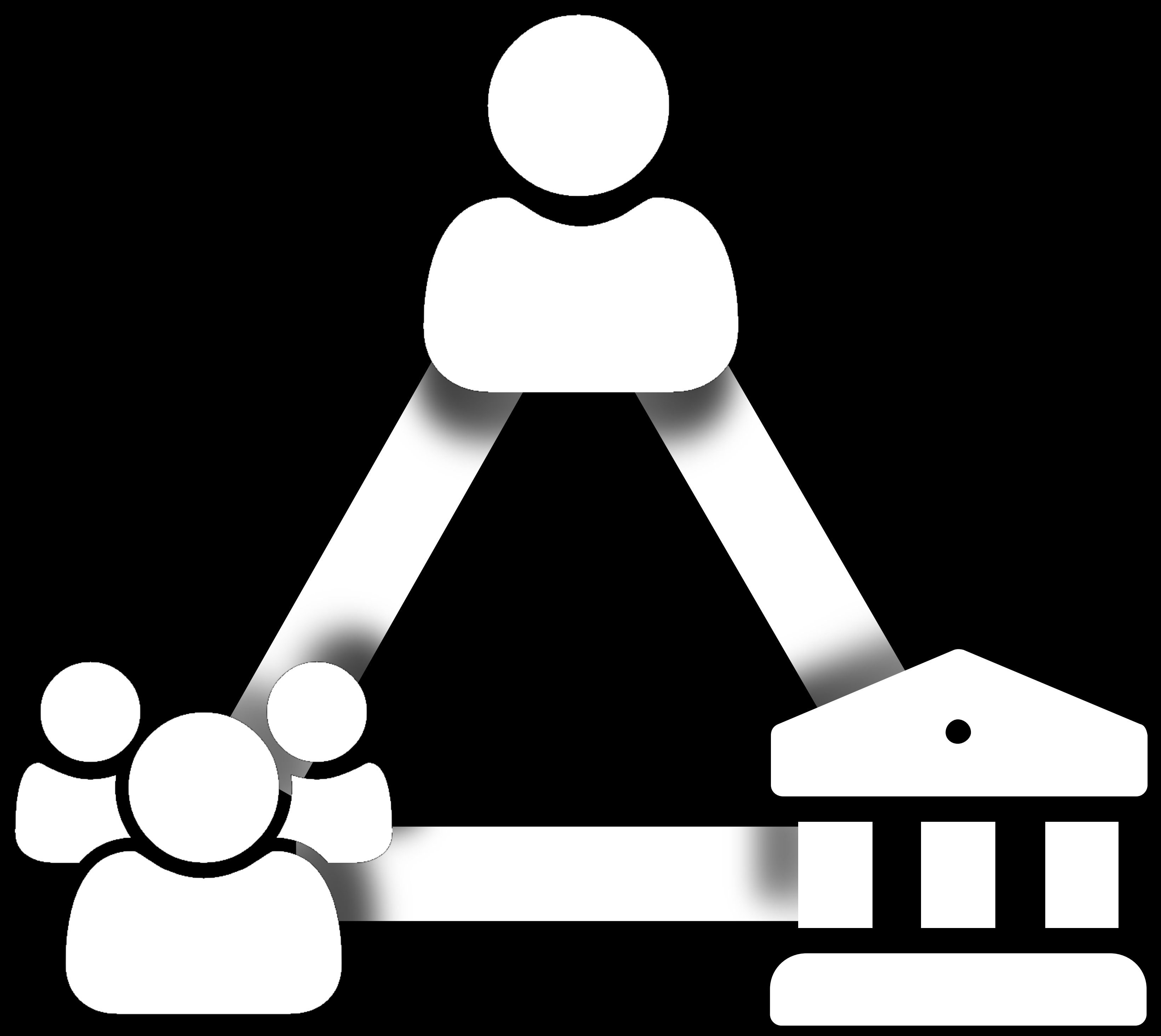
Interactions



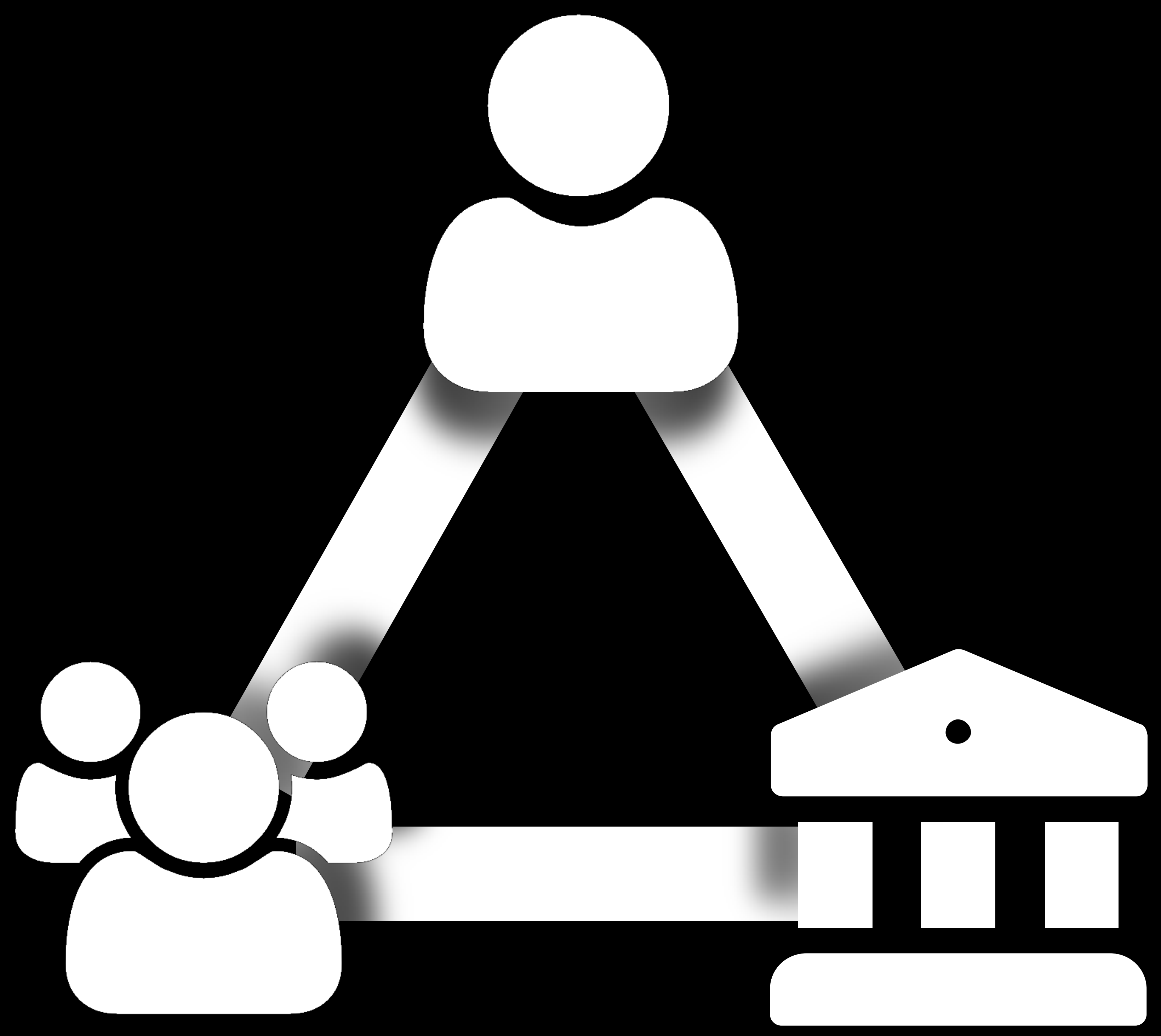
Interactions



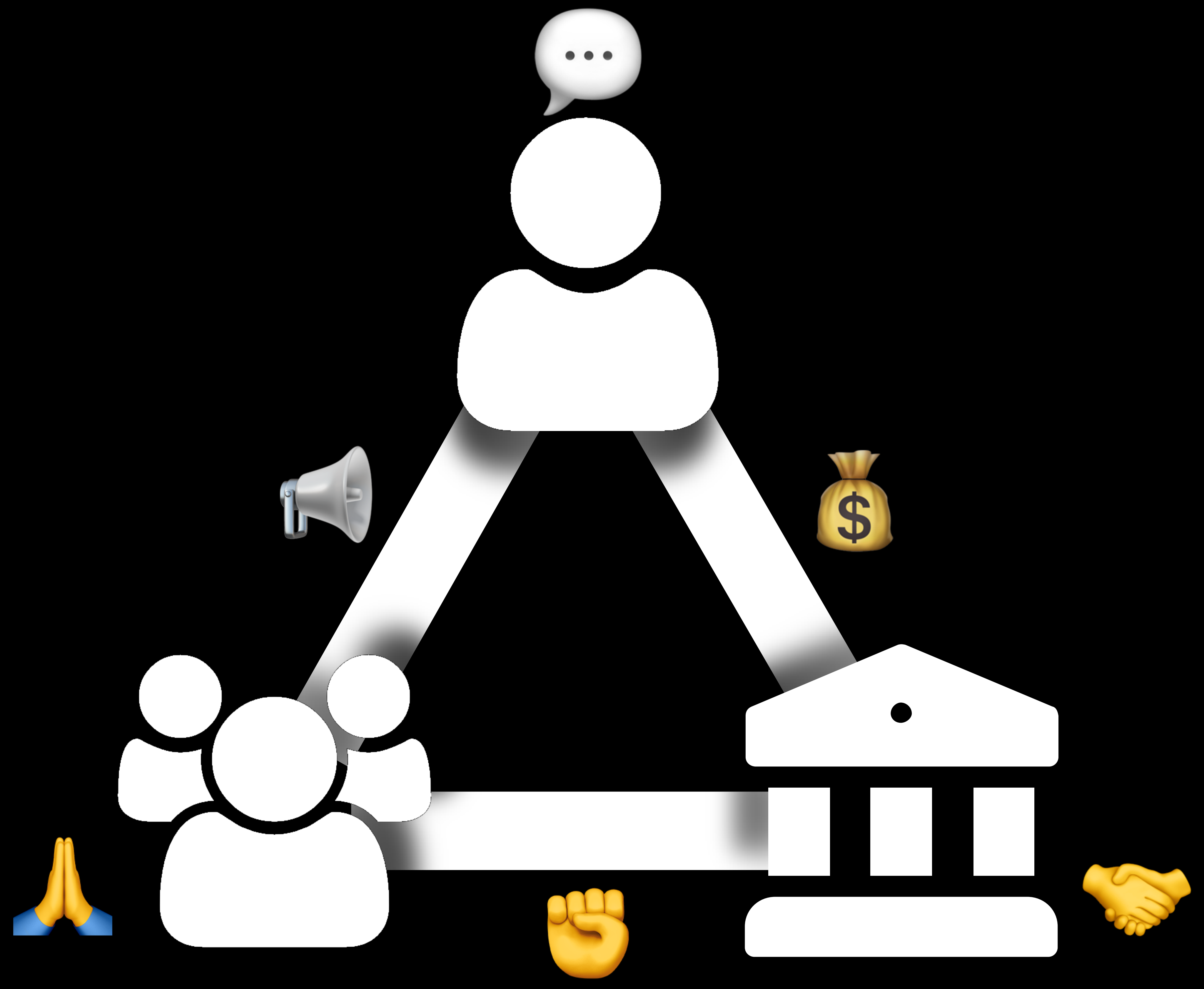
Interactions

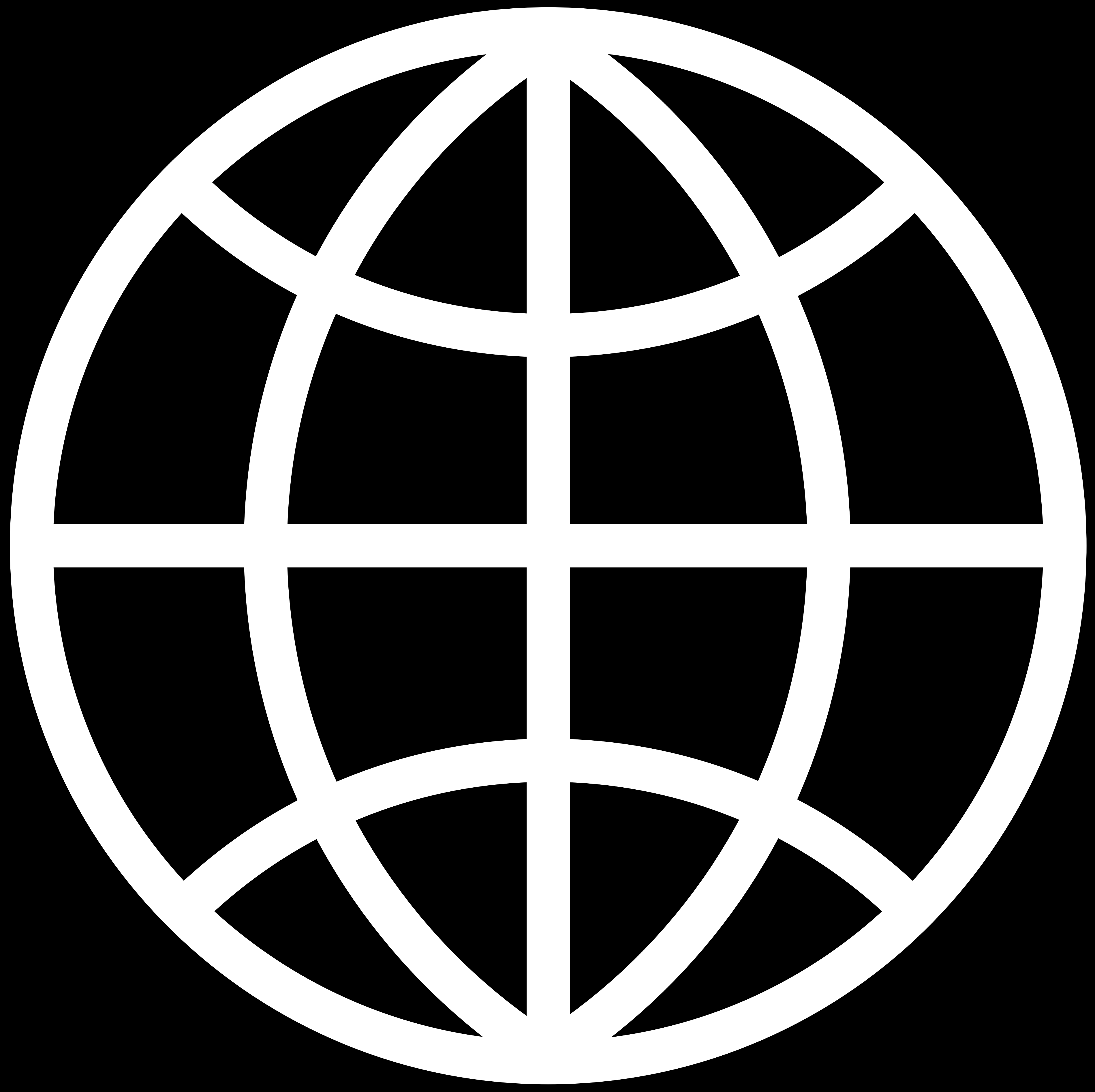


Interactions

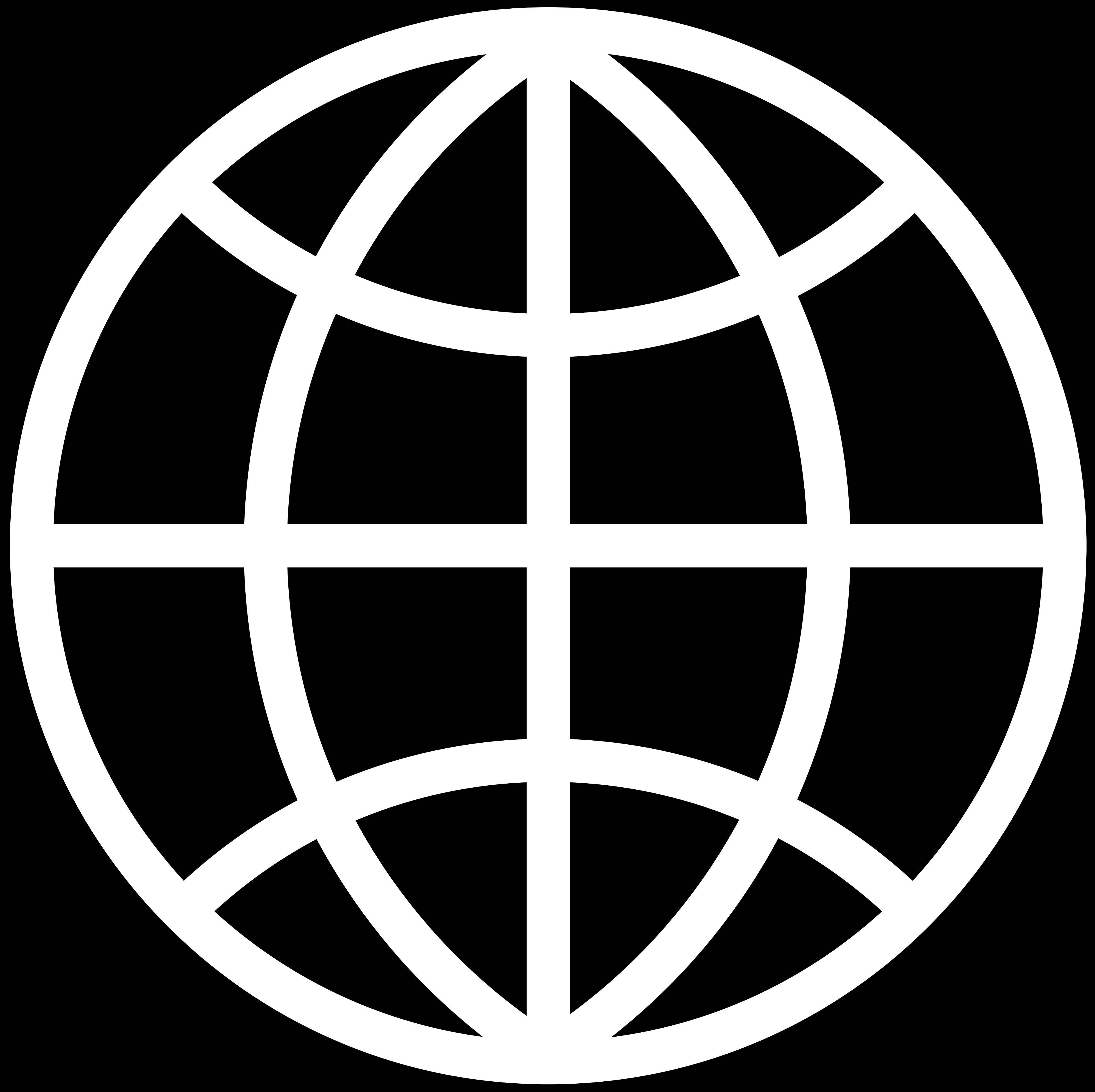


Interactions

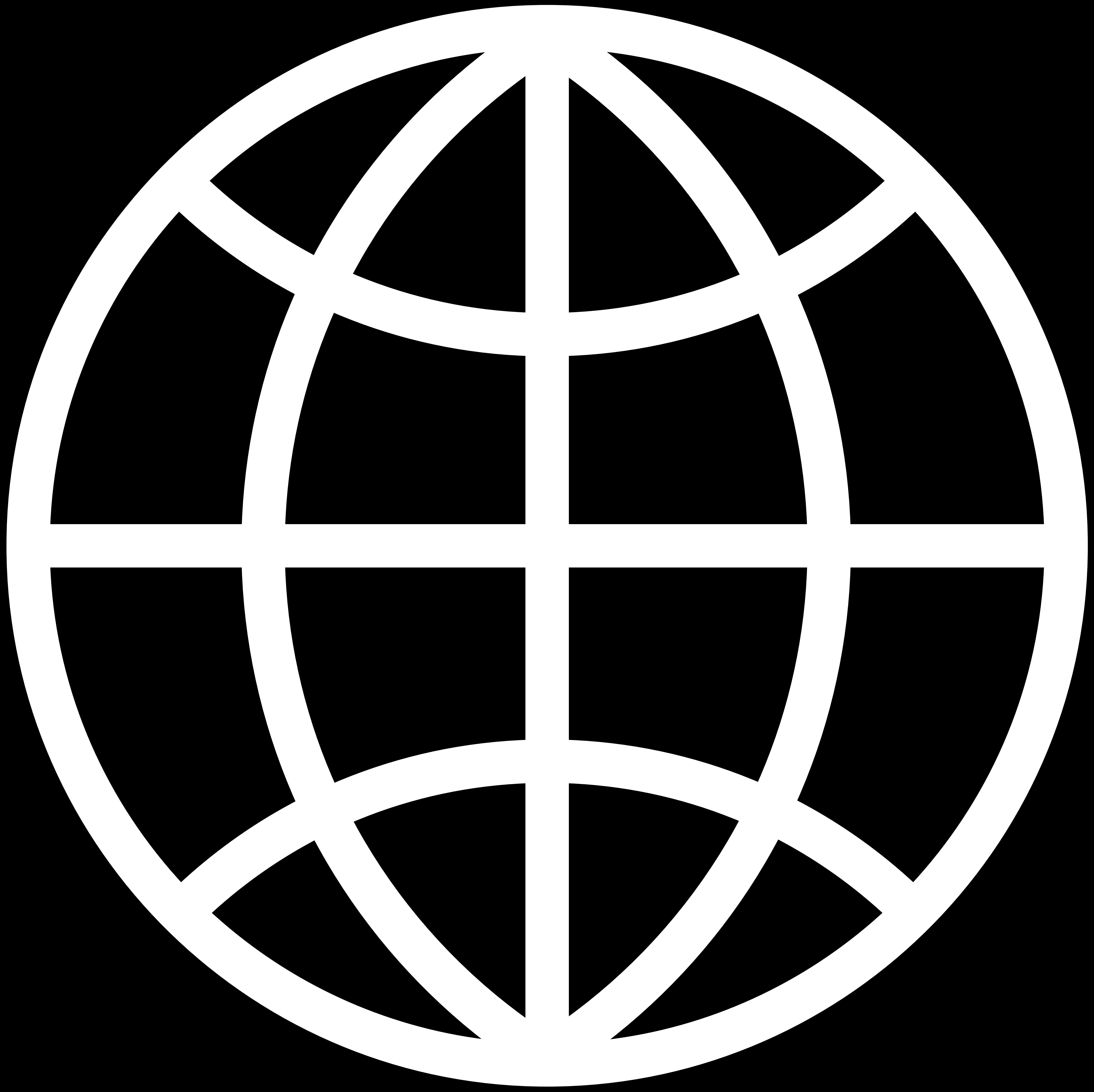




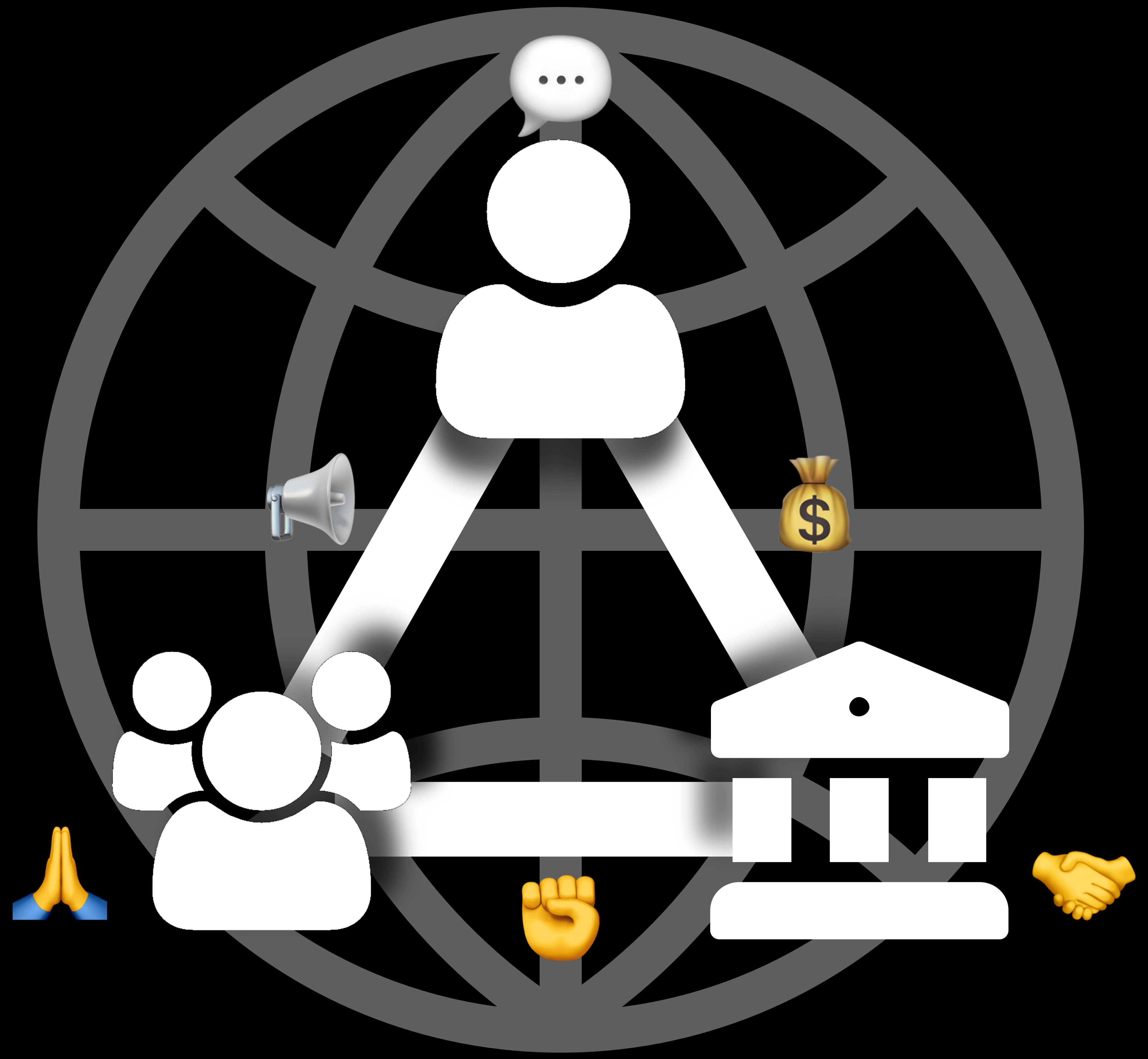
Interactions



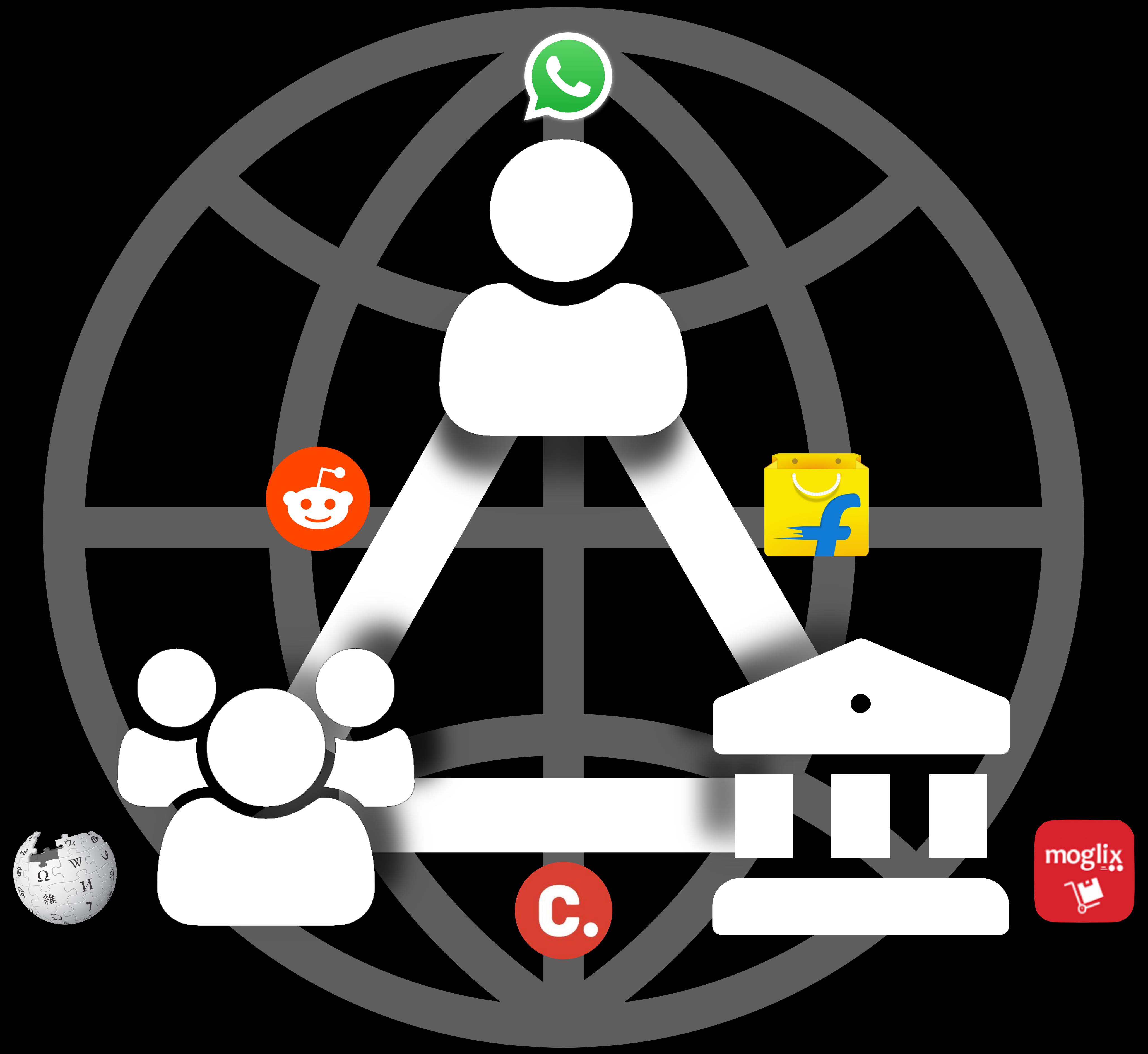
Online Interactions

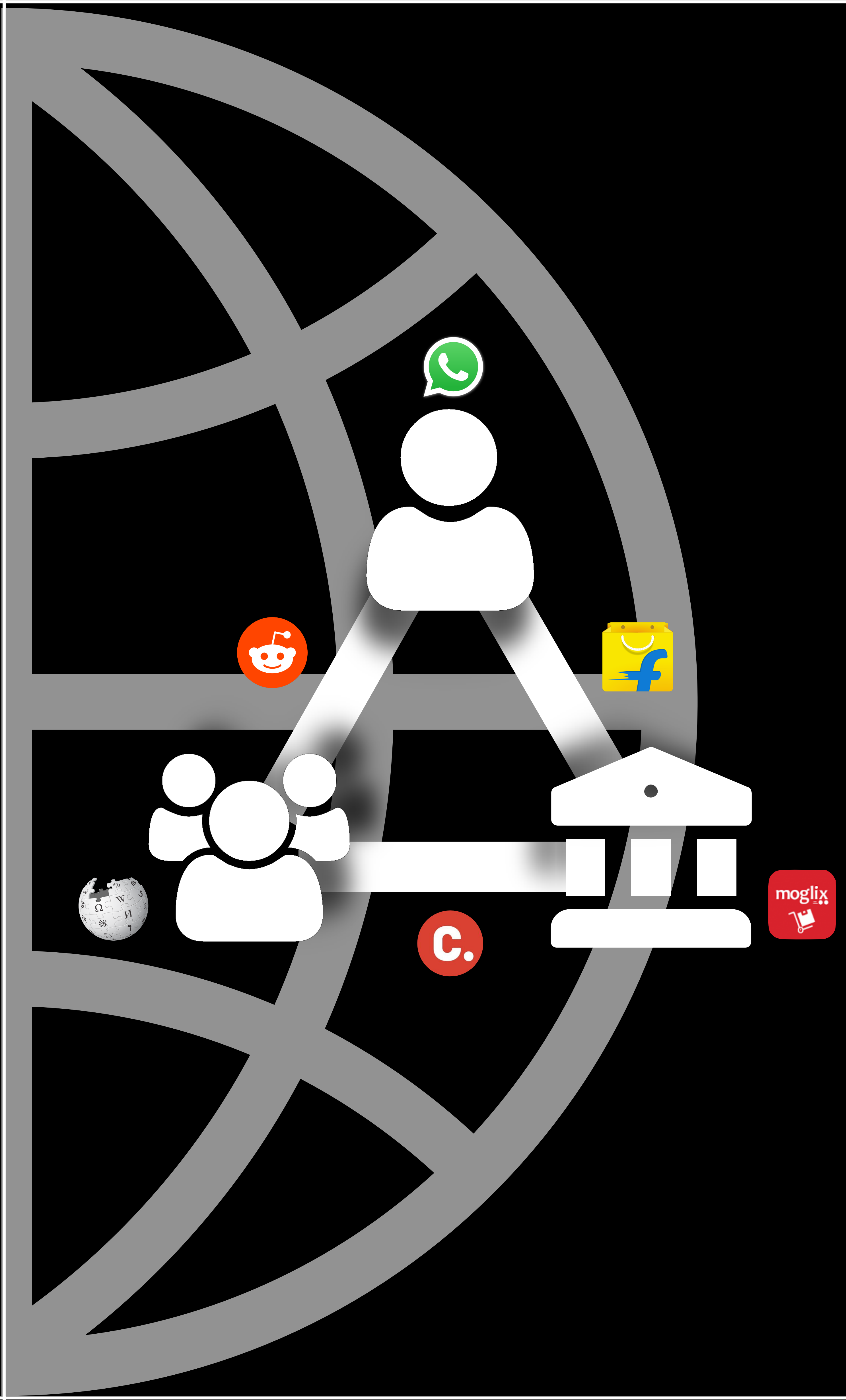


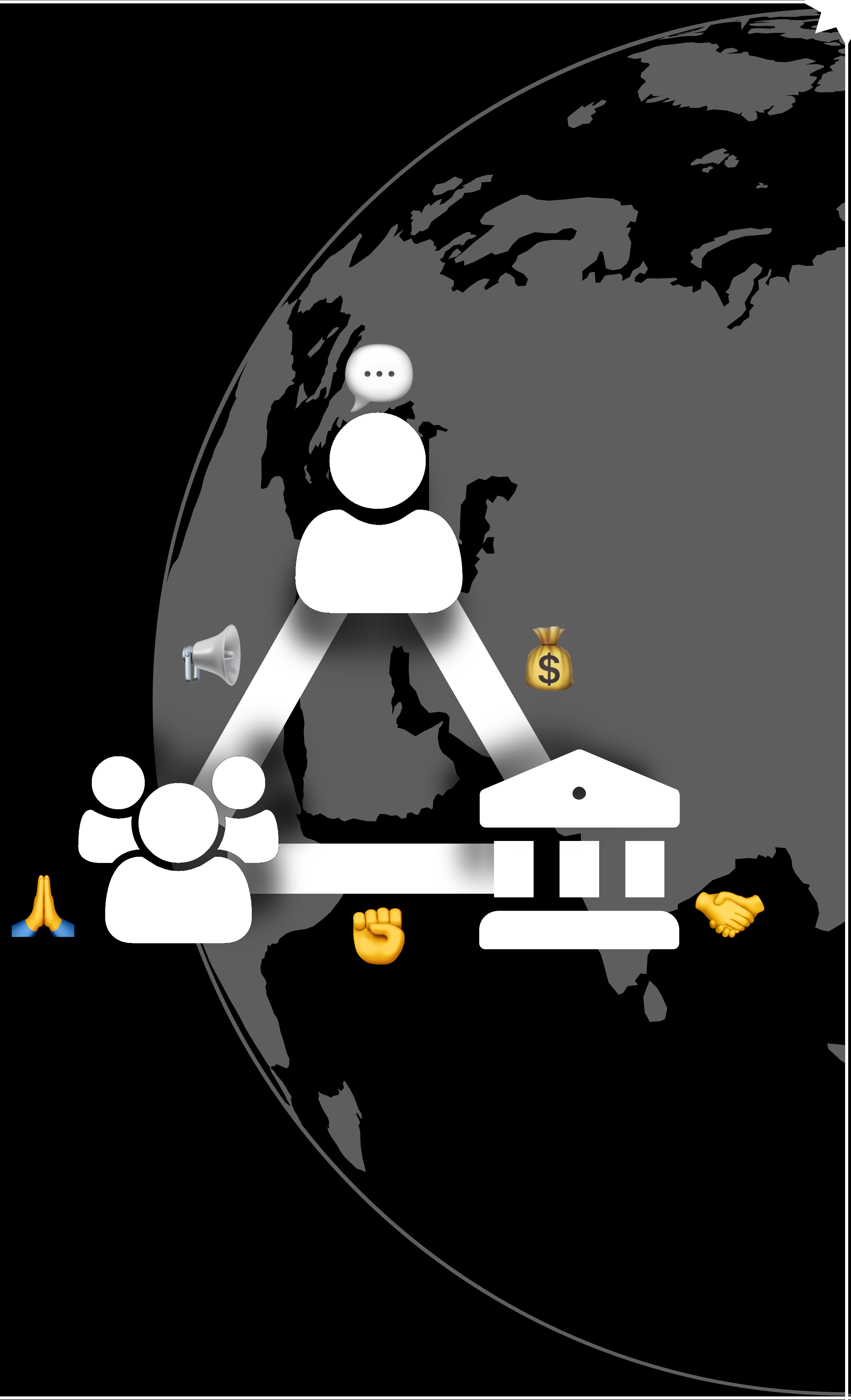
Online Interactions

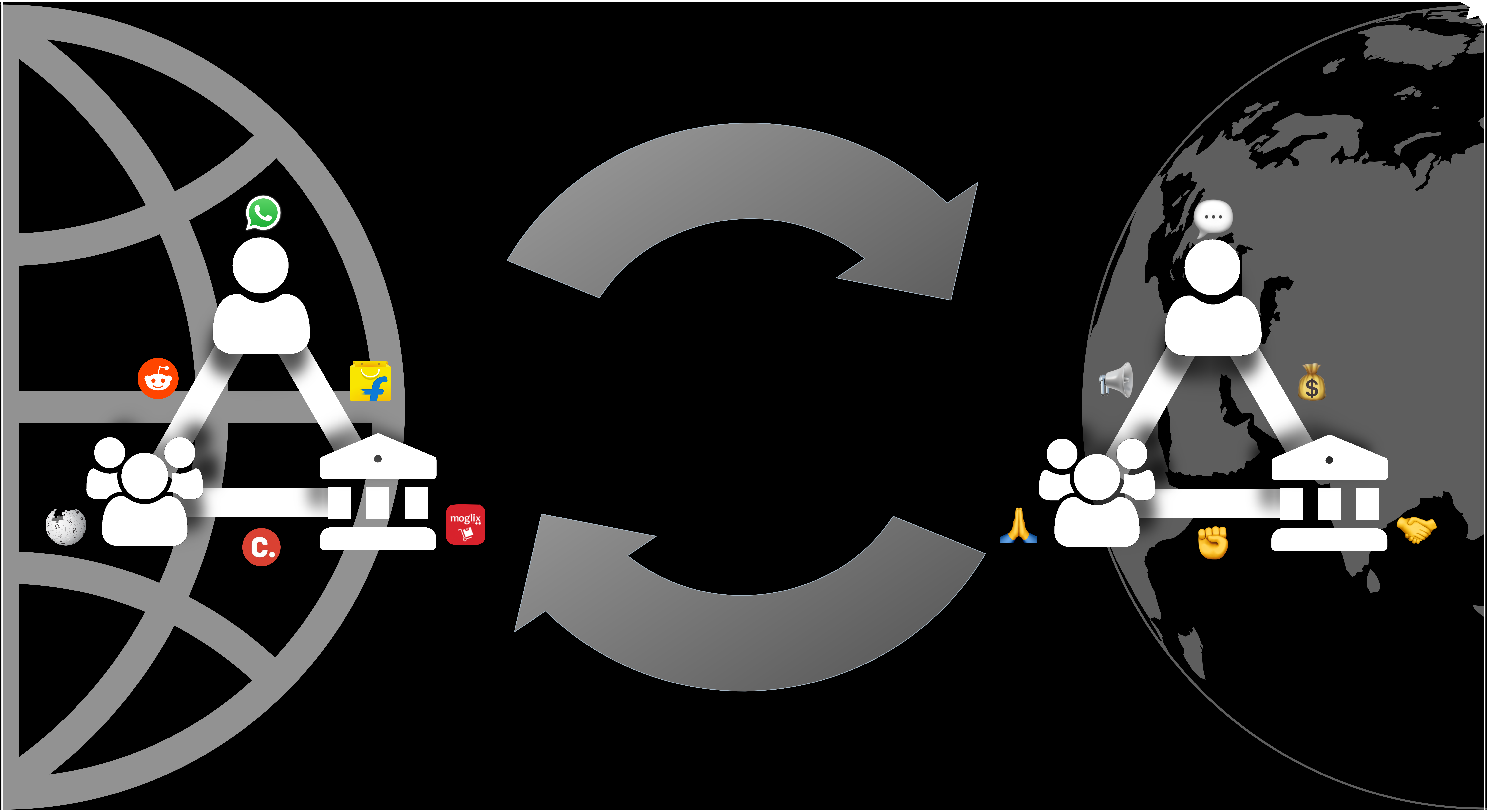


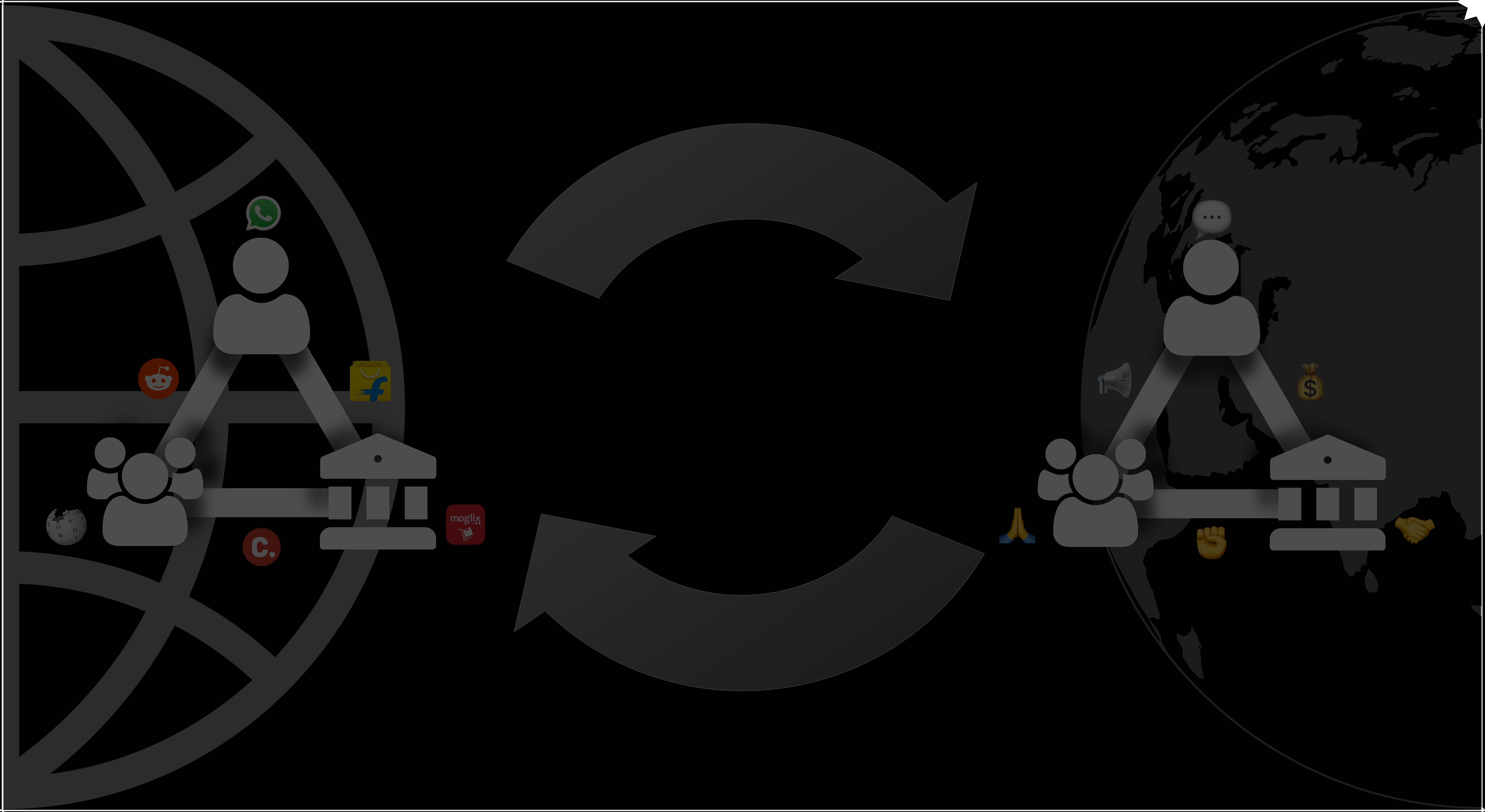
Online Interactions











Overflow

Overflow

Overflow - Positive

Overflow - Positive

Overflow - Positive

Overflow - Positive

Overflow - Negative

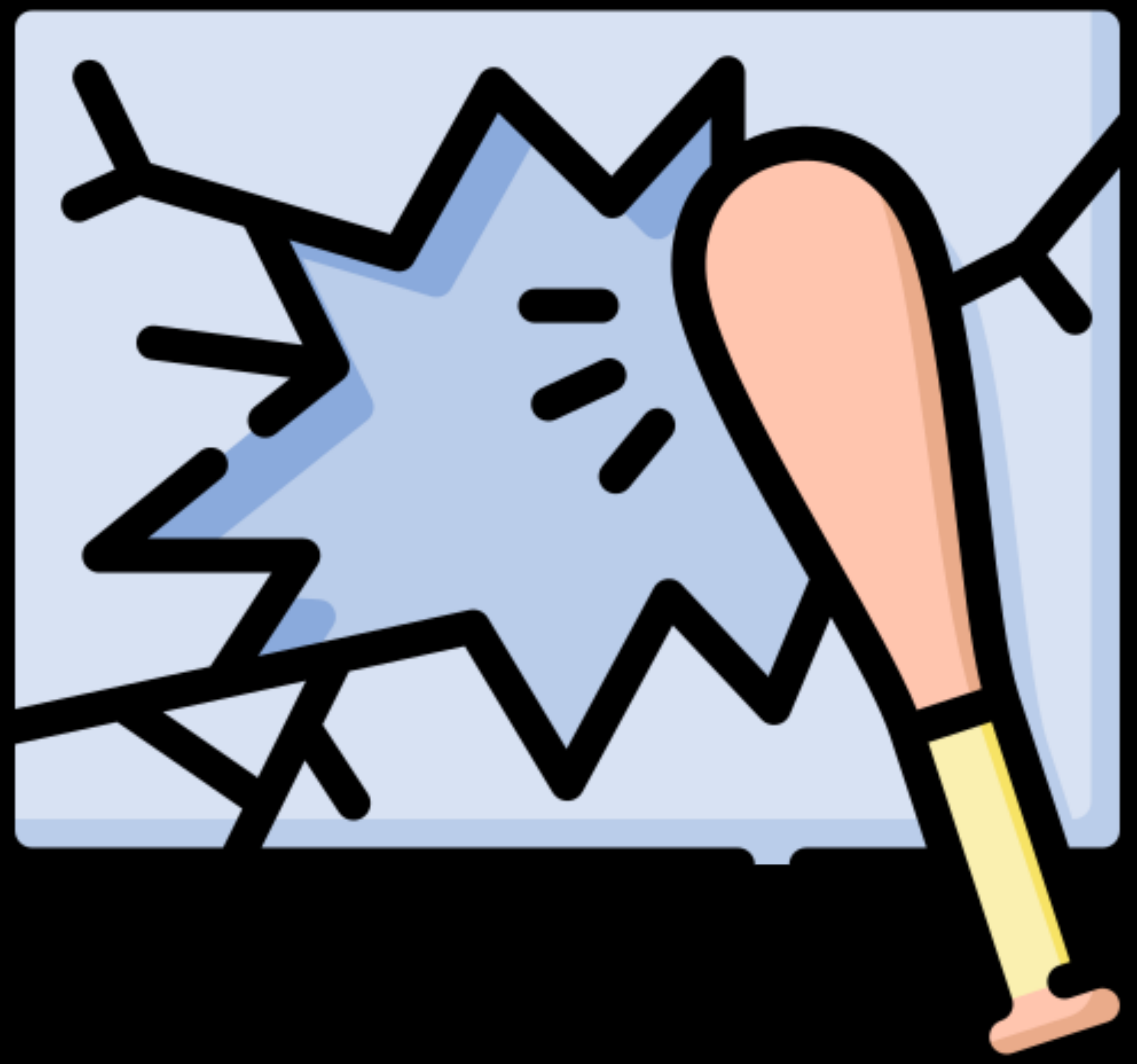
Overflow - Negative

Overflow - Negative

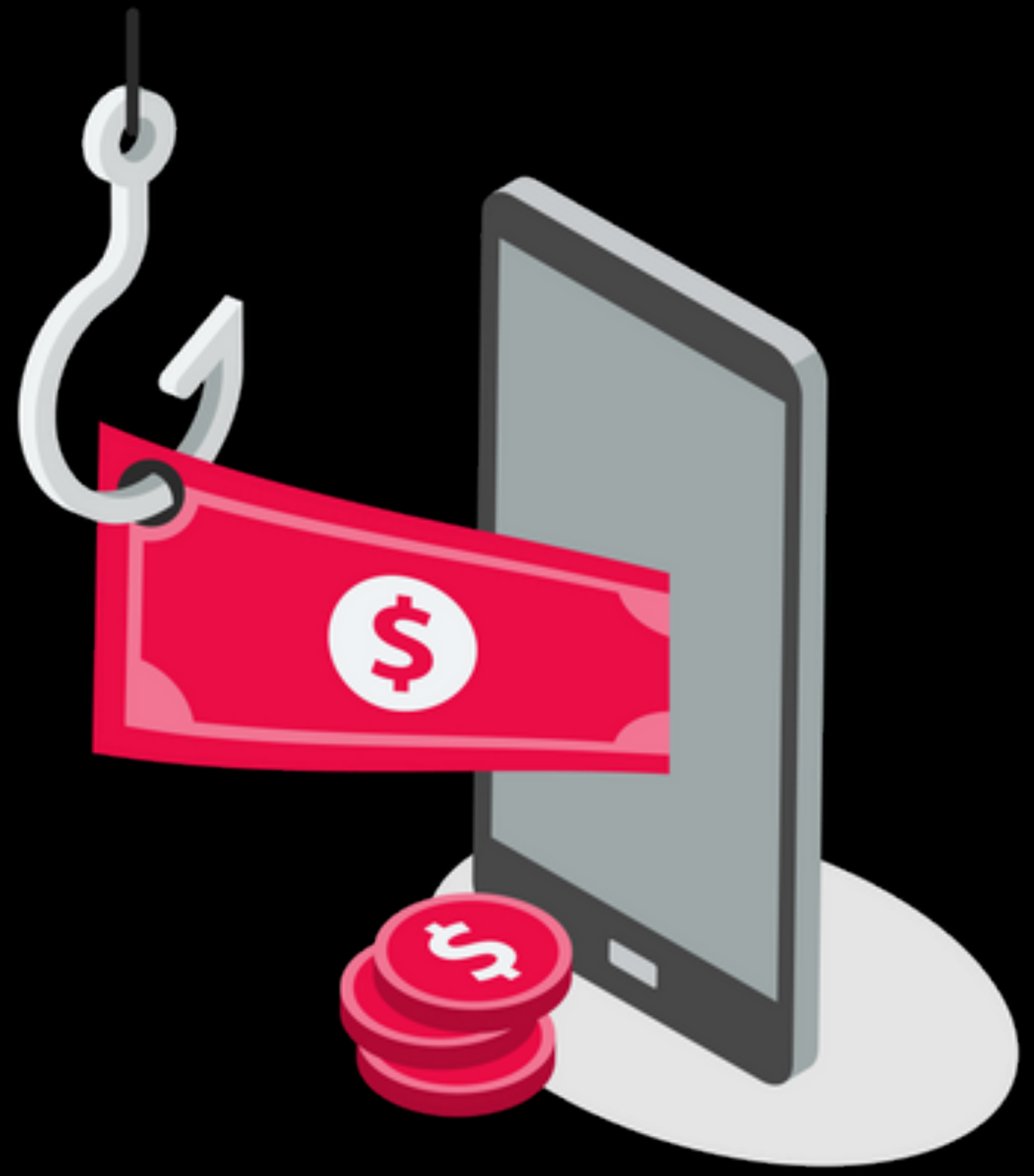
Overflow - Negative

Overflow - Negative

Overflow - Negative



Overflow - Negative



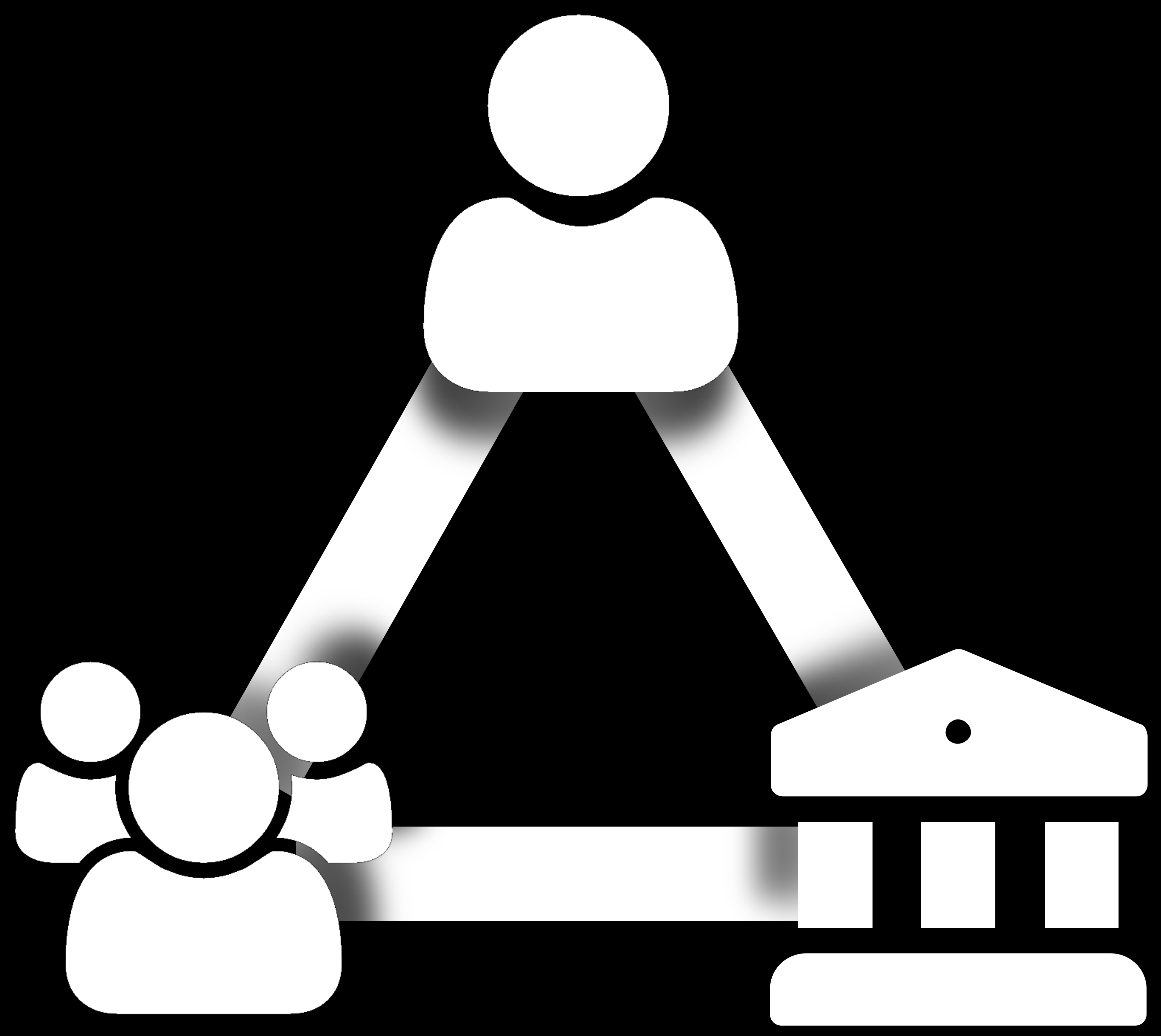
RESEARCH STATEMENT

Quantifying effects of online interactions in offline world and suggest implications

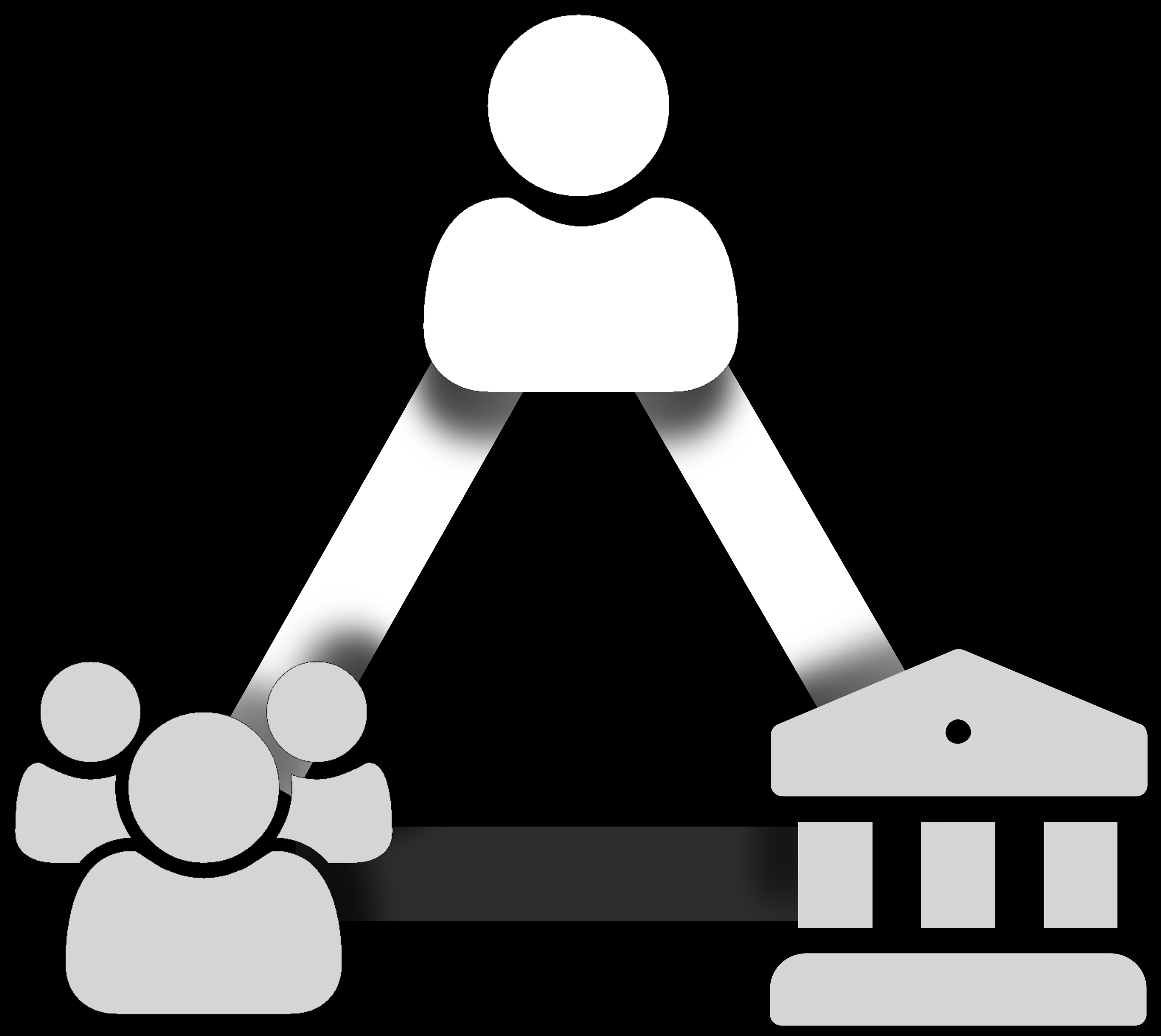
RESEARCH STATEMENT

Quantifying effects of online interactions in offline world and suggest **implications**

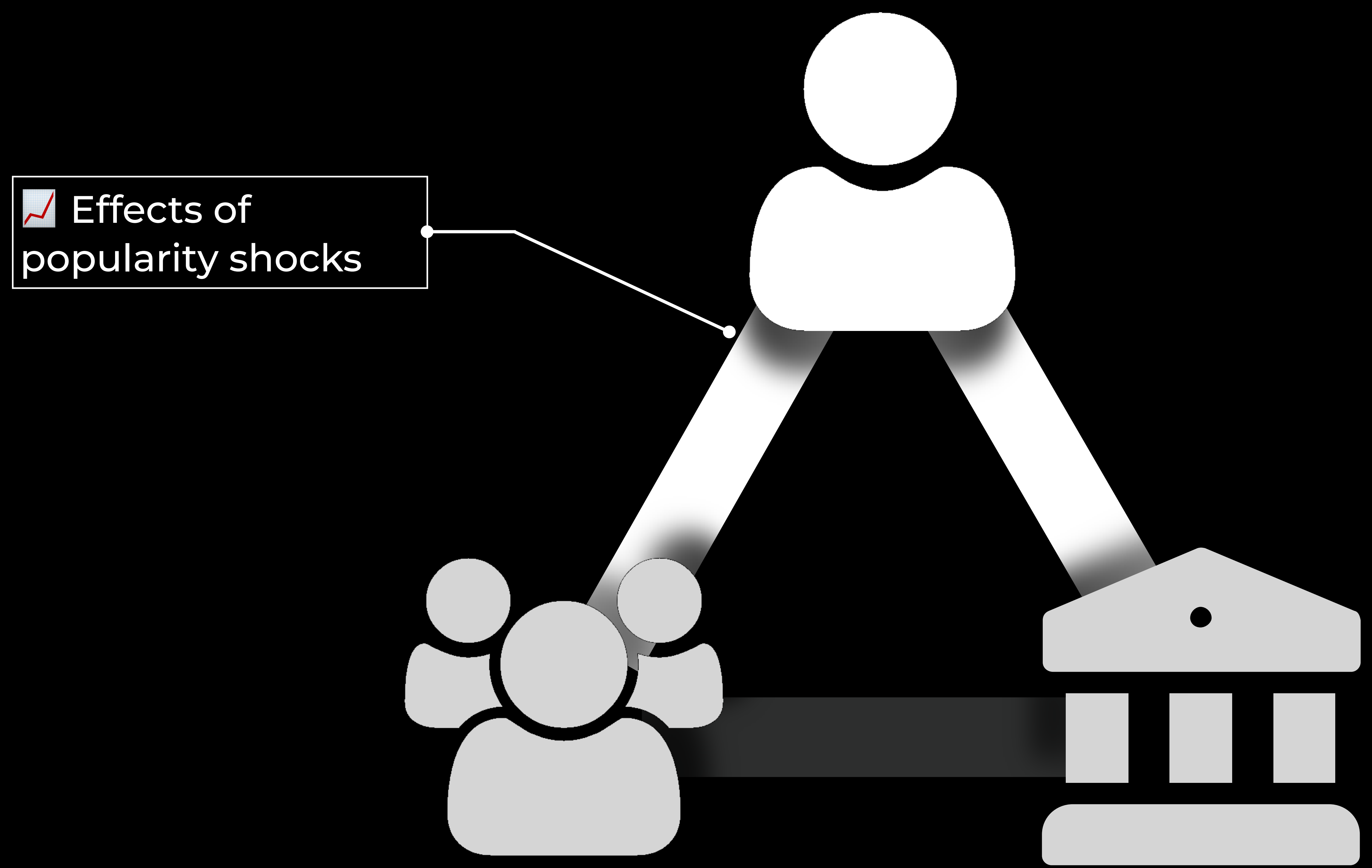
Our Focus



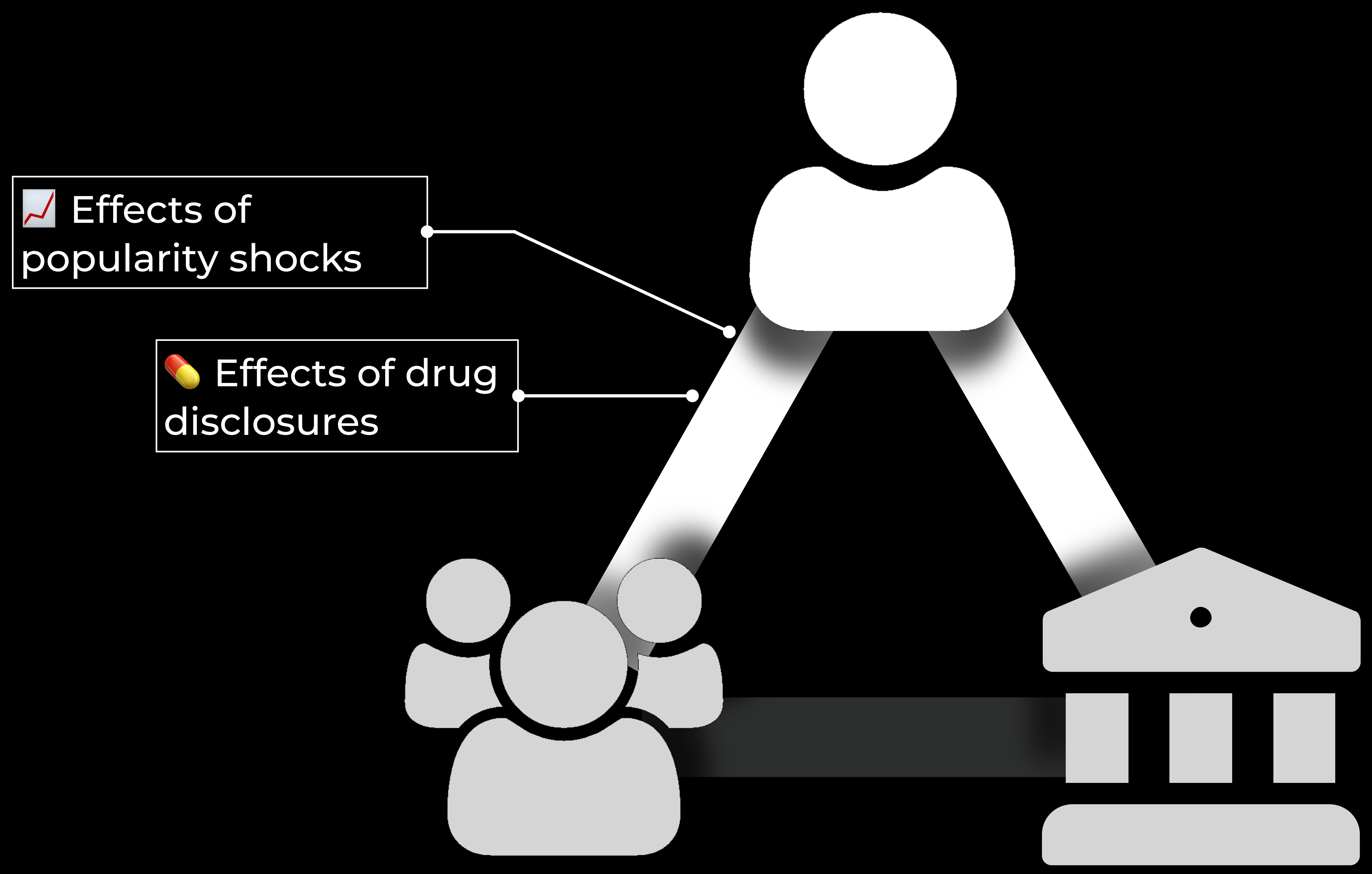
Our Focus



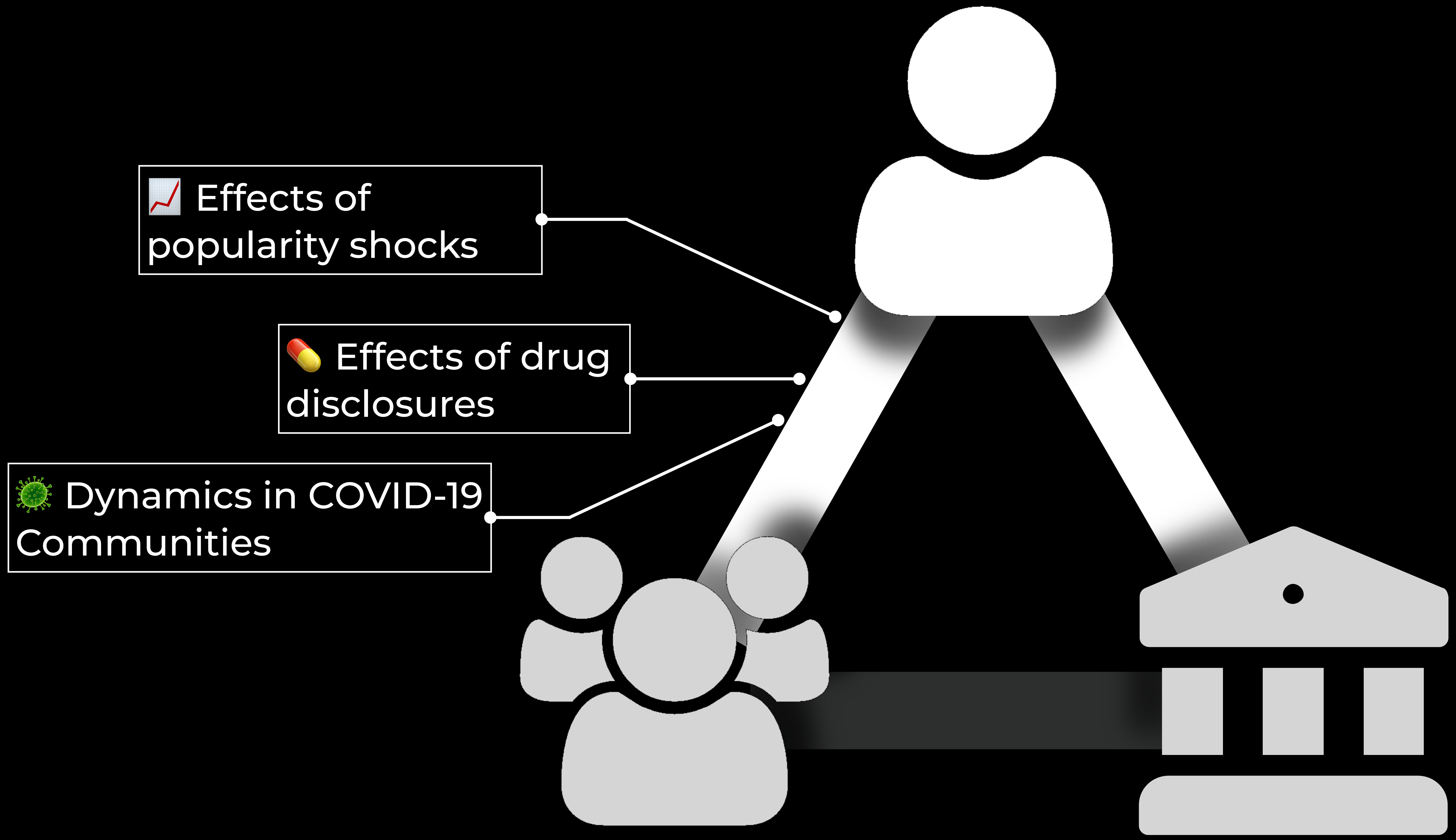
Our Focus



Our Focus

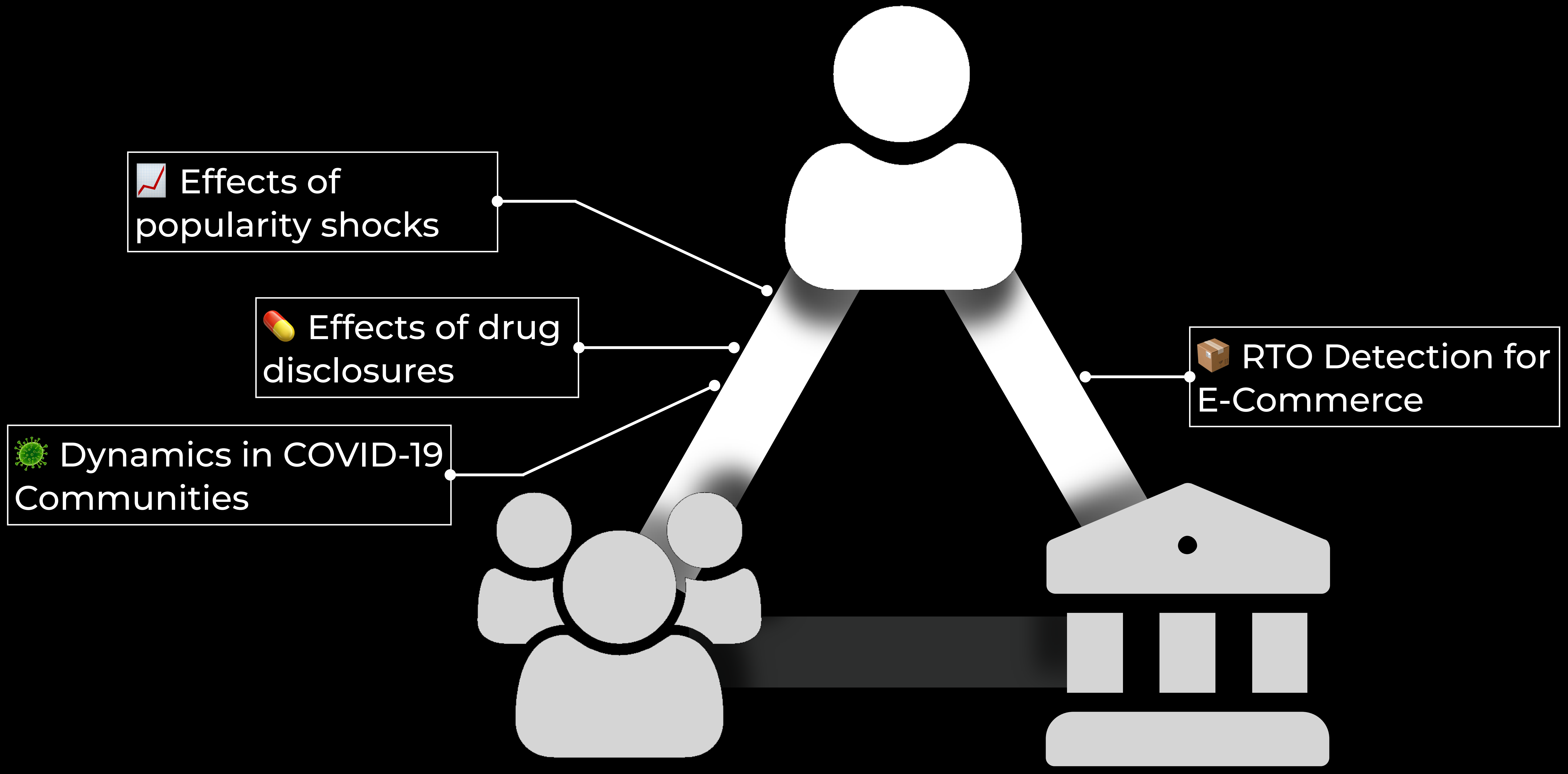


Our Focus

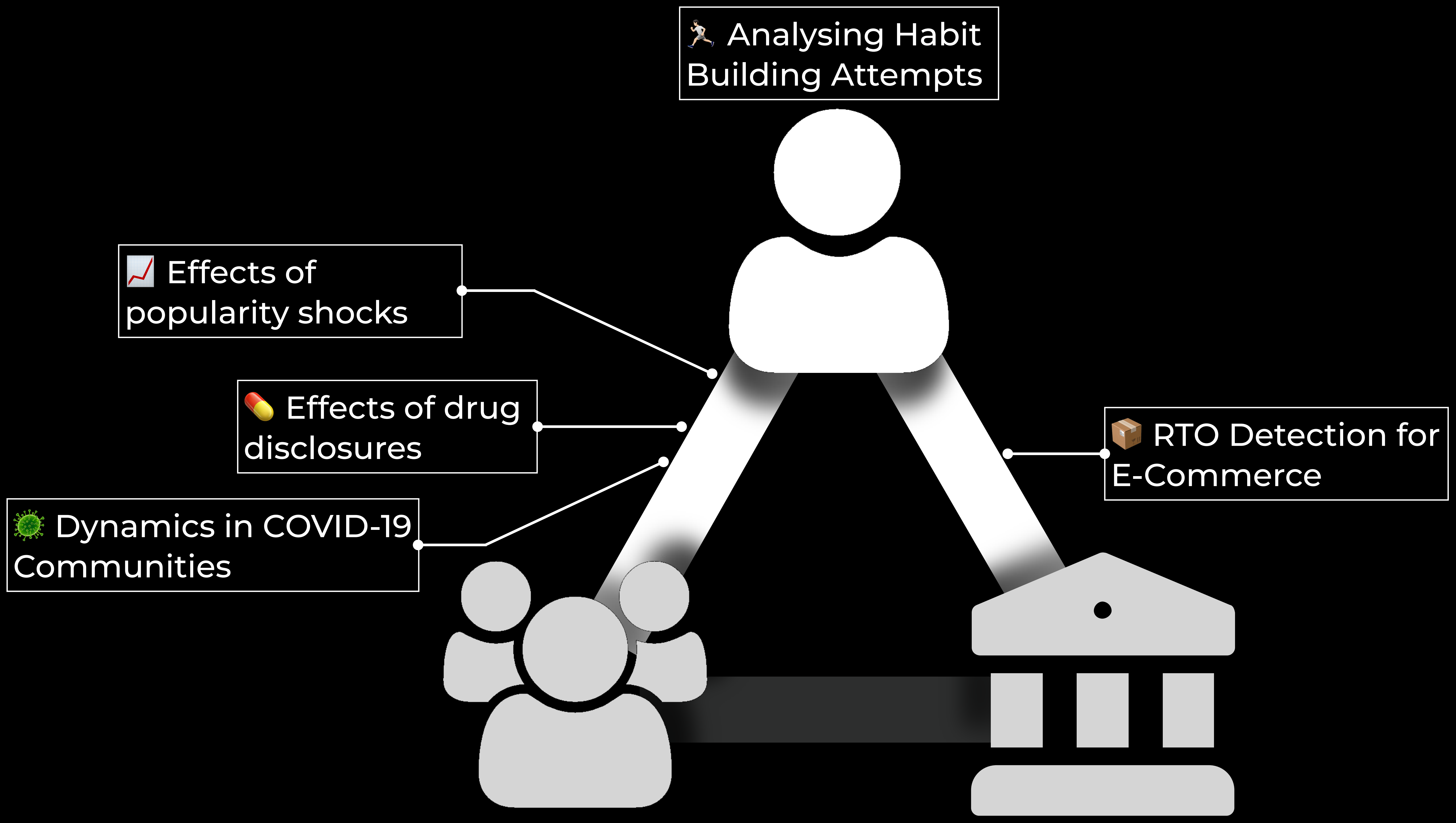


📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities

Our Focus

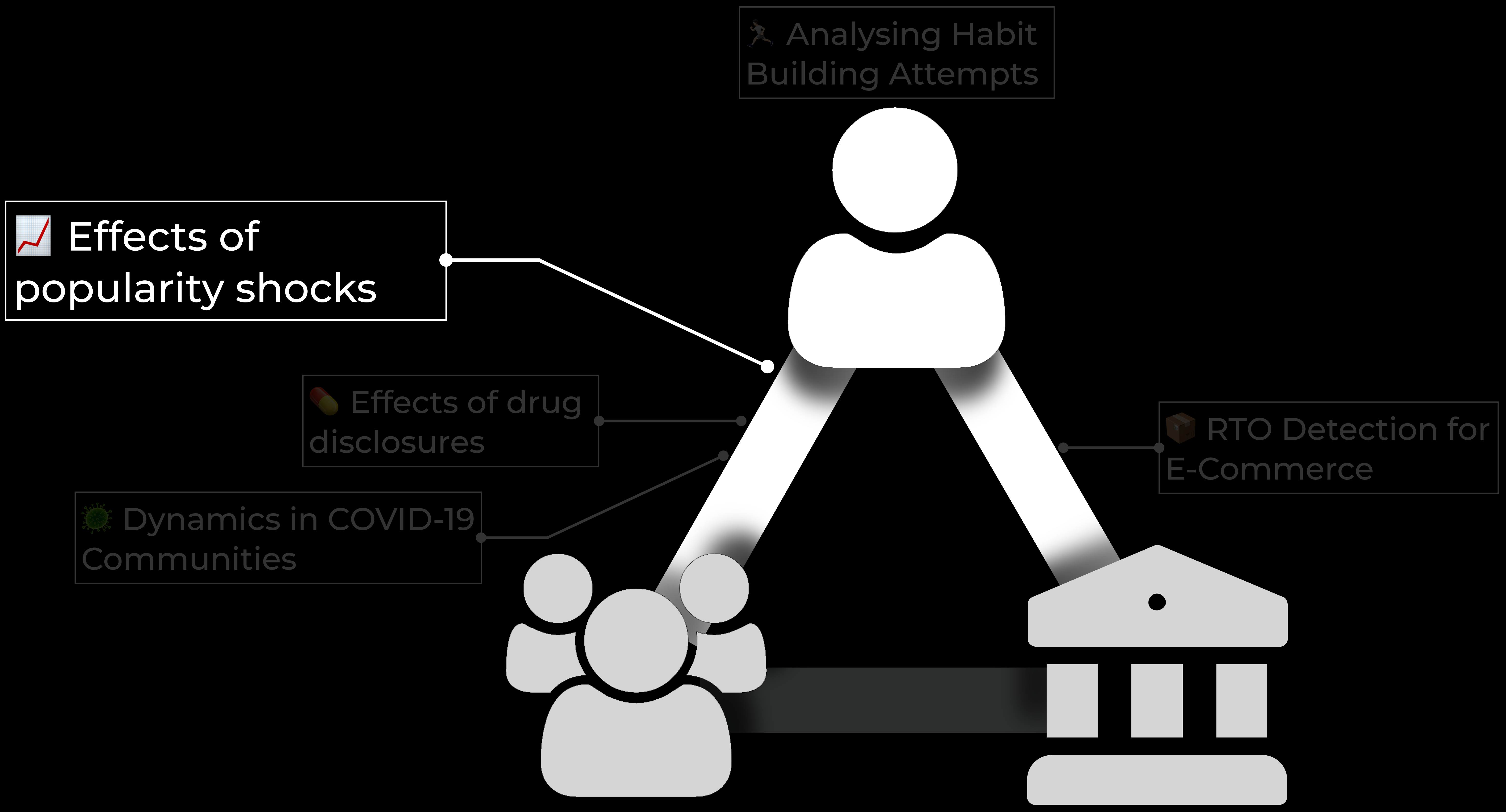

📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

Our Focus



🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

Our Focus


🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

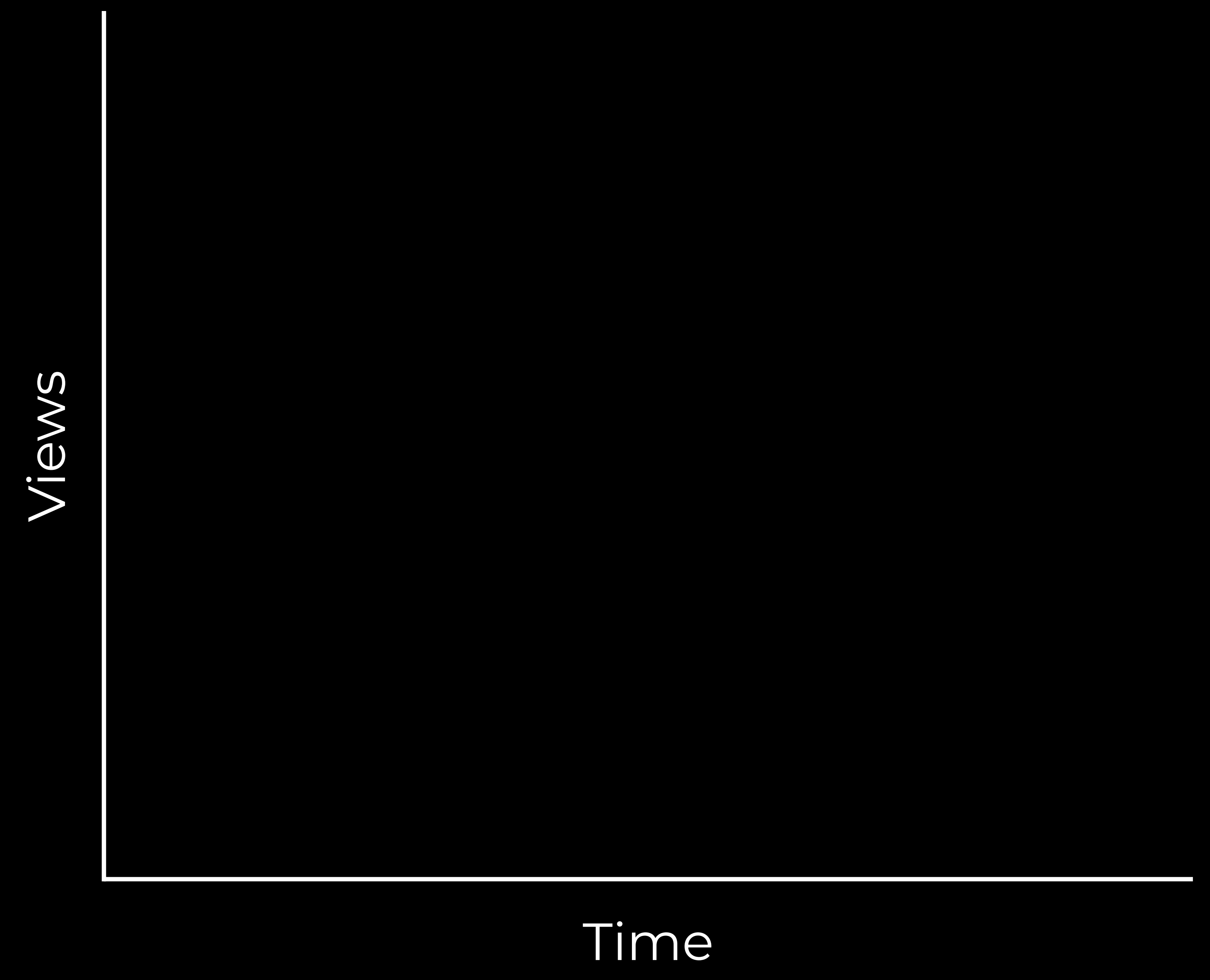
Effect of Popularity Shocks on User Behavior

ICWSM' 22

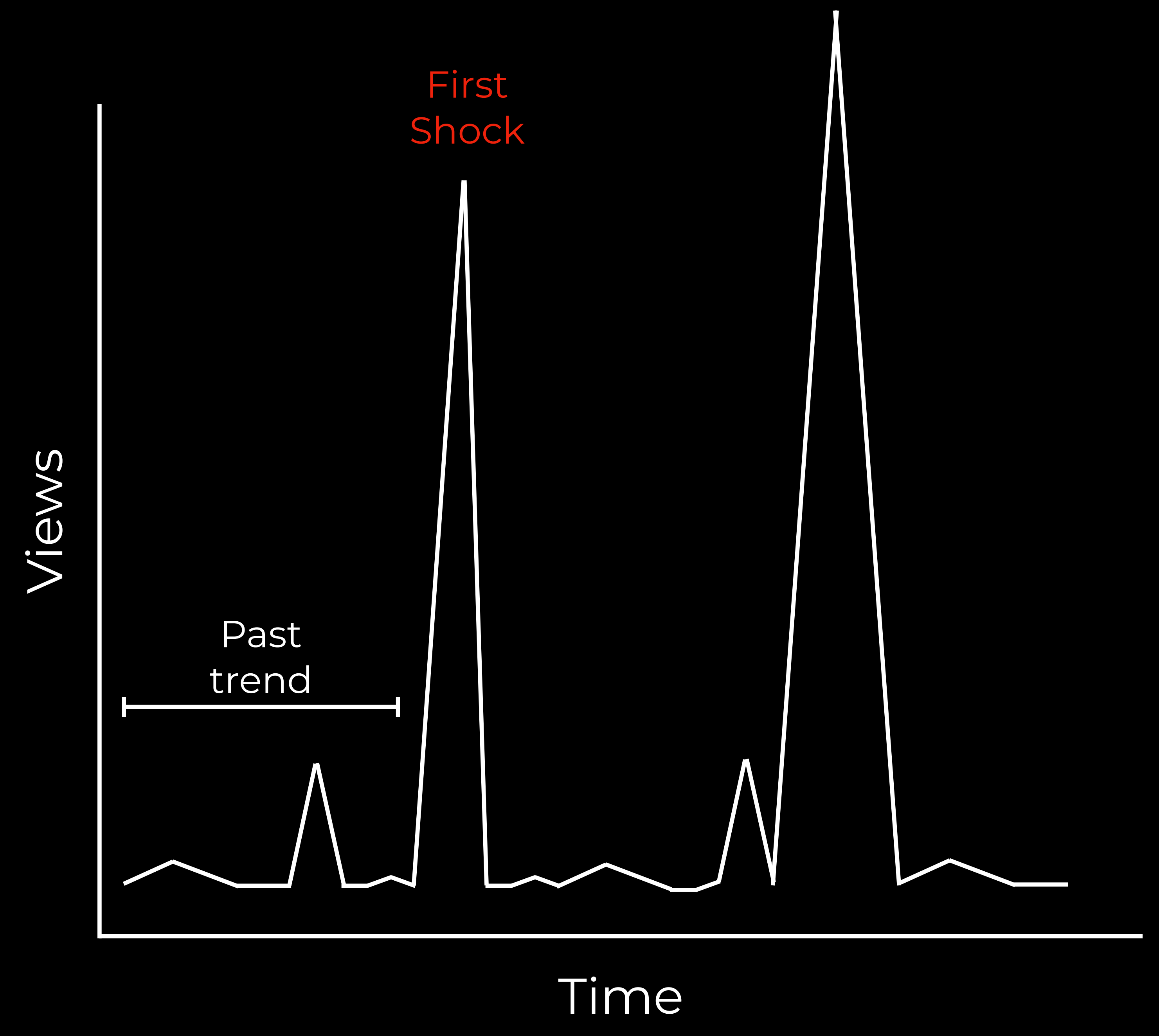
Omkar Gurjar, Tanmay Bansal, **Hitkul**, Hemank Lamba, Ponnurangam Kumaraguru

What is Popularity Shock?

What is Popularity Shock?



What is Popularity Shock?



Why is Shock Interesting?

Why is Shock Interesting?

@yashrajmukhate

Why is Shock Interesting?

@yashrajmukhate

Why is Shock Interesting?

Why is Shock Interesting?

Why is Shock Interesting?

@khaby00

Why is Shock Interesting?

@khaby00

Why is Shock Interesting?

Why is Shock Interesting?

Why is Shock Interesting?

Why is Shock Interesting?

Related Work

Related Work

Social Feedback

Positive reinforcement for improvements

Widely studied offline and online

Few online exceptions

Popularity Shock

Collaborative platforms - Wikipedia, Github

Exogenous shocks

Changes in network structures

Content Virality

Predicting Virality

Characterizing content

Information cascade

Research Questions

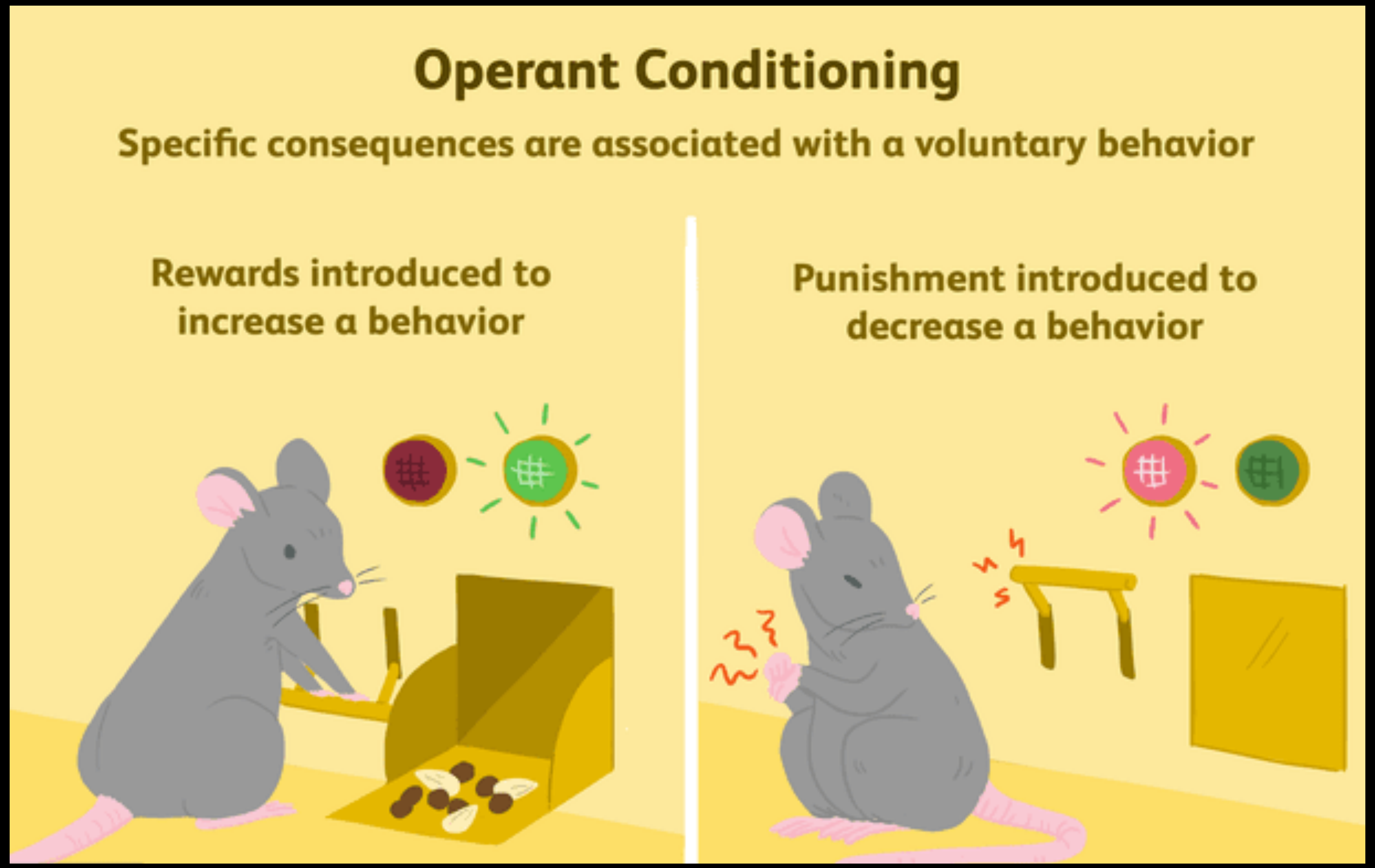
Economies of Online Cooperations
(Kollock et al. 1999)

Research Questions

Economies of Online Cooperations
(Kollock et al. 1999)

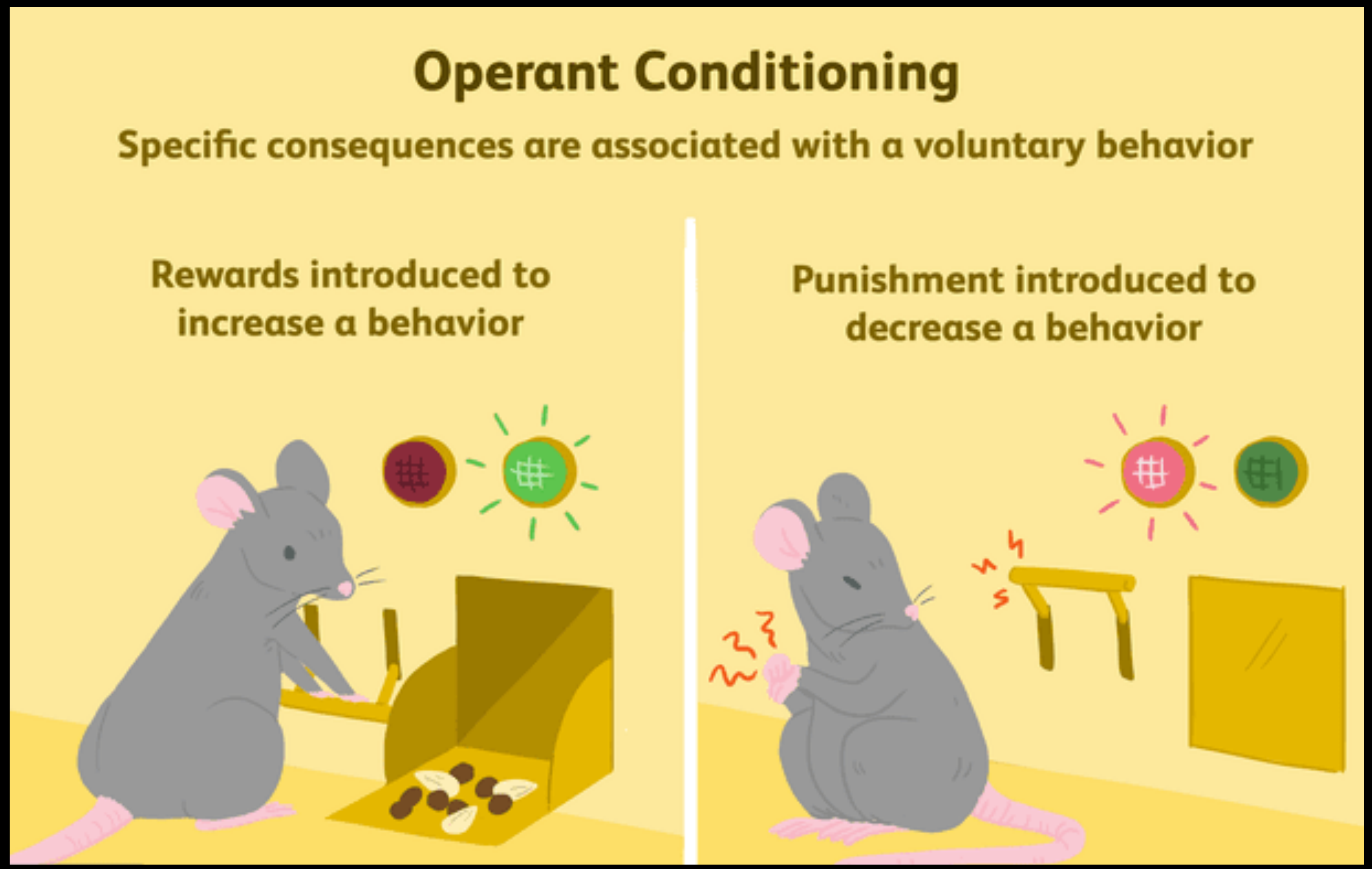
Do users increase their **posting frequency** after receiving popularity shocks?

Research Questions



Operant Conditioning
(Skinner, 1938)

Research Questions



Operant Conditioning
(Skinner, 1938)

Do users alter their **content** in response the Popularity shock?

Research Questions

Identifying Platform Effects in Social Media Data

Momin M. Malik[1] and **Jürgen Pfeffer**[1,2]
[1]Institute for Software Research
School of Computer Science
Carnegie Mellon University
[2]Bavarian School of Public Policy
Technical University of Munich

A Tempest in a Teacup?
Analyzing Firestorms on Twitter

Hemank Lamba*, Momin M. Malik*, and Jürgen Pfeffer
{hlamba,momin.malik,jpfeffer}@cs.cmu.edu
School of Computer Science
Carnegie Mellon University

Longevity in Social Networks

Research Questions

Identifying Platform Effects in Social Media Data

Momin M. Malik[1] and **Jürgen Pfeffer**[1,2]
[1]Institute for Software Research
School of Computer Science
Carnegie Mellon University
[2]Bavarian School of Public Policy
Technical University of Munich

A Tempest in a Teacup?
Analyzing Firestorms on Twitter

Hemank Lamba*, Momin M. Malik*, and Jürgen Pfeffer
{hlamba,momin.malik,jpfeffer}@cs.cmu.edu
School of Computer Science
Carnegie Mellon University

Longevity in Social Networks

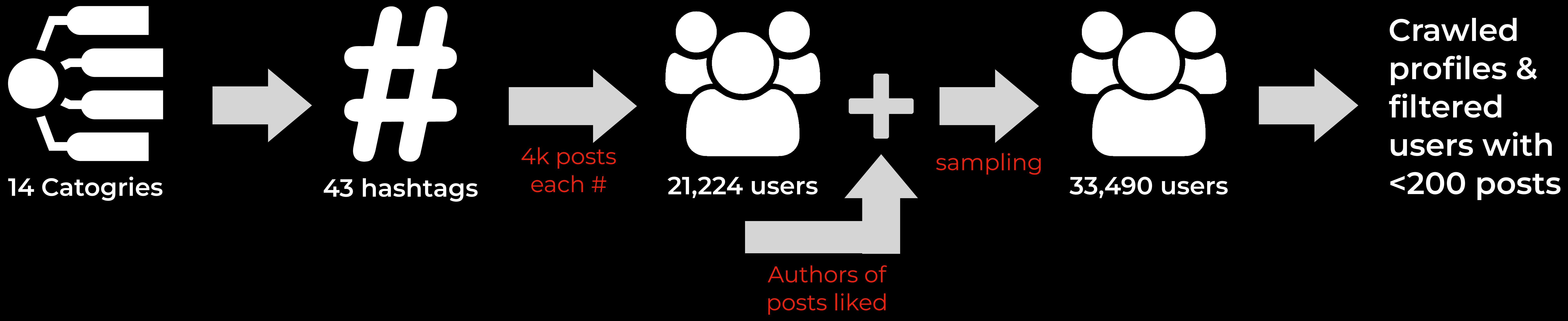
- How long do effect of popularity **shocks last**?
- What leads to **sustainability** of popularity?

Research Questions

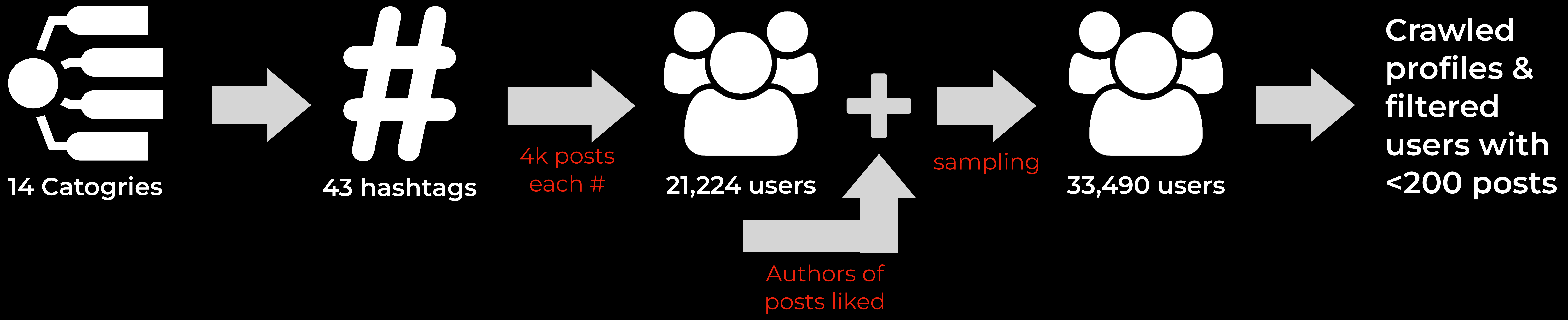
- Do users increase their **posting frequency** after receiving popularity shocks?
- Do users alter their **content** in response the Popularity shock?
- How long do effect of popularity **shocks last**?
- What leads to **sustainability** of popularity?

Data Collection

Data Collection



Data Collection



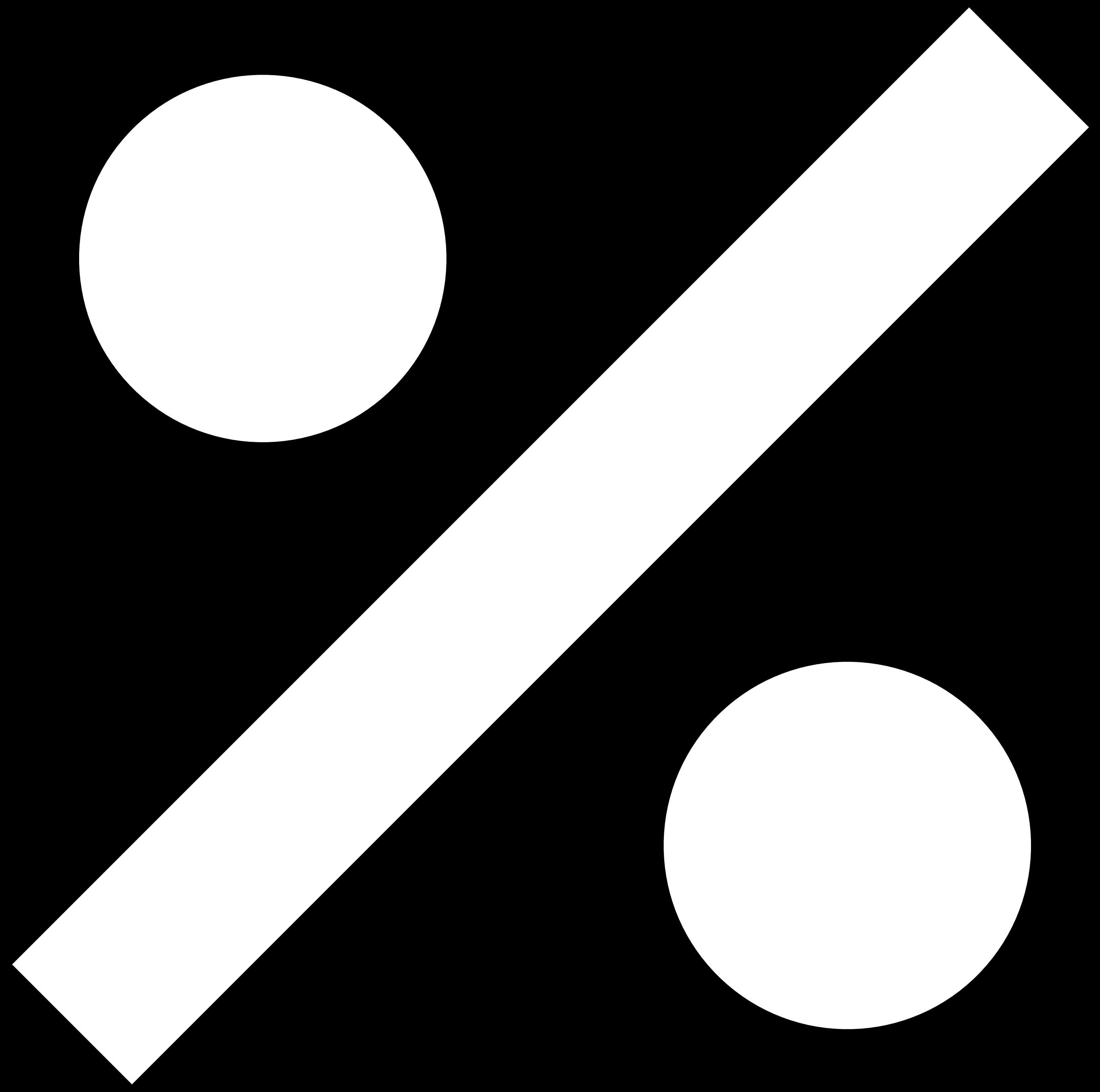
30,969 users

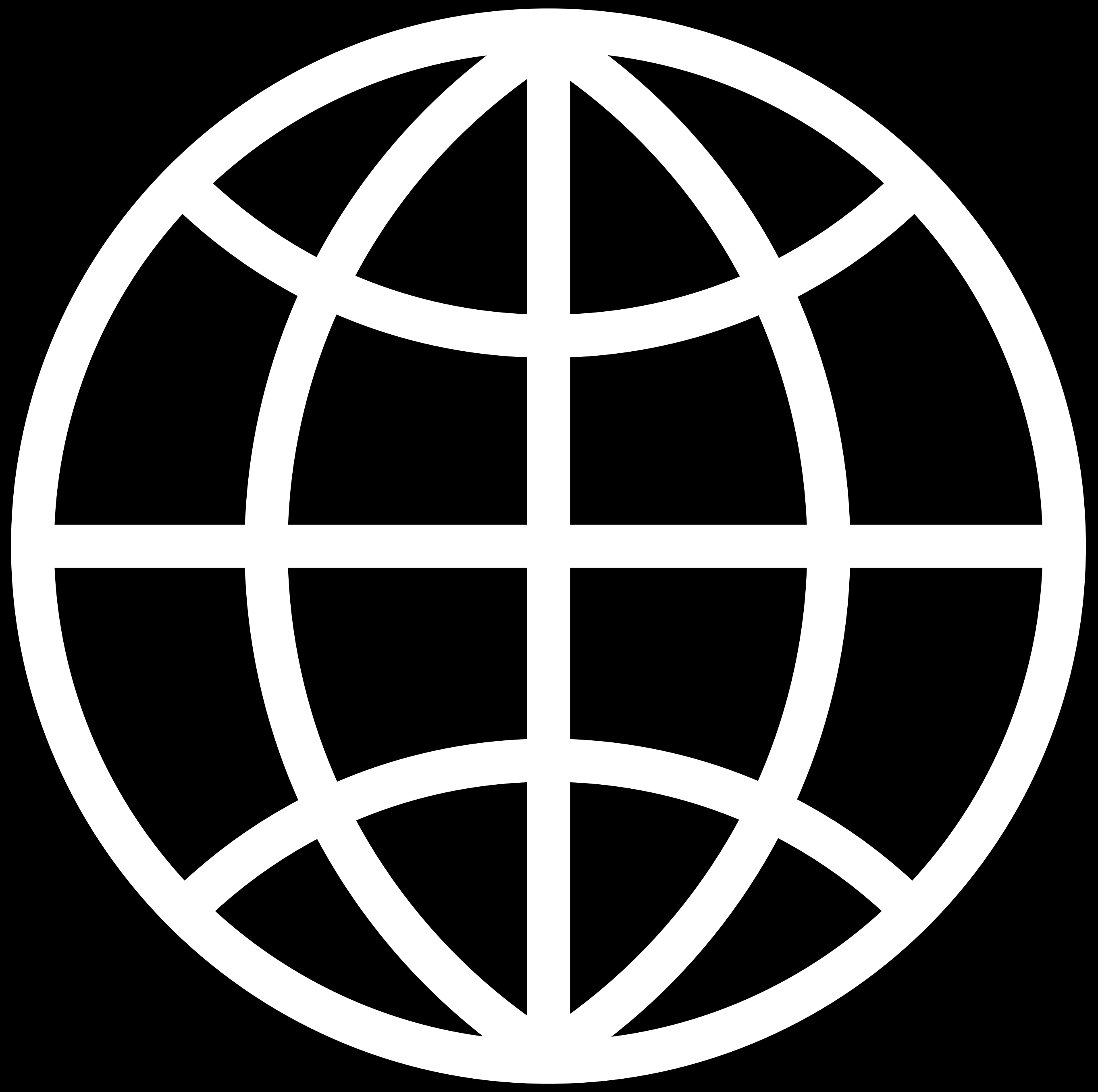
20 million Posts

First post: 07/01/2015
Last post: 31/12/2021

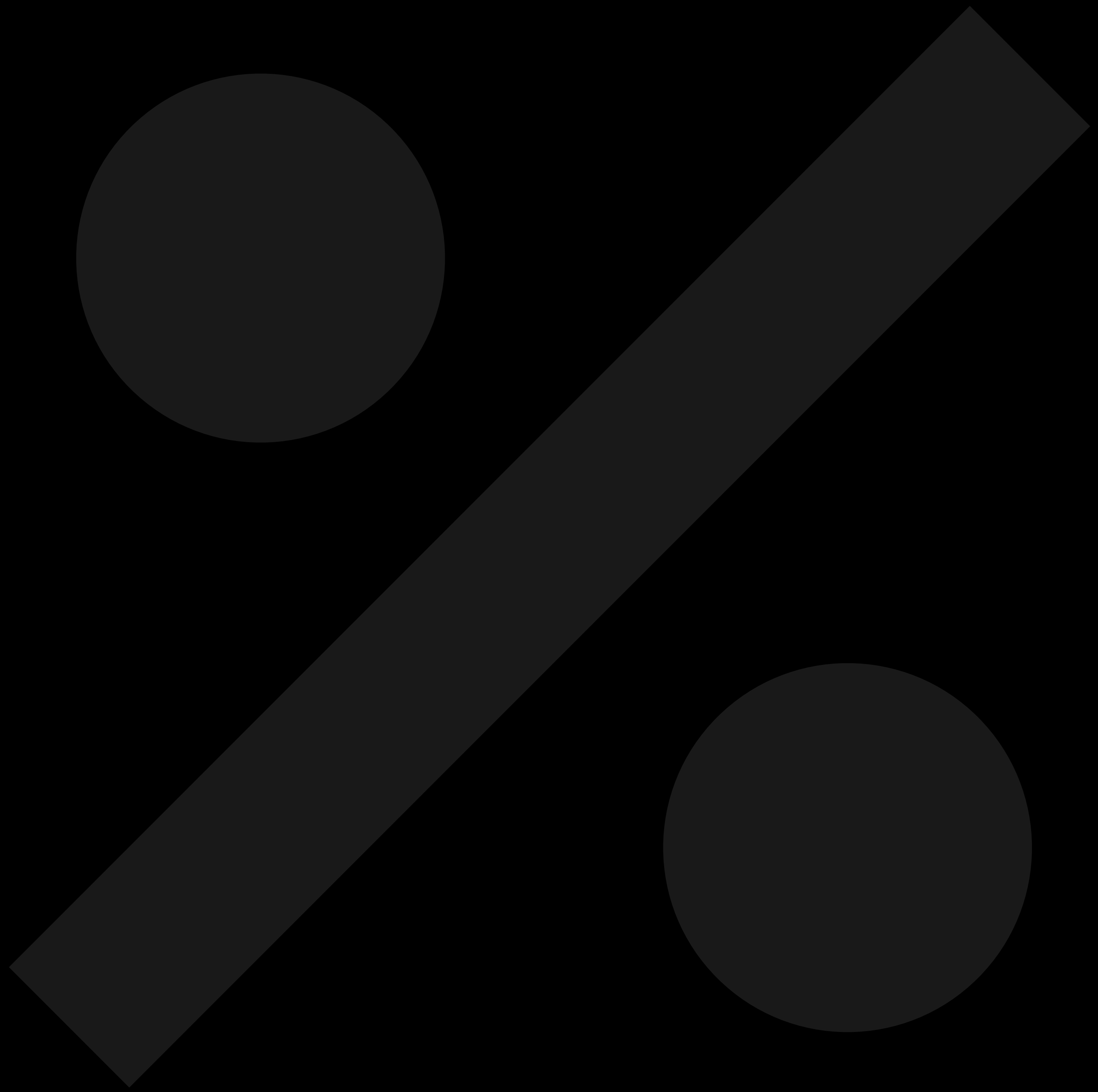
Detecting Shocks

Detecting Shocks





Detecting Shocks



Detecting Shocks

Z-Score

Detecting Shocks

Z-Score

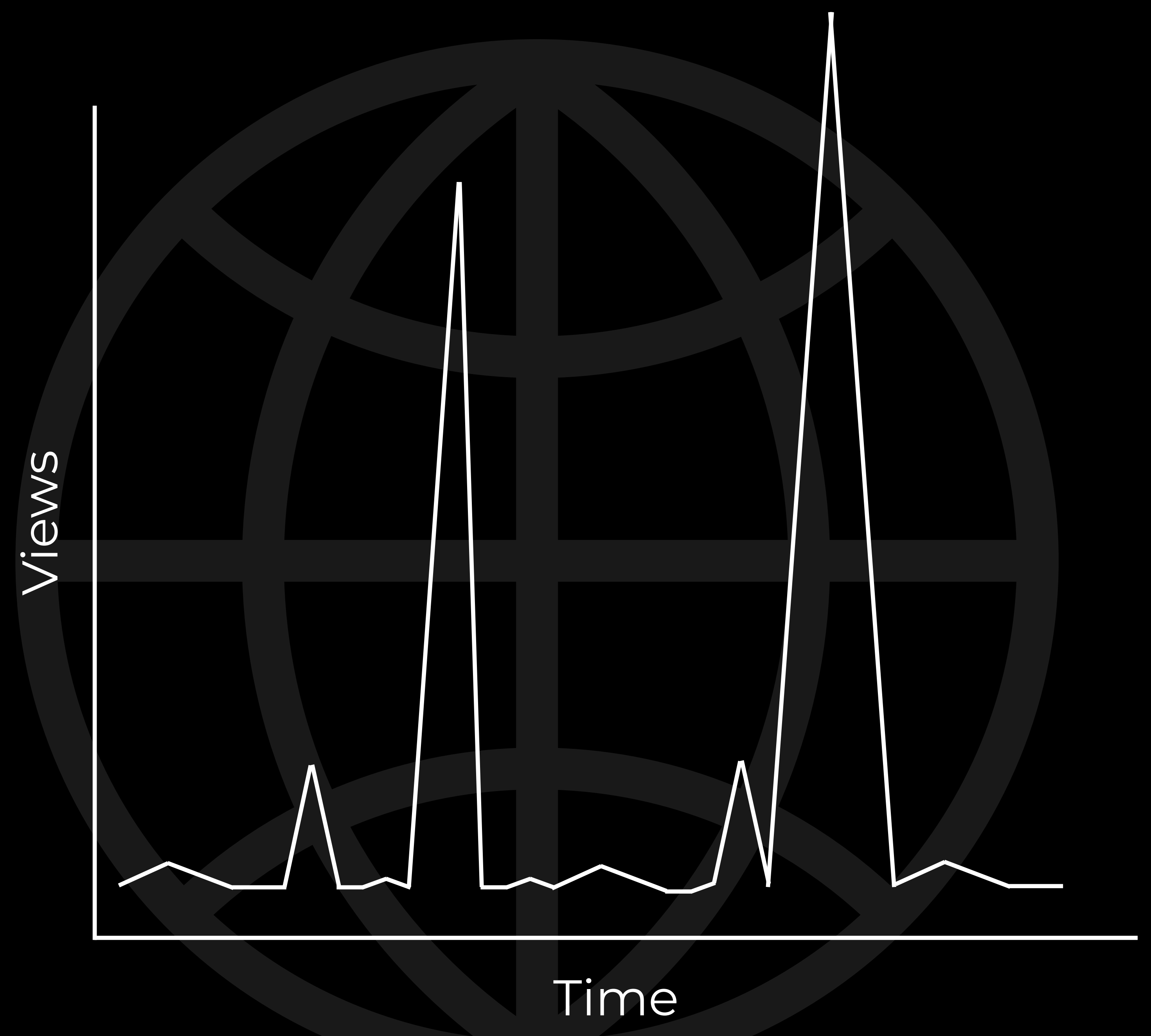
Fb’s Prophet

Detecting Shocks

Z-Score

Fb’s Prophet

Proposed

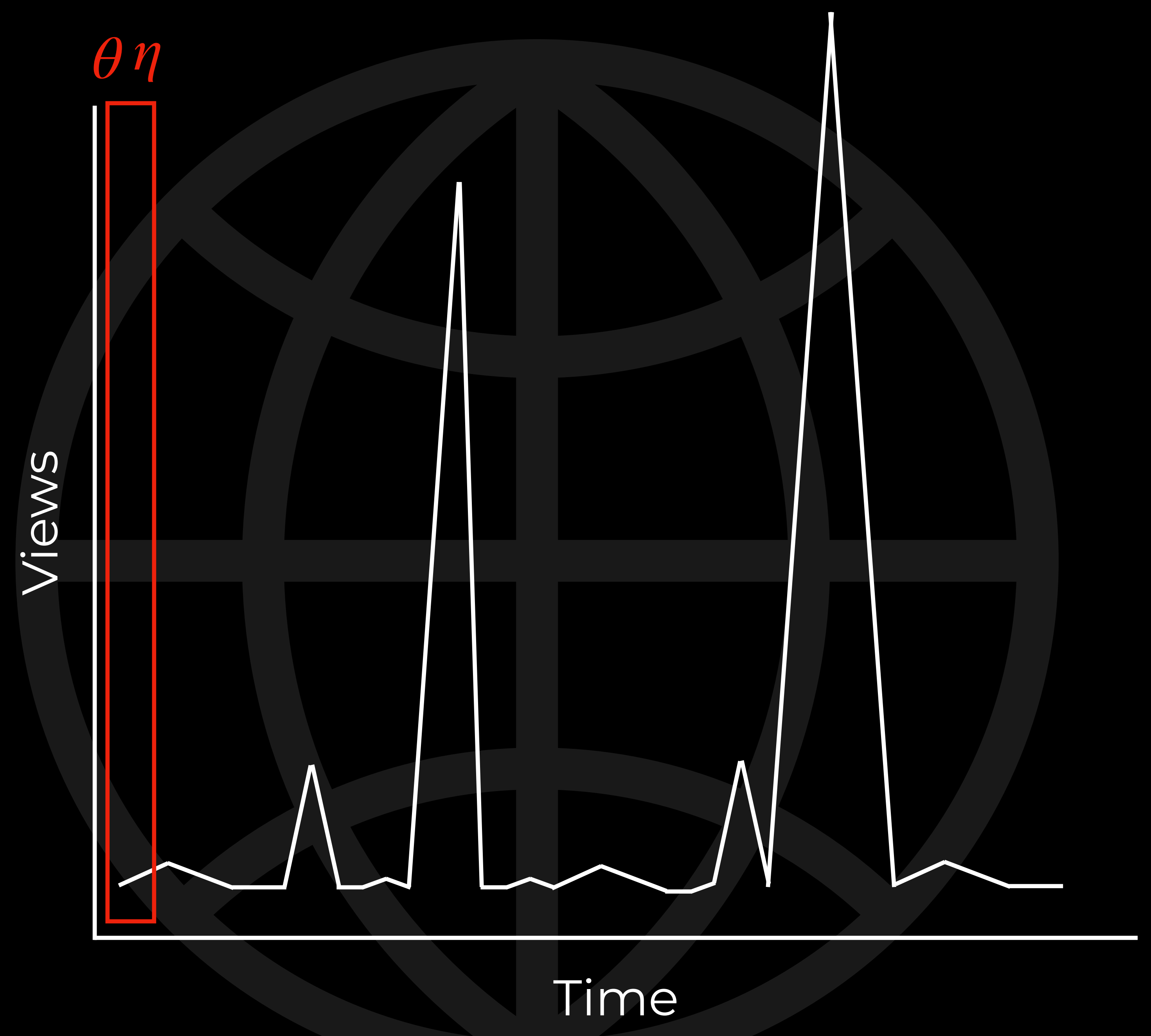


Detecting Shocks

Z-Score

Fb's Prophet

Proposed

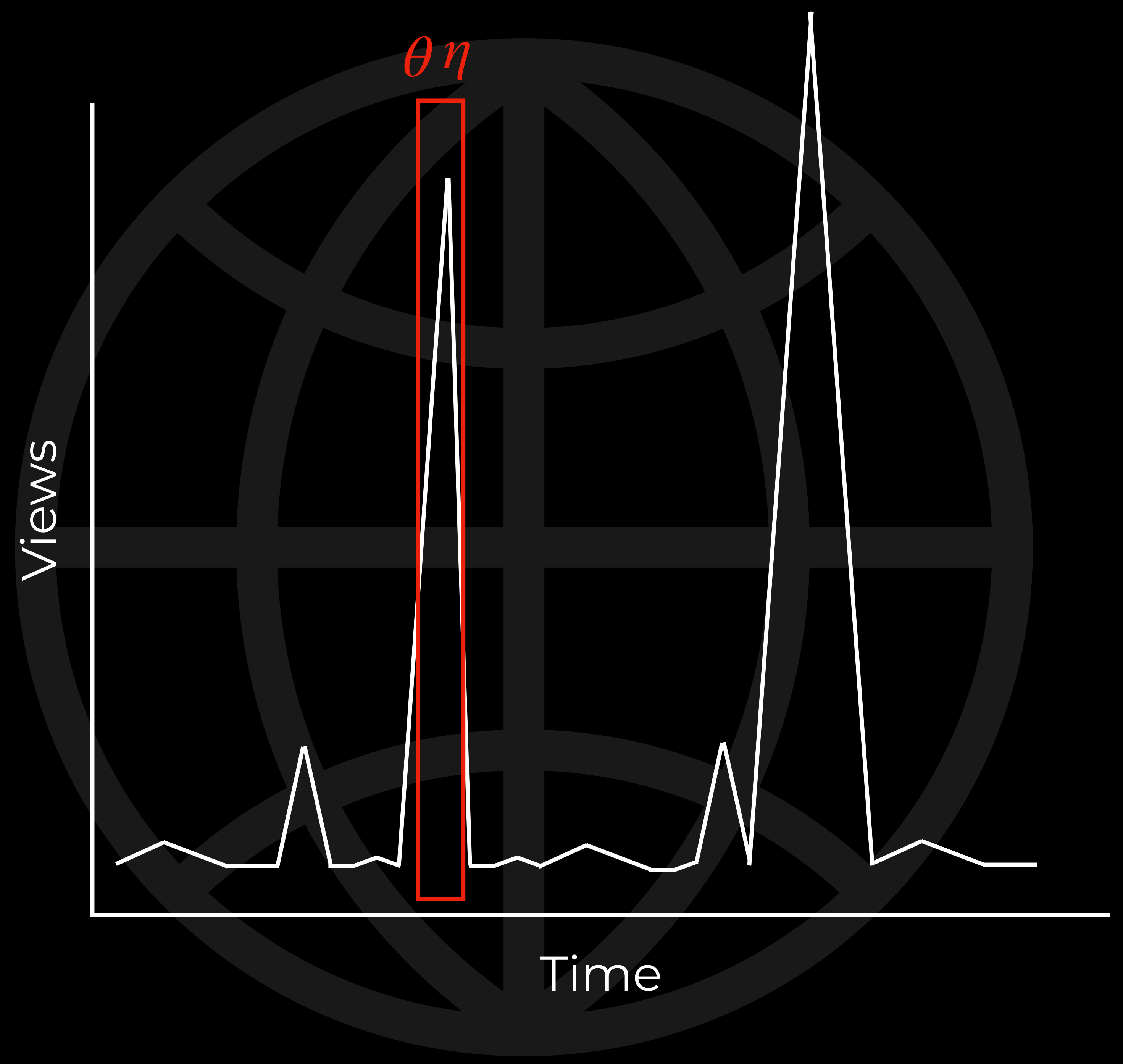


Detecting Shocks

Z-Score

Fb's Prophet

Proposed



Detecting Shocks

Detecting Shocks

1,000

Samples

Detecting Shocks

1,000
Samples

3
Annotators

Detecting Shocks

1,000
Samples

3
Annotators

0.6
Fleiss Kappa

Detecting Shocks

1,000
Samples

3
Annotators

0.6
Fleiss Kappa

Detecting Shocks

Detecting Shocks

Detecting Shocks

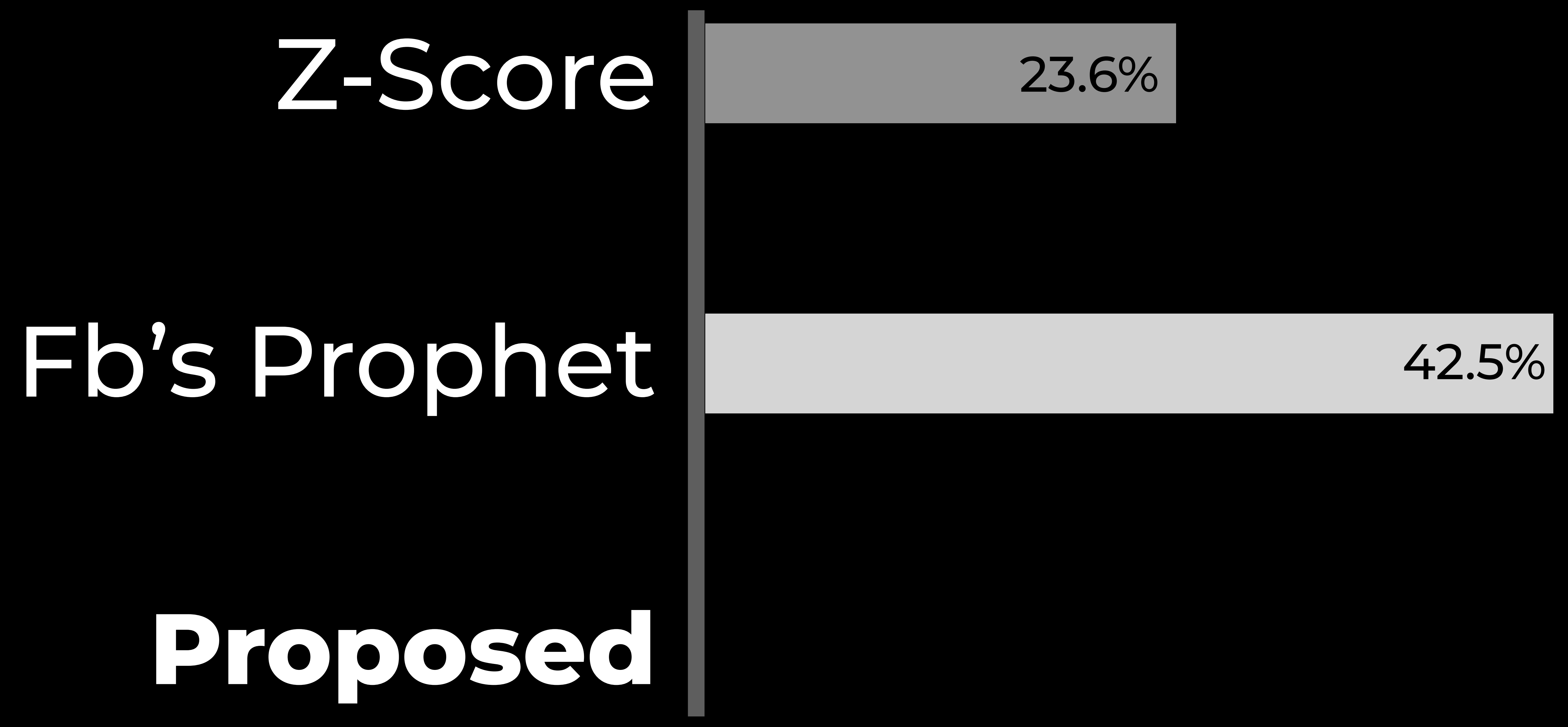
Z-Score

Fb’s Prophet

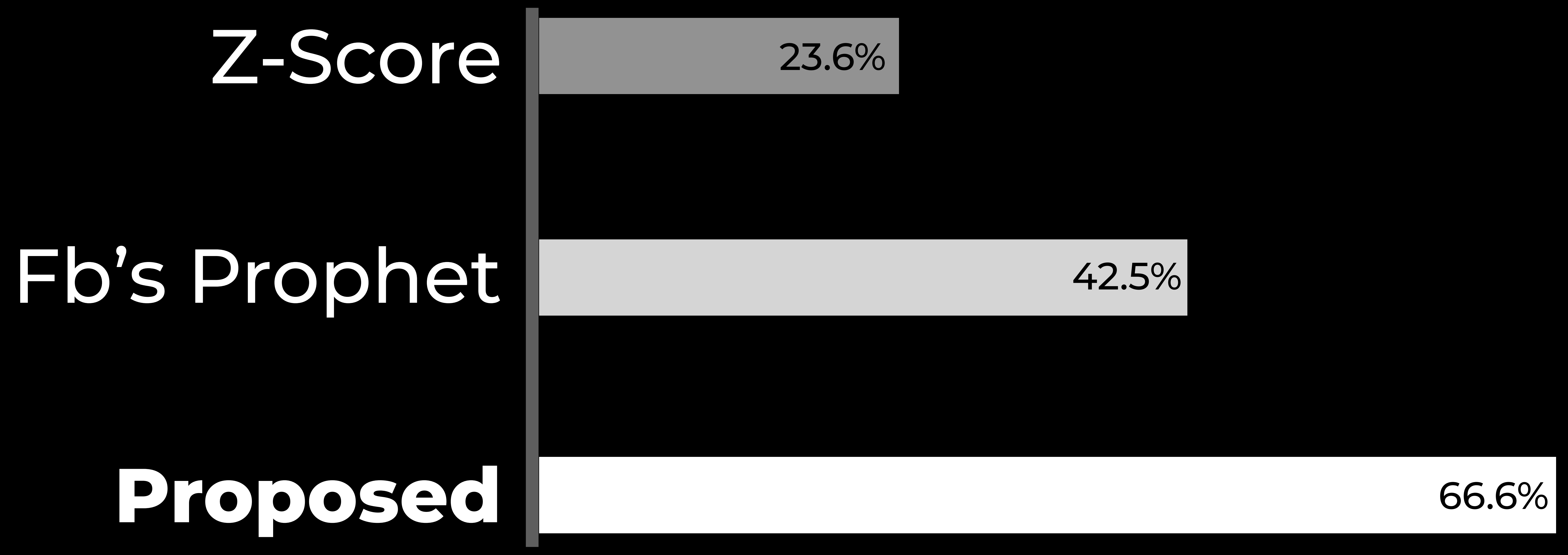
Proposed

Detecting Shocks

Detecting Shocks



Detecting Shocks



Detecting Shocks

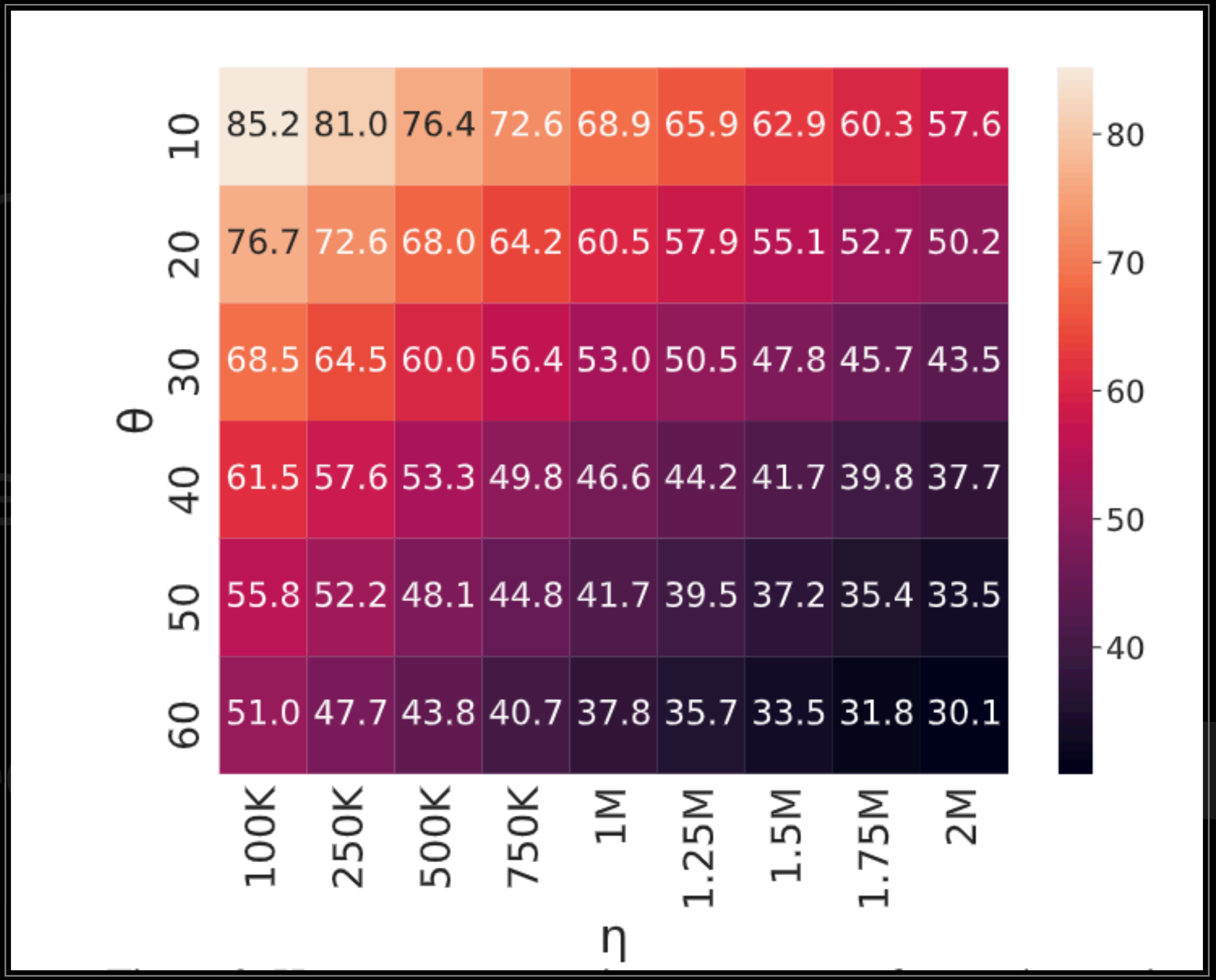
Detecting Shocks

Z-Scor

Fb’s Prophe

Propose

66.6%



Effect on Posting Frequency

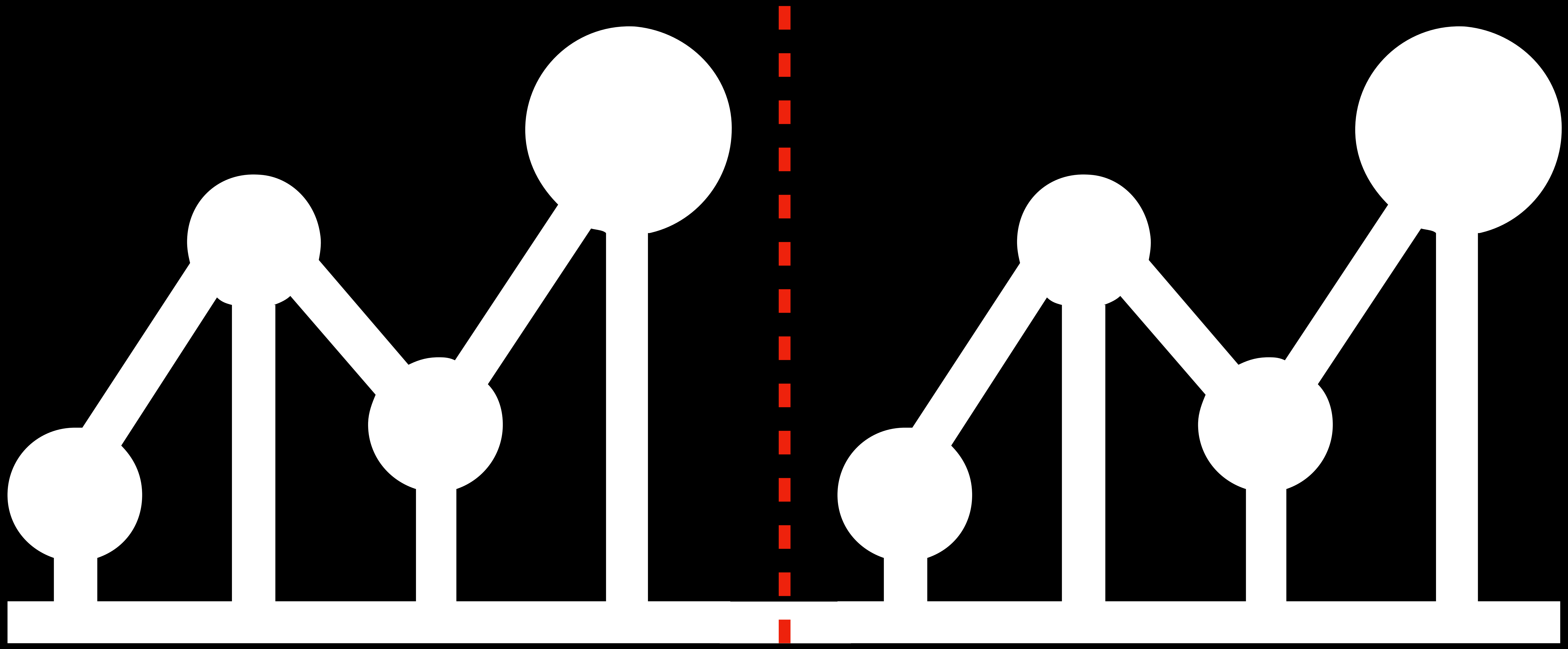
Effect on Posting Frequency

Regression Discontinuity in Time

Effect on Posting Frequency

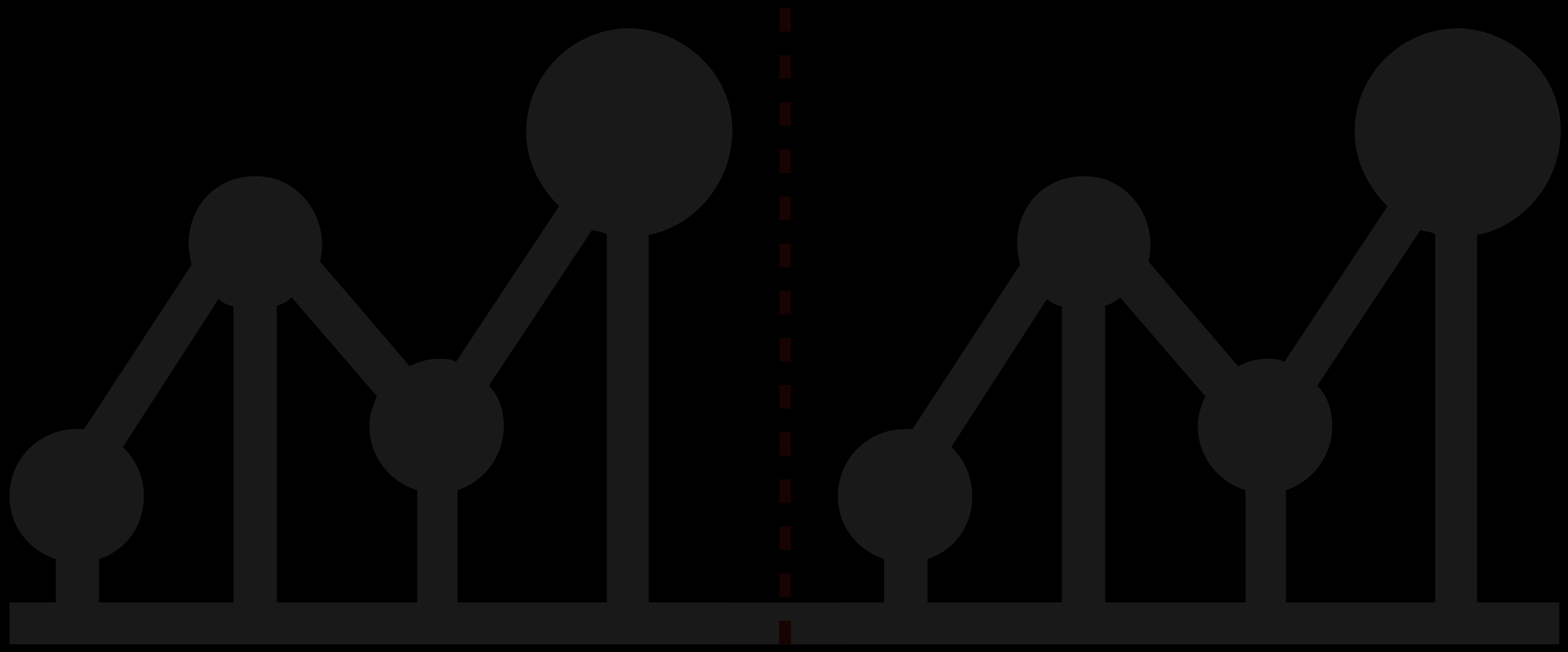
Regression Discontinuity in Time

Effect on Posting Frequency



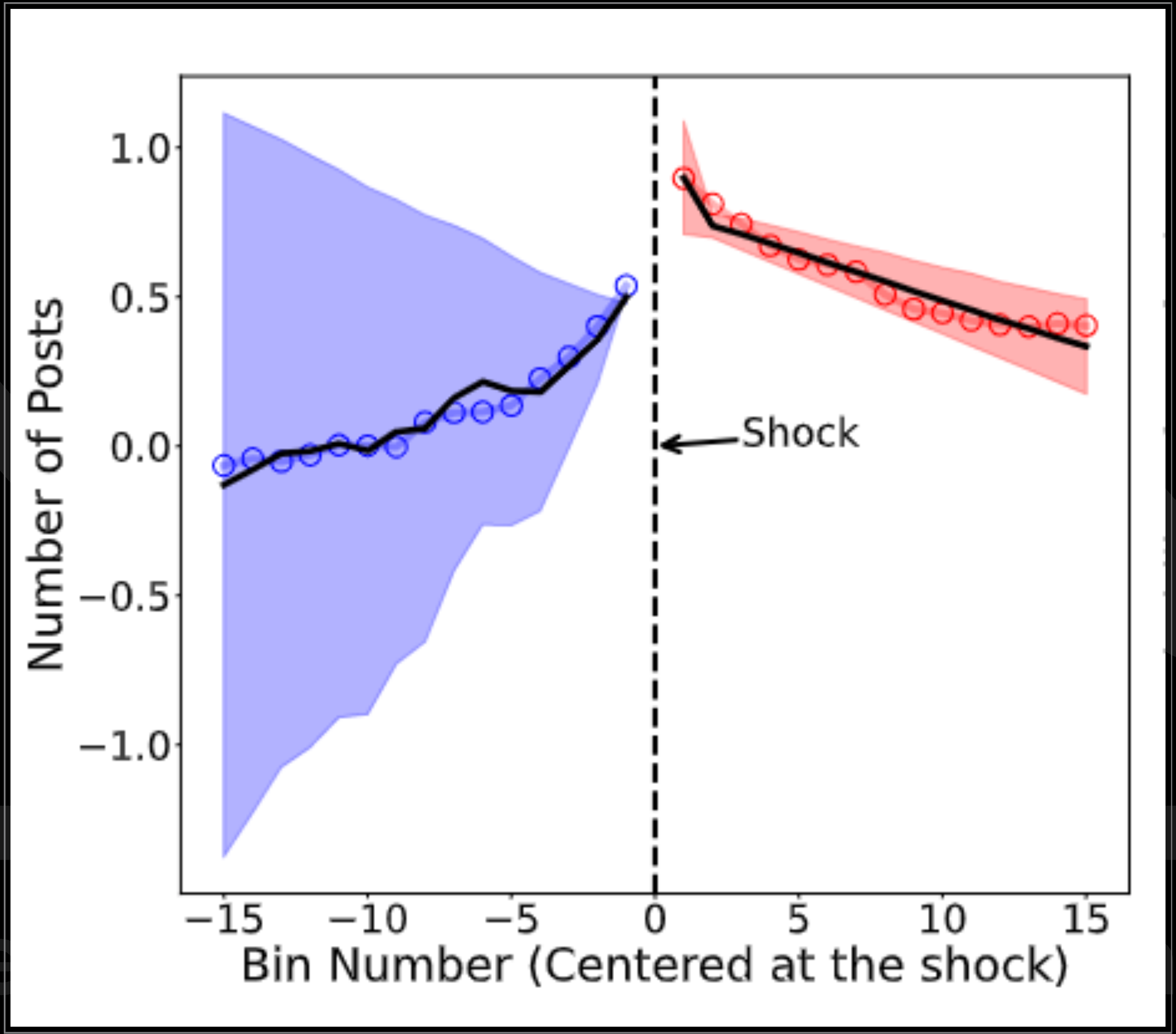
Regression Discontinuity in Time

Effect on Posting Frequency



Regression Discontinuity in Time

Effect on Posting Frequency



Control Variables: Age of User, and Intensity of user

Effect on Content

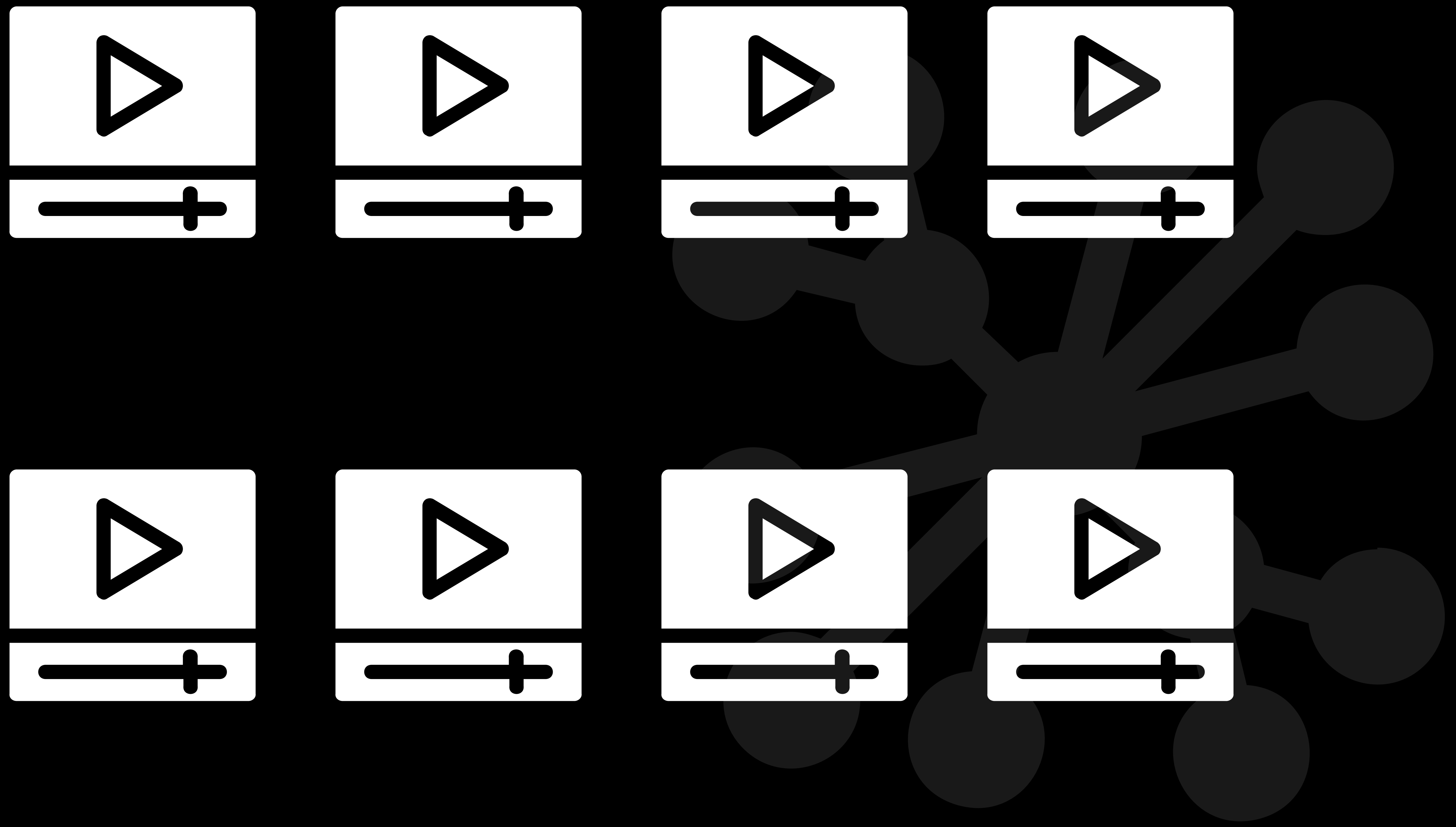
Effect on Content

Custom Doc2vec Embeddings

Effect on Content

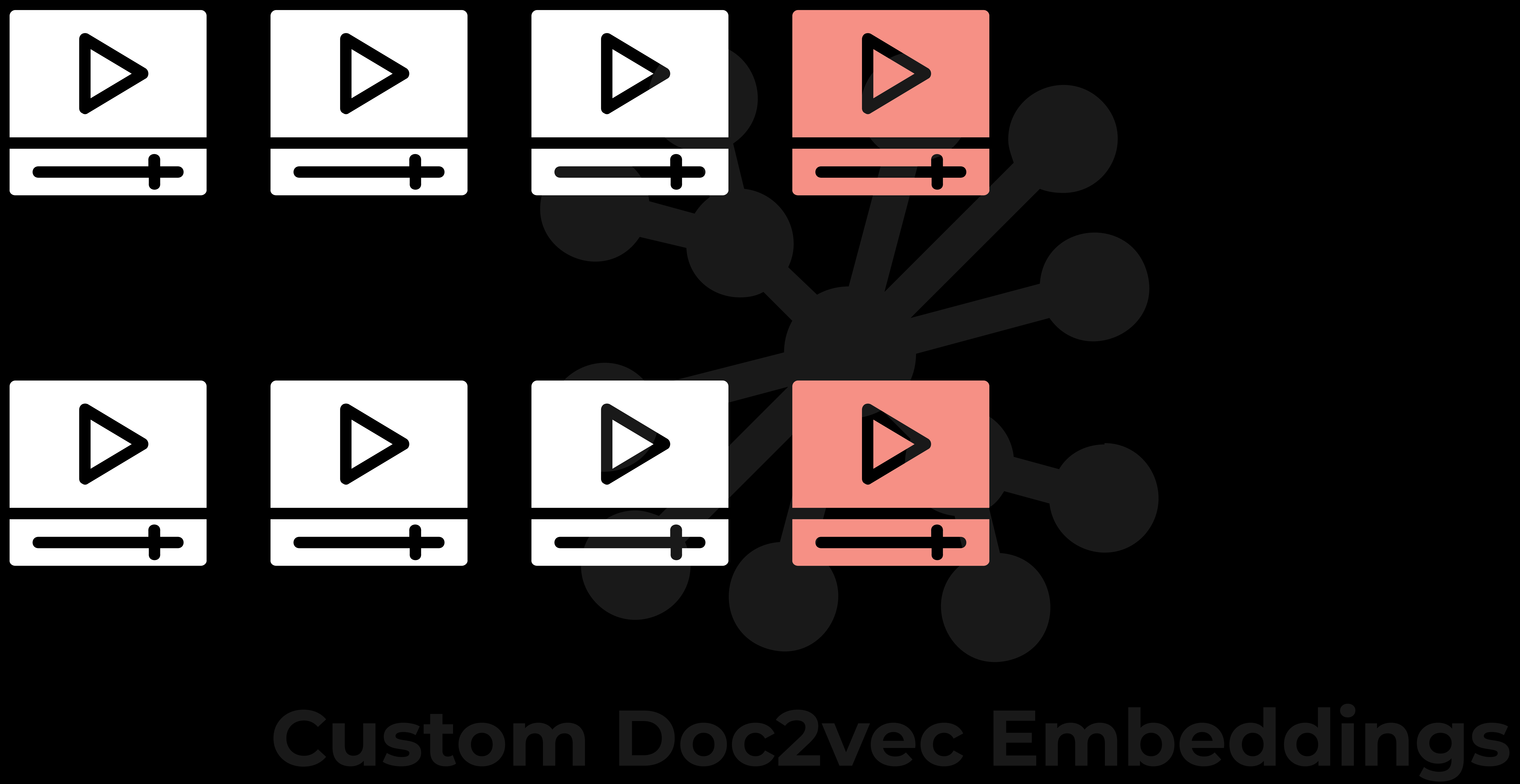
Custom Doc2vec Embeddings

Effect on Content

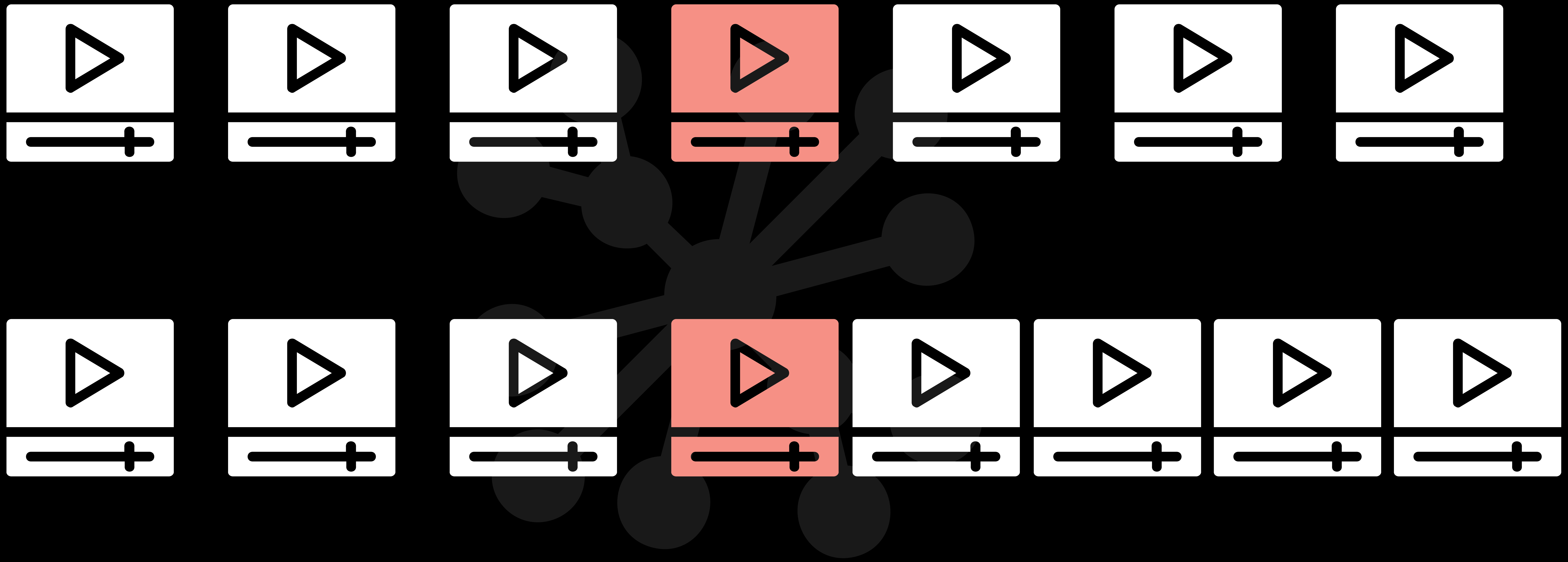


Custom Doc2vec Embeddings

Effect on Content

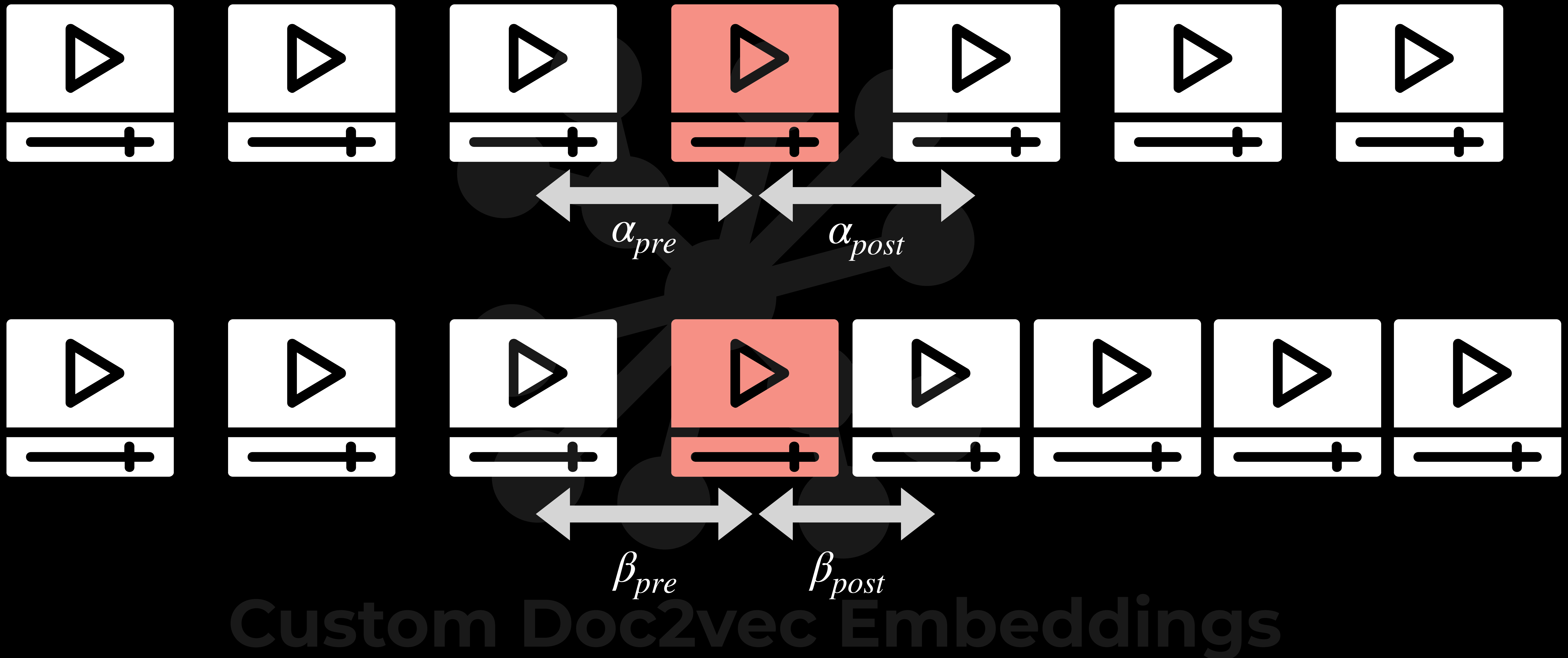


Effect on Content

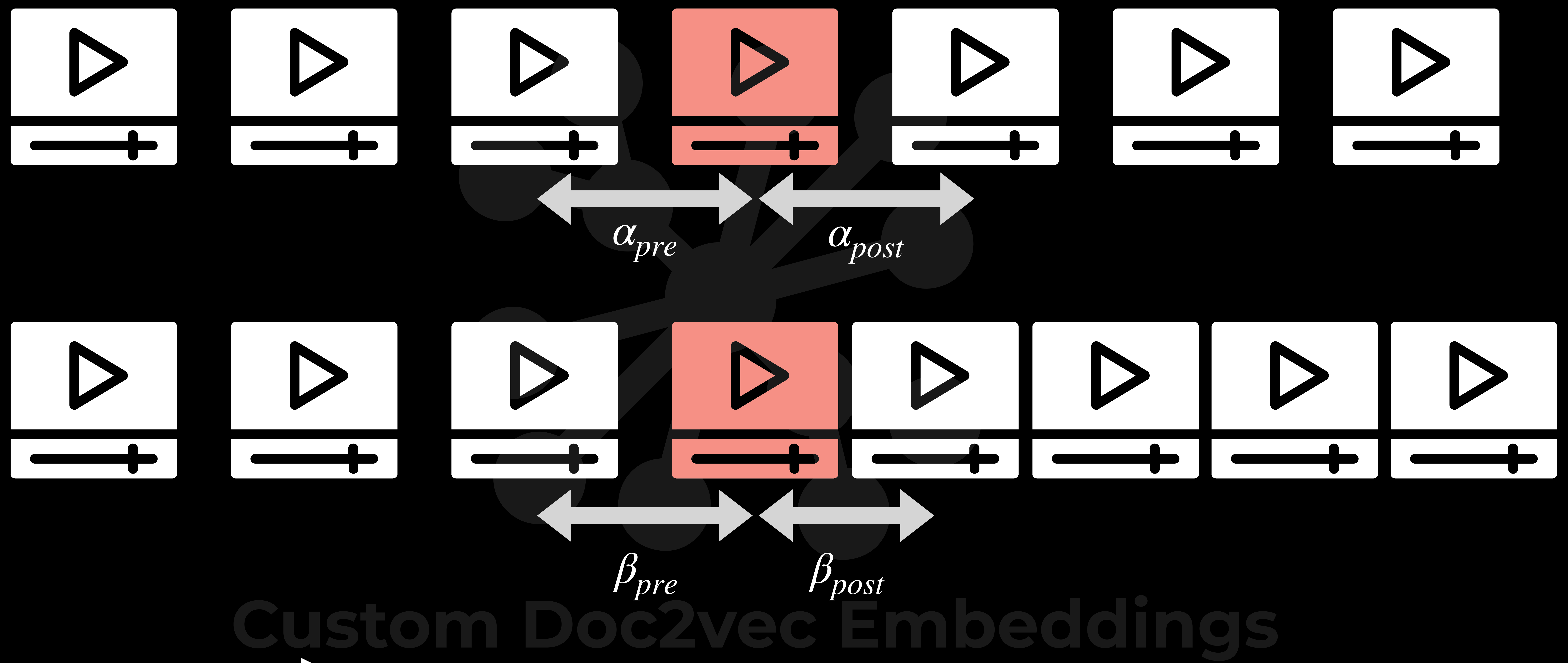


Custom Doc2vec Embeddings

Effect on Content

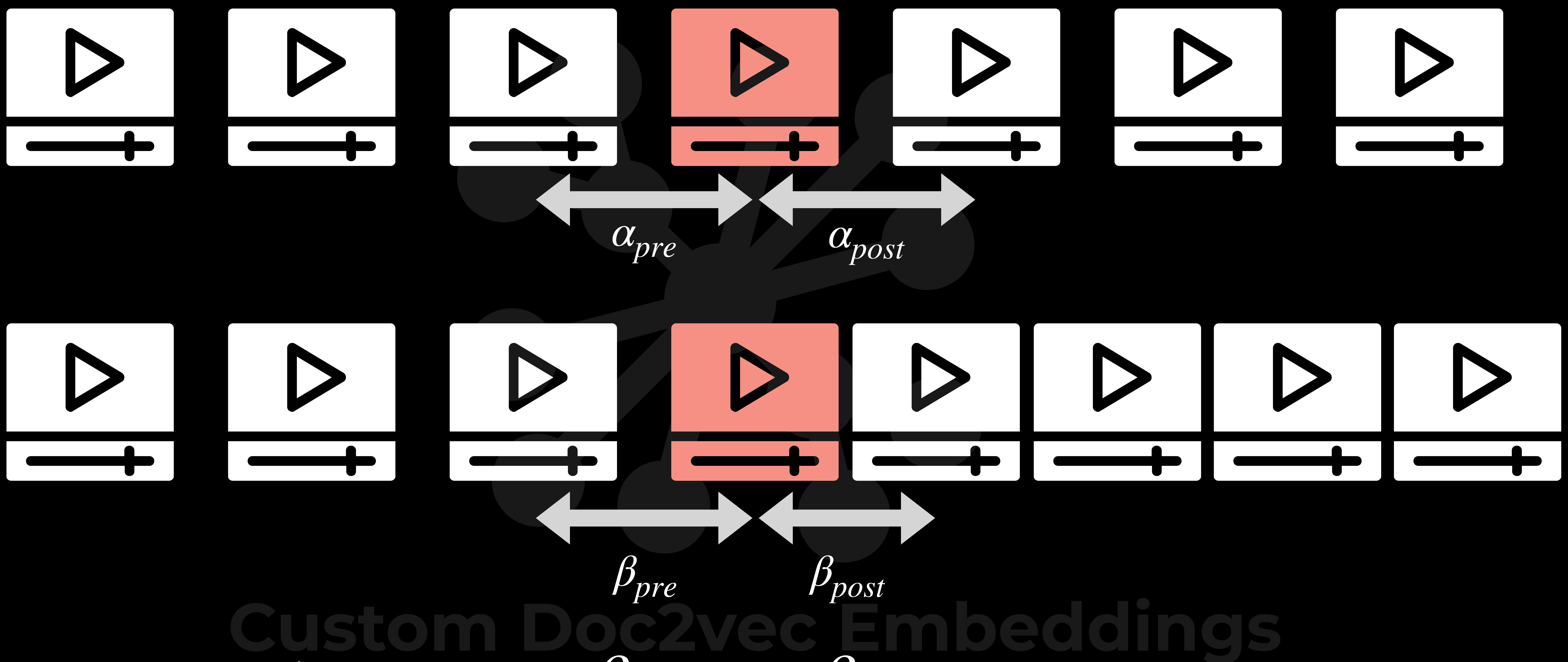


Effect on Content



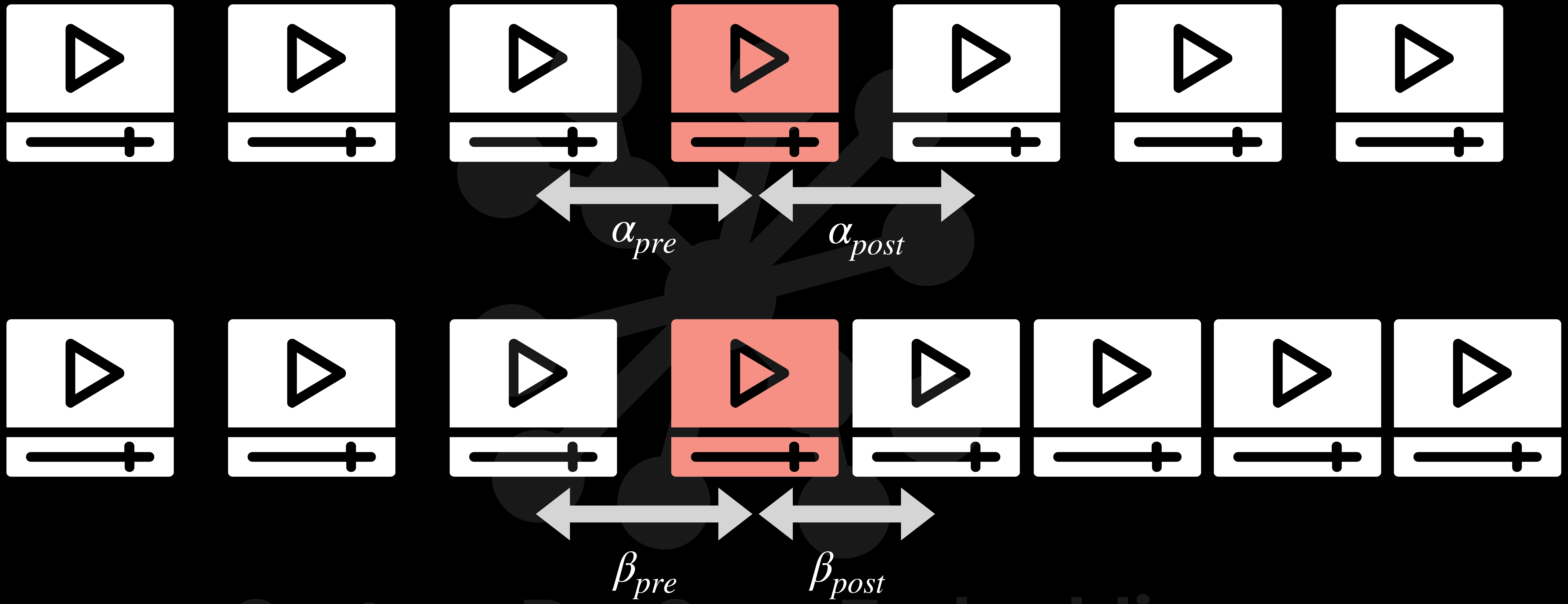
$$\alpha_{post} > \alpha_{pre}$$

Effect on Content



$$\alpha_{post} > \alpha_{pre} \qquad \beta_{post} > \beta_{pre}$$

Effect on Content

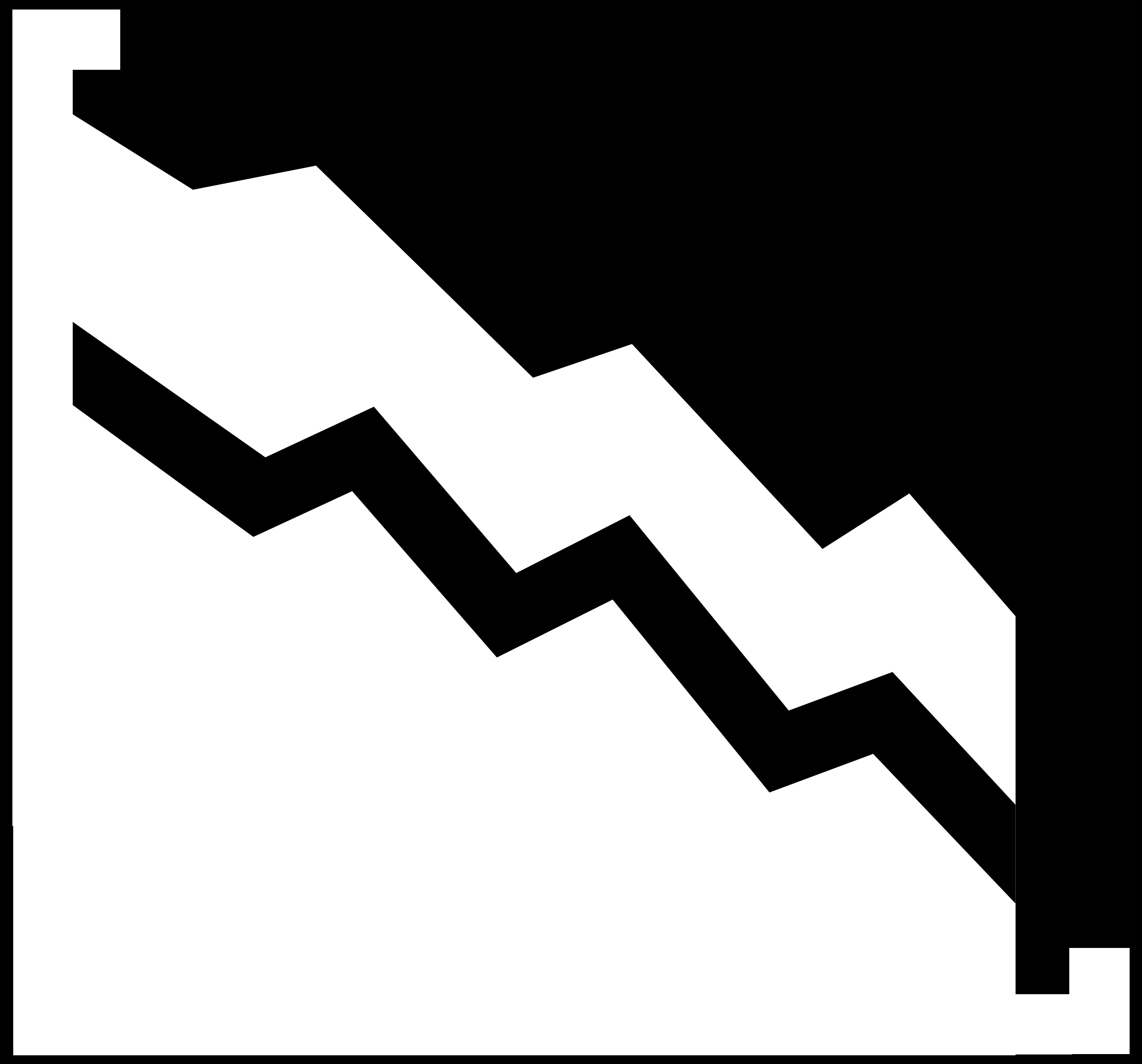


Custom Doc2vec Embeddings

$$\alpha_{post} > \alpha_{pre} \qquad \beta_{post} > \beta_{pre} \qquad \beta_{post} > \alpha_{post}$$

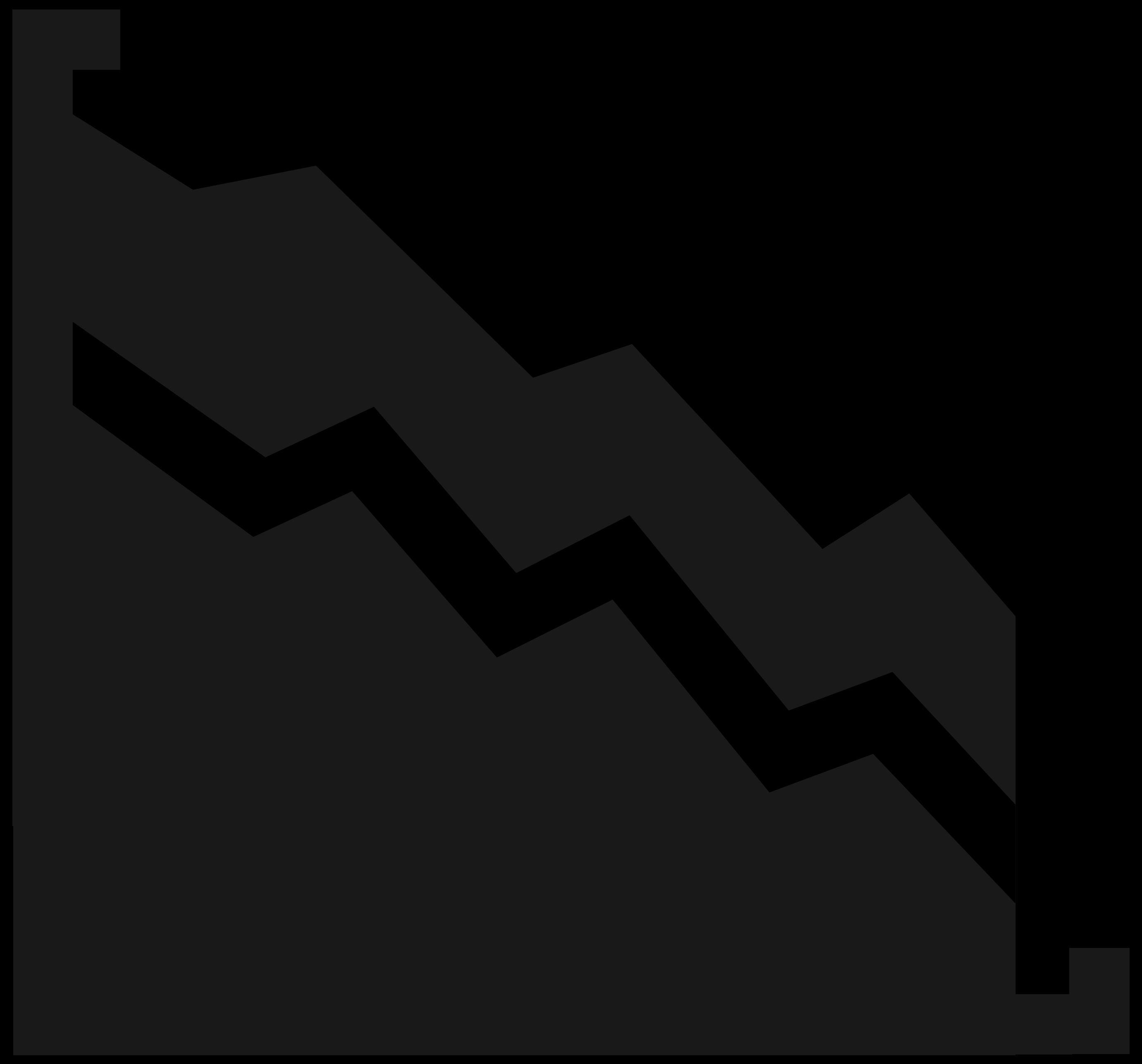
Sustainability of Shock

Sustainability of Shock



Survival Analysis / CoX Regression

Sustainability of Shock



Survival Analysis / CoX Regression

Sustainability of Shock

Popularity is generally short lived

Survival Analysis / CoX Regression

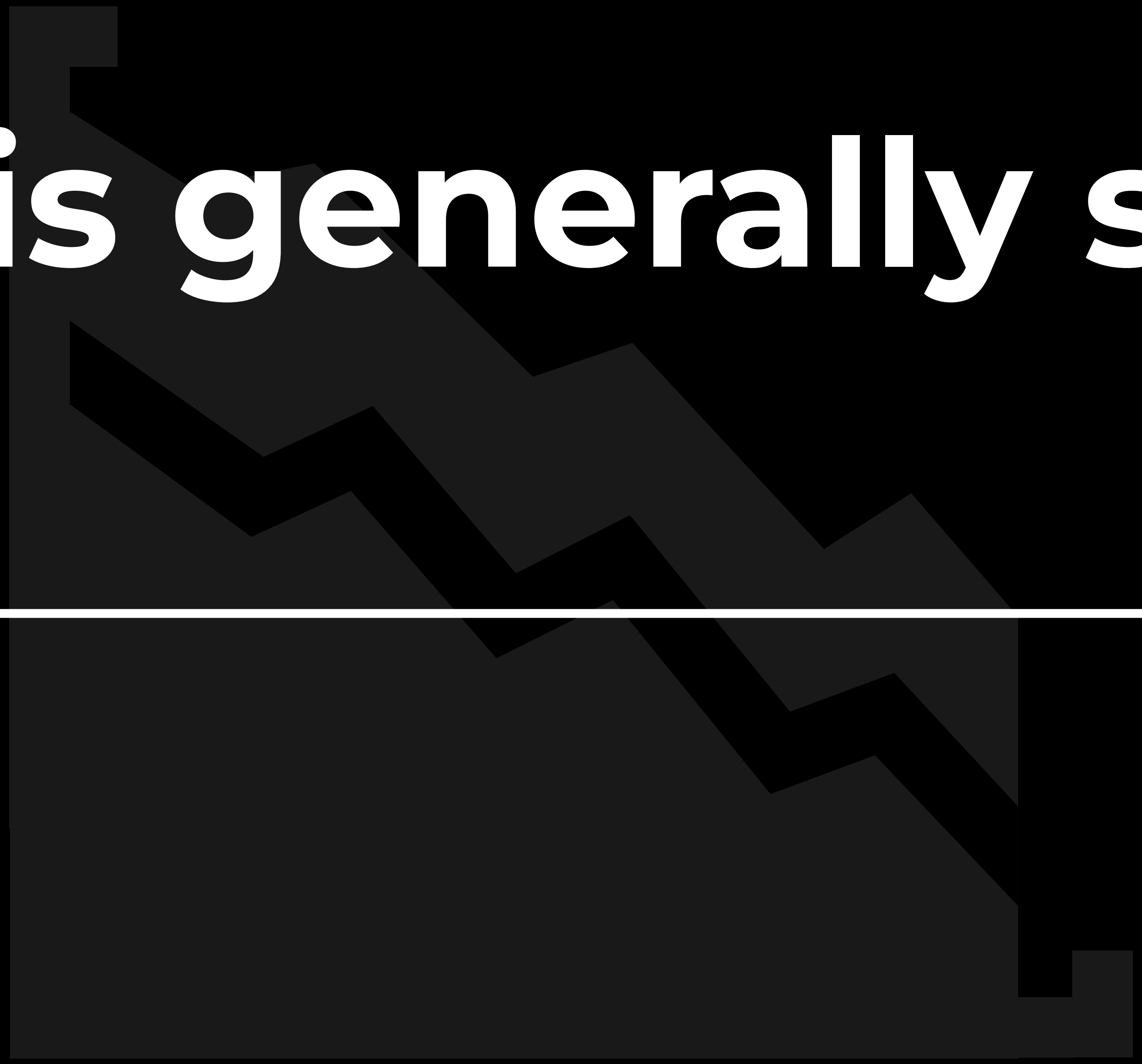
Sustainability of Shock

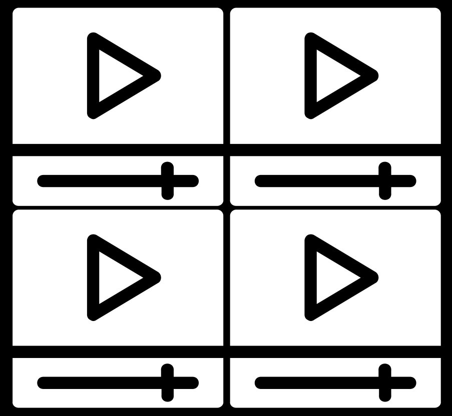
Popularity is generally short lived

Survival Analysis / CoX Regression

Sustainability of Shock

Popularity is generally short lived





Post Frequency

Survival Analysis / CoX Regression

Sustainability of Shock

Popularity is generally short lived

Post Frequency

Community Engagement

Survival Analysis / CoX Regression

Sustainability of Shock

Popularity is generally short lived

Post
Frequency

Community
Engagement

Balance
Content Similarity

Survival Analysis / CoX Regression

Implications

Implications

Implications

Implications

Contributions

Contributions

First paper to study user behaviour after popularity shocks on short video platform

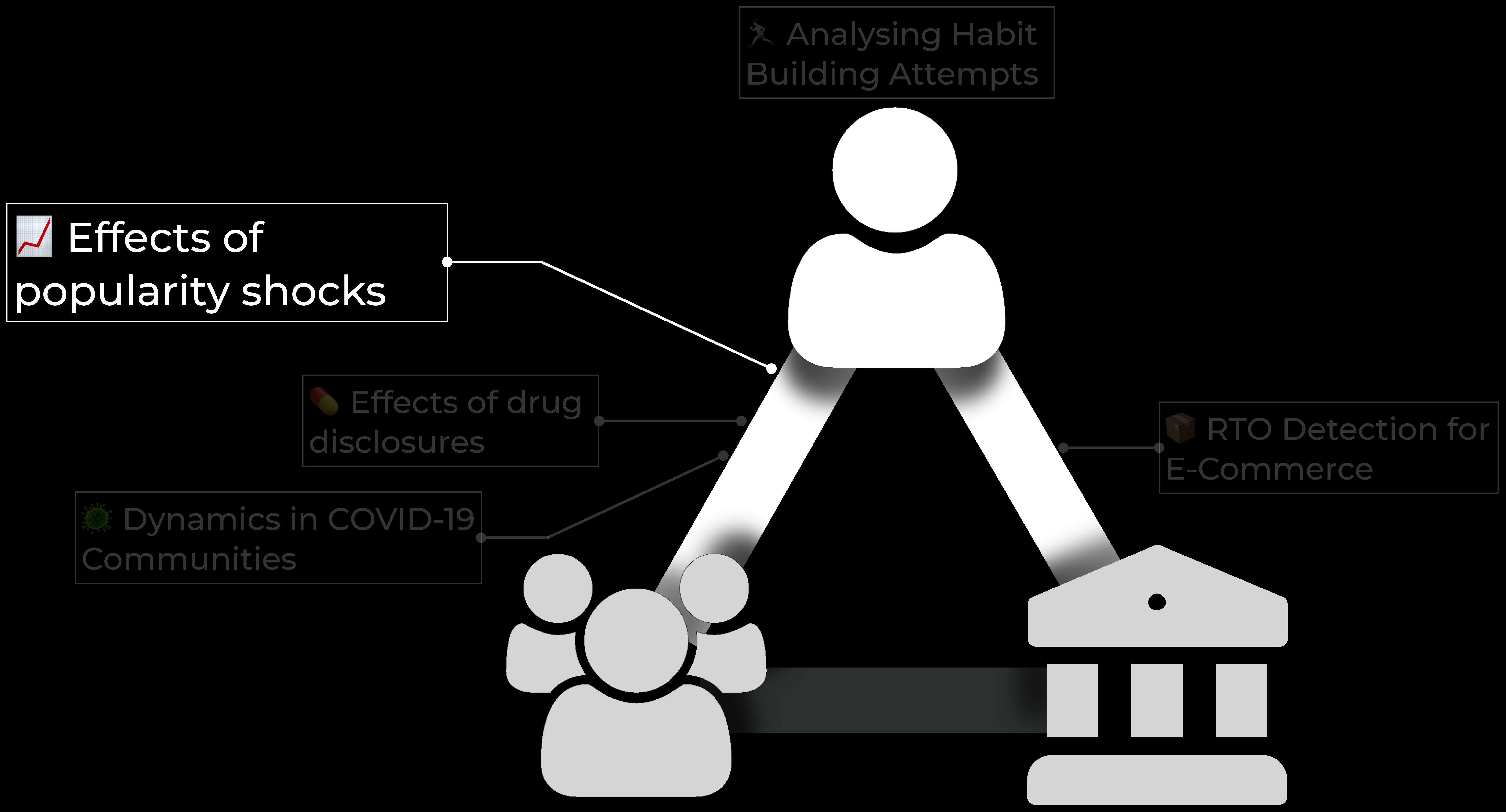
Contributions

- First paper to study user behaviour after popularity shocks on short video platform
- Actionable insights for users to sustain popularity

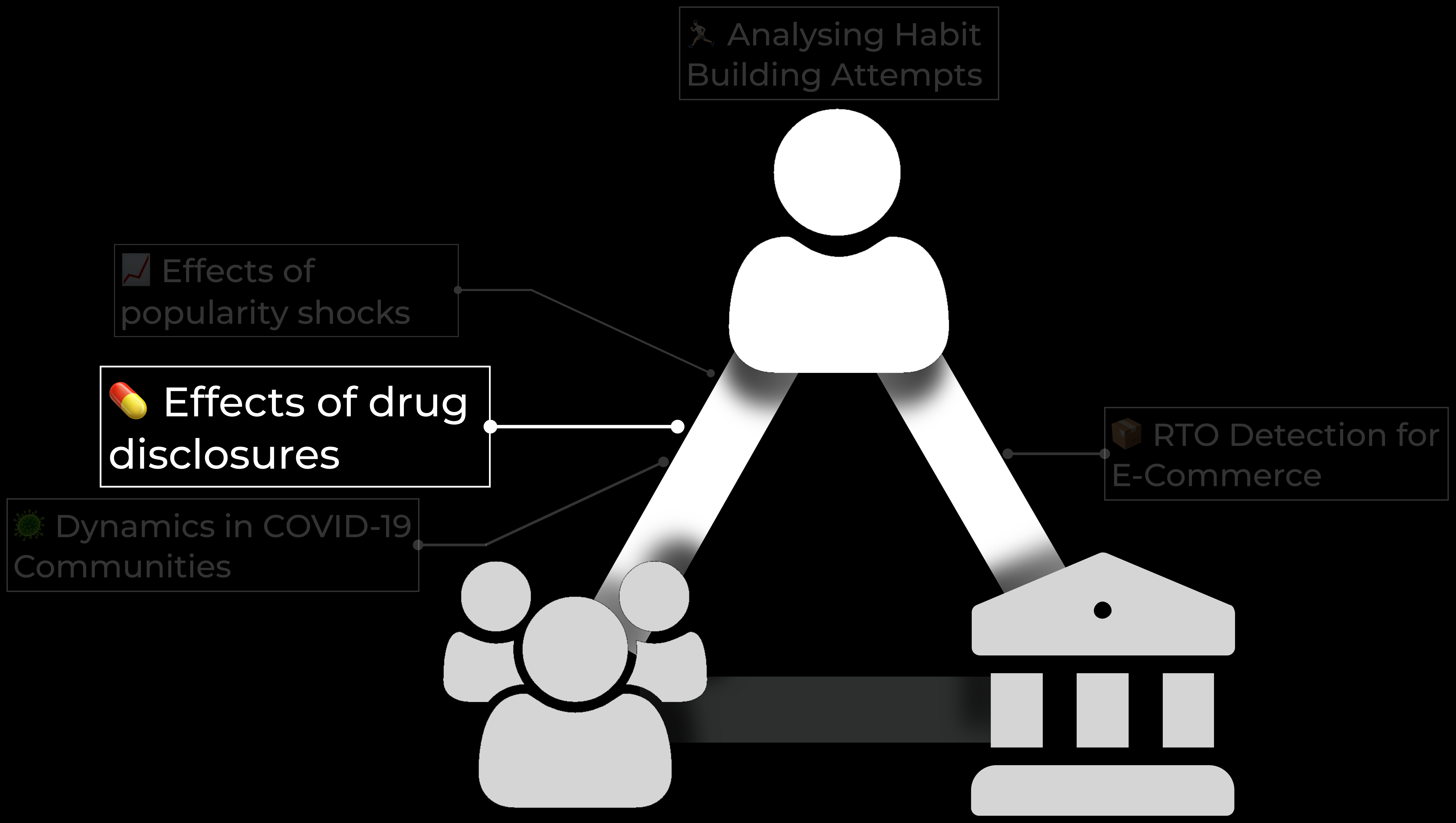
Contributions

- First paper to study user behaviour after popularity shocks on short video platform
- Actionable insights for users to sustain popularity
- First dataset to provide posts and accompanying metadata for a short video platform

Our Focus


🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

Our Focus



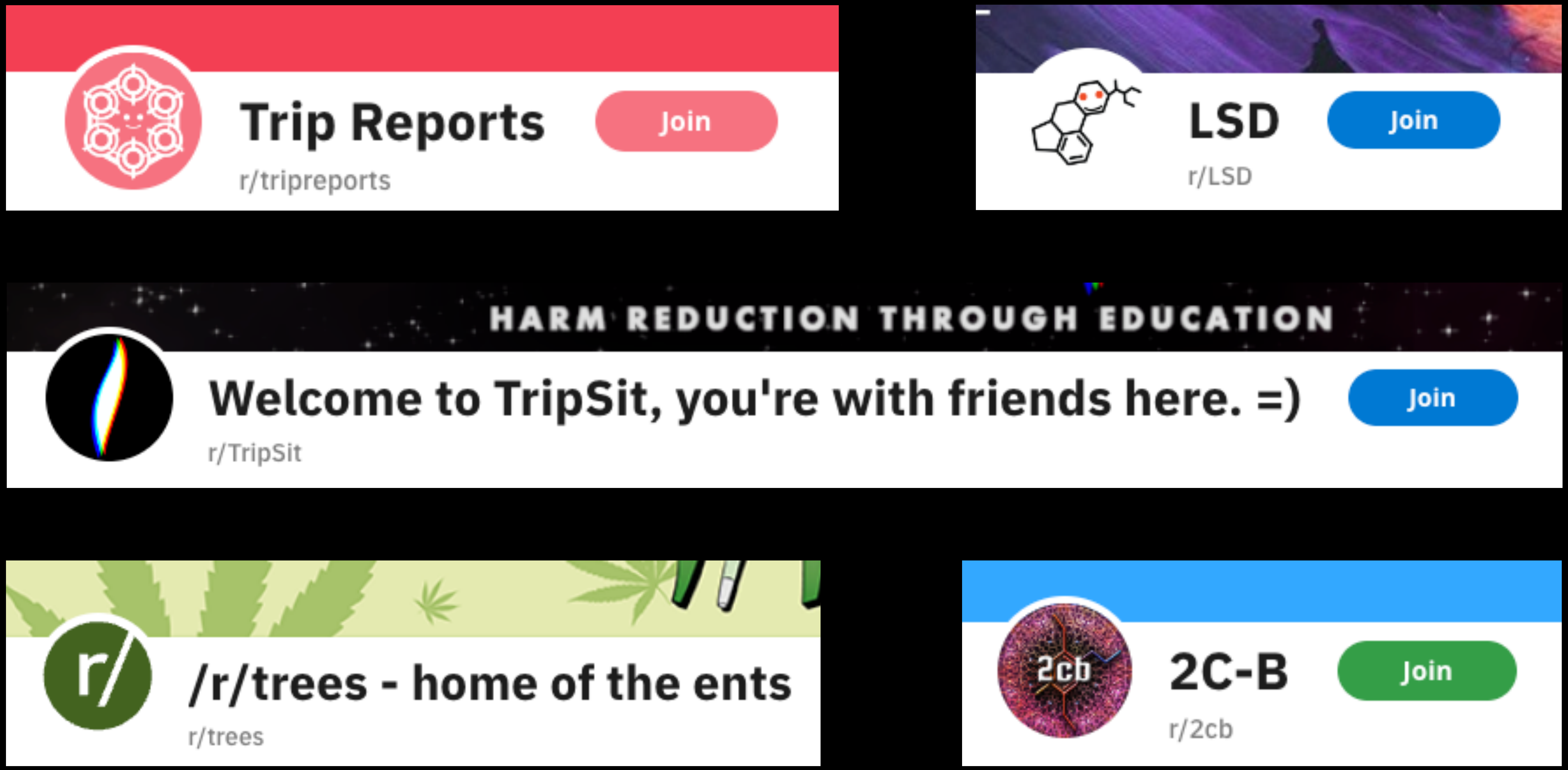
🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

Effect of Feedback on Drug Consumption Disclosures on Social Media

ICWSM’ 23

Hitkul, Rajiv Ratn Shah, Ponnurangam Kumaraguru

Drug Consumption Subreddits



Drug Consumption Disclosure

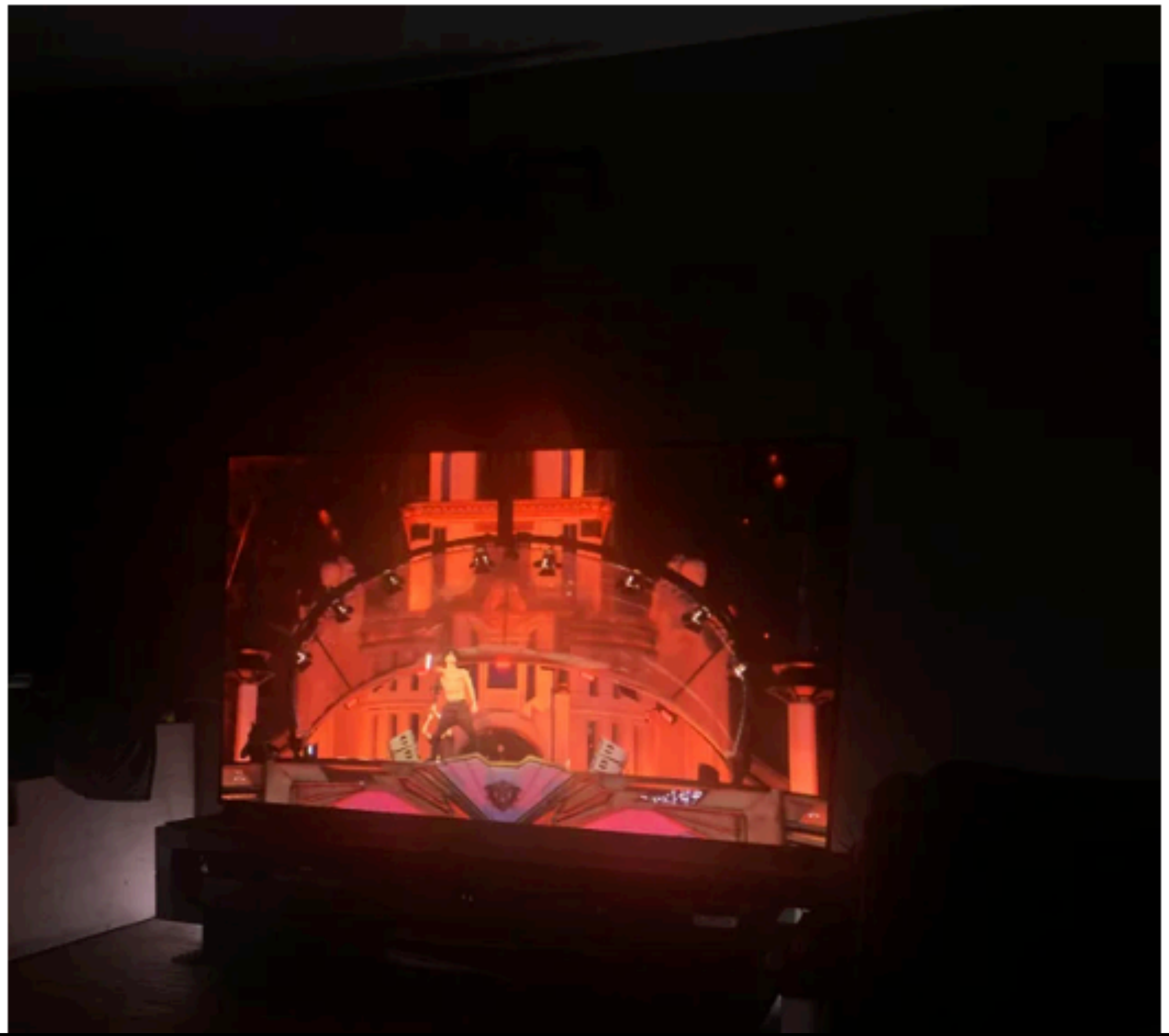
Drug Consumption Disclosure

r/MDMA · Posted by u/we_hella_believe 14 hours ago

16

Rolling Hard. Watching Tomorrowland Videos. Solo. Be safe everyone and stay hydrated.

Trip Report



r/tripreports · Posted by u/Fuzzy_Wrongdoer3246 3 hours ago

2

300 Mg of Dxm HBr + 40 mg of Edibles Trip Repot

DXM

So the other night I decided that I was going to take some Dxm, Ive done it a handful of times just kinda experimenting and fucking around and never more than 200mg which is a second plat for my bodyweight (125 lbs). However this particular night I decided I was going to take my highest dose yet, 300mg witch is the rest that I conveniently had (20-15 Mg gelcaps). After I down all the fuckers I smoke a bowl waiting for them to kick in and after a bit of smoking I remember of some edibles I had on hand, so I decide to down 40 mg of them as well.

As Im playing some bo2 zombies and watching some YouTube I have the normal effects that I've experienced before, body feels heavy and I have little motivation to get out my chair, I almost feel drunk, I want to say I had some slight visuals but nothing remarkable, if I had any at all. This goes on for maybe 2-3 hours until I pass out.

r/MDMA · Posted by u/dubplateer 18 hours ago

51

I'm **roooling balls right now.** I love everything and everyone. Happy new year

MDMA + Pot

Title

r/2cb · Posted by u/Lopsided_Law_1762 7 hours ago

7

2CB is just insane

Did about **30mg yesterday night**, did few beers and i dont even fell messy on my head. Night was beautiful, only positive vibes, a lot of laugh lol. Being able to go to sleep easily after 2CB trip and sleep for 6-7 hours is HUGE. This is literally the best social drug i've ever used.

Related Work

Related Work

Drug on
social media

Causal inference
on social data

Self-harm behavior
on social media

Related Work

Drug on social media

Addiction detection

Trend analysis

Forecast drug overdose, abuse

Causal inference on social data

Simulating RCT using online social data

Applications have been shown in weight loss, alcohol consumption, content quality

Self-harm behavior on social media

Self-harm to improve social standing

Online challenges e.g. KiKi

Demographic features

Research Questions

Impression Management

(Goffman 1959)
(Leary et al. 1994)
(Hogan 2010)

Research Questions

Impression Management
(Goffman 1959)
(Leary et al. 1994)
(Hogan 2010)

What is the **extent** of content indicating offline drug consumption?

Research Questions

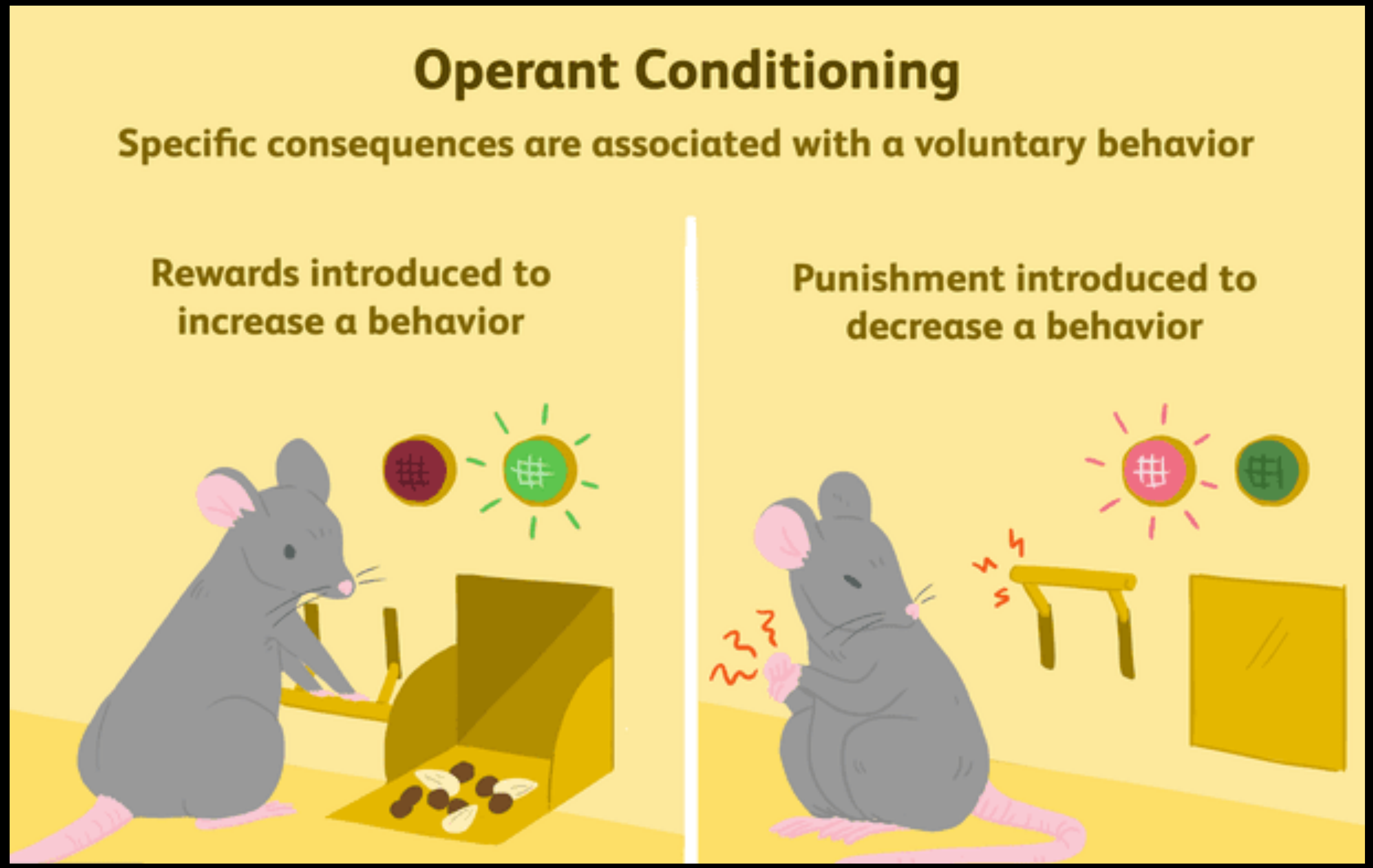
Primacy Effect
(Asch 1946)
(Murdock Jr 1962)

Research Questions

Primacy Effect
(Asch 1946)
(Murdock Jr 1962)

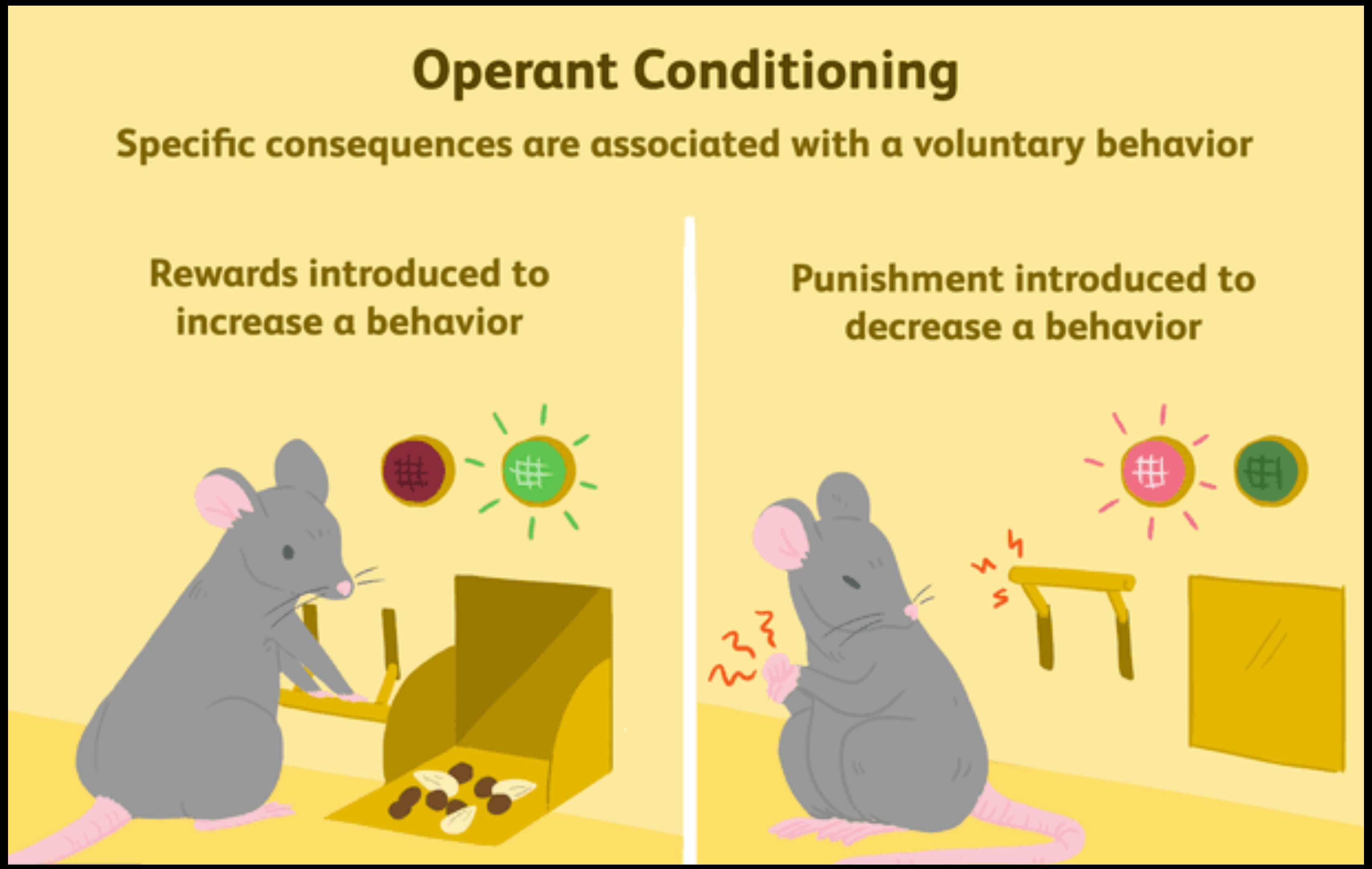
How does the community feedback on **first drug consumption** post affect users' future drug consumption?

Research Questions



Operant Conditioning
(Skinner, 1938)

Research Questions



Operant Conditioning
(Skinner, 1938)

How does **continuous positive community feedback** affect users' future drug consumption?

Research Questions

- What is the **extent** of content indicating offline drug consumption?
- How does the community feedback on **first drug consumption** post affect users' future drug consumption?
- How does **continuous positive community feedback** affect users' future drug consumption?
- Can we use Reddit textual data to **classify** between drug consumption and non-drug consumption content?

Data

Data

10 Subreddits

Data

10 Subreddits

826,905 Posts
6.6M Comments

Data

10 Subreddits

826,905 Posts
6.6M Comments

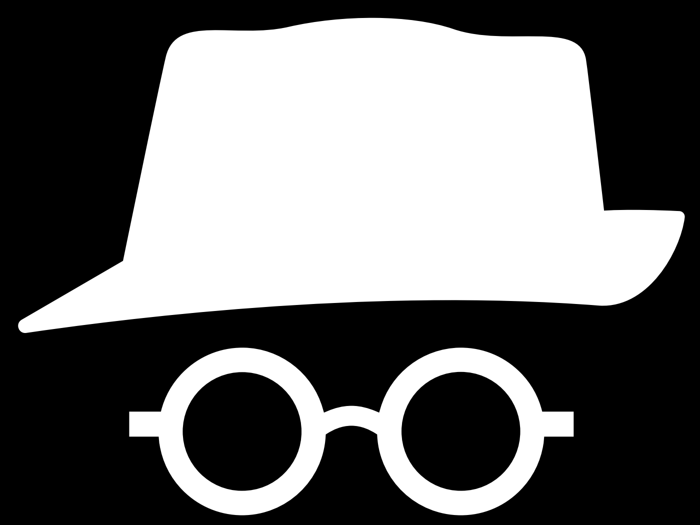
493,906 Users

Data - User Study

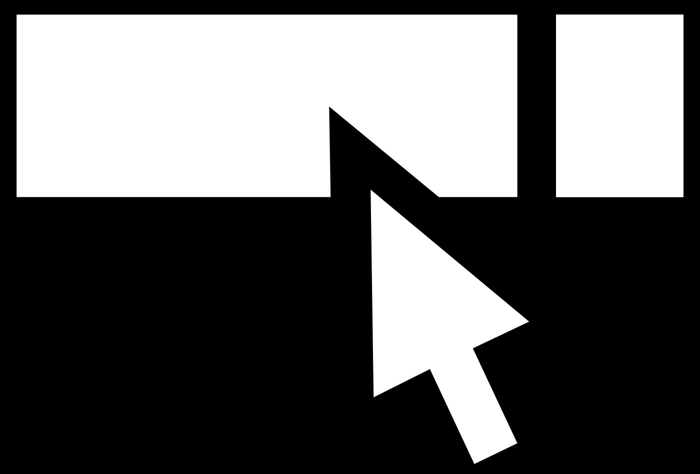
Data - User Study

Anonymous

Data - User Study



Anonymous



Likert Scale

Data - User Study

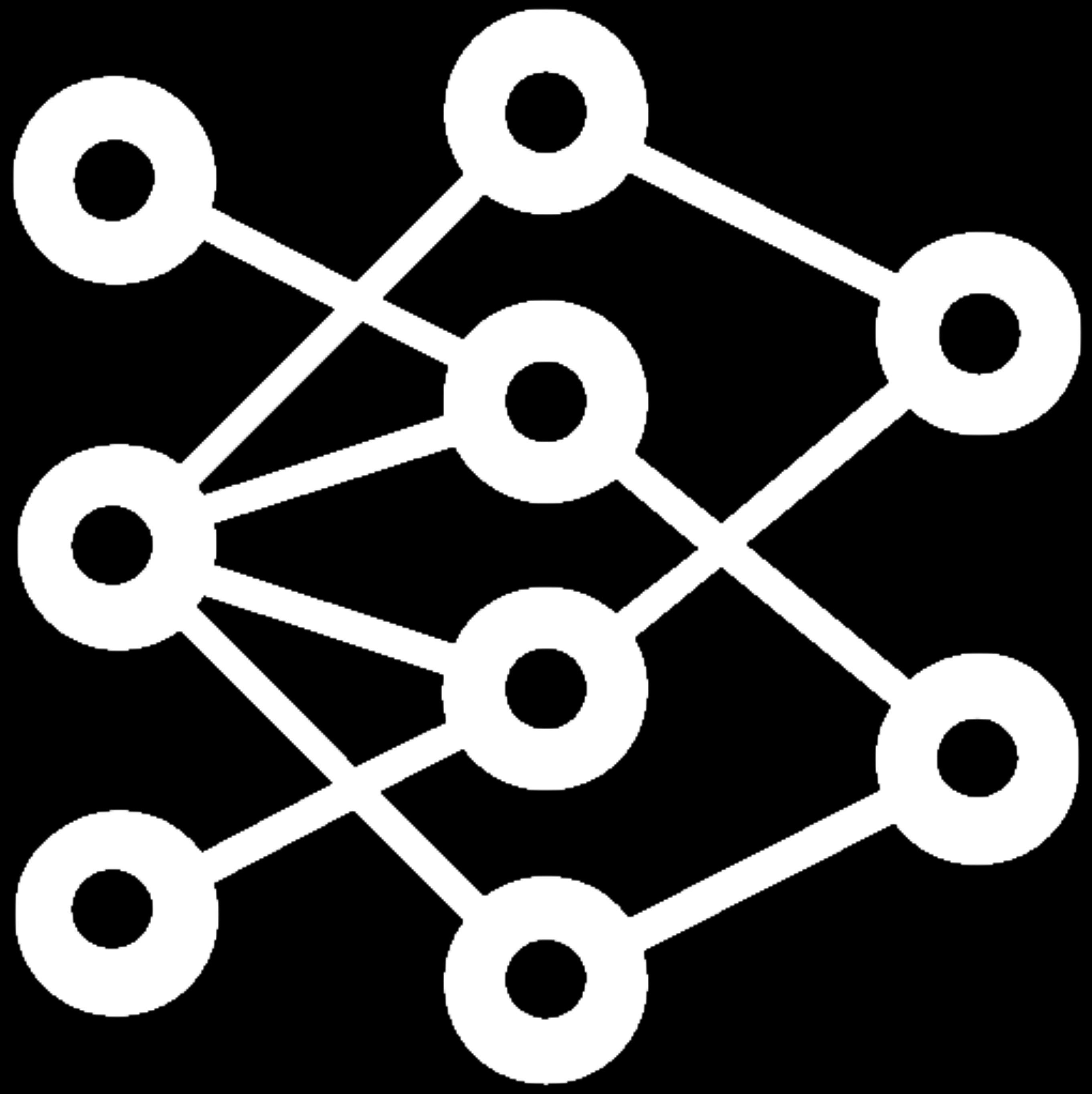
Anonymous

Likert Scale

45 Participants

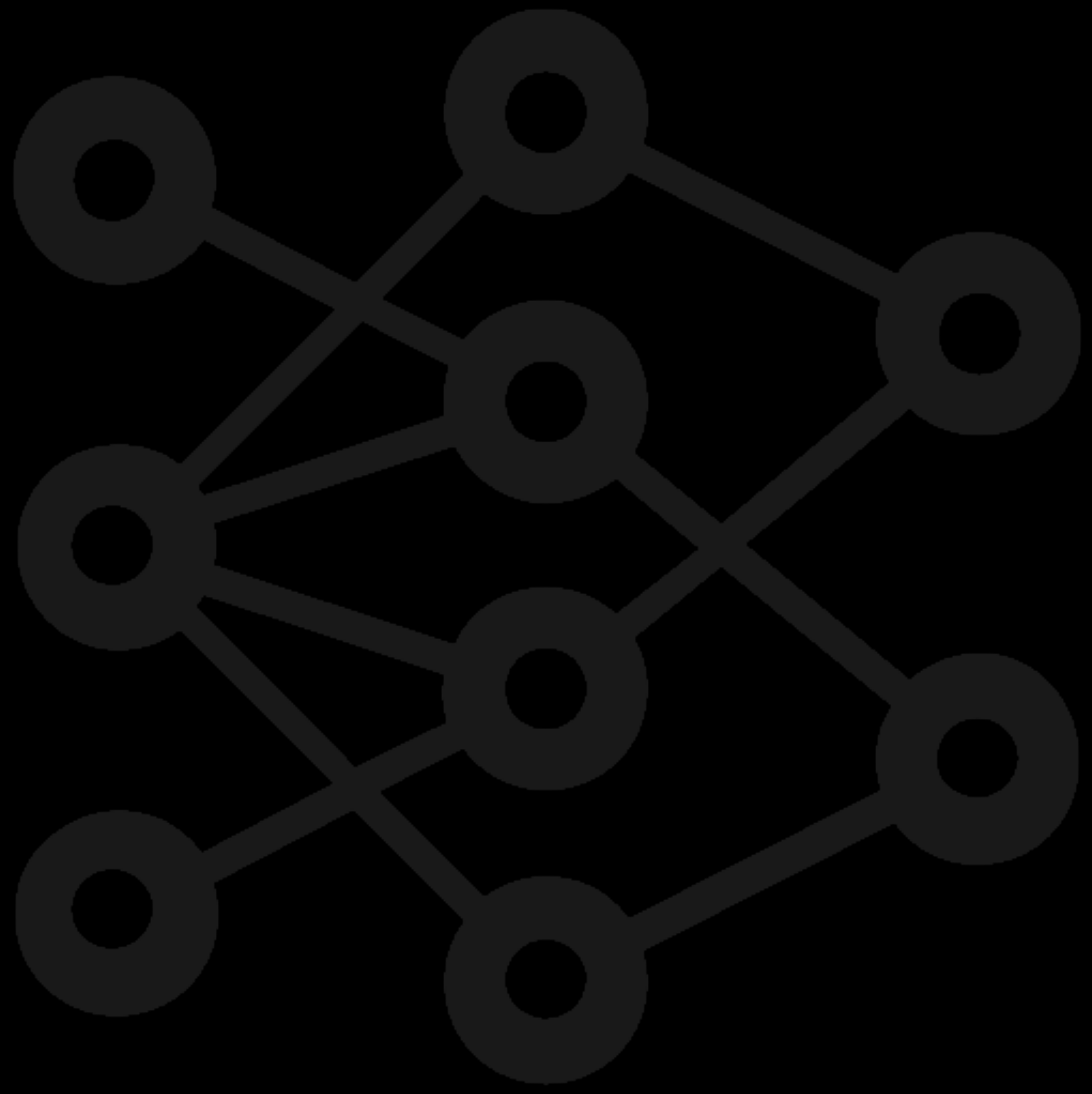
Detecting Drug Disclosure

Detecting Drug Disclosure



Deep Learning Classifier

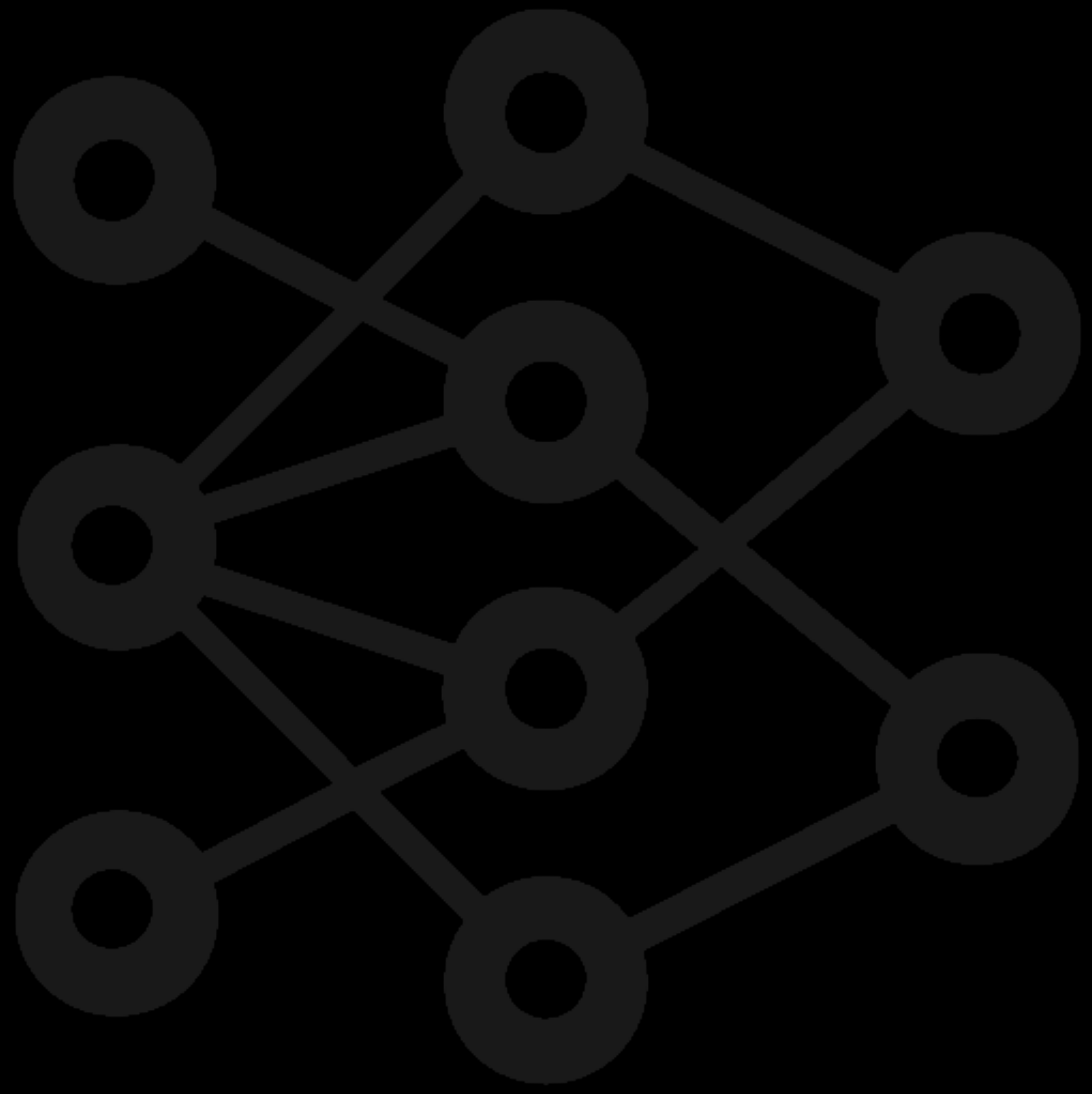
Detecting Drug Disclosure



Deep Learning Classifier

Detecting Drug Disclosure

4,000
Samples



Deep Learning Classifier

Detecting Drug Disclosure

4,000
Samples

3
Annotators

Deep Learning Classifier

Detecting Drug Disclosure

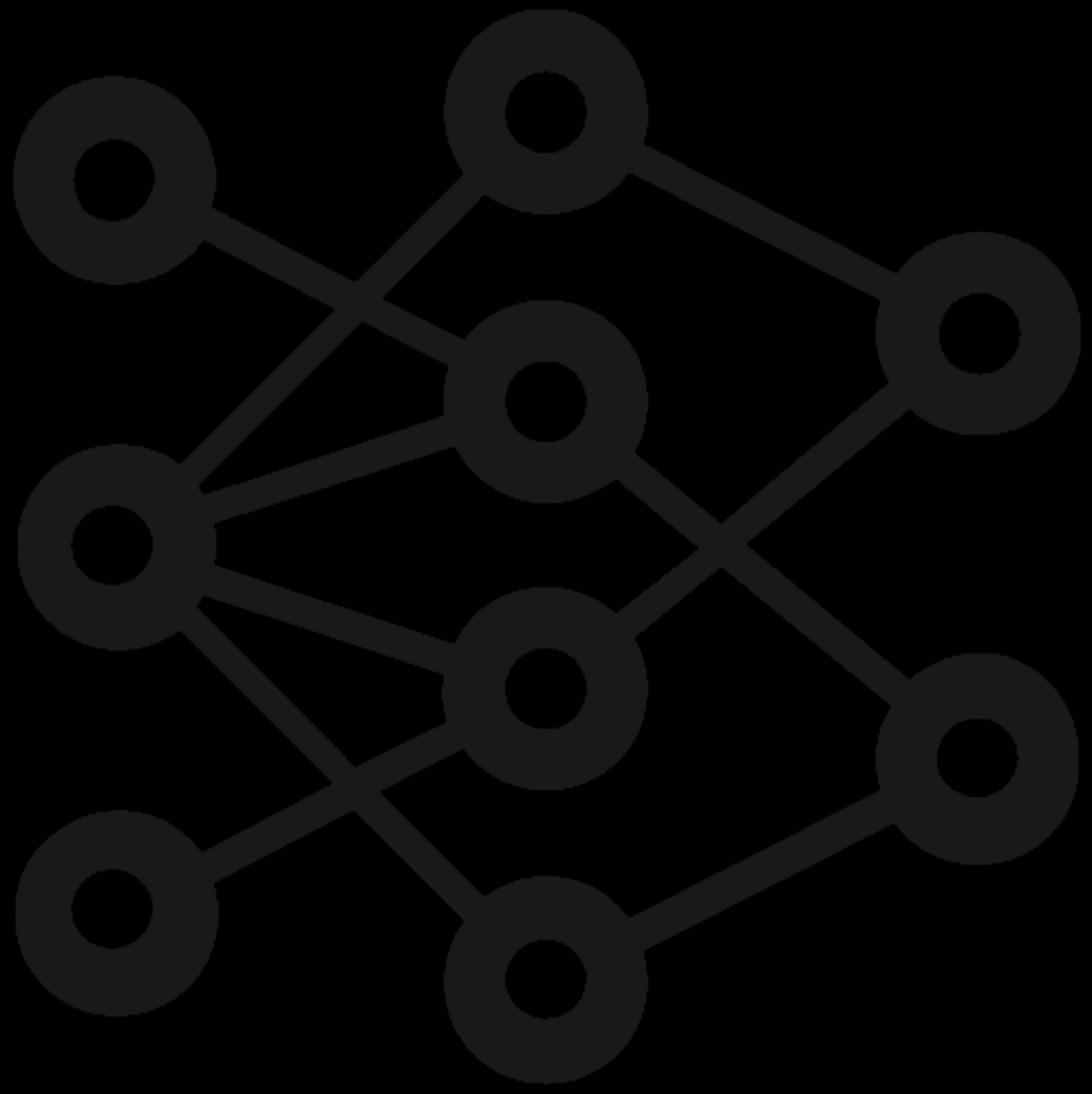
4,000
Samples

3
Annotators

0.7
Fleiss Kappa

Deep Learning Classifier

Detecting Drug Disclosure



Deep Learning Classifier

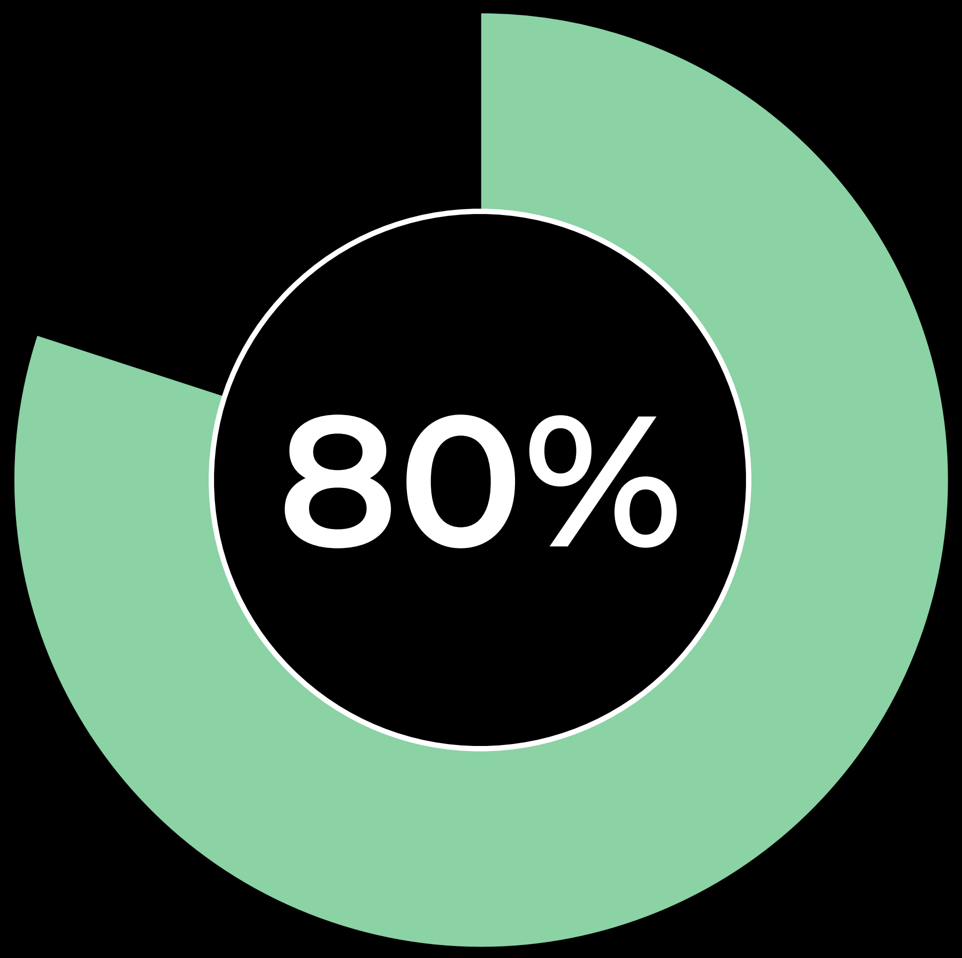
Detecting Drug Disclosure

Model	Accuracy	Macro Precision	Macro Recall	Macro F1
5-fold cross validation				
Text CNN	73.51 ± 3.10	78.79 ± 1.82	63.28 ± 5.78	62.34 ± 7.28
BERT	83.79 ± 1.03	82.13 ± 1.15	82.22 ± 1.29	82.14 ± 1.16
RoBERTa	83.16 ± 0.95	81.72 ± 1.32	80.96 ± 1.15	81.22 ± 0.98
Test set				
Text CNN	78.65	77.52	73.71	74.89
BERT	81.27	79.51	78.67	79.05
RoBERTa	**81.90**	**80.43**	**78.89**	**79.54**

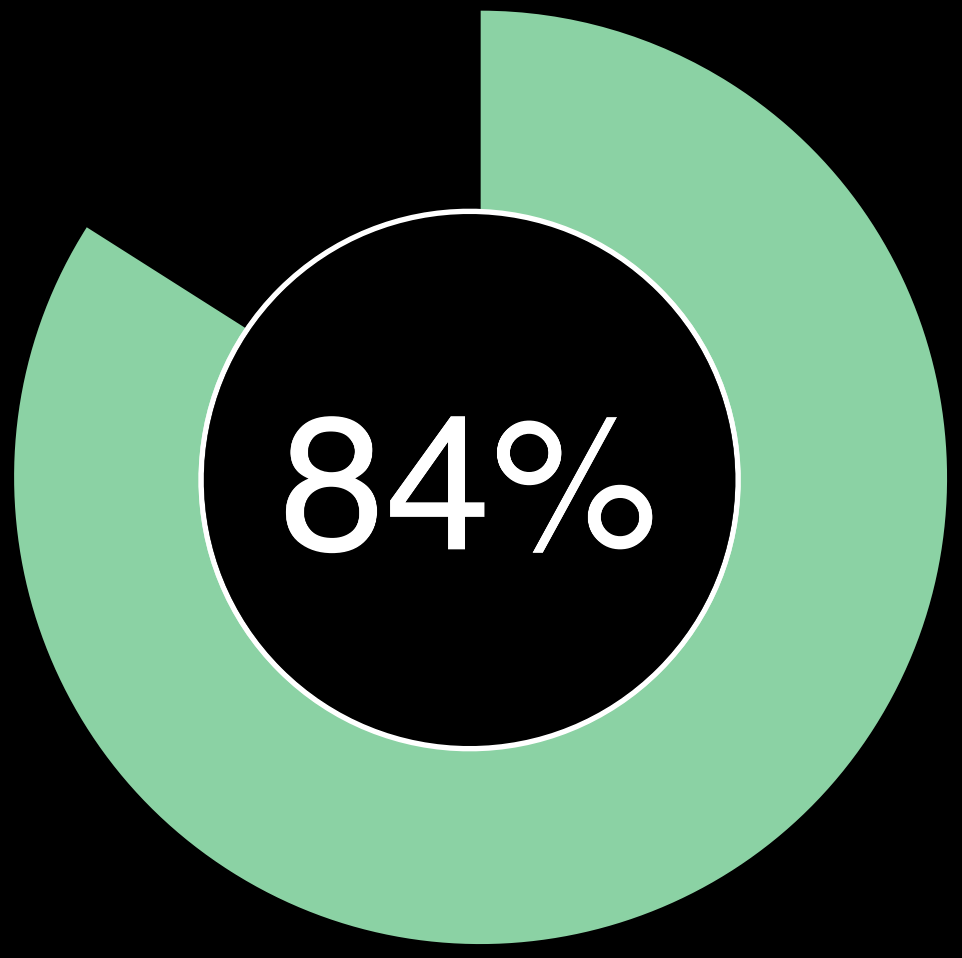
Individual subreddit performance equally good

Extent of Drug Consumption

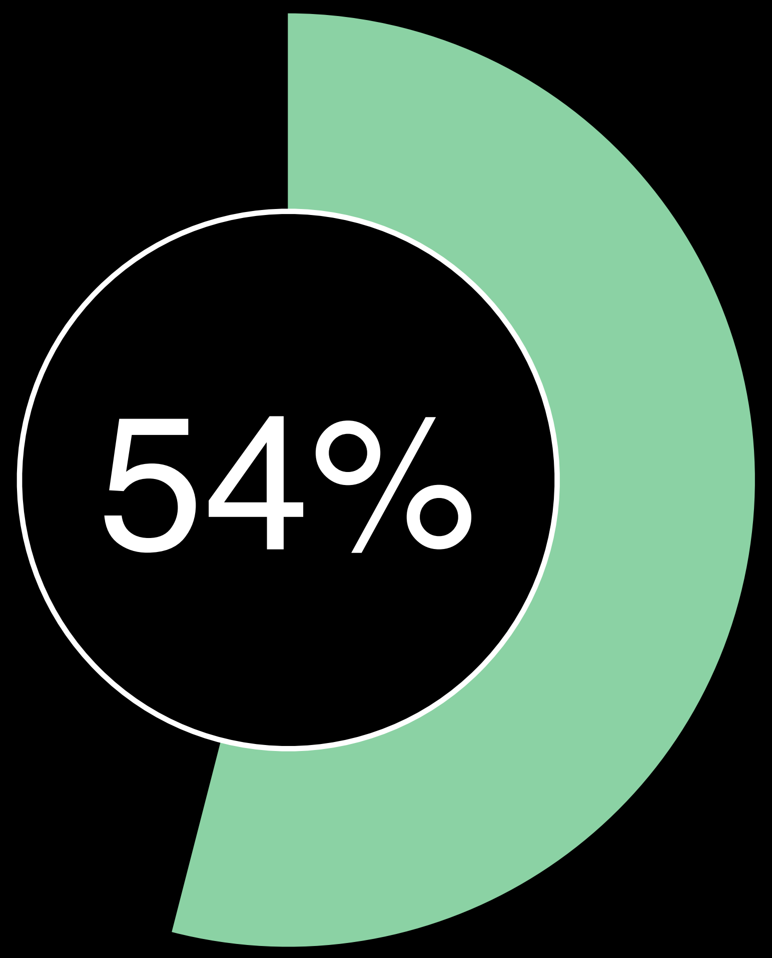
Extent of Drug Consumption



Users
indicate drug consumption
(84% in user study)



Posts
indicate drug consumption



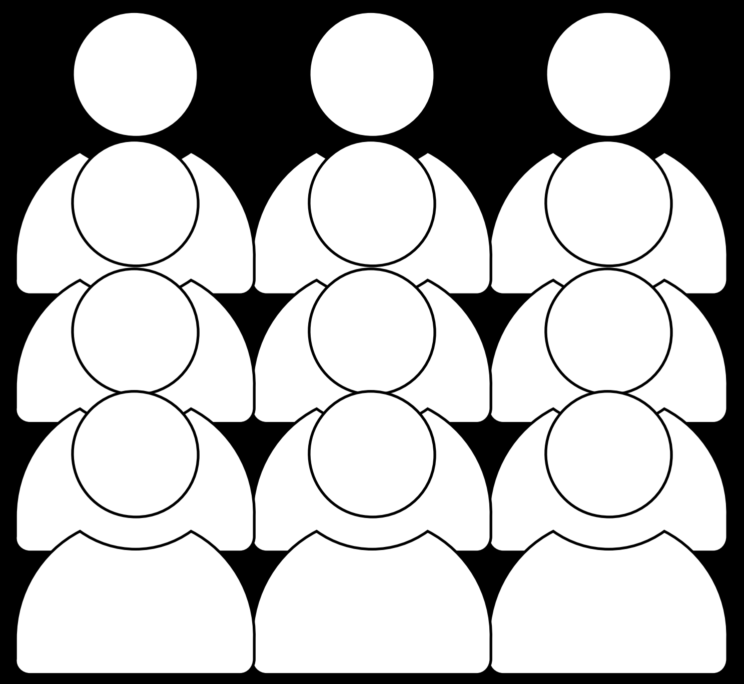
Comments
indicate drug consumption

Effect of Community Feedback

Effect of Community Feedback

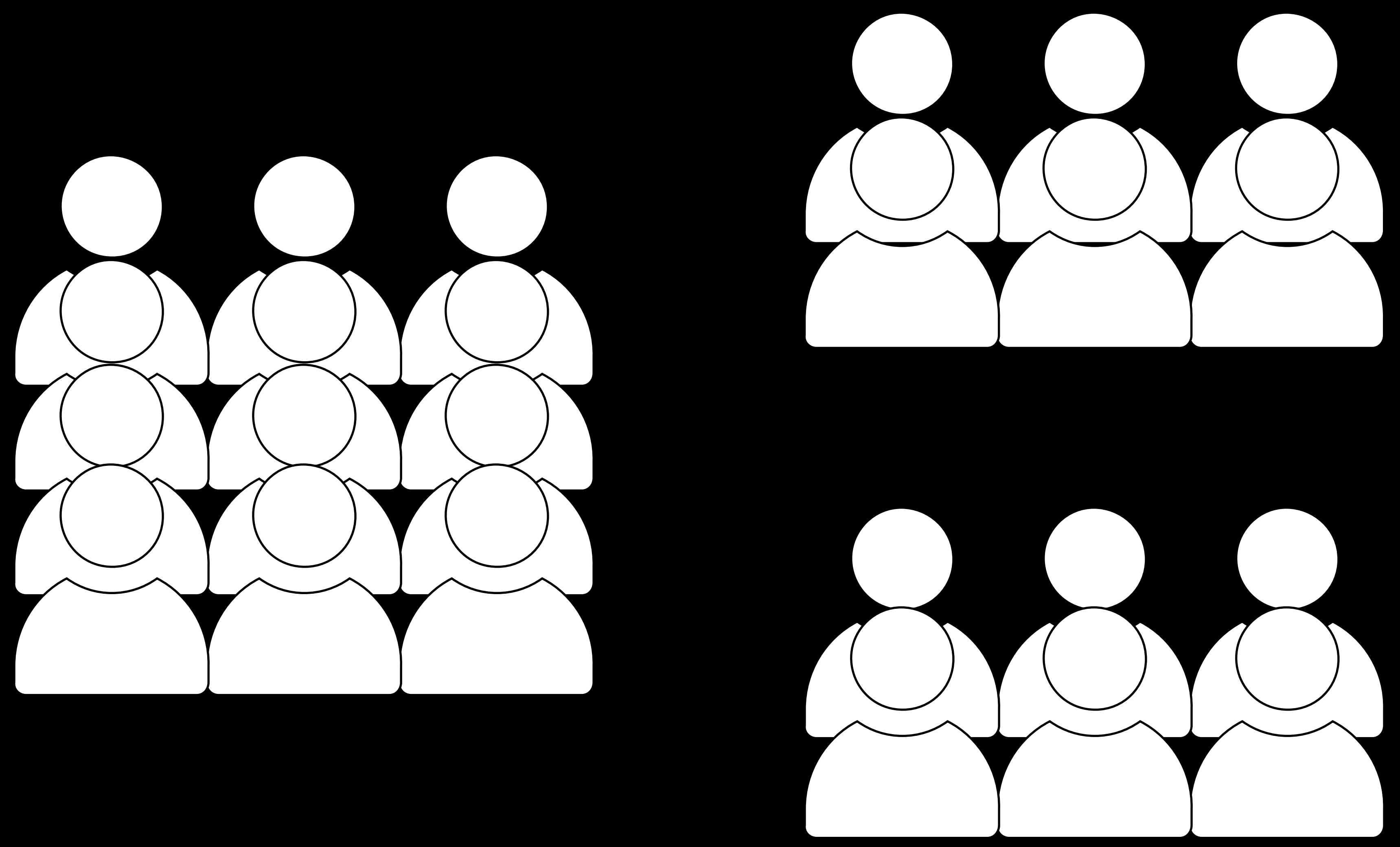
Propensity Score Matching

Effect of Community Feedback



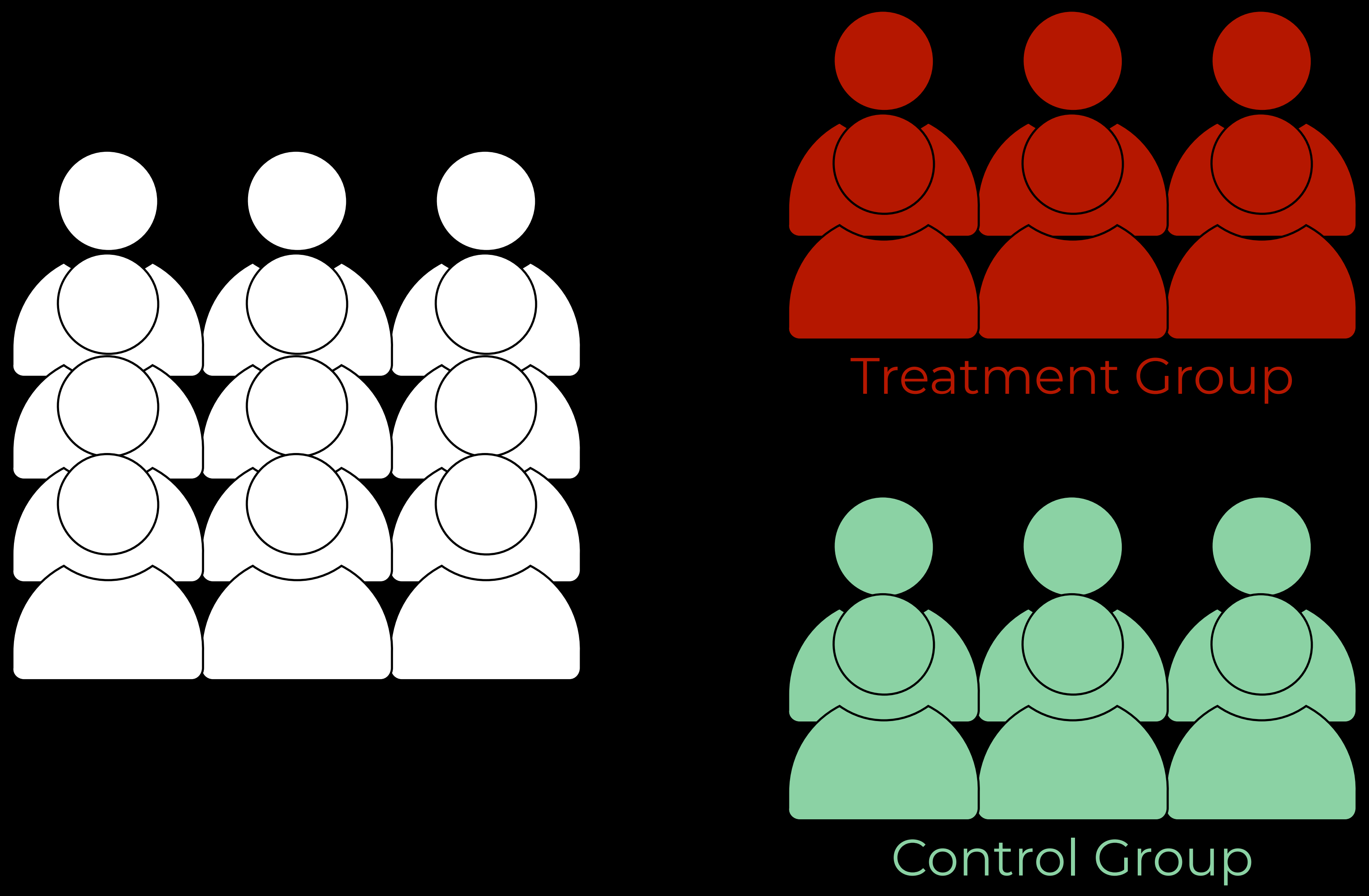
Propensity Score Matching

Effect of Community Feedback



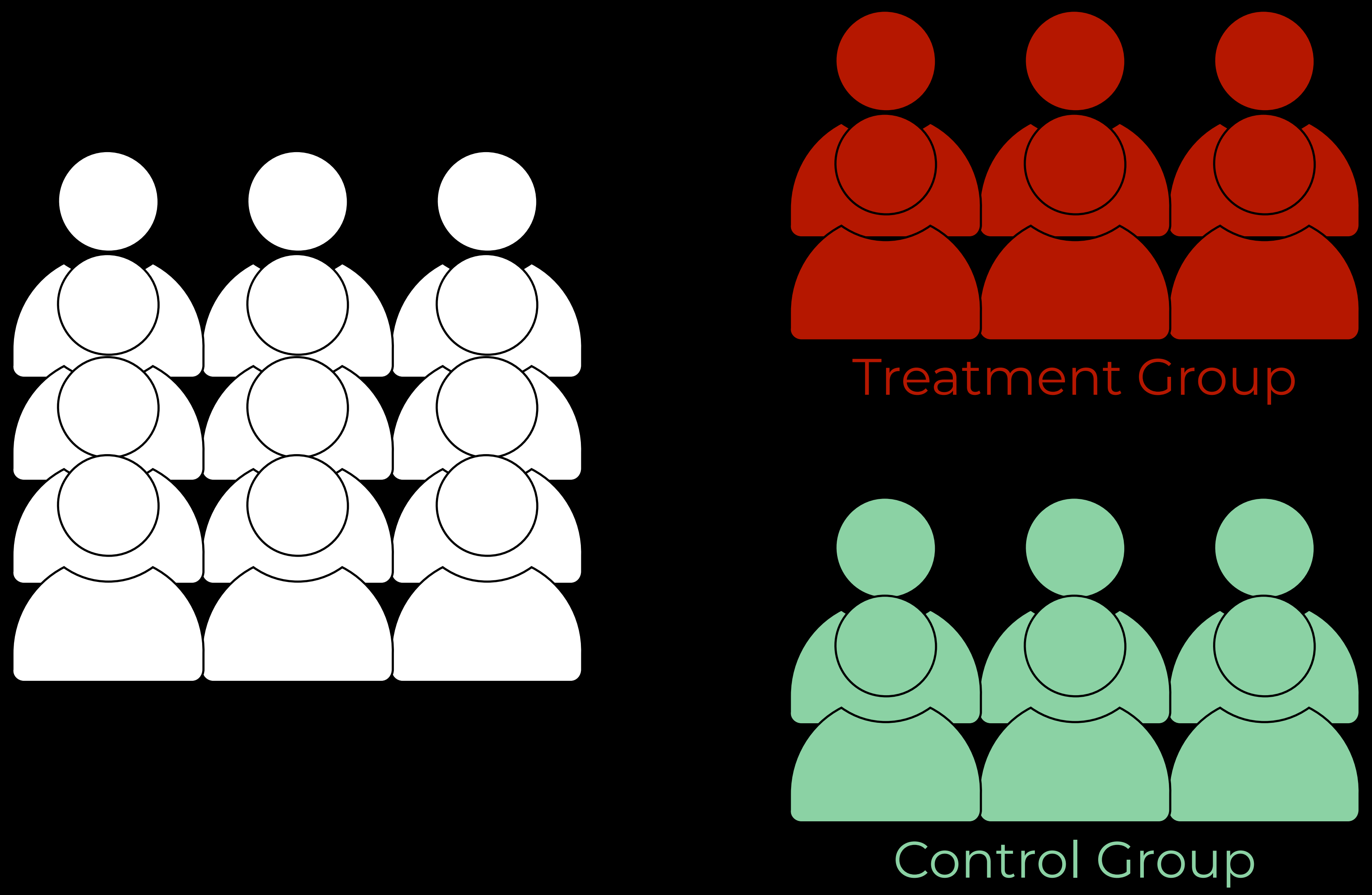
Propensity Score Matching

Effect of Community Feedback



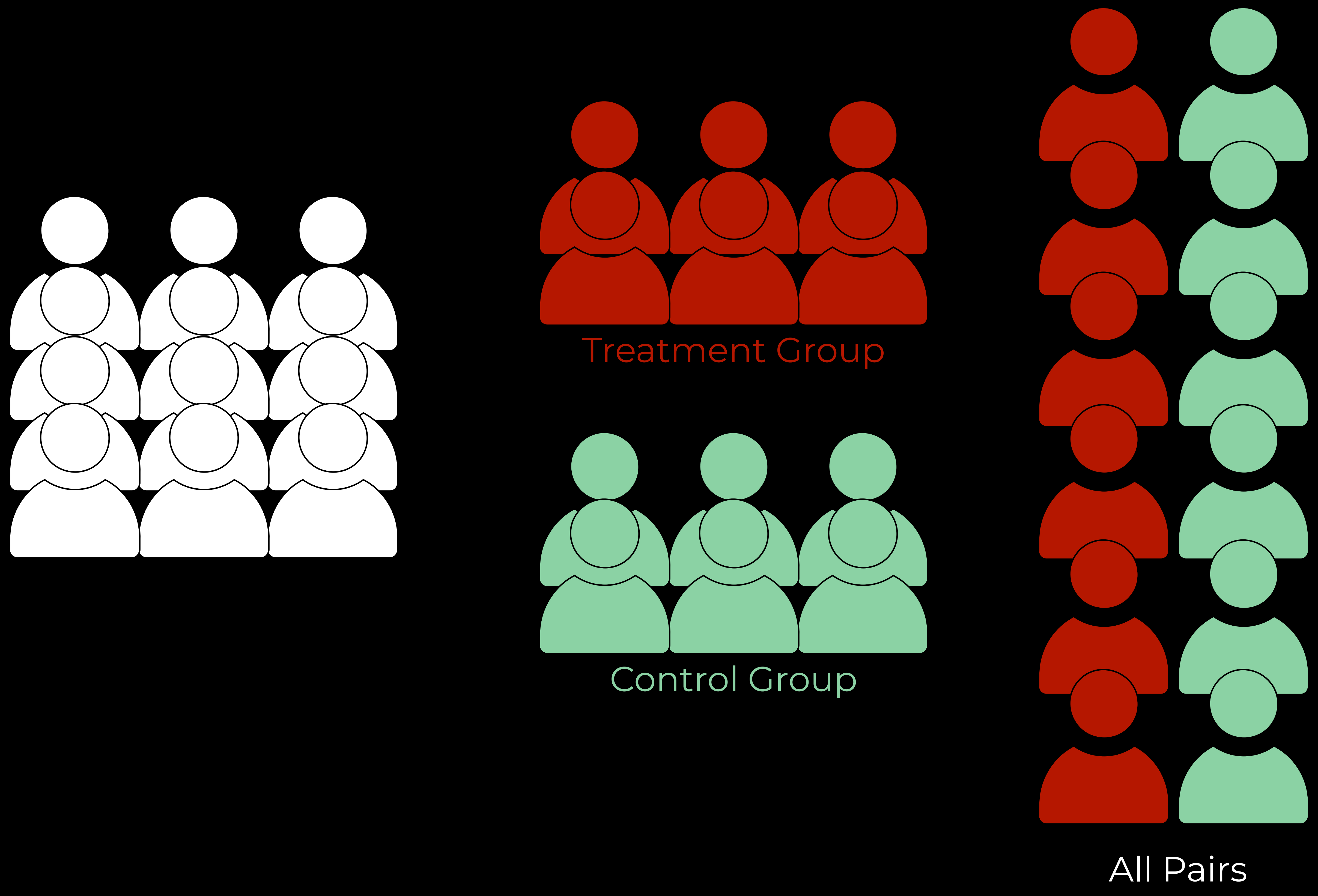
Propensity Score Matching

Effect of Community Feedback



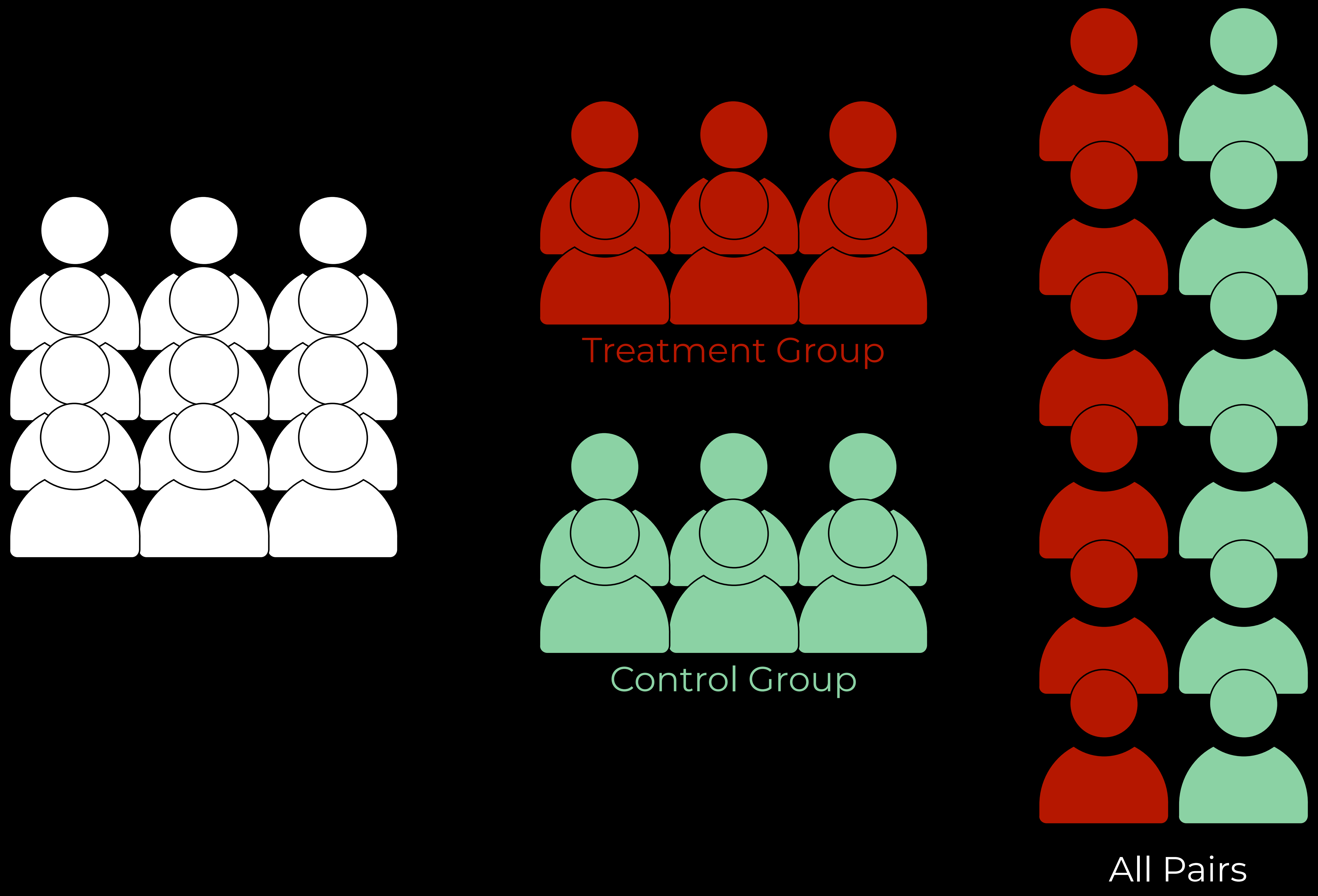
Propensity Score Matching

Effect of Community Feedback



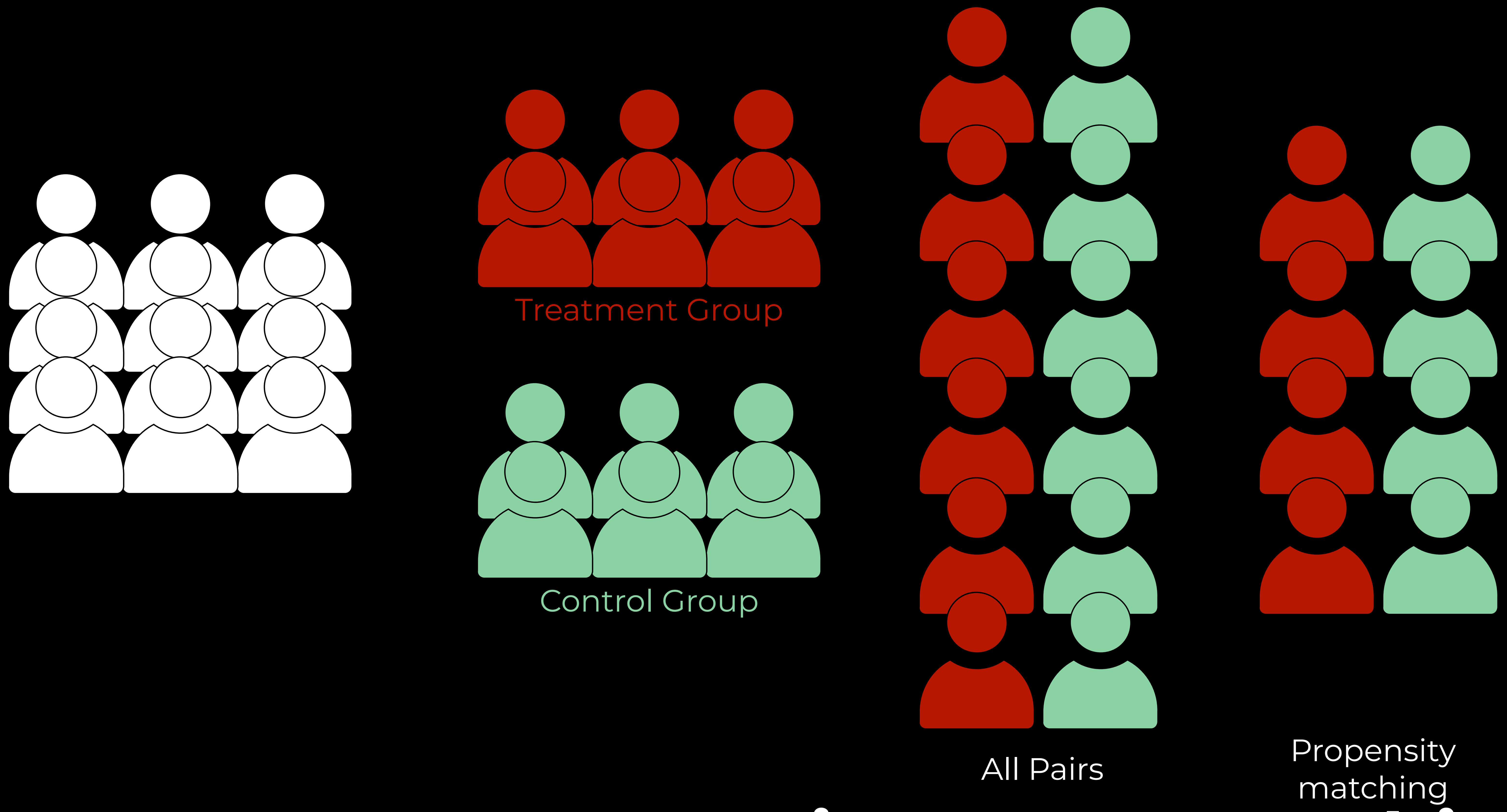
Propensity Score Matching

Effect of Community Feedback



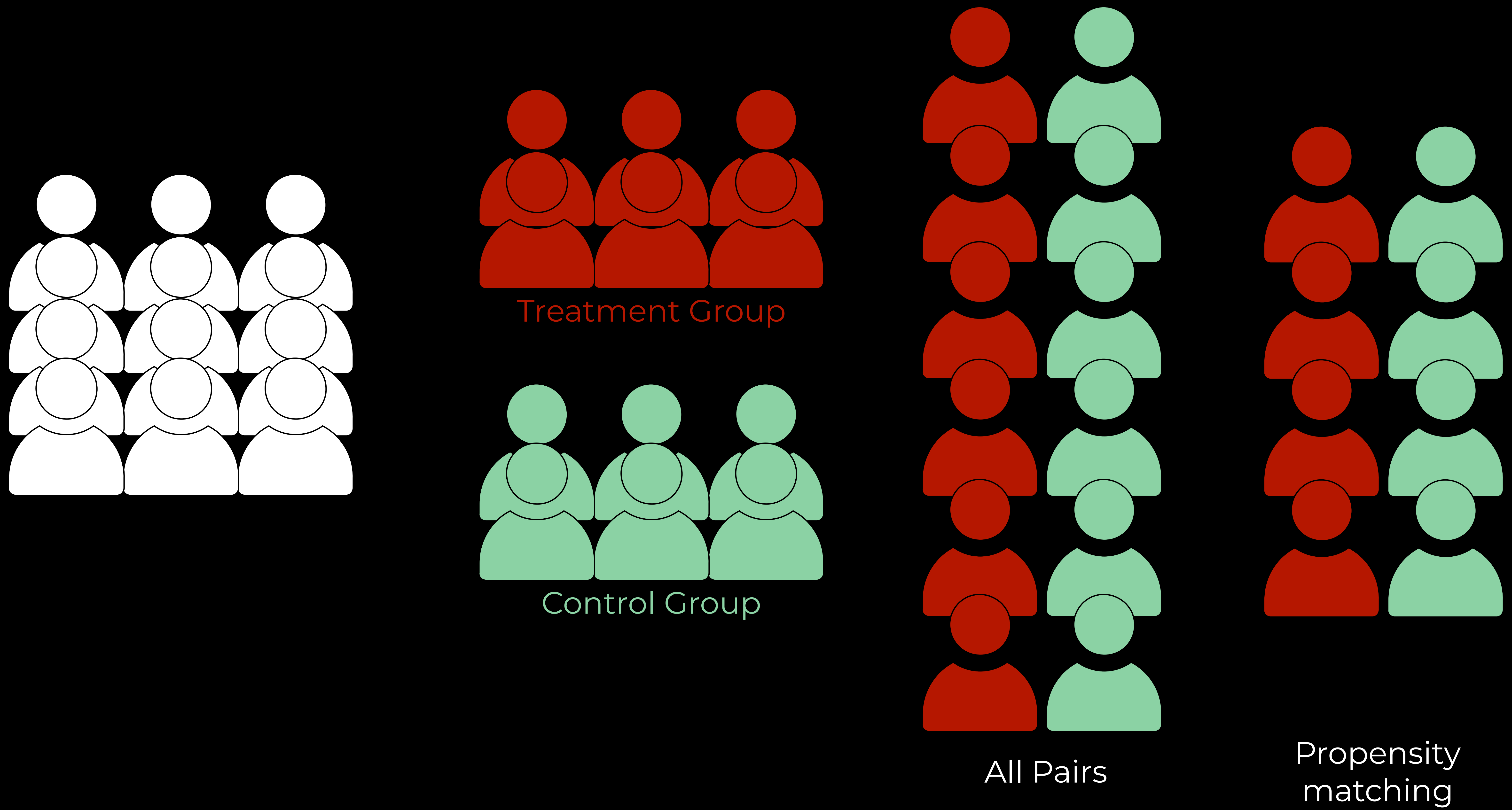
Propensity Score Matching

Effect of Community Feedback



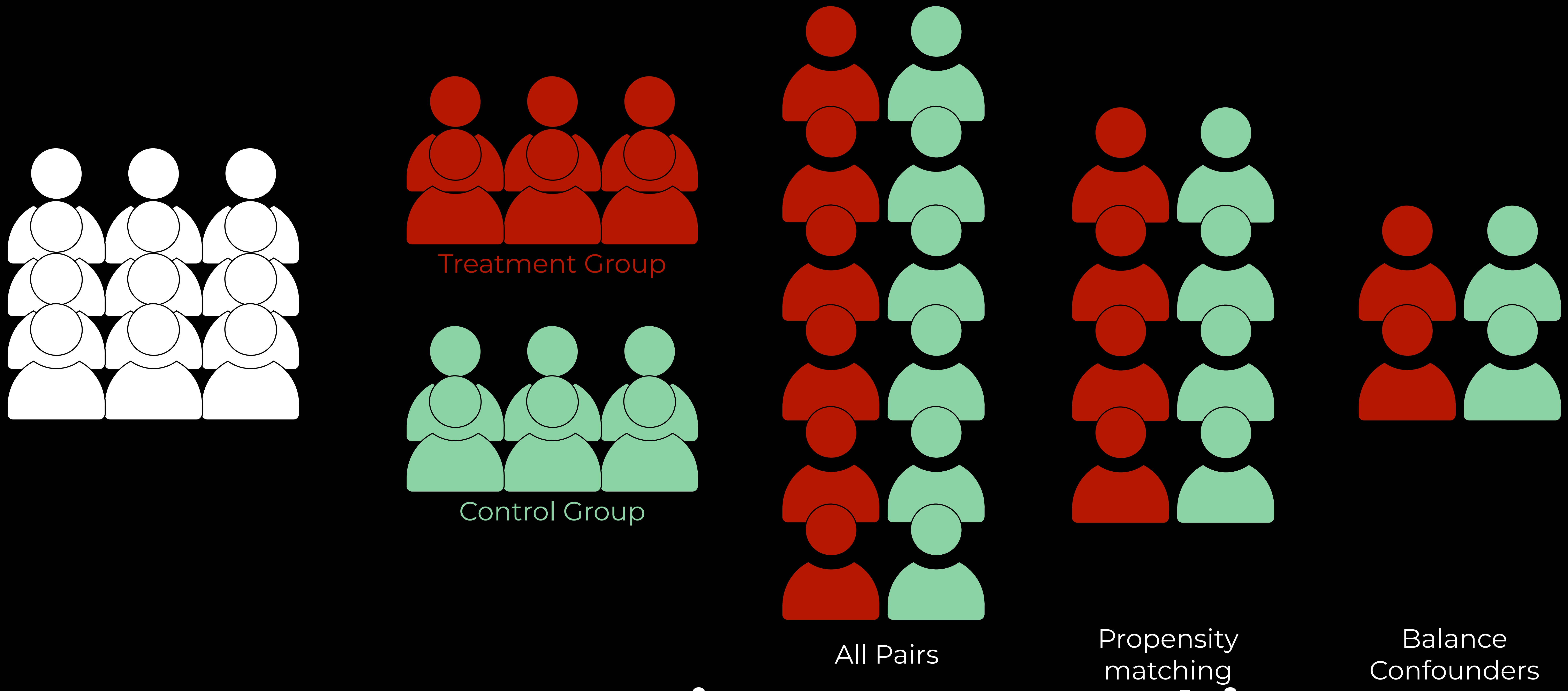
Propensity Score Matching

Effect of Community Feedback



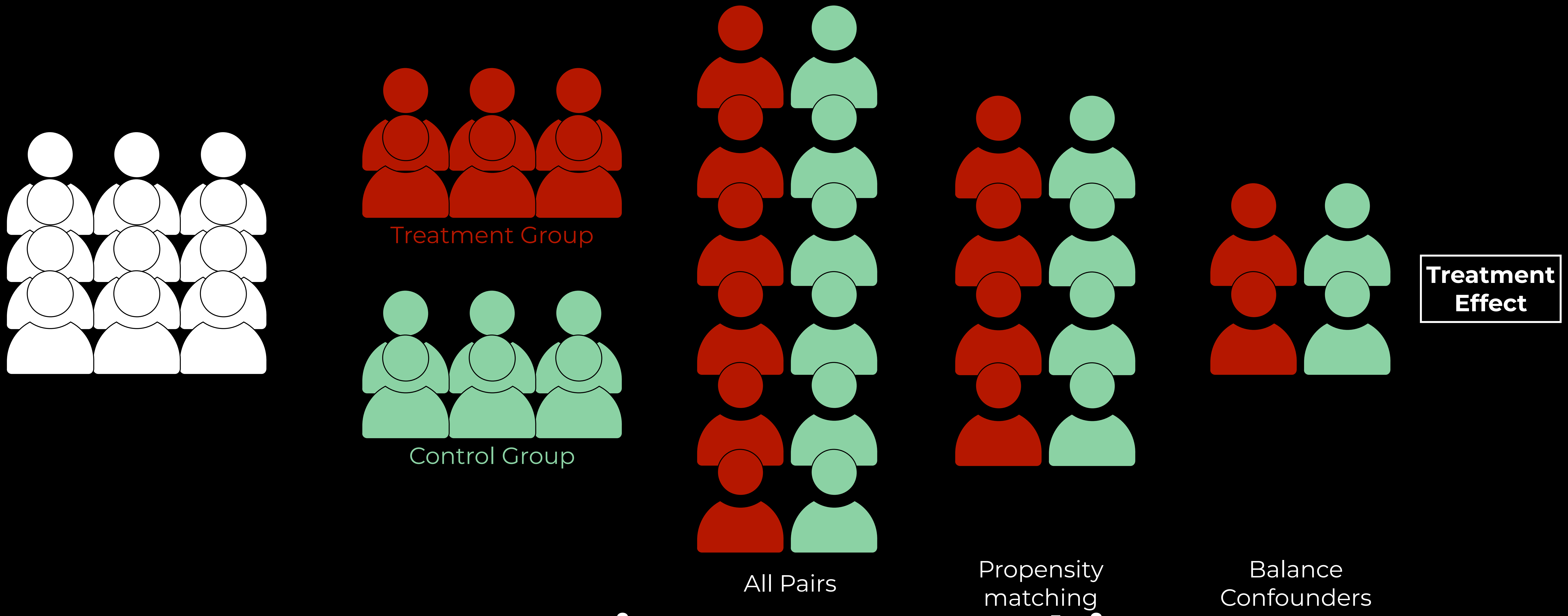
Propensity Score Matching

Effect of Community Feedback



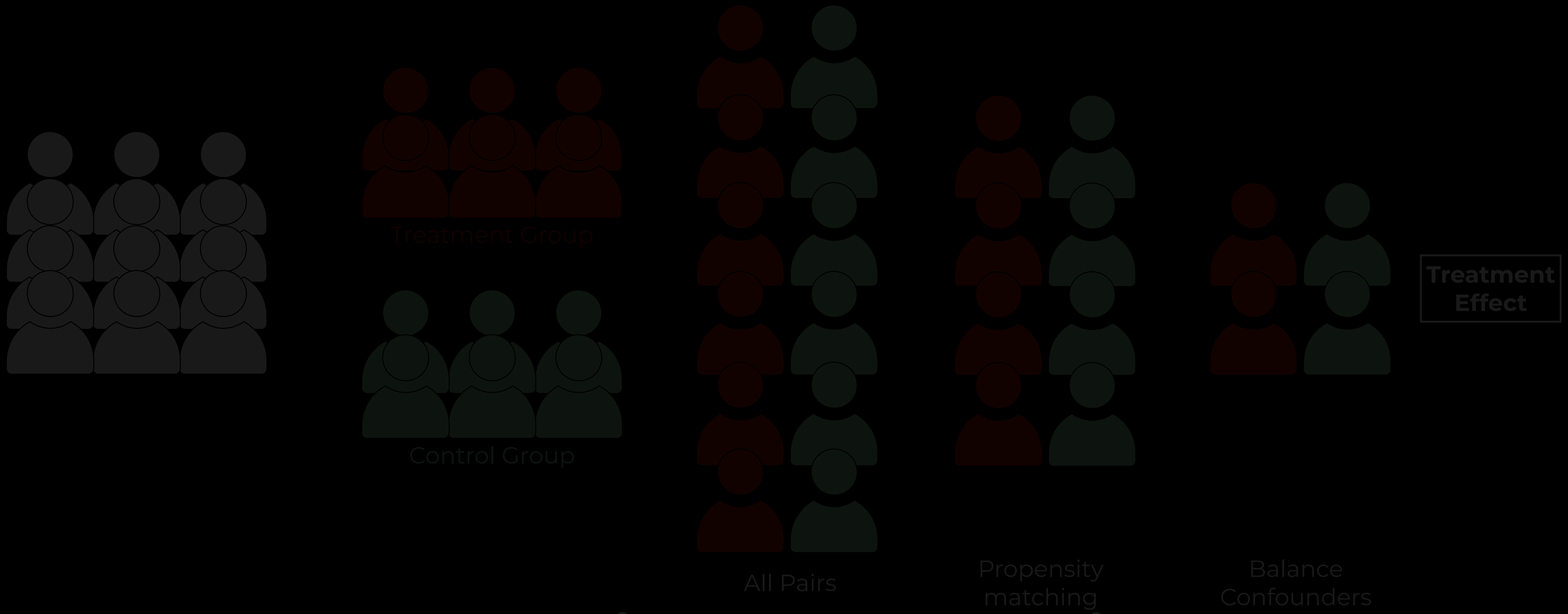
Propensity Score Matching

Effect of Community Feedback



Propensity Score Matching

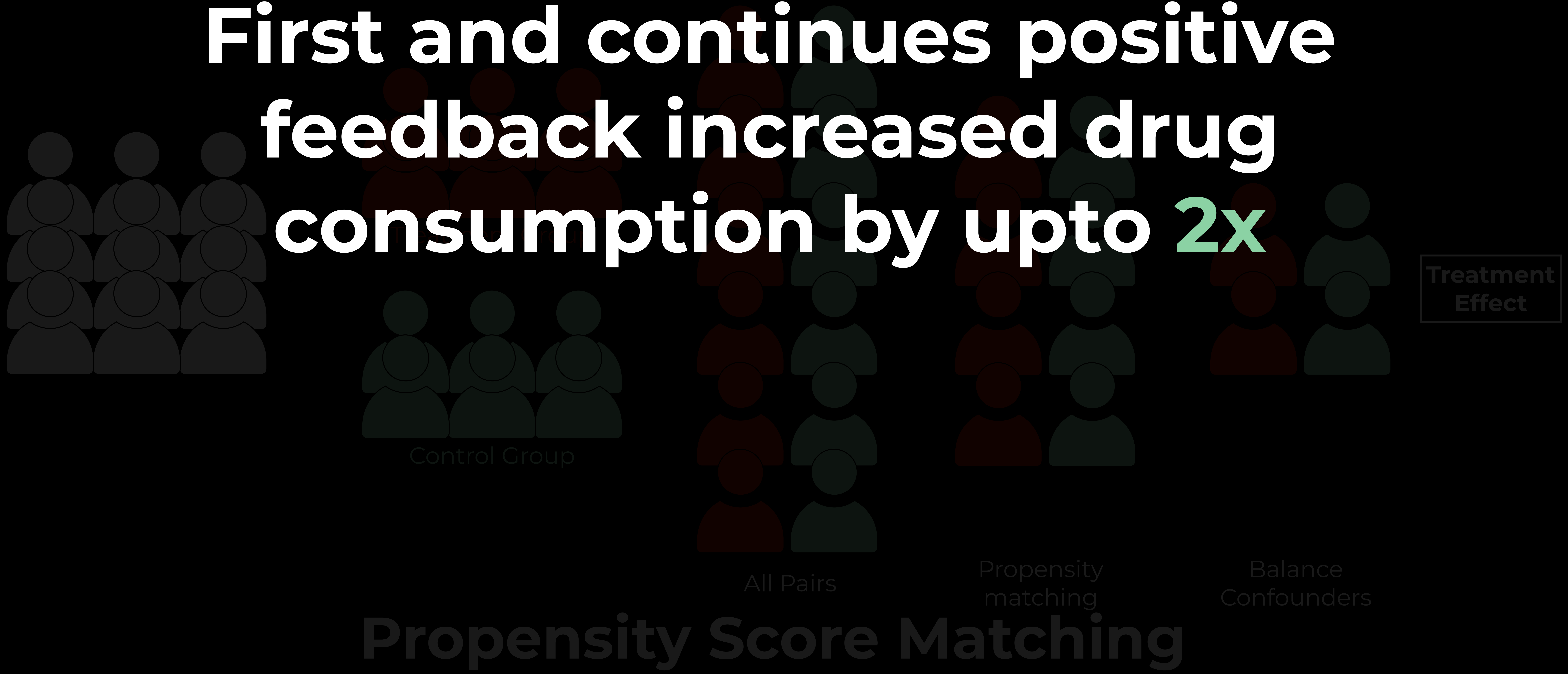
Effect of Community Feedback



Propensity Score Matching

Effect of Community Feedback

First and continues positive feedback increased drug consumption by upto 2x



Effect of Community Feedback

First and continues positive feedback increased drug consumption by upto 2x

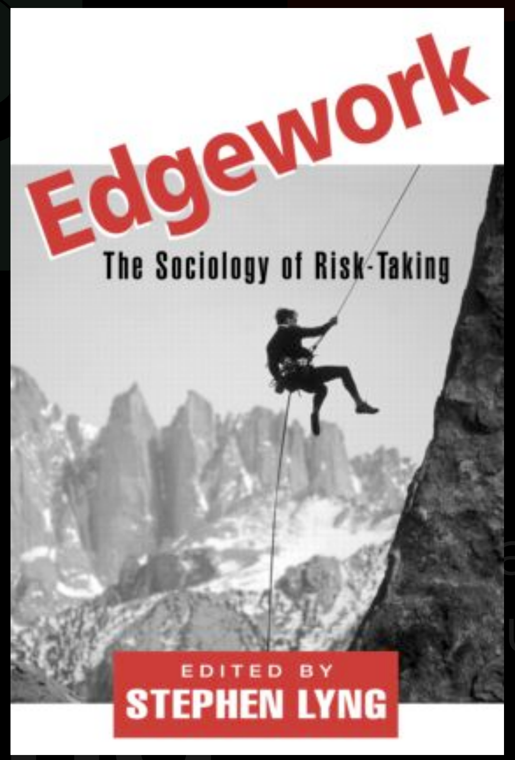
User belief

Scores 2.2/5

Comments 2.5/5

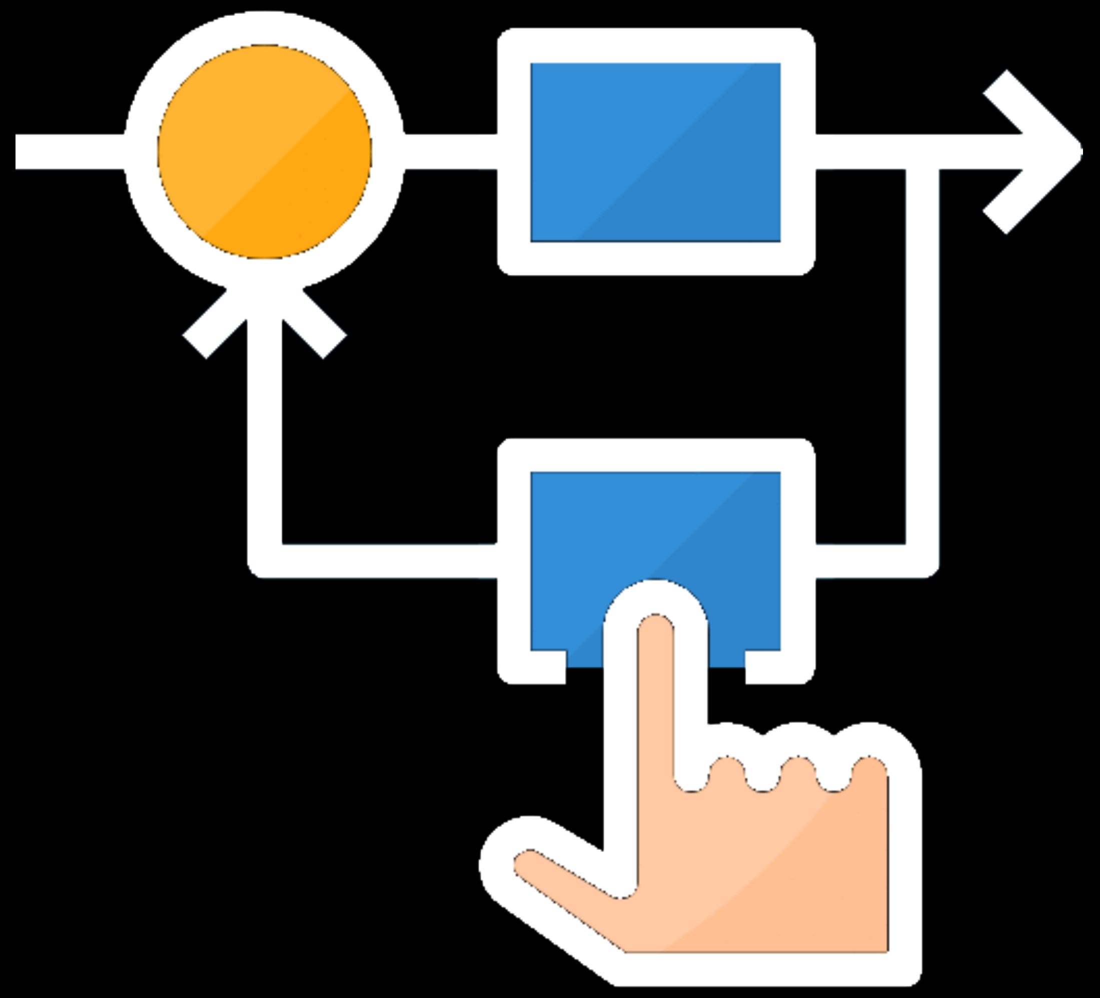
(Little to Moderate impact)

“Illusion of control”
(Lyng 1990)



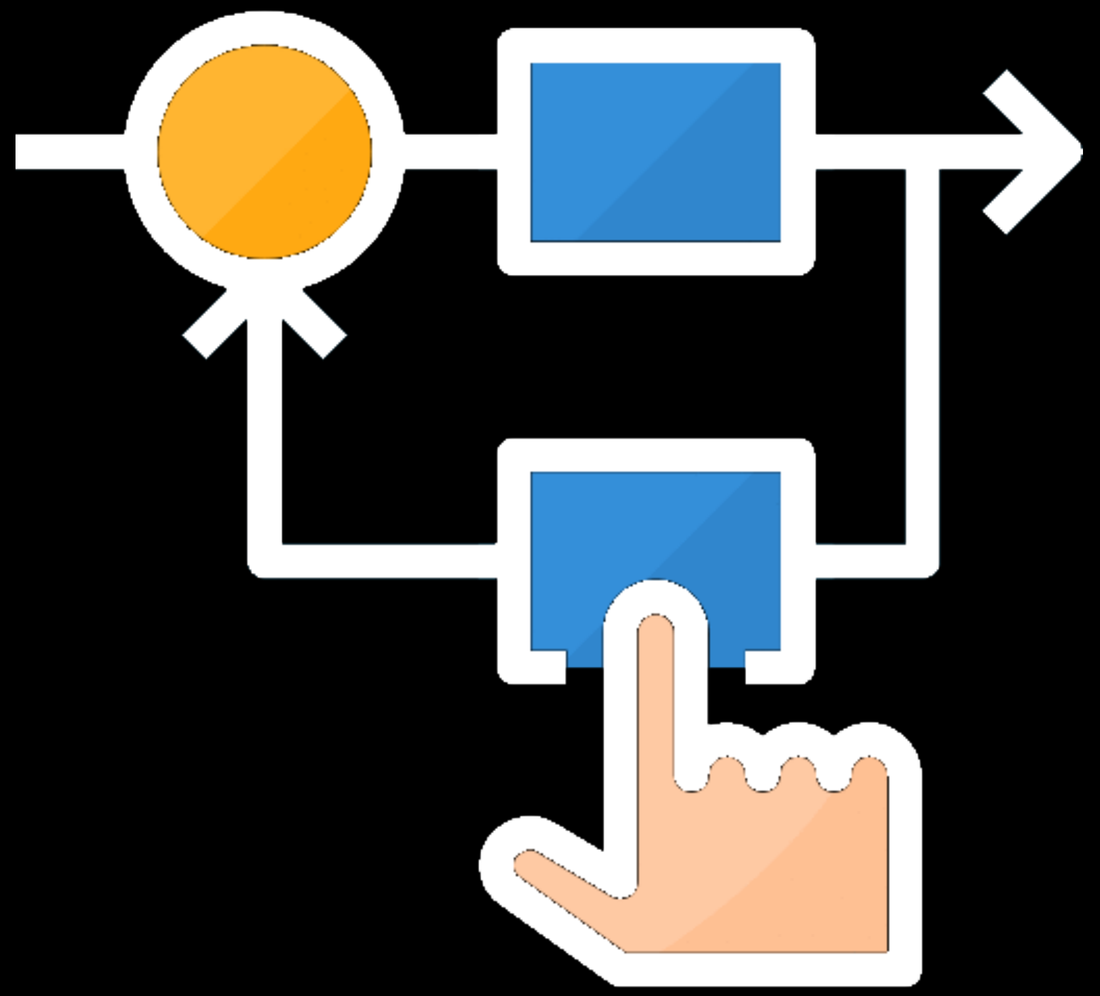
Implications

Implications



Feedback mechanisms

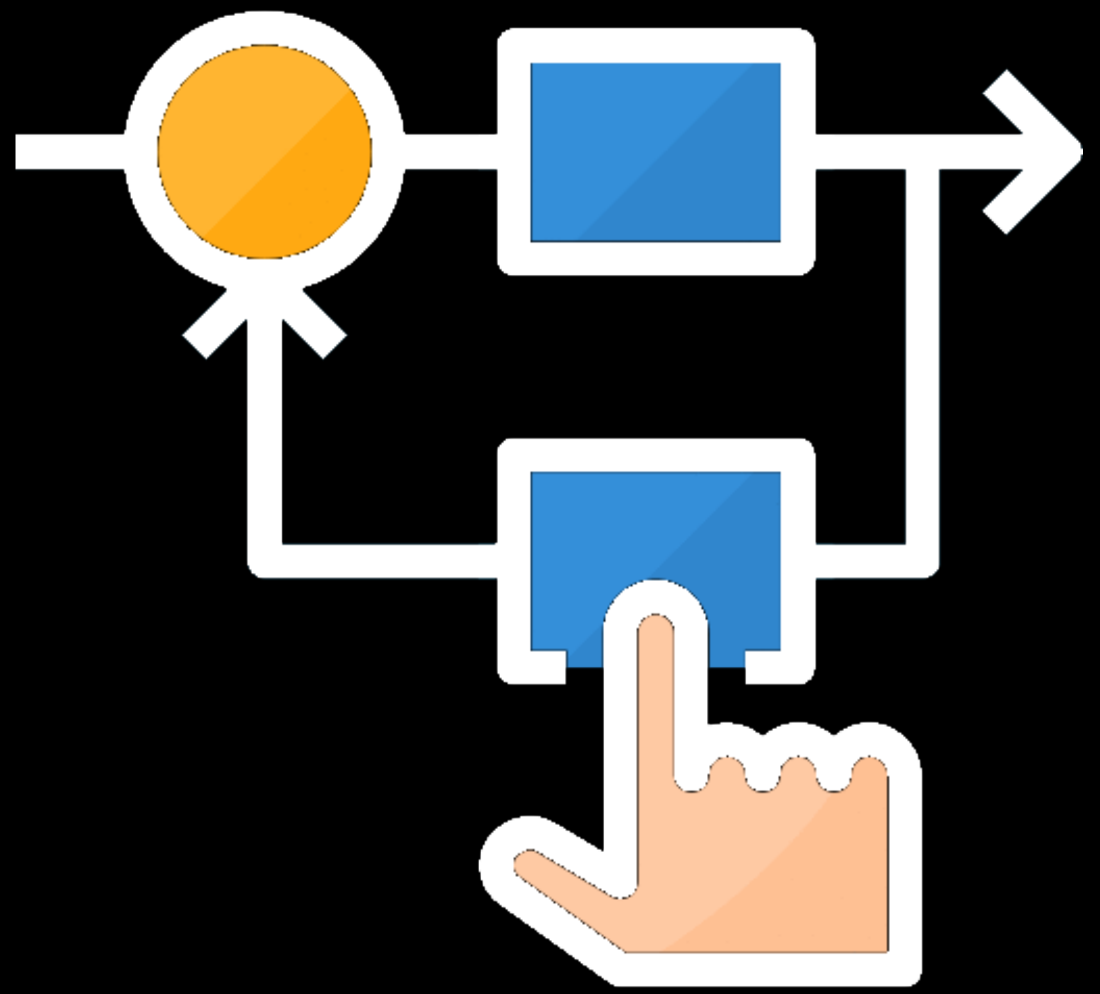
Implications



Feedback mechanisms

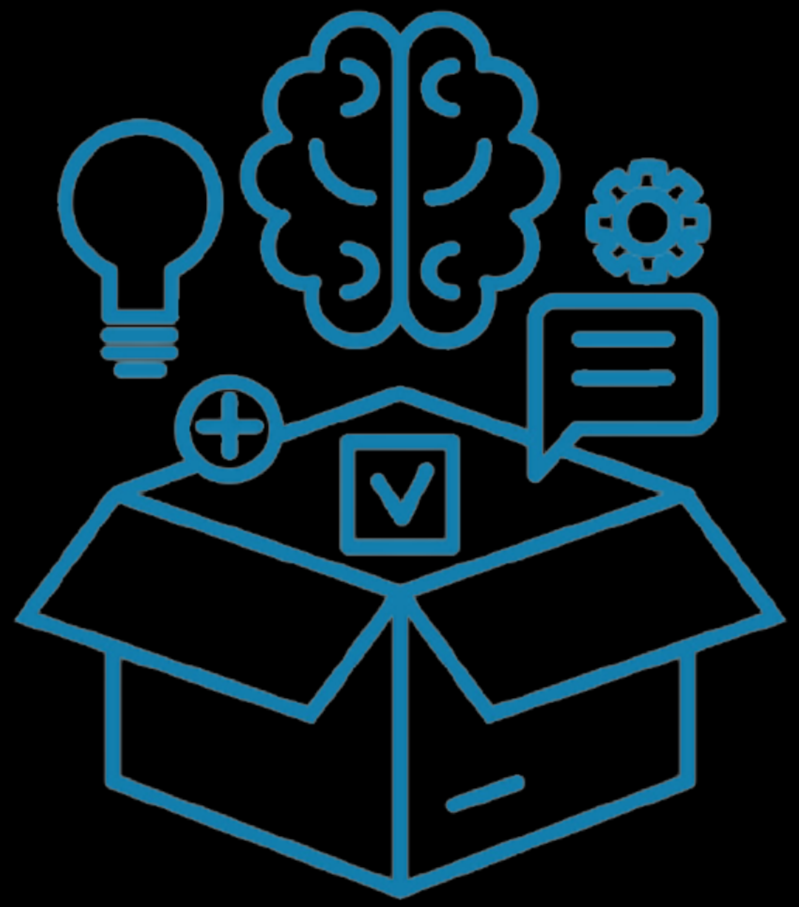
Intervention

Implications



Feedback mechanisms

Intervention



Resource

Contributions

Contributions

First paper to study effect of community feedback on drug consumption disclosure

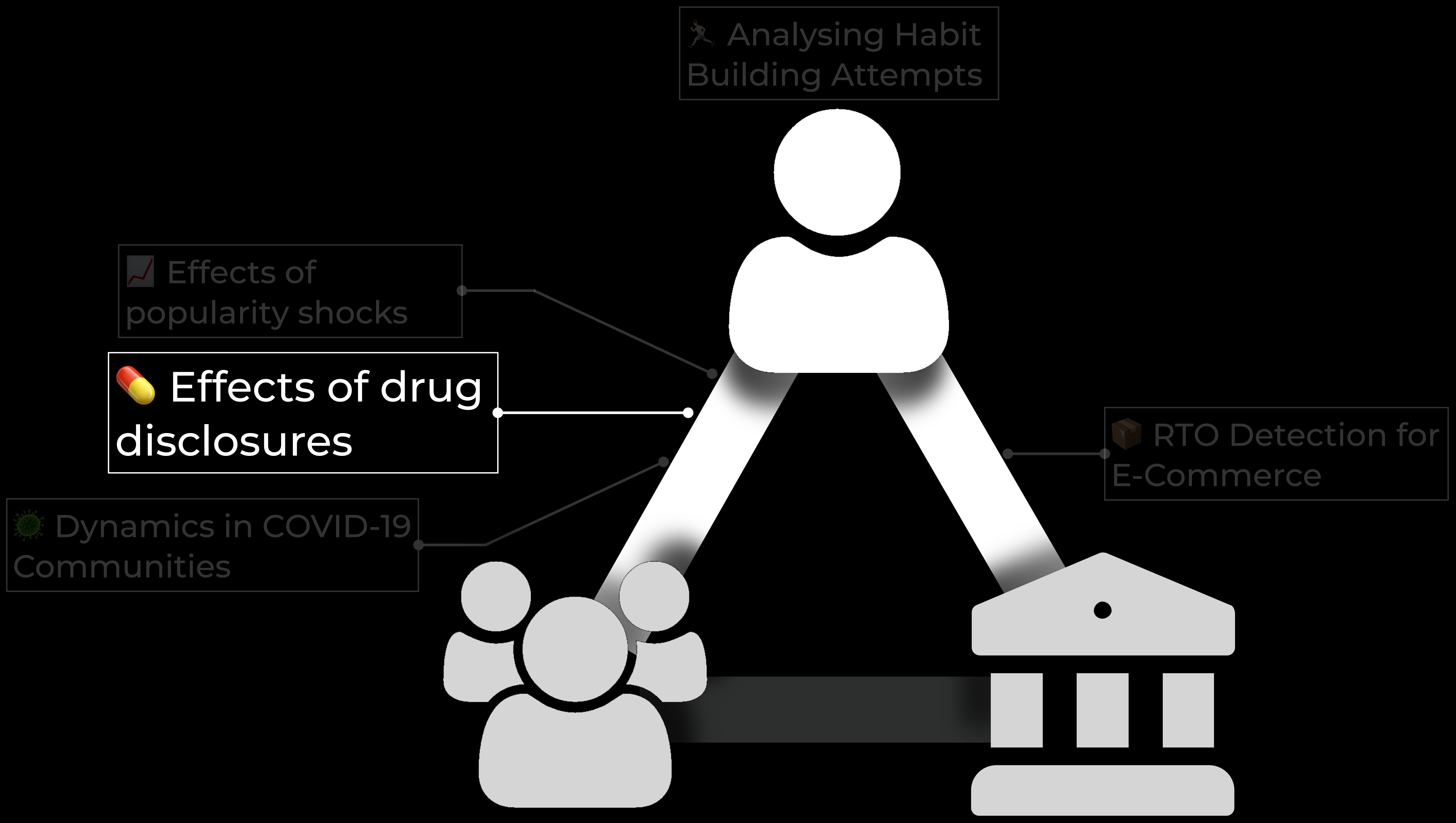
Contributions

- First paper to study effect of community feedback on drug consumption disclosure
- Actionable insights for moderators and platforms

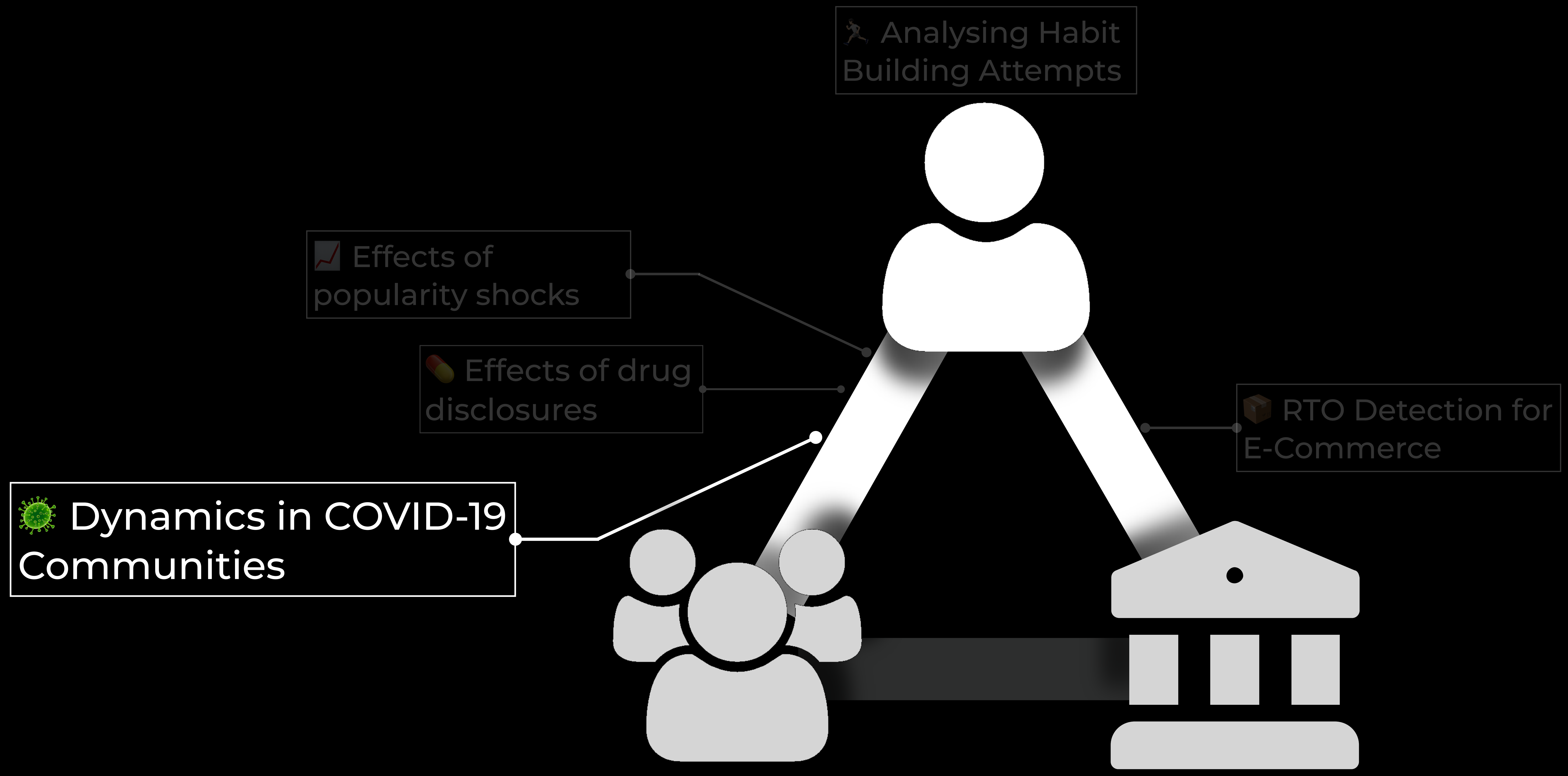
Contributions

- First paper to study effect of community feedback on drug consumption disclosure
- Actionable insights for moderators and platforms
- First dataset to provide drug consumption indicative posts and model checkpoints

Our Focus


🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

Our Focus



🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
📦 RTO Detection for E-Commerce
🦠 Dynamics in COVID-19 Communities

Together Apart: Decoding Support Dynamics in Online COVID-19 Communities

ASONAM' 23

Hitkul, Tanisha Pandey, Sonali Singhal, Pranjal Kandhari,
Aryamann Tomar, Ponnurangam Kumaraguru

Why?

Why?

Why?

Necessary to understand user dynamics to help moderators
and platform designers increase accessibility

Data

Data

r/COVID19positive

Data

r/COVID19positive

93,576 Posts
993k Comments

Data

r/COVID19positive

93,576 Posts
993k Comments

104,818 Users

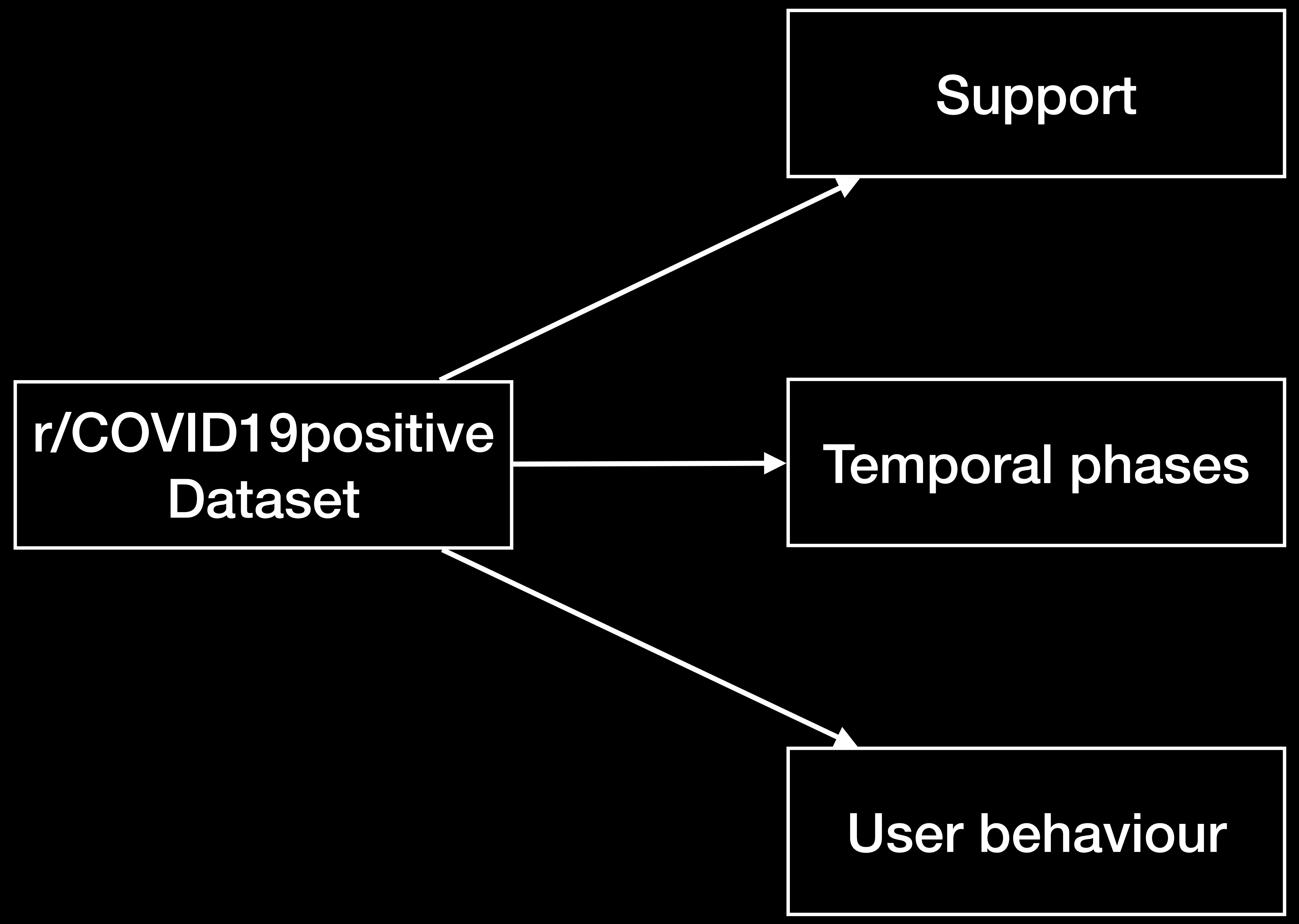
Data

Data Classification

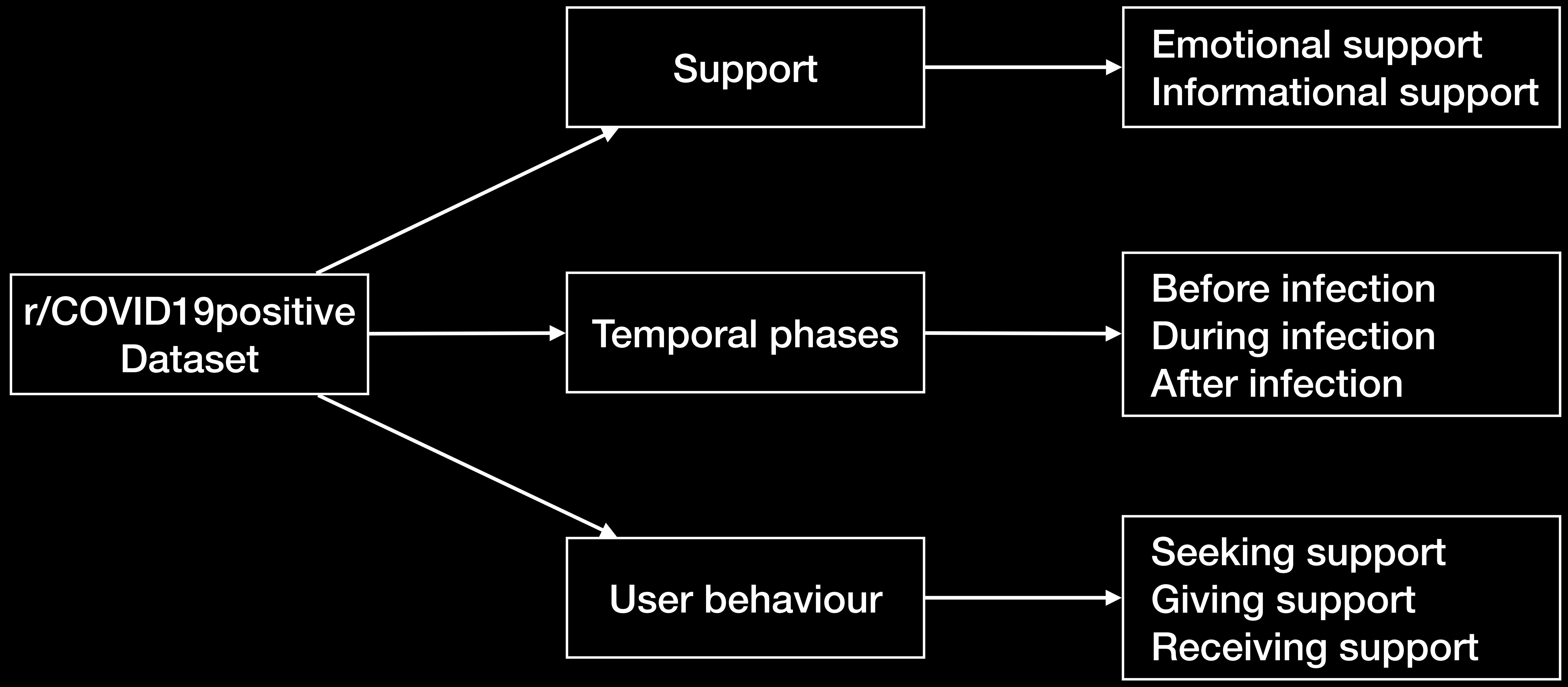
Data Classification

r/COVID19positive Dataset

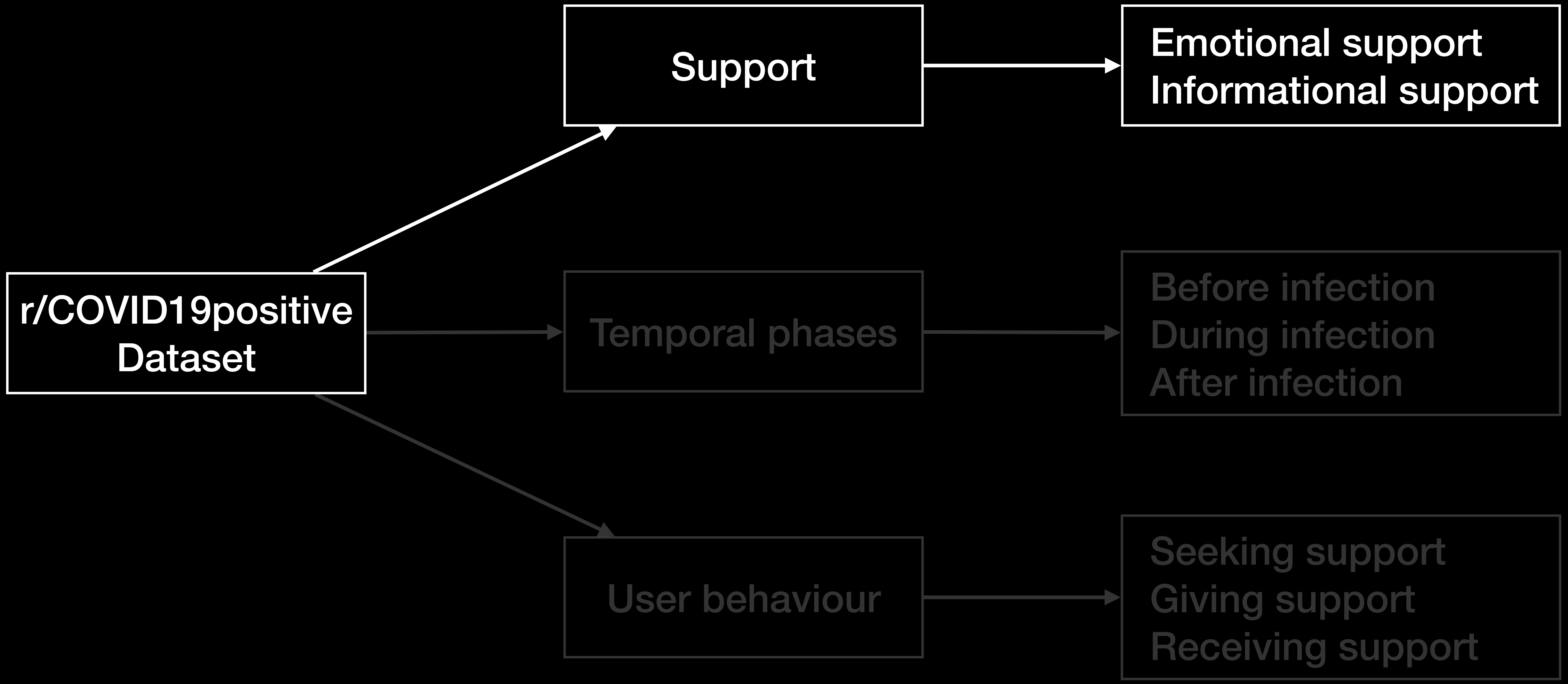
Data Classification



Data Classification

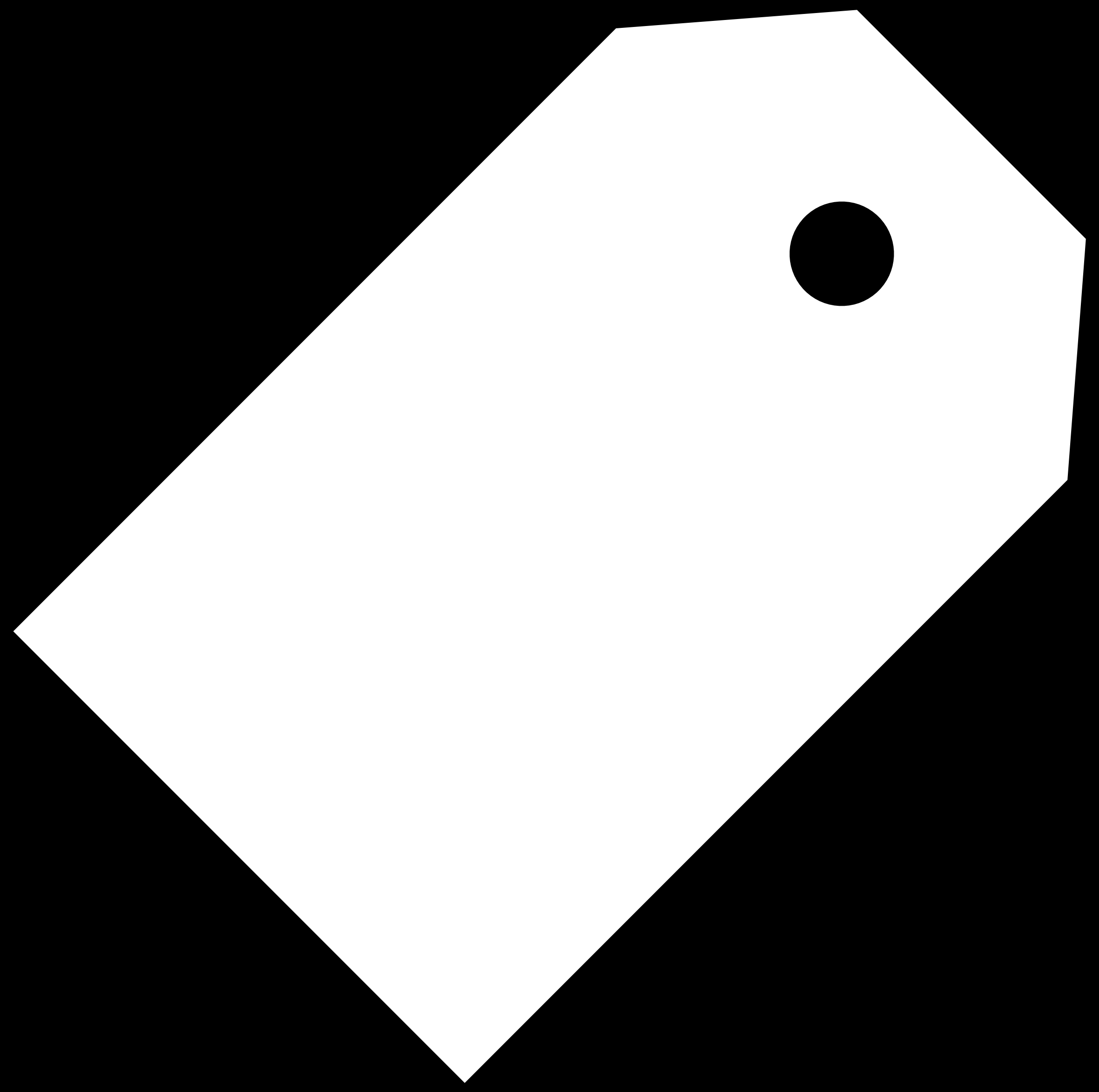


Data Classification



Support Classification

Support Classification



Flair Based Classification

Support Classification

Flair Based Classification

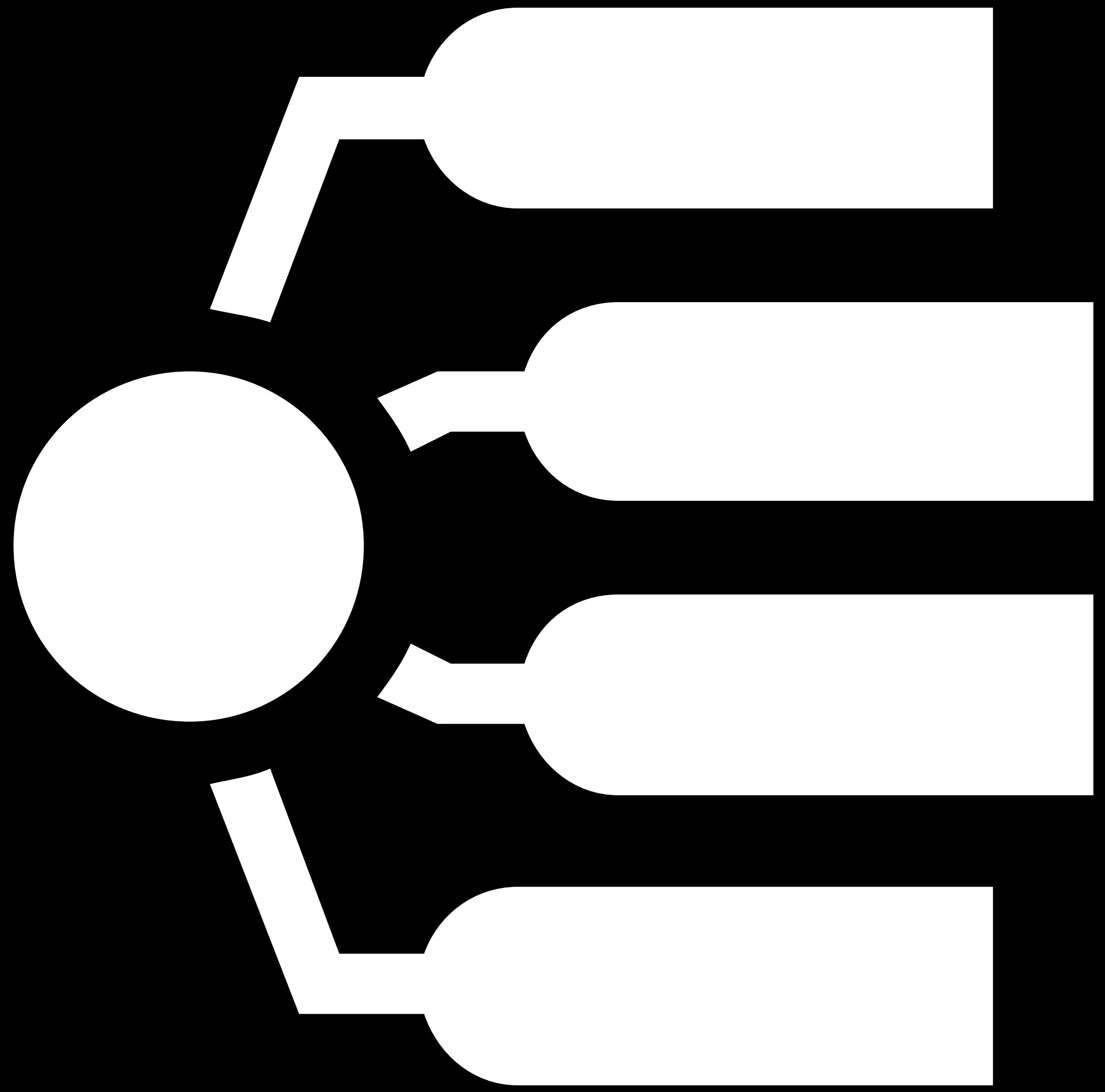
Support Classification

	Flair	Posts	Comments
Emotional flair	TP	4,219 (4.51%)	46,302 (4.95%)
	TP - Me	26,975 (28.82%)	266,806 (28.53%)
	TP - Family	7,452 (7.96%)	85,243 (9.12%)
	TP - Friends	1,575 (1.68%)	15,588 (1.67%)
	TP - LongHauler	785 (0.84%)	6,860 (0.73%)
	TP - Unvaccinated	269 (0.29%)	3,237 (0.35%)
	TP - Breakthrough	1,034 (1.1%)	10,431 (1.11%)
Informational flair	Verified Research	183 (0.19%)	723 (0.07%)
	Question - to those who tested positive	25,439 (27.18%)	212,748 (22.75%)
	Question - for medical research	4,311 (4.61%)	38,055 (4.07%)

TABLE I
TOTAL POSTS AND COMMENTS FOR DIFFERENT FLAIRS ON THE COVID19POSITIVE DATASET. TP = TESTED POSITIVE.

Support Classification - Validation

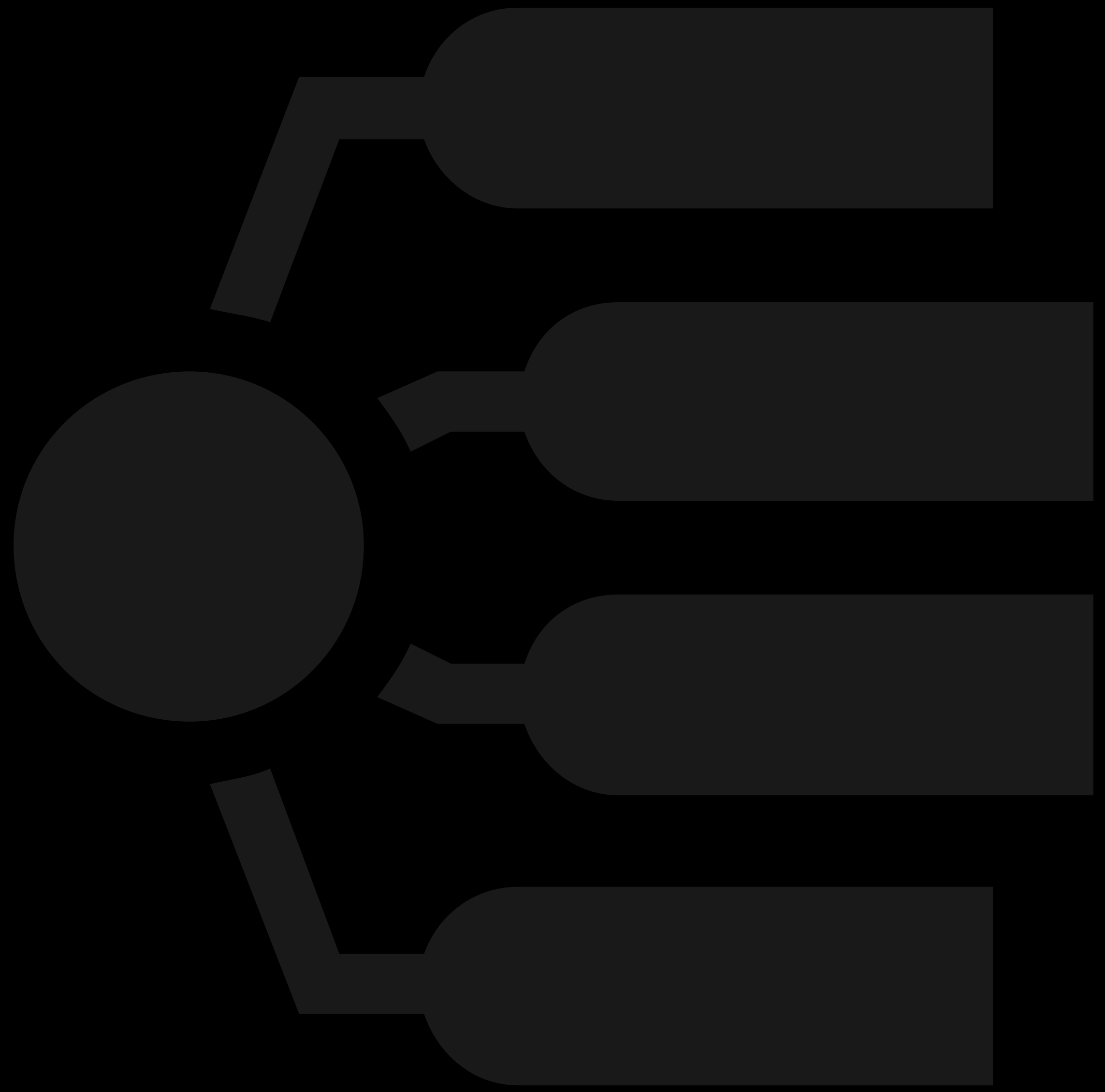
Support Classification - Validation



Topic Modeling

(BERTopic + TF-IDF + Odds Ratio)

Support Classification - Validation

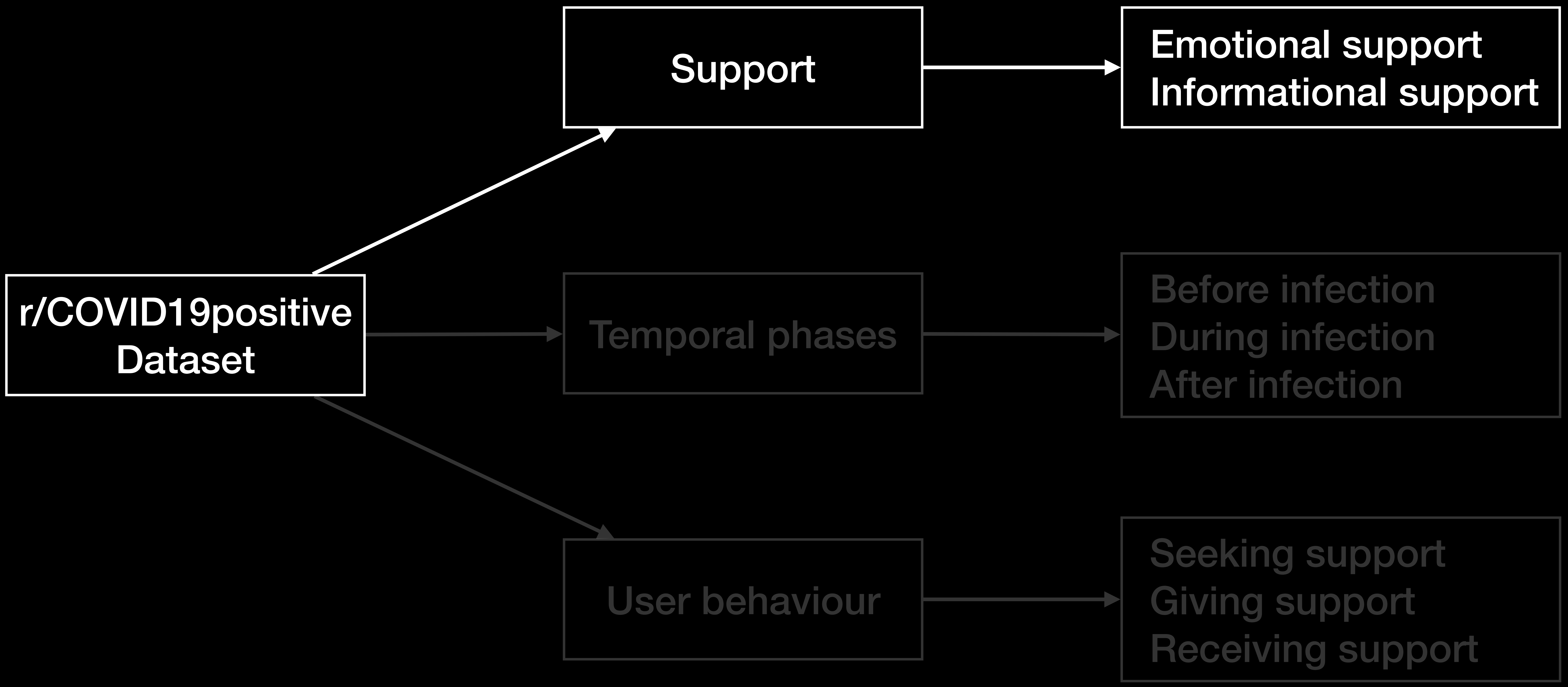


Topic Modeling
(BERTopic + TF-IDF + Odds Ratio)

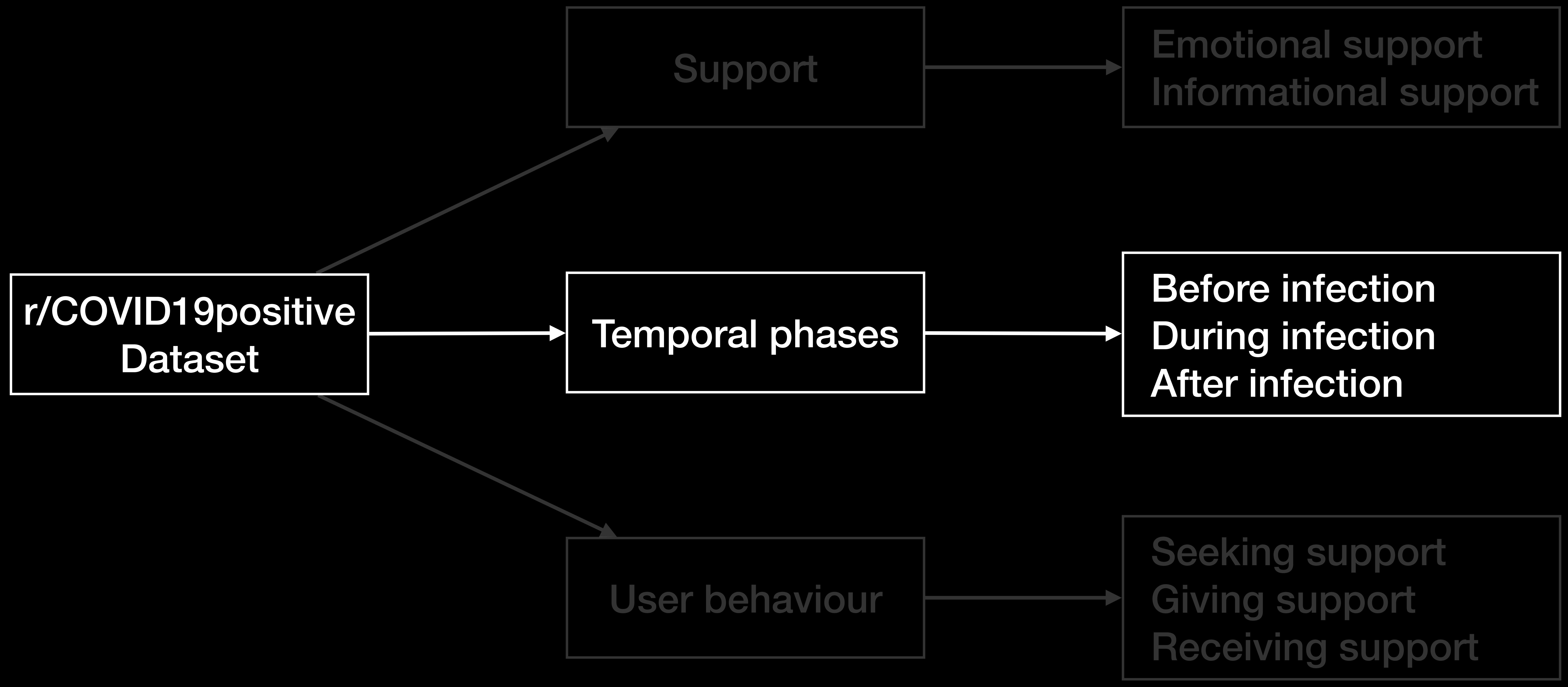
Support Classification - Validation

Support seeking				Support giving			
Informational		Emotional		Informational		Emotional	
Topic	**Words**	**Topic**	**Words**	**Topic**	**Words**	**Topic**	**Words**
Symptoms	smell, taste, body, temp, headache, breathe, shortness, asthma, insomnia	Sickness and Family	dad, mom, father, hospital, baby, son, kid, cough, daughter, husband, fever, toddler, family	Infection	swab, allergy, response, nose, itchy, immune, system, fever, breath, shortness, fatigue, headache	Love and Care	love, hug, hugs, sending, virtual, glad, feel, sorry, loss, grief supportive, healthcare
Nutritional and Lifestyle advice	appetite, eat, weight, exercise, take, taking, ivermectin, zinc, quercetin, paxlovid, vitamin, supplement	Recovery	contagious, still, positive, quarantine, resting, test, tested, exercise, run, workout, walk, recovered, longer, isolate, isolation, day	Wellness and Rest	chicken, soup, fruit, immunity, immune, taking, b12, daily, salt, appetite, gargle, honey, electrolyte, hydrated, exercise, workout	Coping Strategies	pfizer, moderna, shot, steroid, antibiotic, netflix, watch, binge, watching, game, podcasts, book, tv, show, green, tea, honey, ginger, lemon, water, cayenne, manuka
Test	test, tested, positive, pulse, levels, oxygen, oximeter, day, vaccinated, antibody, blood	Mental health and Anxiety	anxiety, feel, pain, fear, panic, scared, nausea, anyone, brain, fog, memory, focus	Research and Facts	covid19, science, study, studies, data, evidence, research, scientific, scientist, theory	Gratitude	please, thank, thanks, you, so, much, contribution
Other related topics	menstrual cycle, periods, urination, bladder, alcohol, smoke, smoking, dog, cats, pets	Personal and Health Concerns	fever, cough, taste, fatigue, breath, job, loss, pay, employer, manager, living	COVID19 related topics	pcr, antigen, mask, air, n95, quarantine, omicron, delta, variant, body, response, hair	Pray and Hope	wish, speedy, recovery, better, pray, sending, you, strong, hope, hopeful, crossed, fingers, miracles

Data Classification

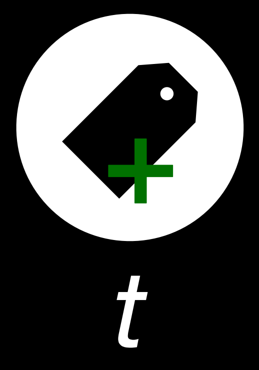


Data Classification



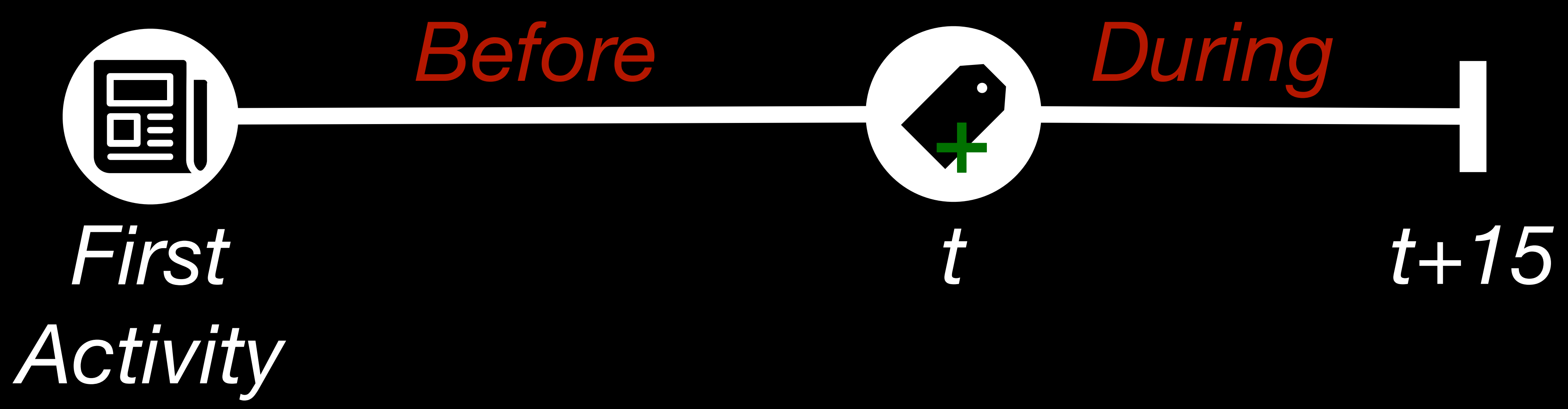
Temporal Phases

Temporal Phases

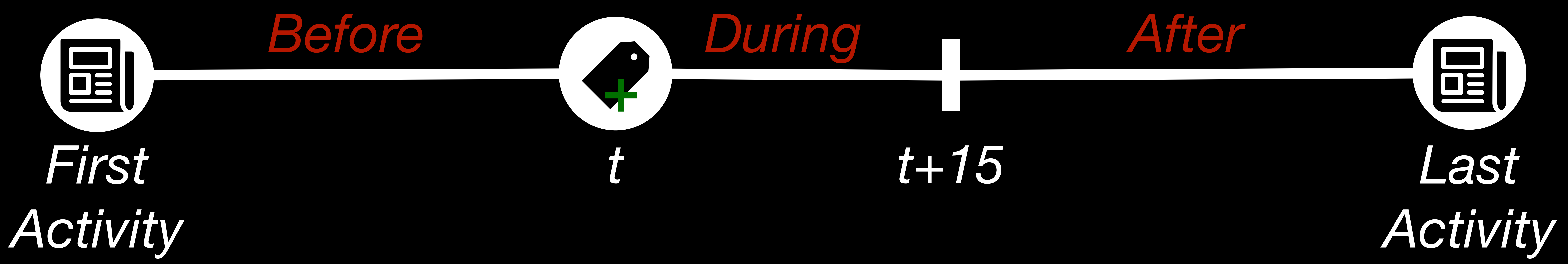


Temporal Phases

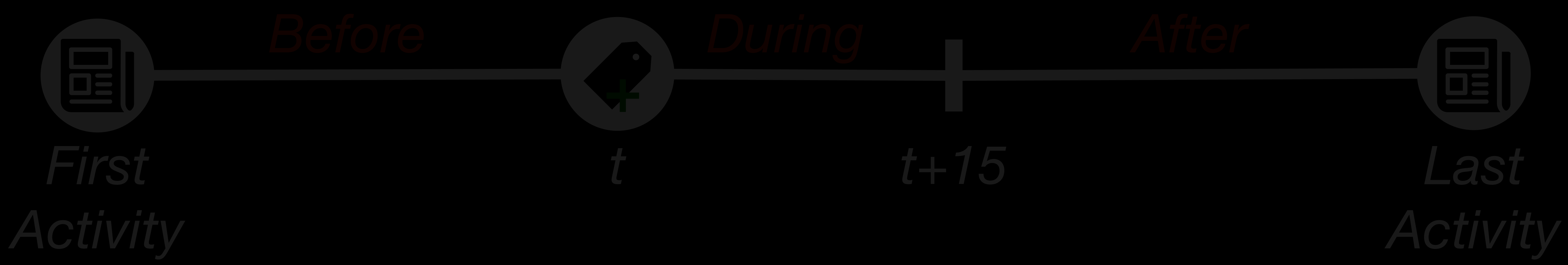
Temporal Phases



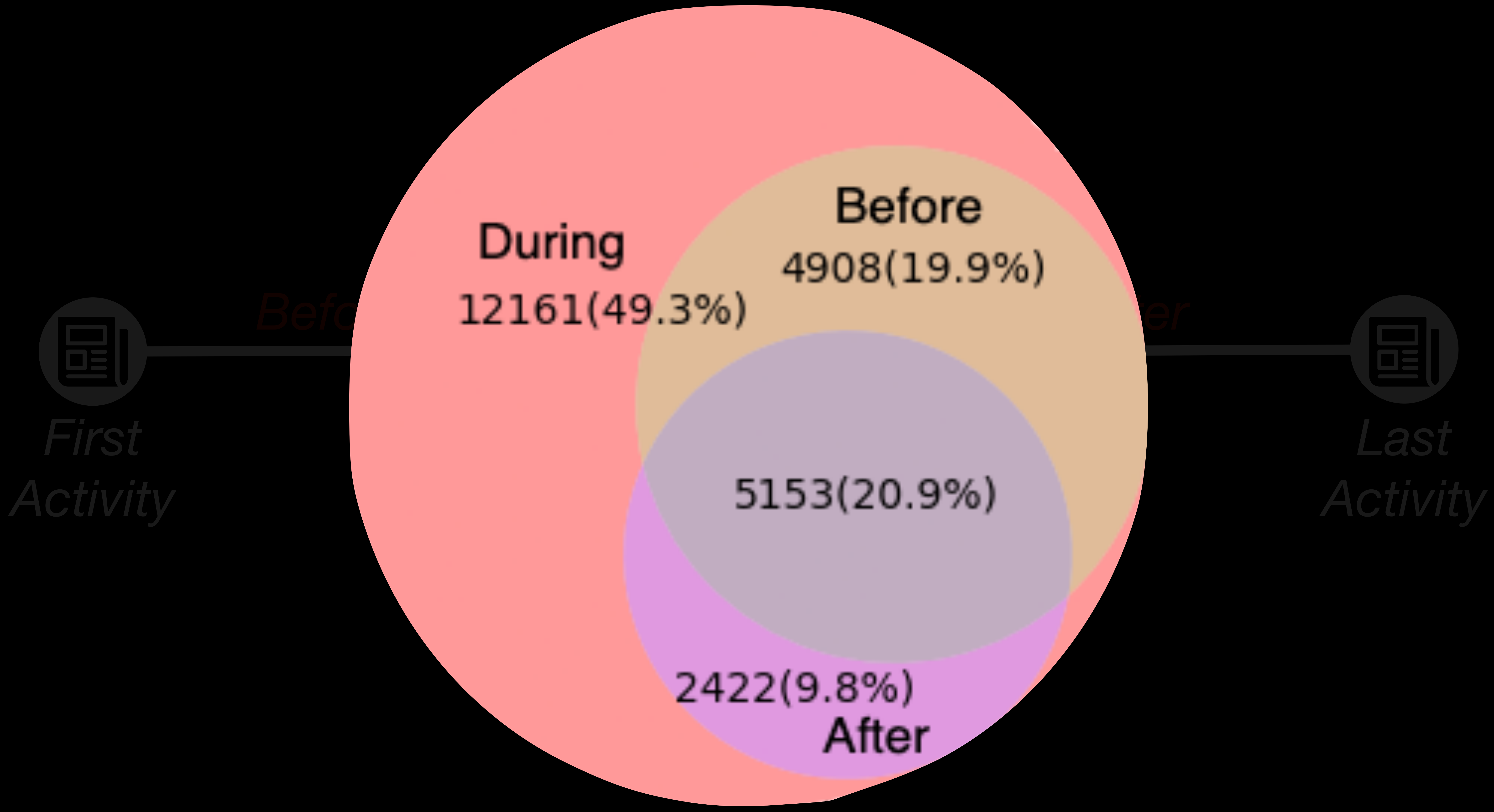
Temporal Phases



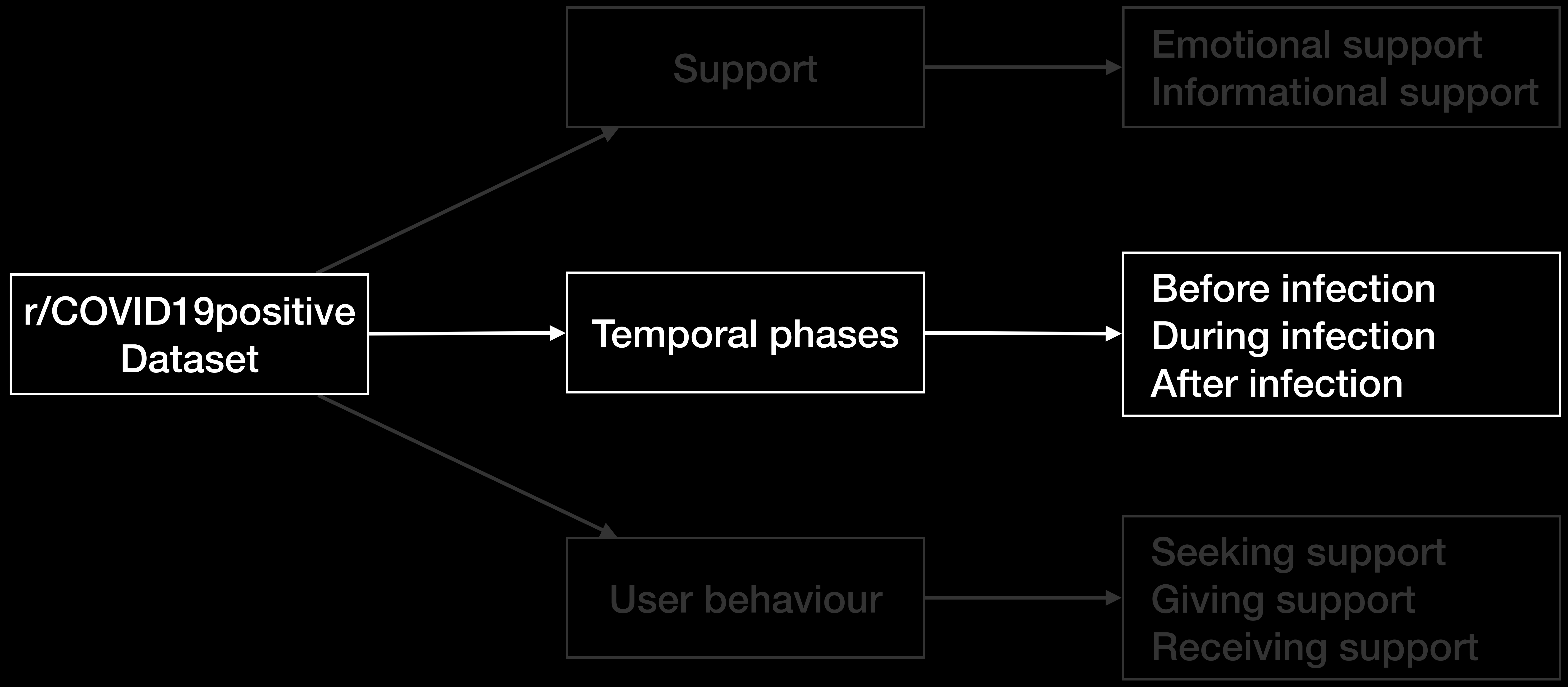
Temporal Phases



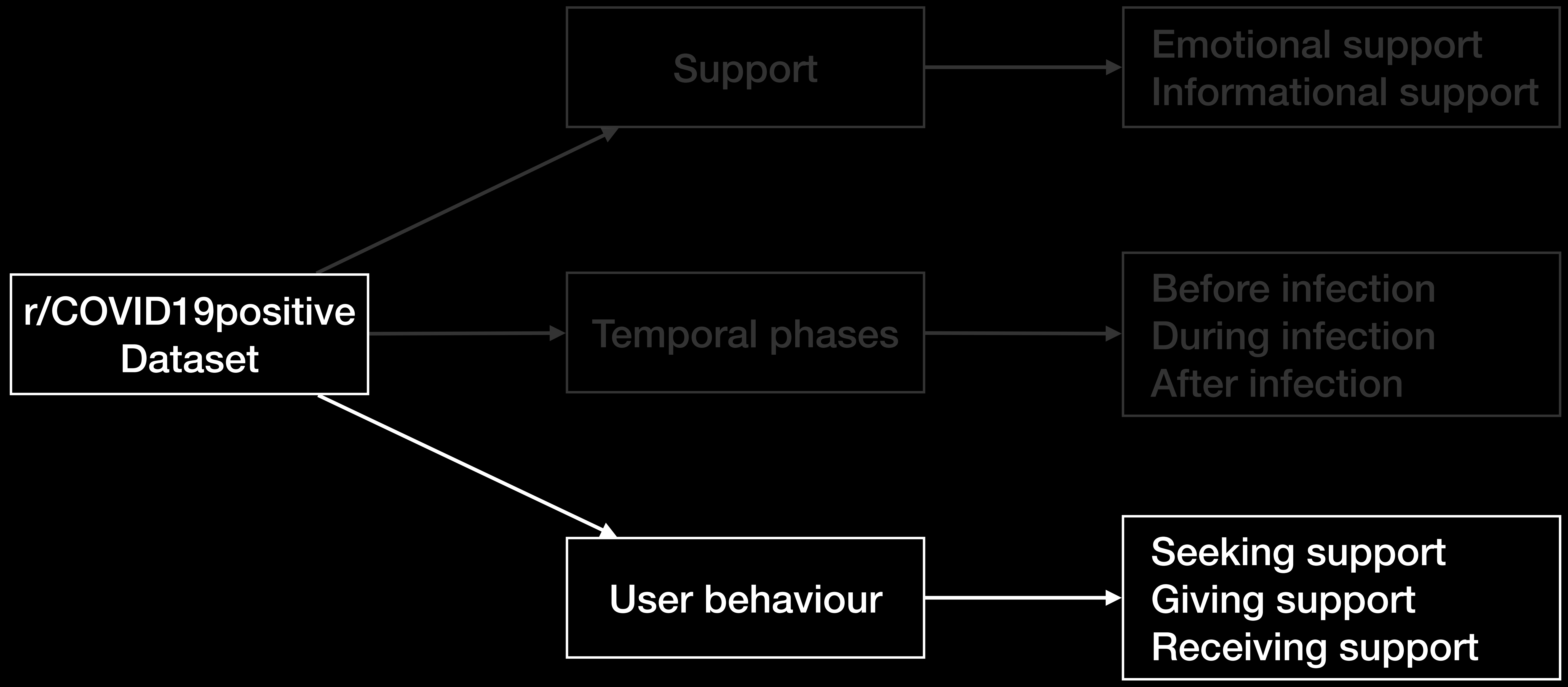
Temporal Phases



Data Classification



Data Classification

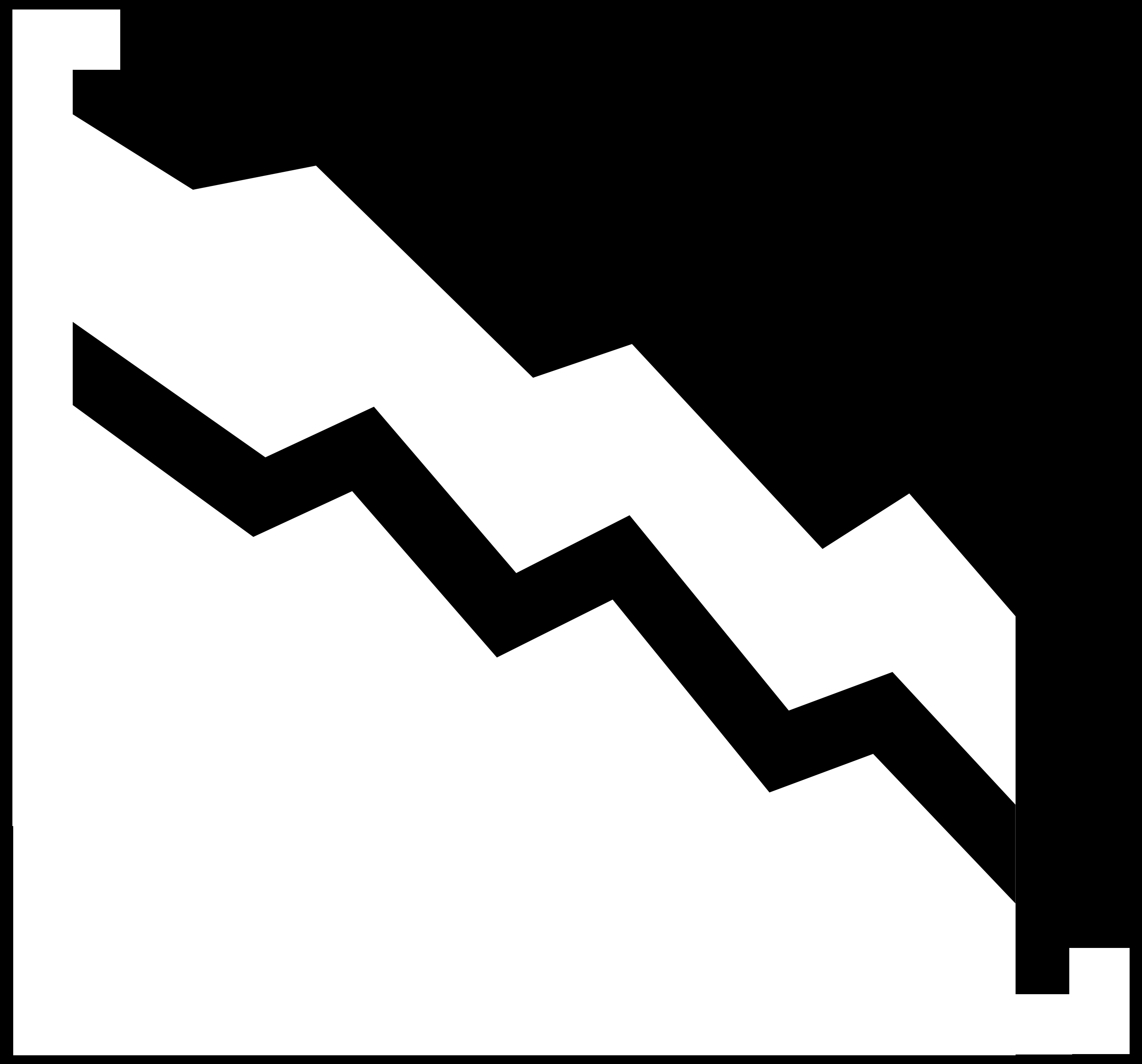


User Behaviour

Seeking	Writing posts
Receiving	Comments received
Giving	Commenting on other user’s post

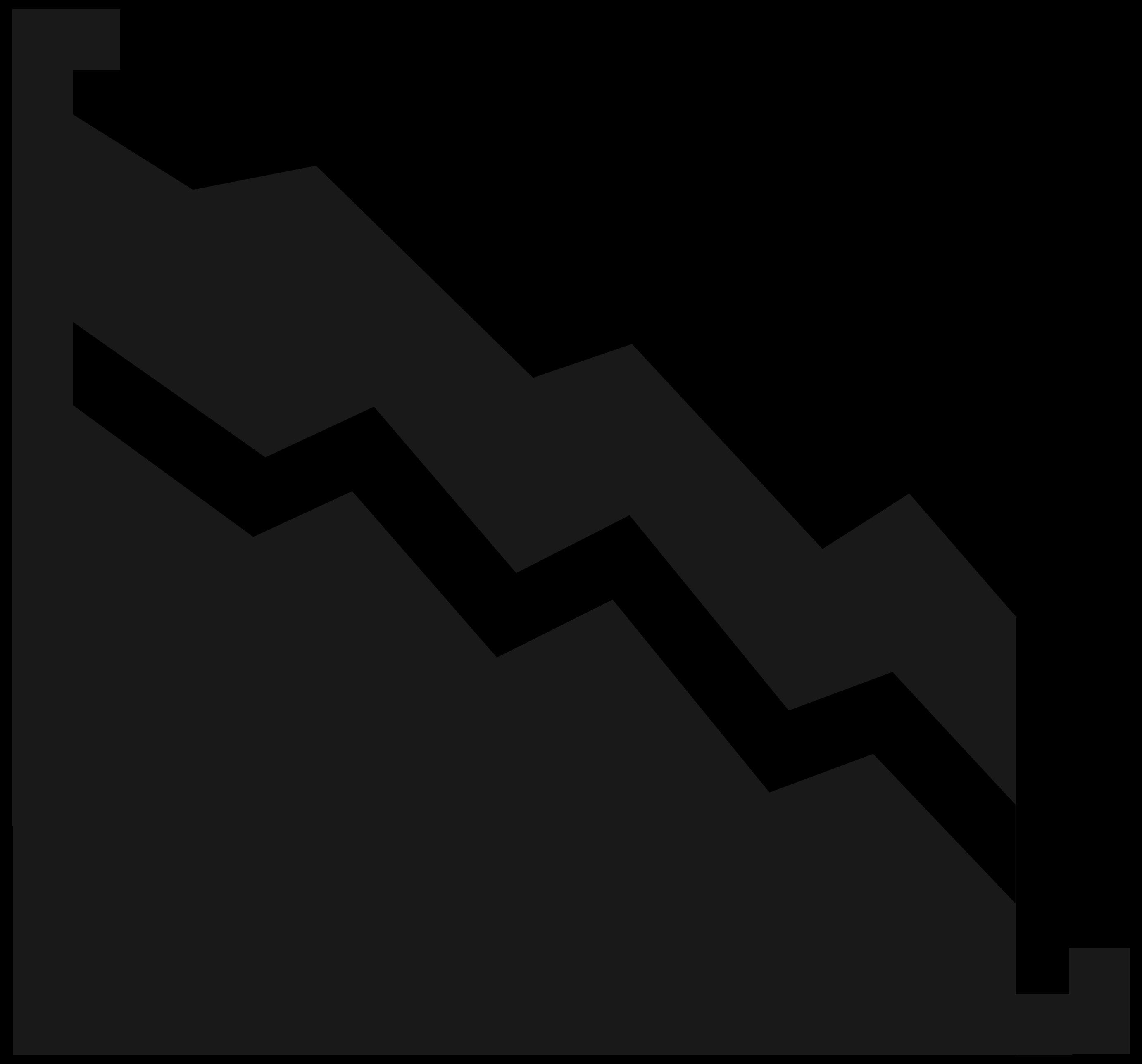
Longevity of a User

Longevity of a User



Survival Analysis / CoX Regression

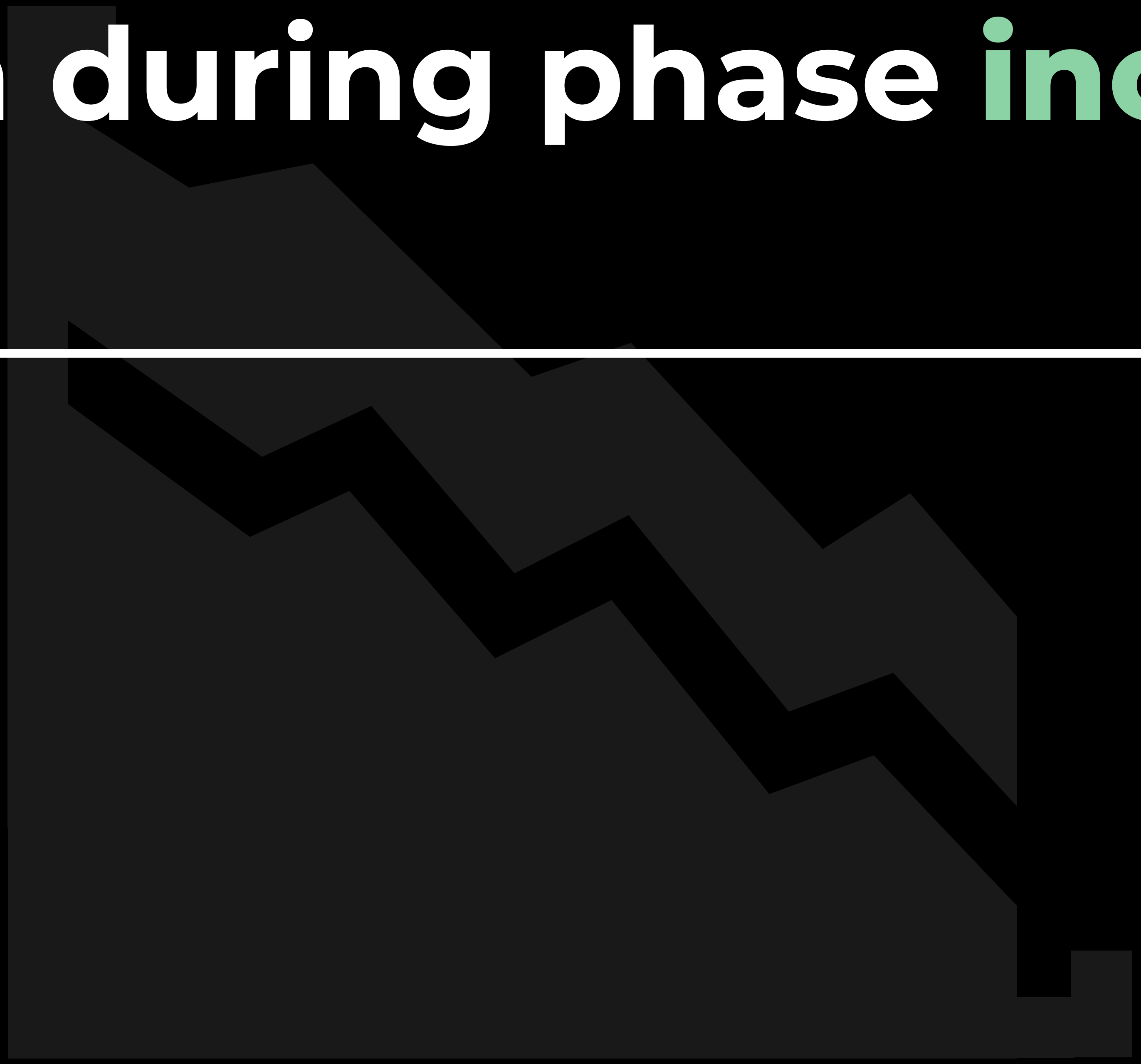
Longevity of a User



Survival Analysis / CoX Regression

Longevity of a User

Seeking support in during phase increases longevity



Survival Analysis / CoX Regression

Longevity of a User

Seeking support in during phase increases longevity

Giving support in early phase increases longevity

Survival Analysis / CoX Regression

Longevity of a User

Seeking support in during phase **increases** longevity

Giving support in early phase **increases** longevity

Receiving support has **no effect** on longevity

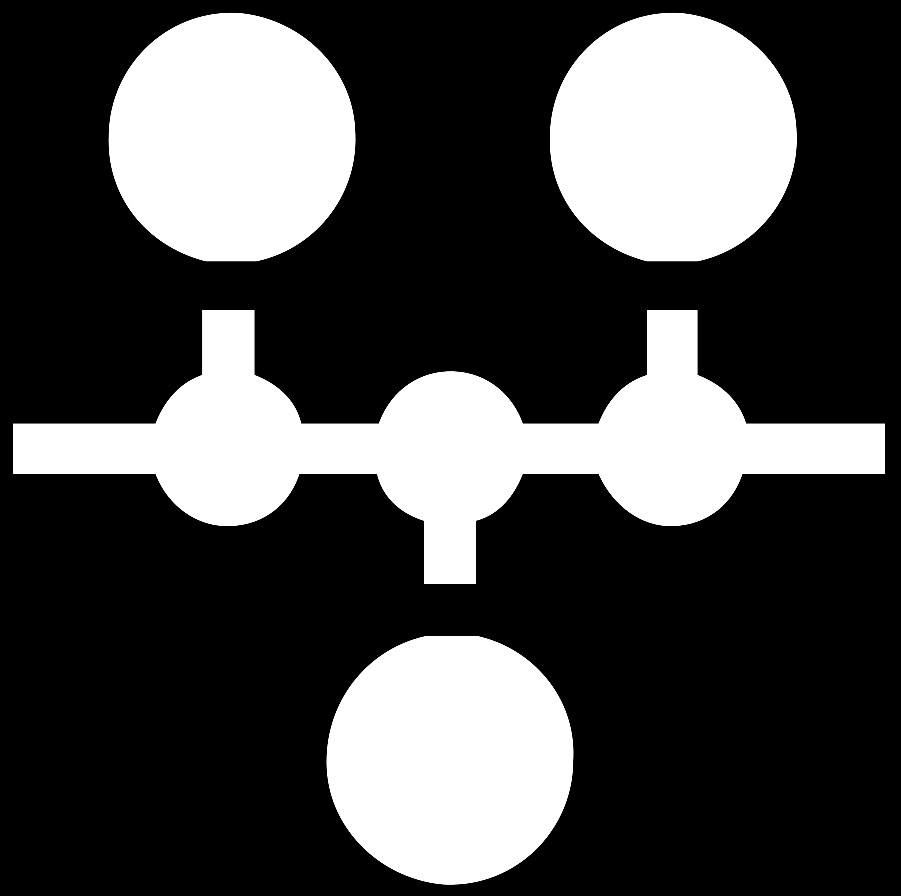
Limitations

Limitations

Time Delta

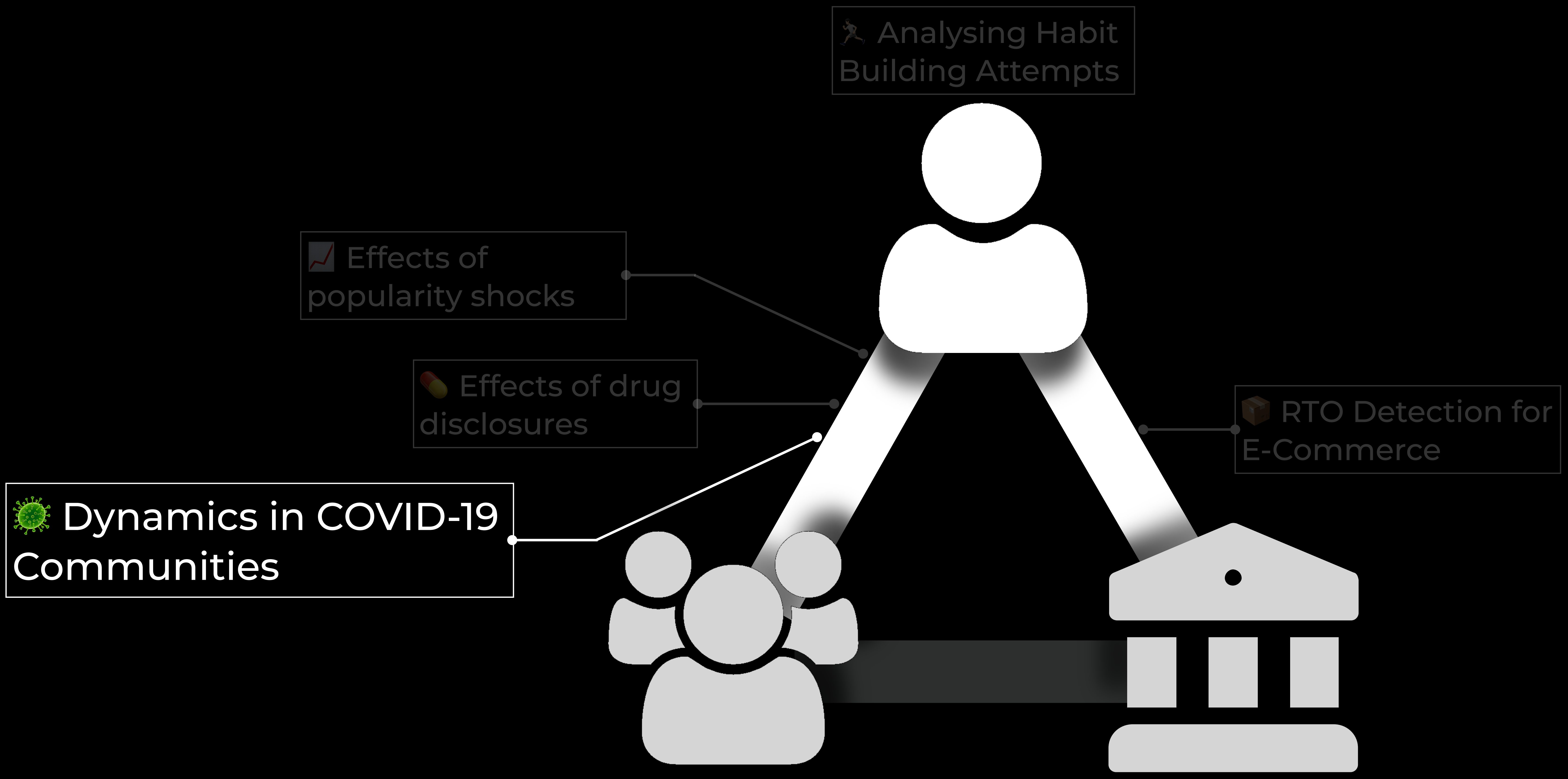
Limitations

Time Delta



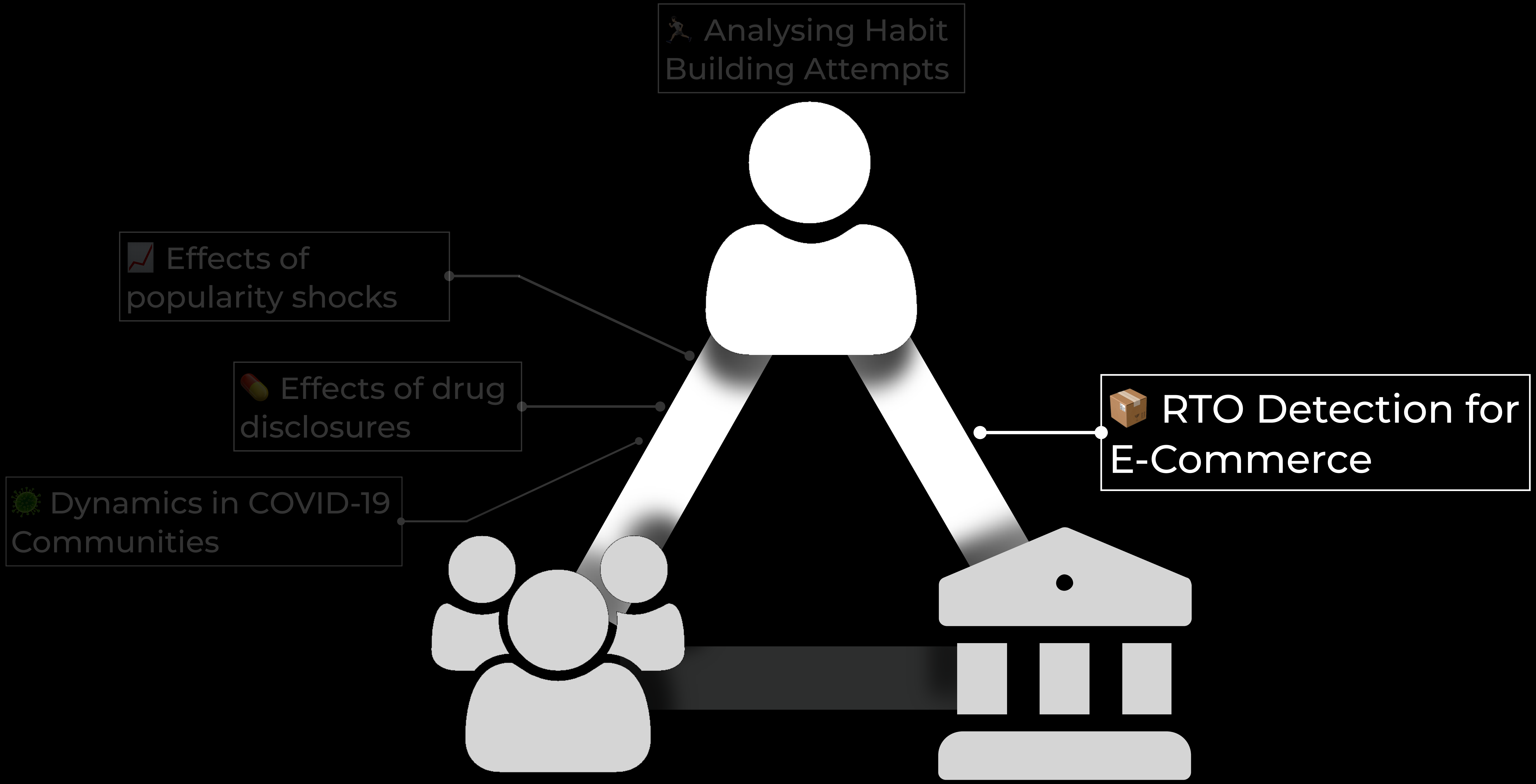
Multiple Infections

Our Focus



🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

Our Focus



🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

Social Re-Identification Assisted RTO Detection for E-Commerce

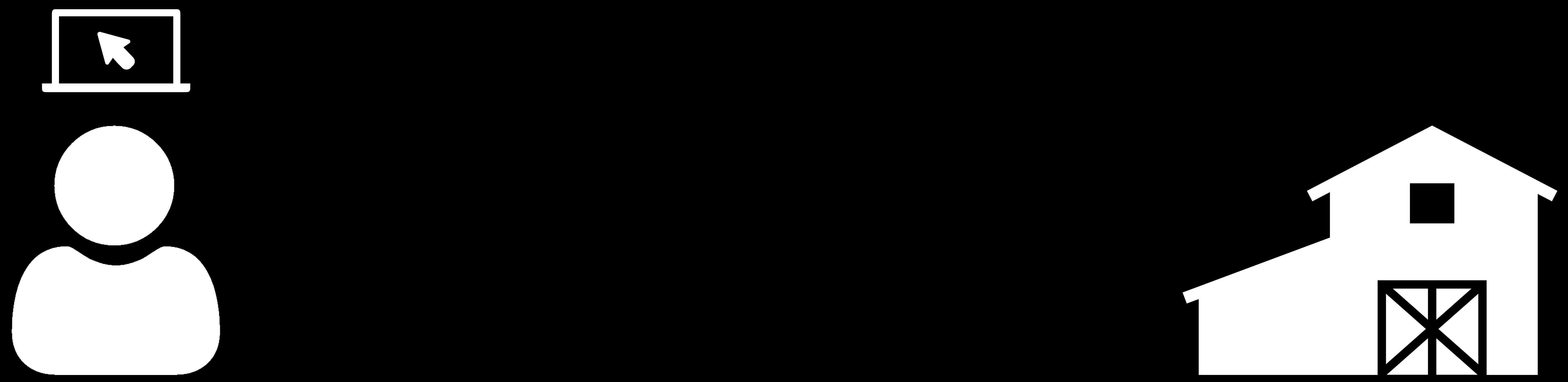
WWW' 23 (Companion)

Hitkul, Abinaya, Soham Saha, Satyajit Banerjee, M. Chellaiah, Ponnurangam Kumaraguru

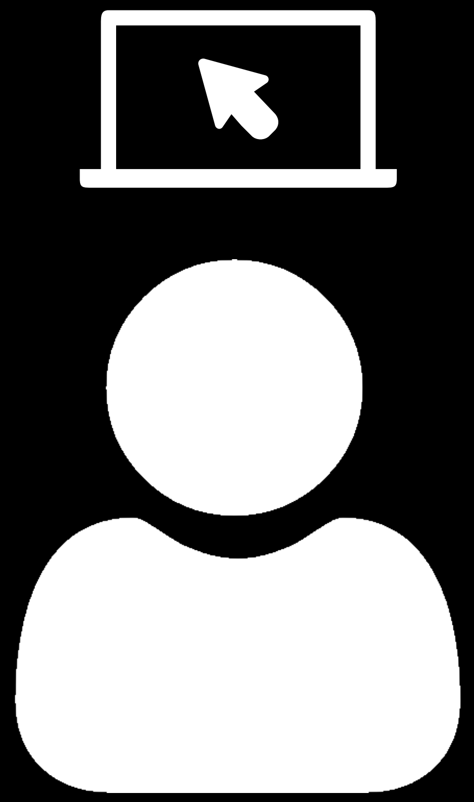
In collaboration with

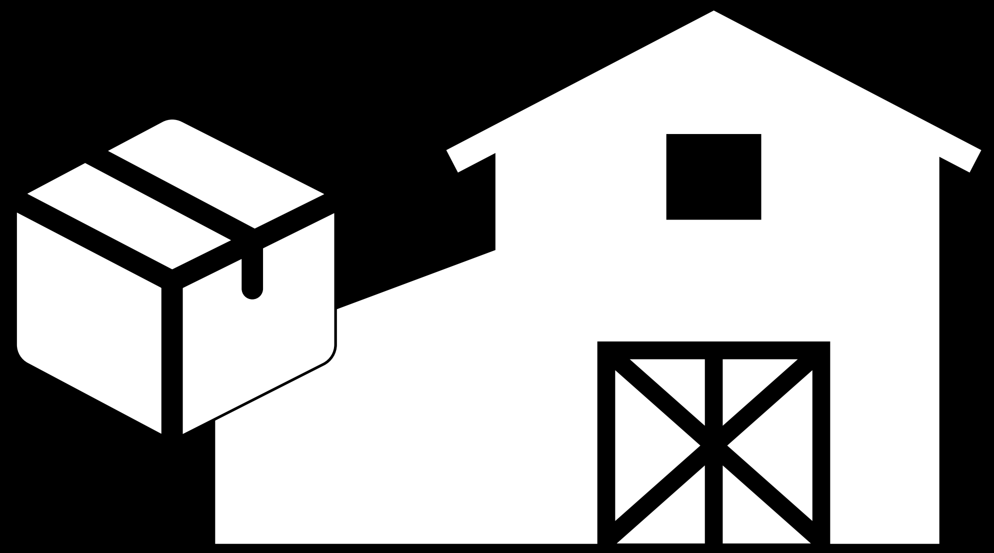
Return to Origin (RTO)

Return to Origin (RTO)

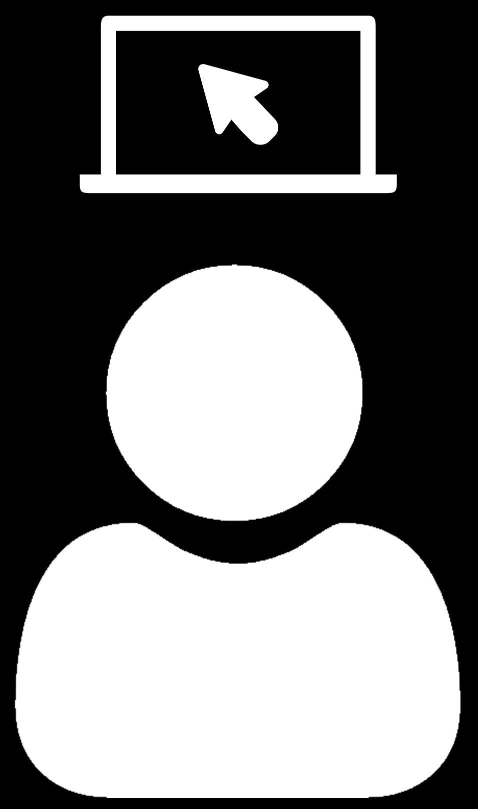


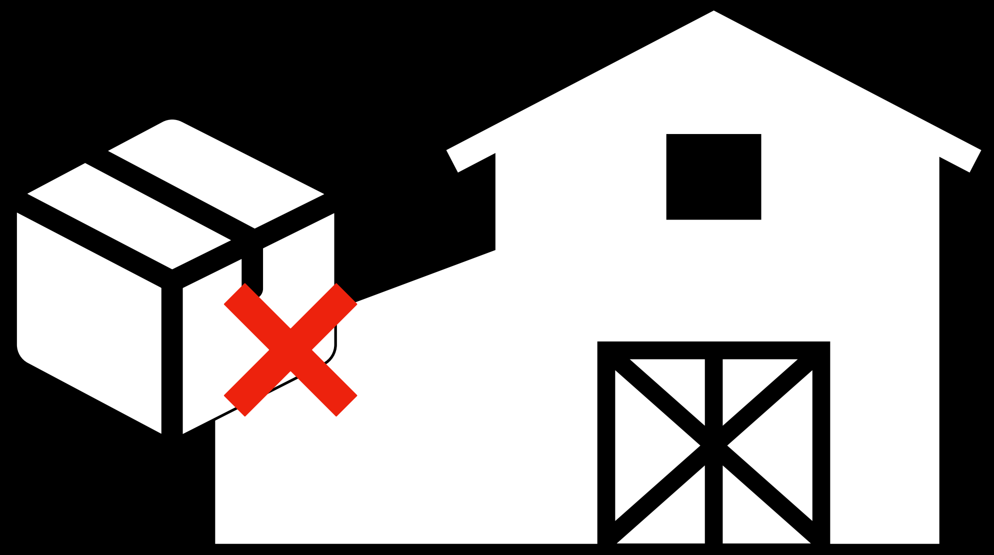
Return to Origin (RTO)





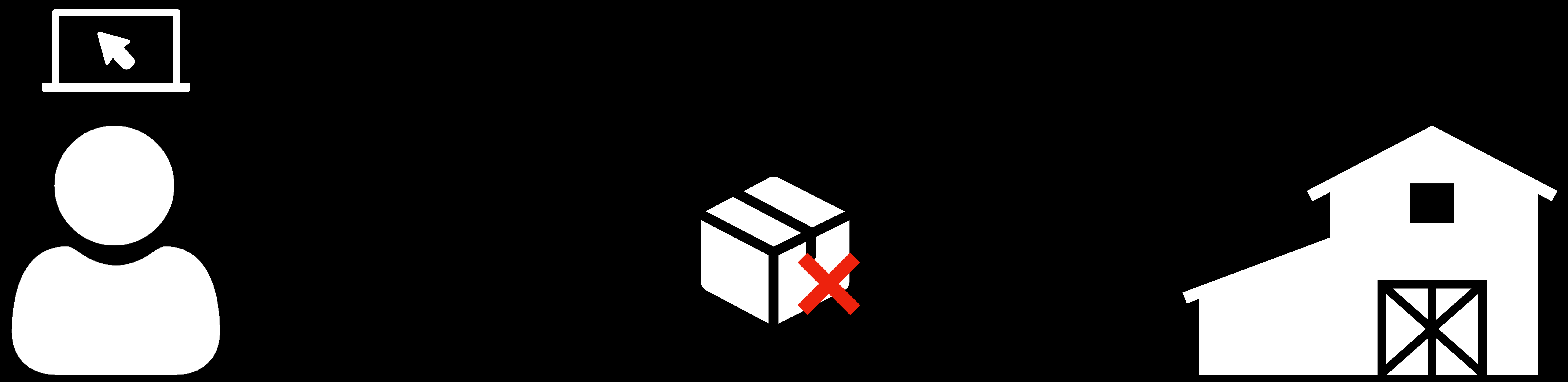
Return to Origin (RTO)





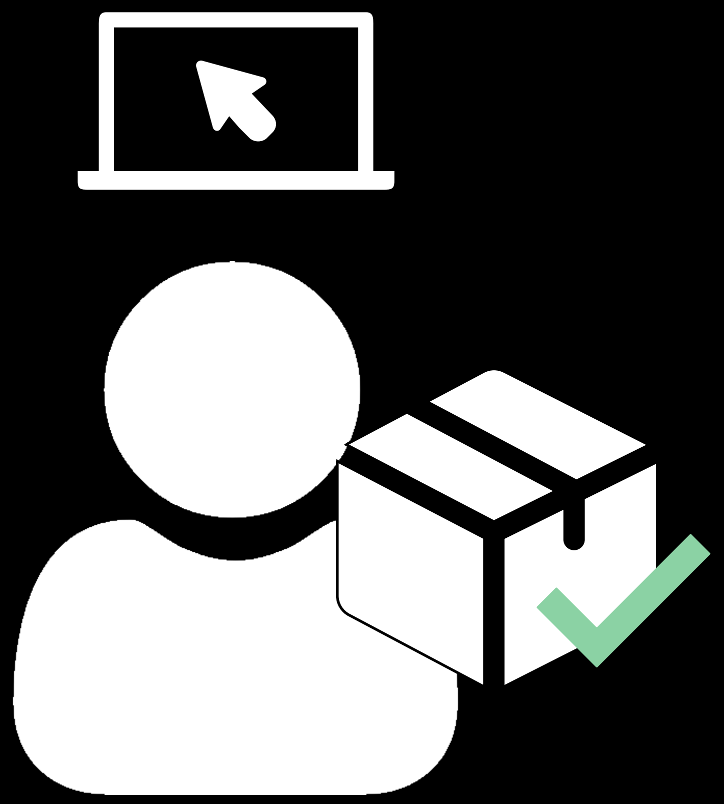
Order Cancelled

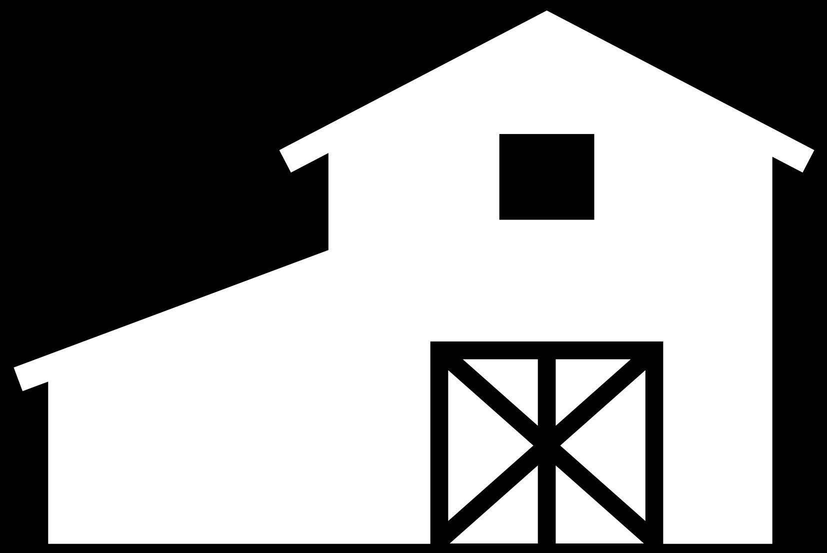
Return to Origin (RTO)



Order Return to Origin

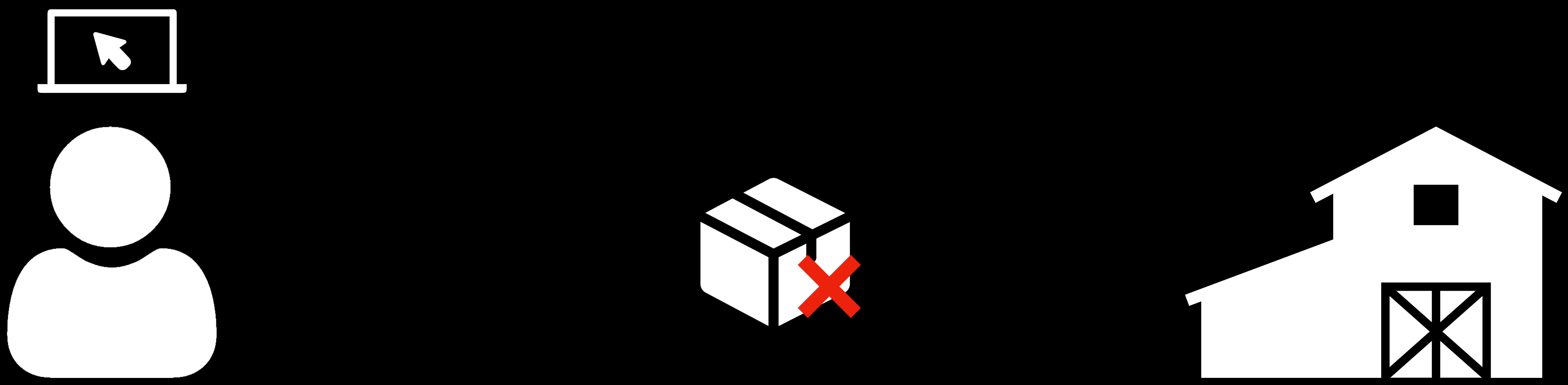
Return to Origin (RTO)





Order Delivered

Return to Origin (RTO)



Order Return to Origin

Why Study RTO?

Why Study RTO?

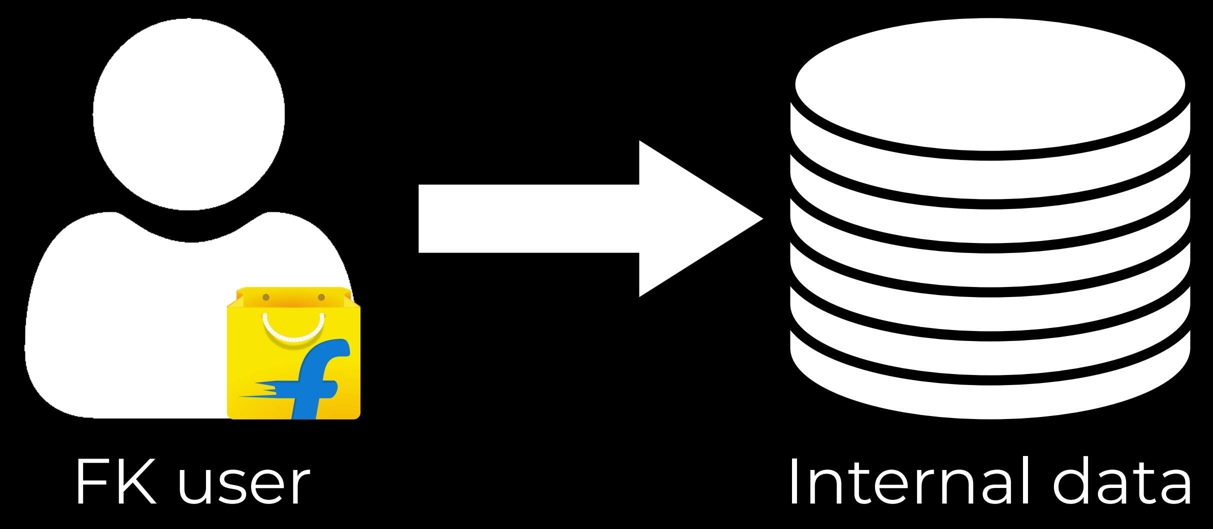
₹15 (2¢) lost per RTO

Hypothesis

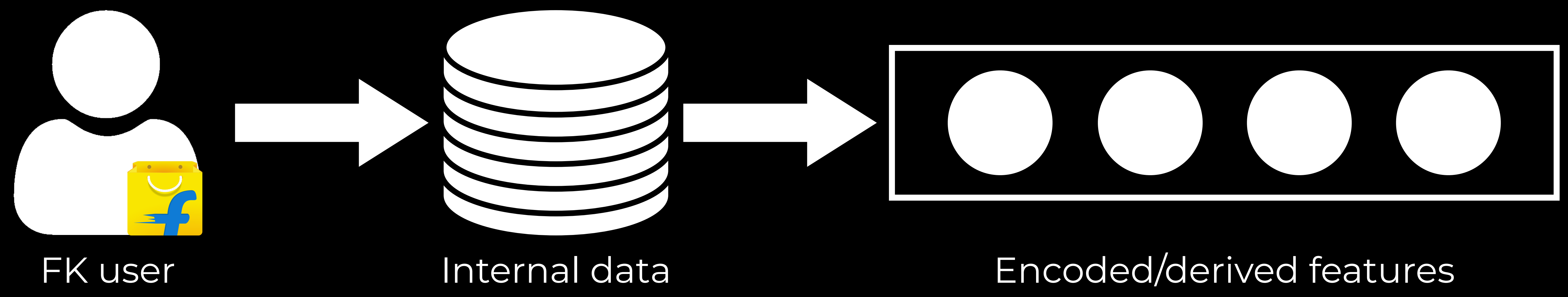
Hypothesis

FK user

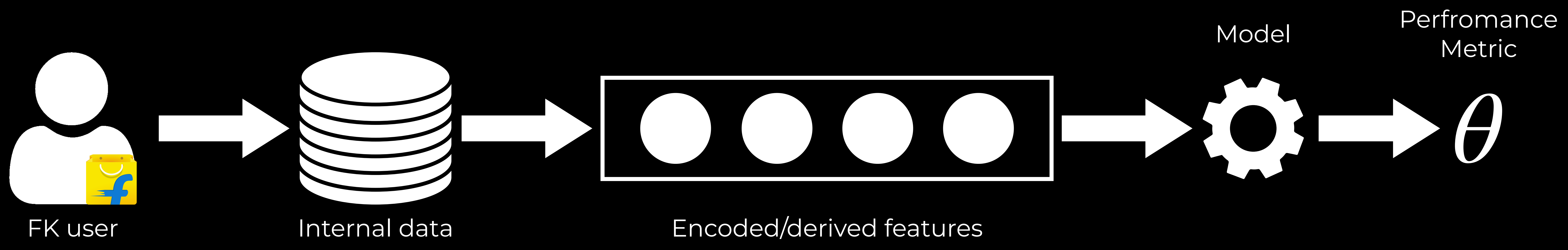
Hypothesis



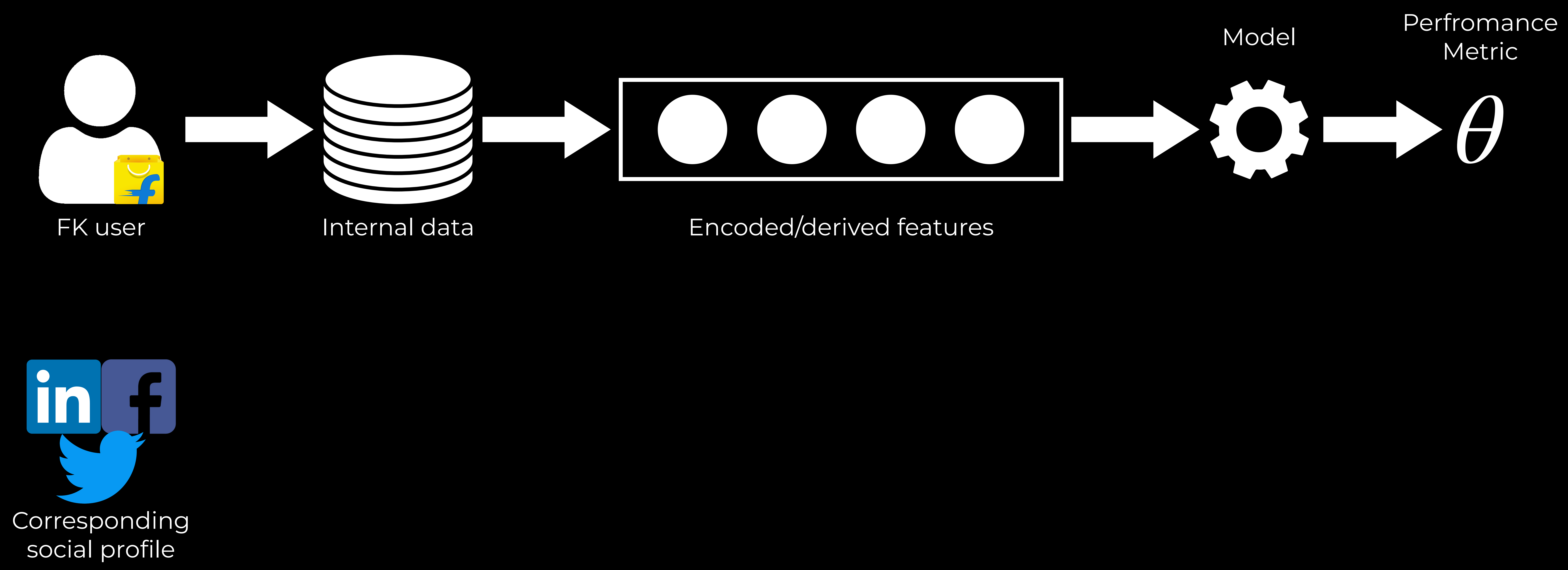
Hypothesis



Hypothesis

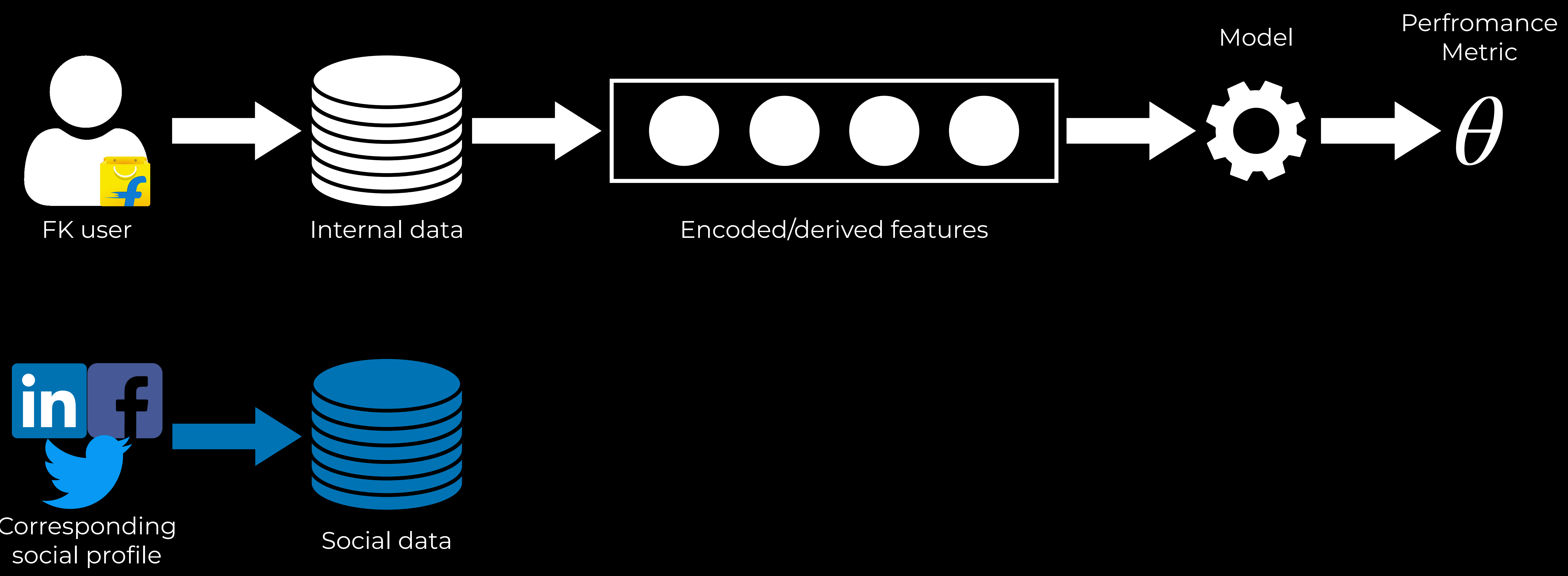

Model
Perfromance
Metric
θ
FK user
Internal data
Encoded/derived features

Hypothesis



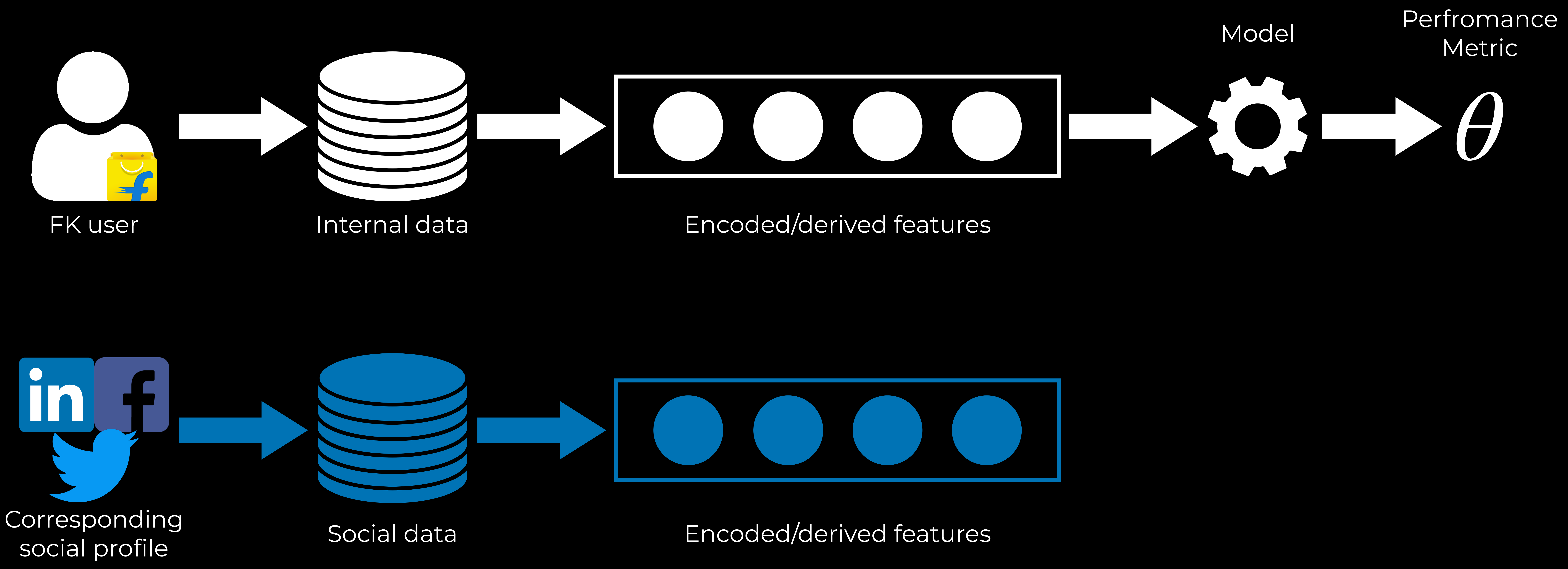
Model
Perfromance
Metric
FK user
Internal data
Encoded/derived features
Corresponding
social profile

Hypothesis

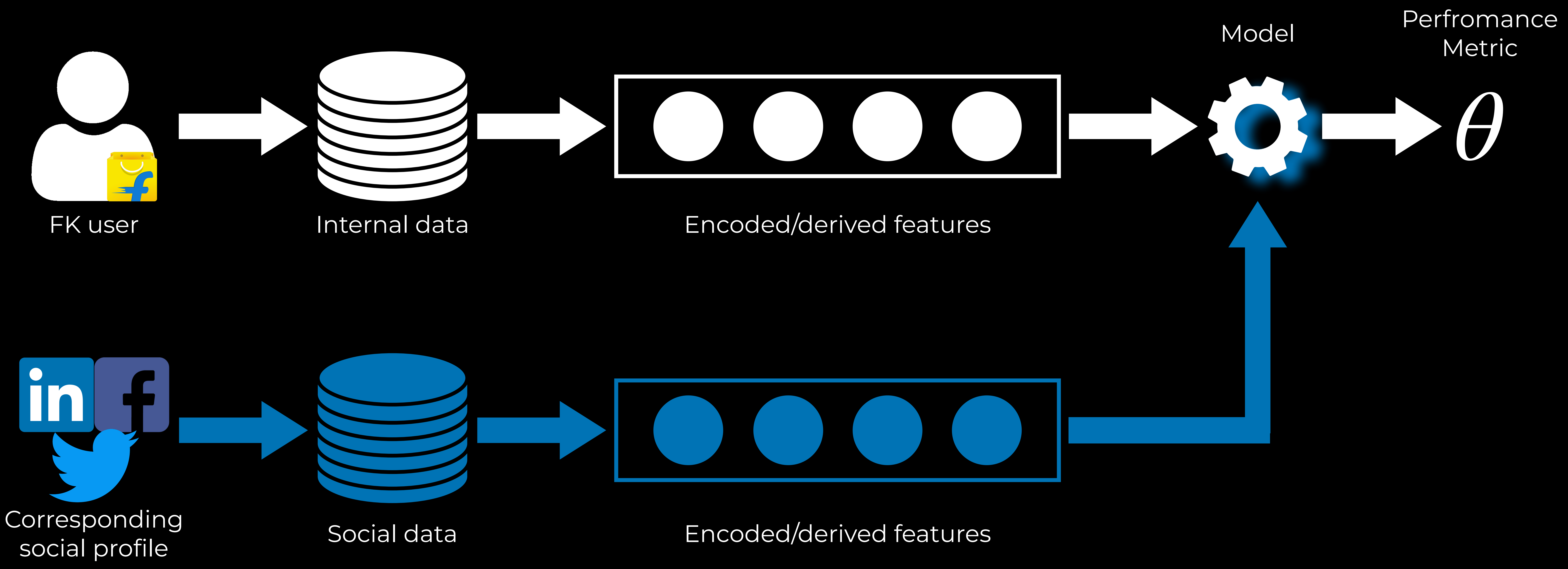


FK user
Internal data
Encoded/derived features
Model
Perfromance
Metric
θ
Corresponding
social profile
Social data

Hypothesis

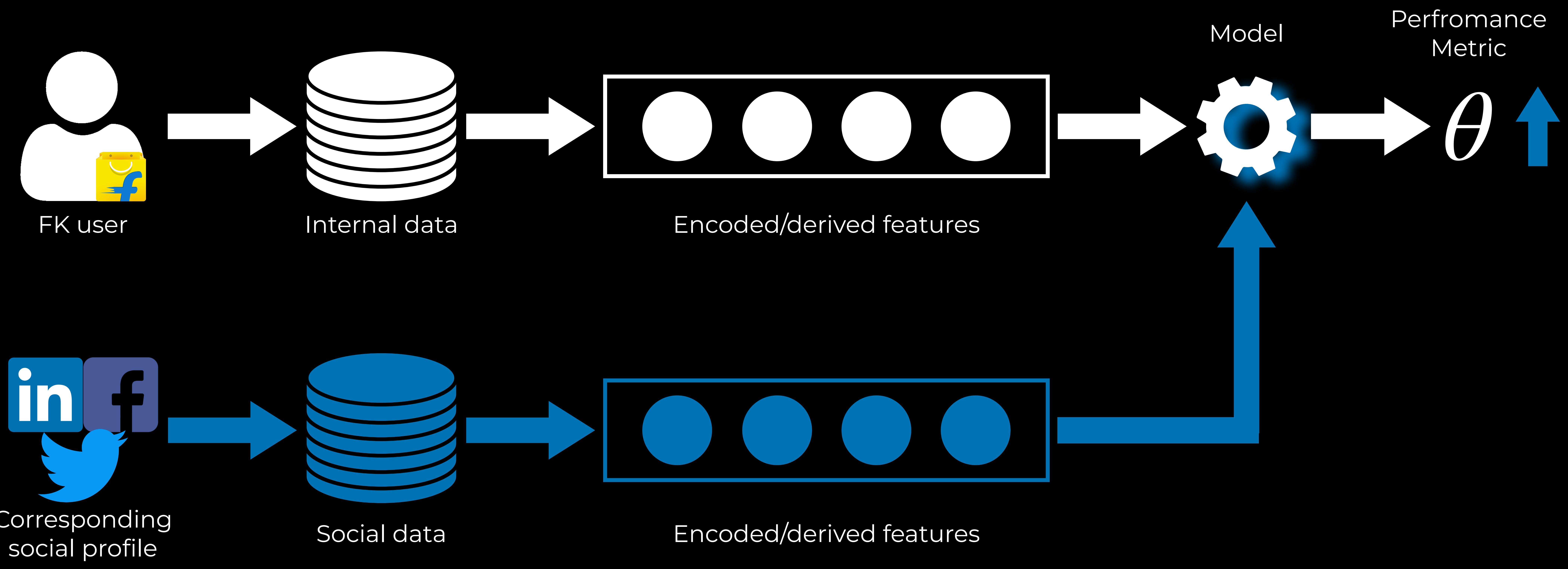

Model
Perfromance
Metric
θ
FK user
Internal data
Encoded/derived features
Corresponding
social profile
Social data
Encoded/derived features

Hypothesis



Model
Perfromance Metric
θ
FK user
Internal data
Encoded/derived features
Corresponding social profile
Social data
Encoded/derived features

Hypothesis



Model
Perfromance
Metric
θ
FK user
Internal data
Encoded/derived features
Corresponding
social profile
Social data
Encoded/derived features

Motivation - Sociology

Motivation - Sociology

Fraud detection in low-voltage electricity consumers using socio-economic indicators and billing profile in smart grids

IET Journals
The Institution of Engineering and Technology
ISSN 2515-0855
doi: 10.1049/oap-cired.2017.1258
www.ietdl.org

Jonatas Pulz[1] ✉, Renan B. Muller[1], Fabio Romero[1], André Meffe[1], Álvaro F. Garcez Neto[2], Aldo S. Jesus[2]

[1]Department of Research, Development & Innovation in Enginnering, Daimon Engineering and Systems, São Paulo, Brazil
[2]Center of Distribution Planning and Fraud Detection of Sulgipe, Sulgipe, Brazil
✉ *E-mail: jonatas.pulz@daimon.com.br*

Motivation - Sociology

Fraud detection in low-voltage electricity consumers using socio-economic indicators and billing profile in smart grids

IET Journals
The Institution of Engineering and Technology
ISSN 2515-0855
doi: 10.1049/oap-cired.2017.1258
www.ietdl.org

Jonatas Pulz[1] ✉,
Álvaro F. Garcez
[1]*Department of Researc*
Systems, São Paulo, Br
[2]*Center of Distribution*
✉ *E-mail: jonatas.pulz@*

Economic institutions and individual ethics: A study of consumer attitudes toward insurance fraud

Sharon Tennyson*

Department of Insurance and Risk Management, University of Pennsylvania, 3641 Locust Walk, Philadelphia, PA 19104, USA

Received 18 July 1994; received in revised form 26 March 1996

Motivation - Sociology

Fraud detection in low-voltage electricity consumers using socio-economic indicators and billing profile in smart grids

IET Journals
The Institution of Engineering and Technology
ISSN 2515-0855
doi: 10.1049/oap-cired.2017.1258
www.ietdl.org

Jonatas Pulz[1] ✉,
Álvaro F. Garcez
[1]Department of Researc
Systems, São Paulo, Br
[2]Center of Distribution
✉ E-mail: jonatas.pulz@

Economic institutions and individual ethics: A study of consumer attitudes toward insurance fraud

Sharon Tennyson*

MORTGAGE FRAUD: A RISK FACTOR ANALYSIS OF AFFECTED COMMUNITIES

Andrew T. Carswell, Douglas C. Bachtel*

Motivation - CSS

Motivation - CSS

TweetCred: Real-Time Credibility Assessment of Content on Twitter

Aditi Gupta[1], Ponnurangam Kumaraguru[1], Carlos Castillo[2], and Patrick Meier[2]

[1] Indraprastha Institute of Information Technology, Delhi, India
{aditig, pk}@iiitd.ac.in
[2] Qatar Computing Research Institute, Doha, Qatar
chato@acm.org, pmeier@qf.org.qa

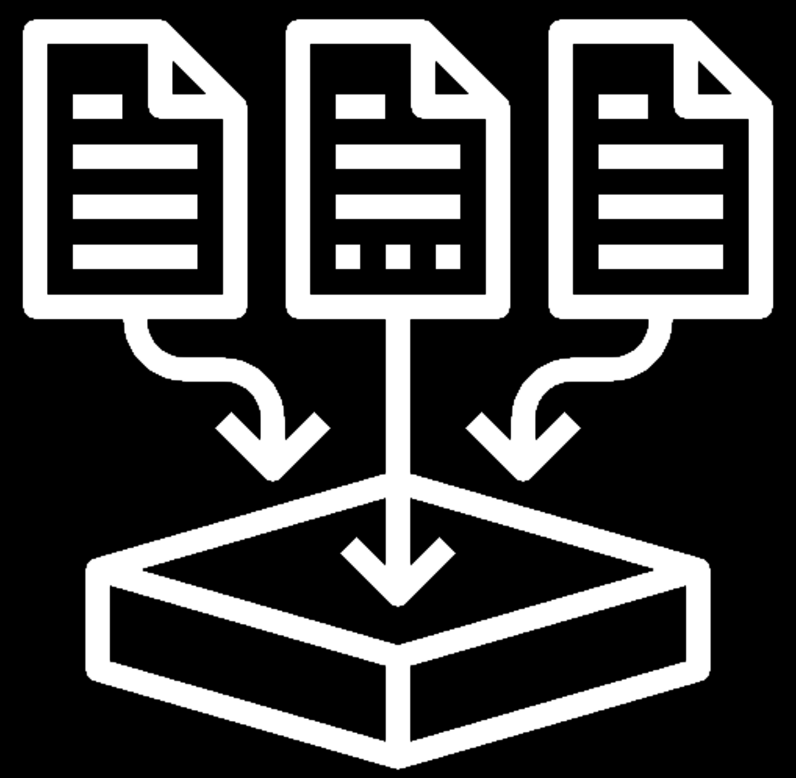
Motivation - CSS

TweetCred: Real-Time Credibility Assessment of Content on Twitter

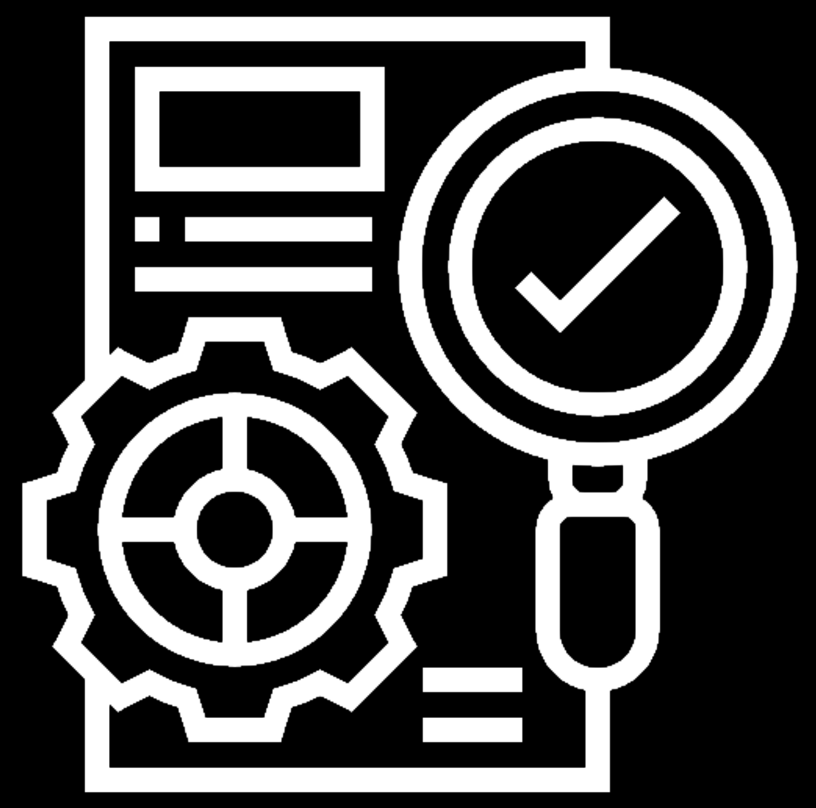
An Efficient Data Enrichment Scheme for Fraud Detection Using Social Network Analysis

Soheil Jamshidi, Mahmoud Reza Hashemi
Internet Fraud Risk Assessment and Ubiquitous Detection Laboratory (iFRAUD)
School of Electrical and Computer Engineering
College of Engineering, University of Tehran, Tehran, Iran
{s.jamshidi, rhashemi}@ut.ac.ir

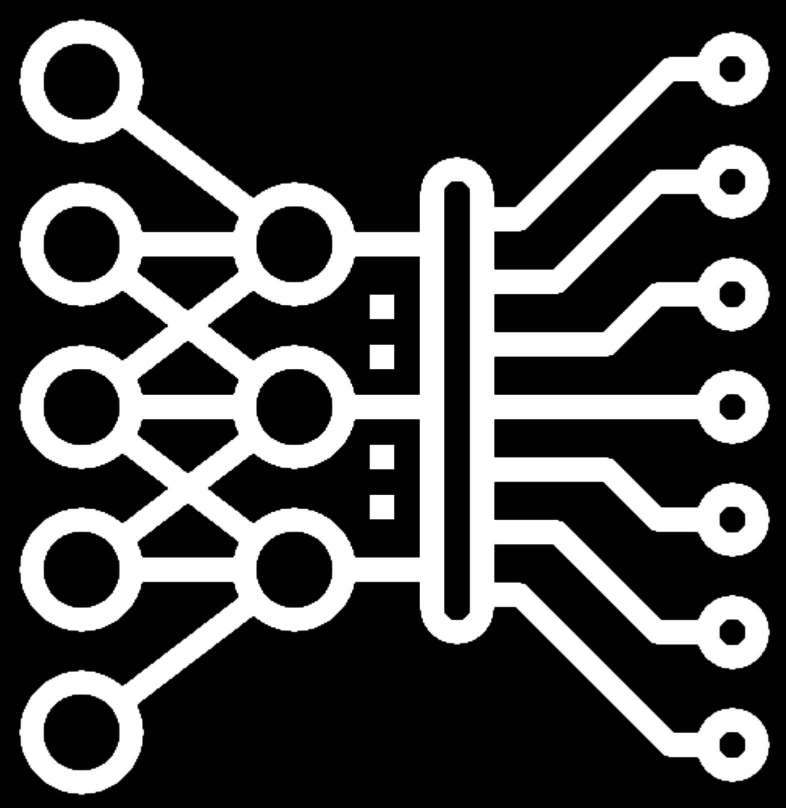
Proposed Flow



Data collection



Data validation



Modeling

Proposed Flow

Data collection

Data validation

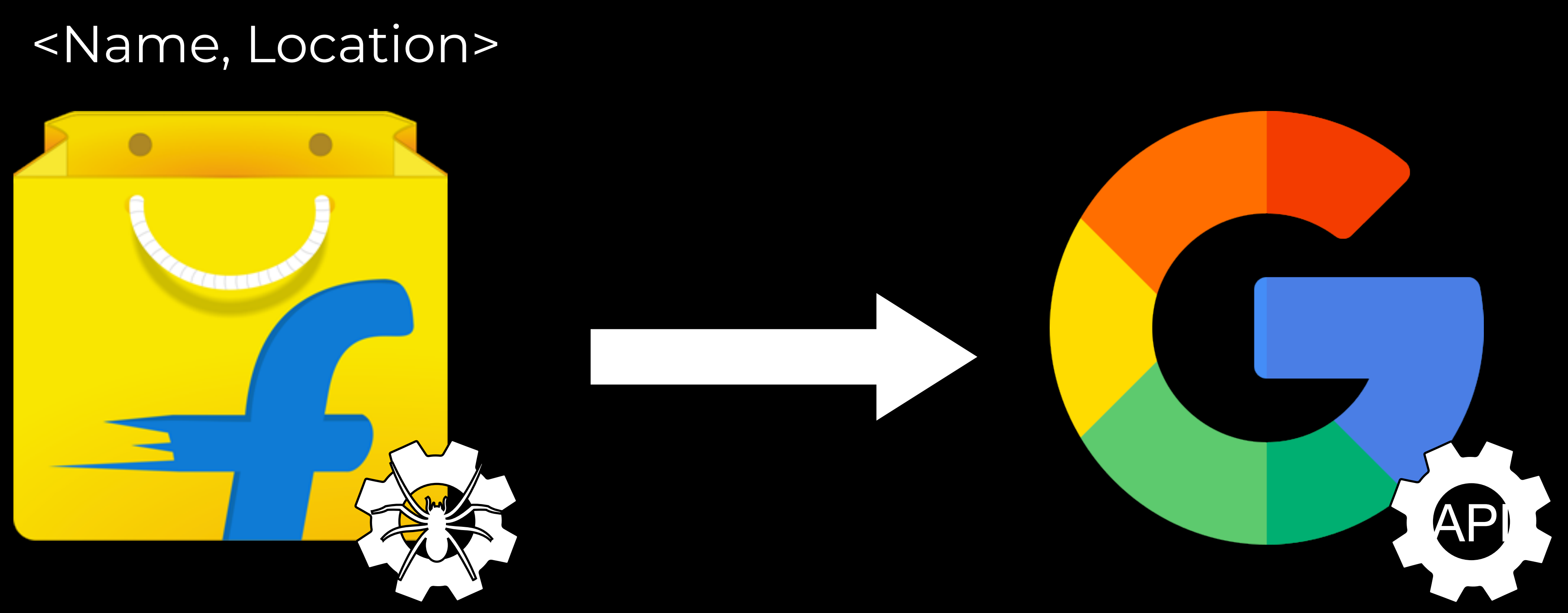
Modeling

Data Collection

Data Collection

<Name, Location>

Data Collection



Data Collection

Data Collection

<Name, Location>

Potential PII, platform violations

Data Collection

Potential PII, platform violations

Only public data
Academic use only

Data Collection

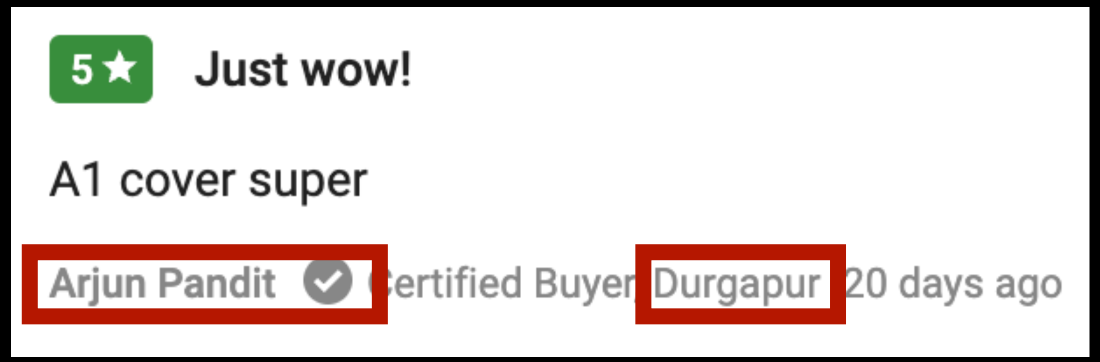
* November 2021 to April 2022, Only Lifestyle org

Data Collection

RTO Users

* November 2021 to April 2022, Only Lifestyle org

Data Collection

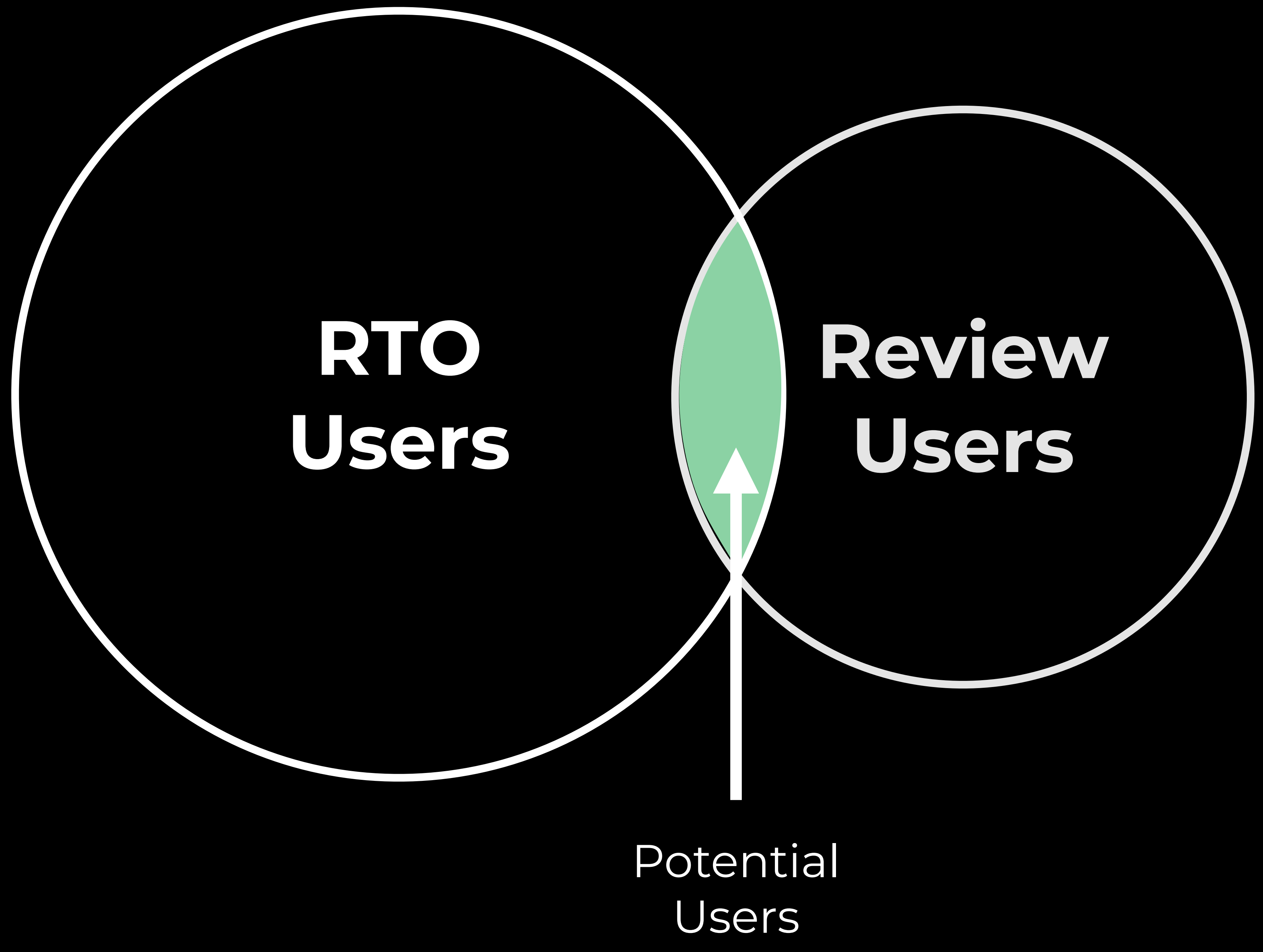


RTO Users

Review Users

* November 2021 to April 2022, Only Lifestyle org

Data Collection



* November 2021 to April 2022, Only Lifestyle org

Data Collection

Data Collection

25,725 rows
6,052 users

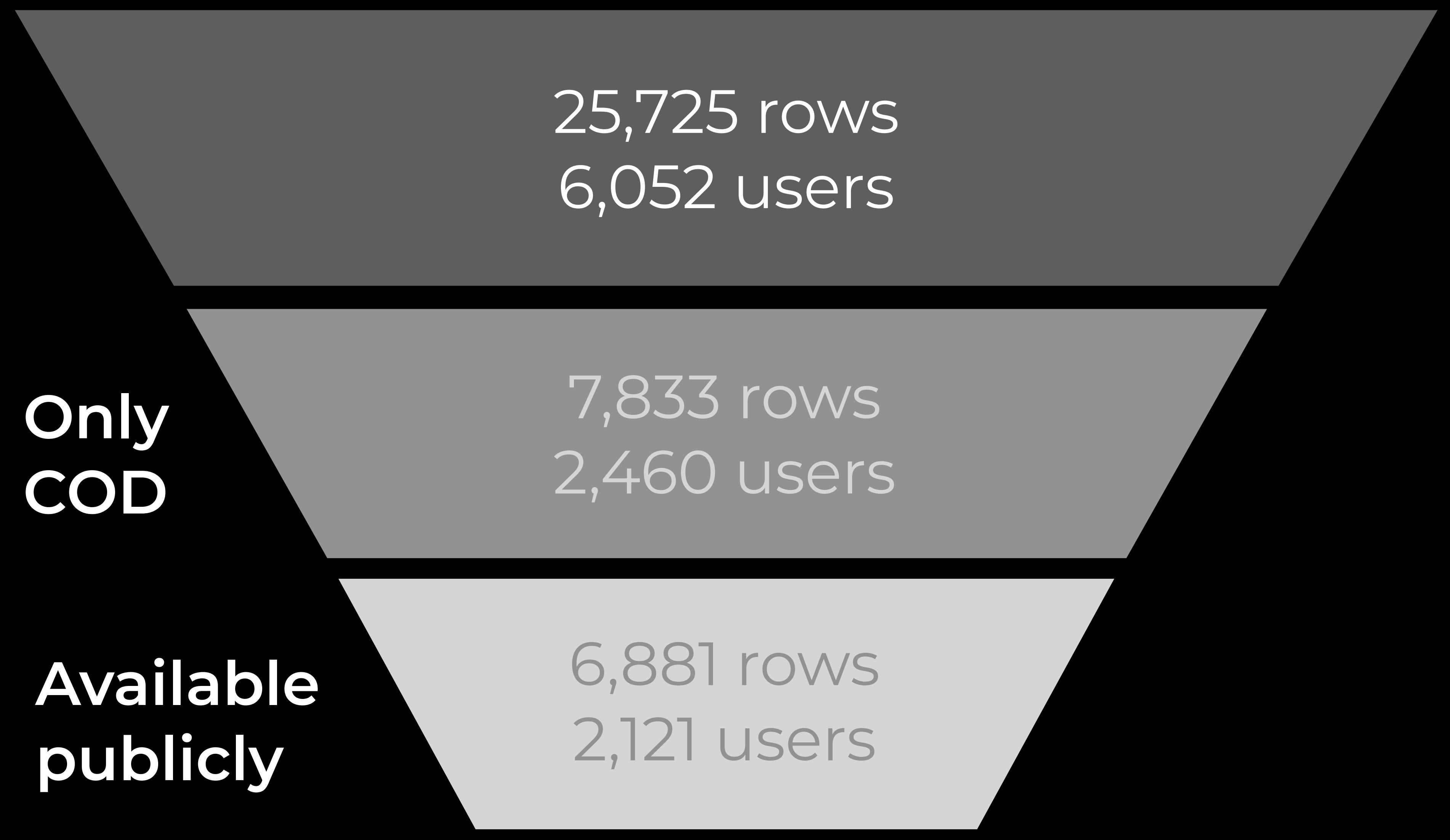
Data Collection

25,725 rows
6,052 users

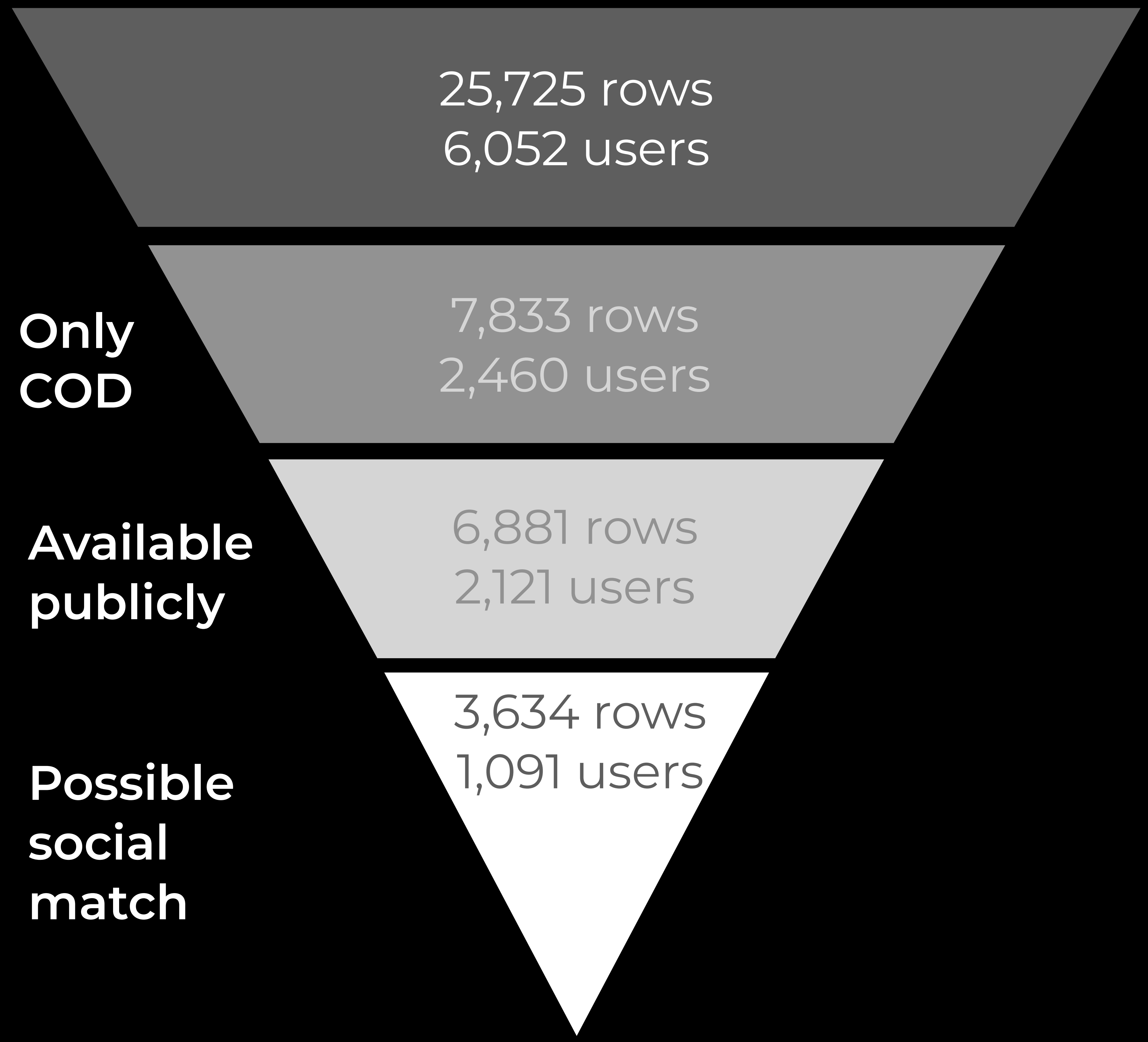
Only COD

7,833 rows
2,460 users

Data Collection



Data Collection



Data Source

80.89%

30.69%

10.37%

8.96%

Data Source

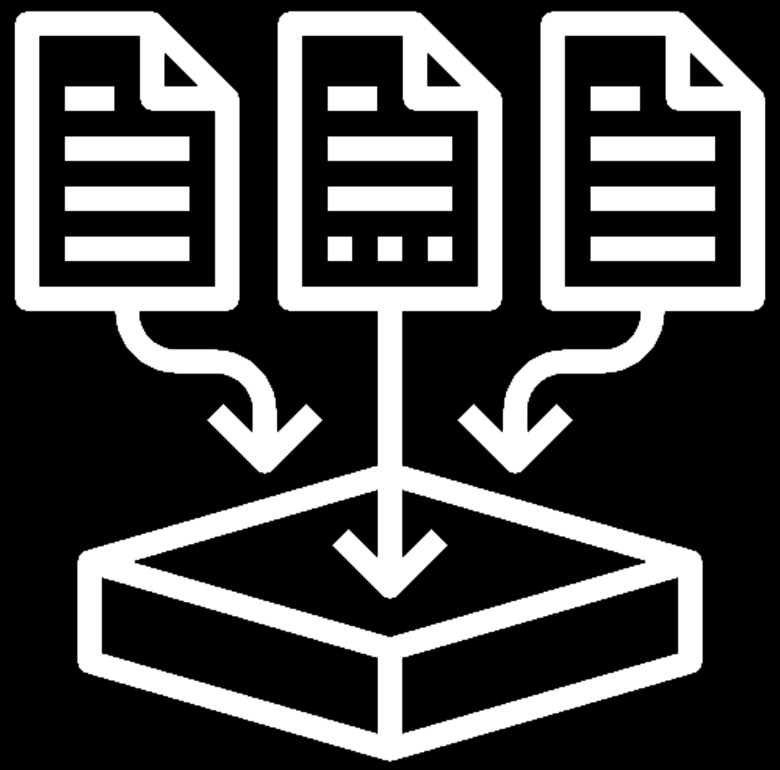
80.89%

30.69%

10.37%

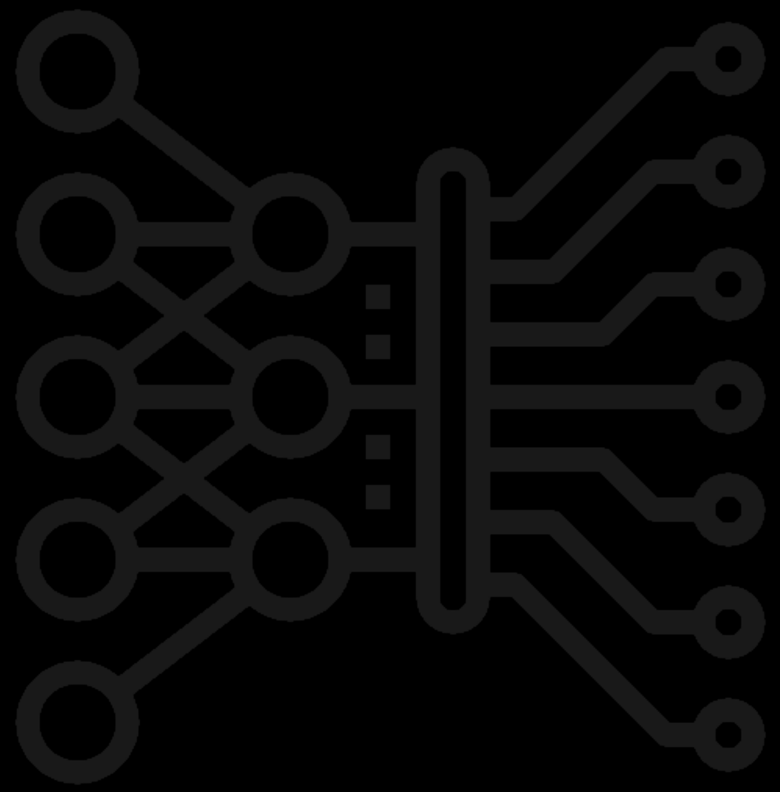
8.96%

Proposed Flow



Data collection

Data validation



Modeling

Proposed Flow

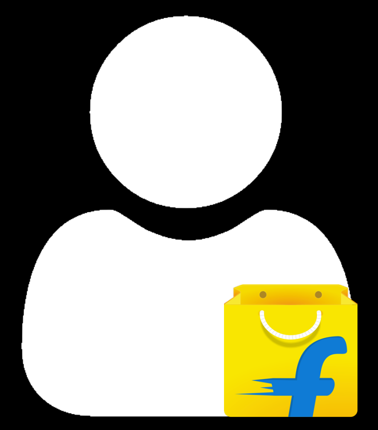
Re-Identification Validation

Re-Identification Validation

City intersection to validate matches

Re-Identification Validation

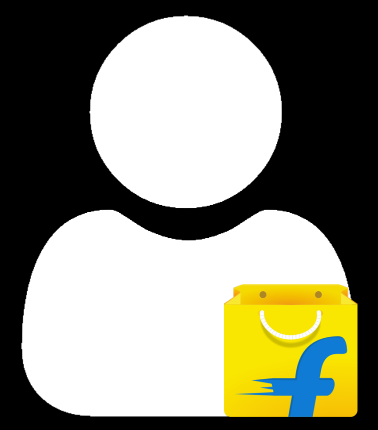
Re-Identification Validation



FK user

Re-Identification Validation

{Bengaluru, Delhi,
Rewari, Jamshedpur}



FK user

Re-Identification Validation

Experience

Data Science Intern
Flipkart
Jun 2022 - Present · 5 months
Bengaluru, Karnataka, India

Researcher Phd Candidate
Precog @ IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Researcher Phd Candidate
MIDAS-IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

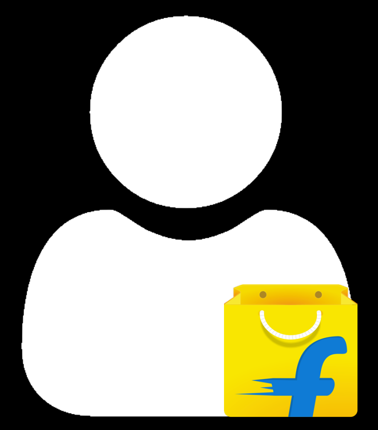
Research Intern
Goldman Sachs
May 2021 - Aug 2021 · 4 months
Bengaluru, Karnataka, India

Research Intern
National Institute of Informatics
Dec 2018 - Jun 2019 · 7 months
Tokyo, Japan

Research Intern
Indraprastha Institute of Information Technology, Delhi
Jan 2018 - May 2018 · 5 months
Delhi Area, India

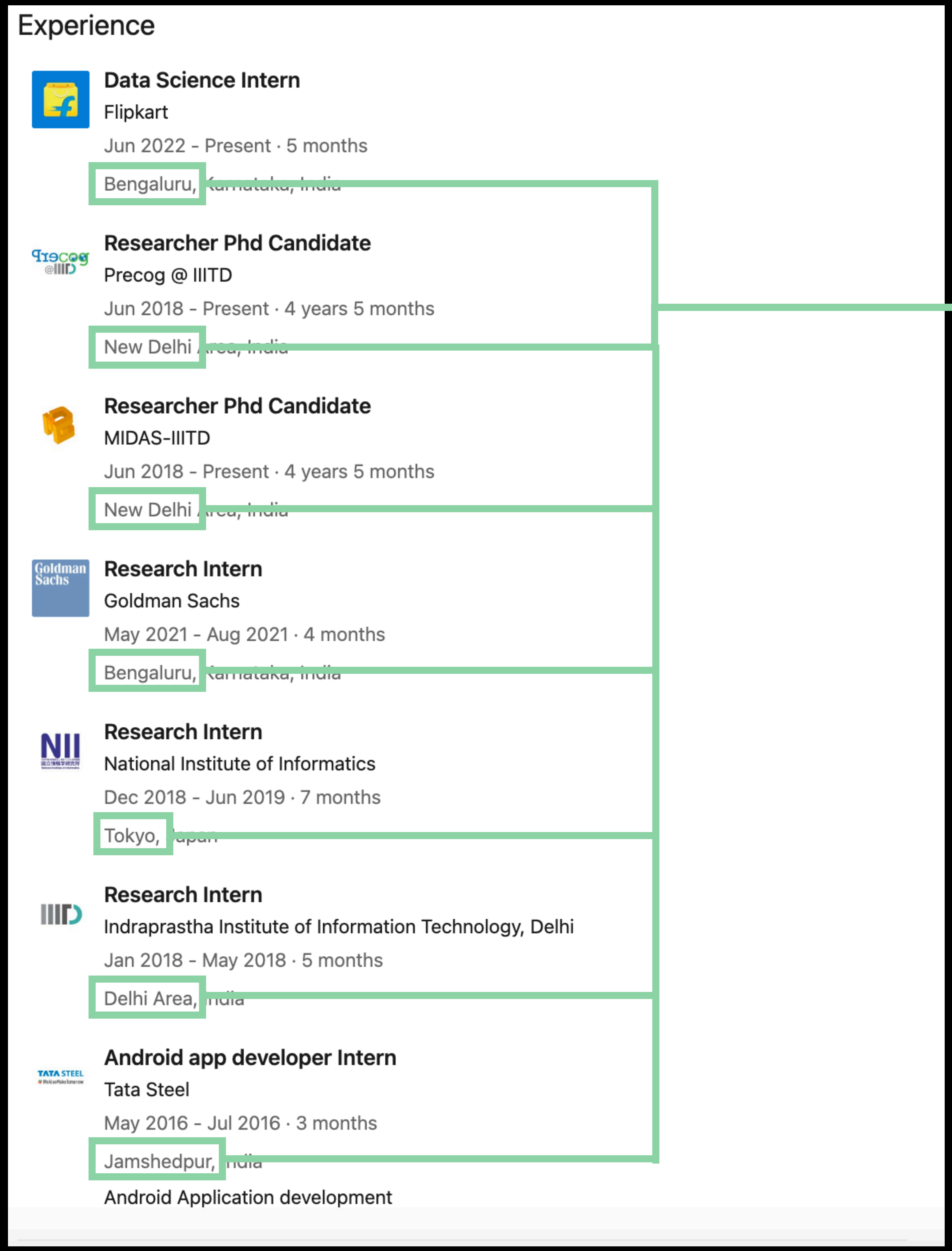
Android app developer Intern
Tata Steel
May 2016 - Jul 2016 · 3 months
Jamshedpur, India
Android Application development

{Bengaluru, Delhi, Rewari, Jamshedpur}



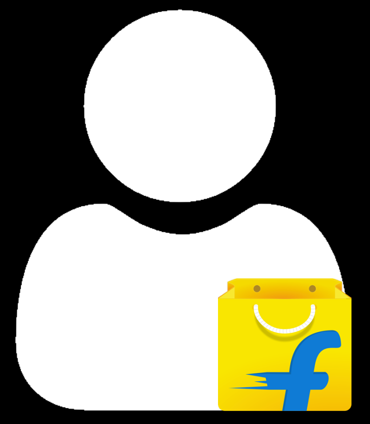
FK user

Re-Identification Validation



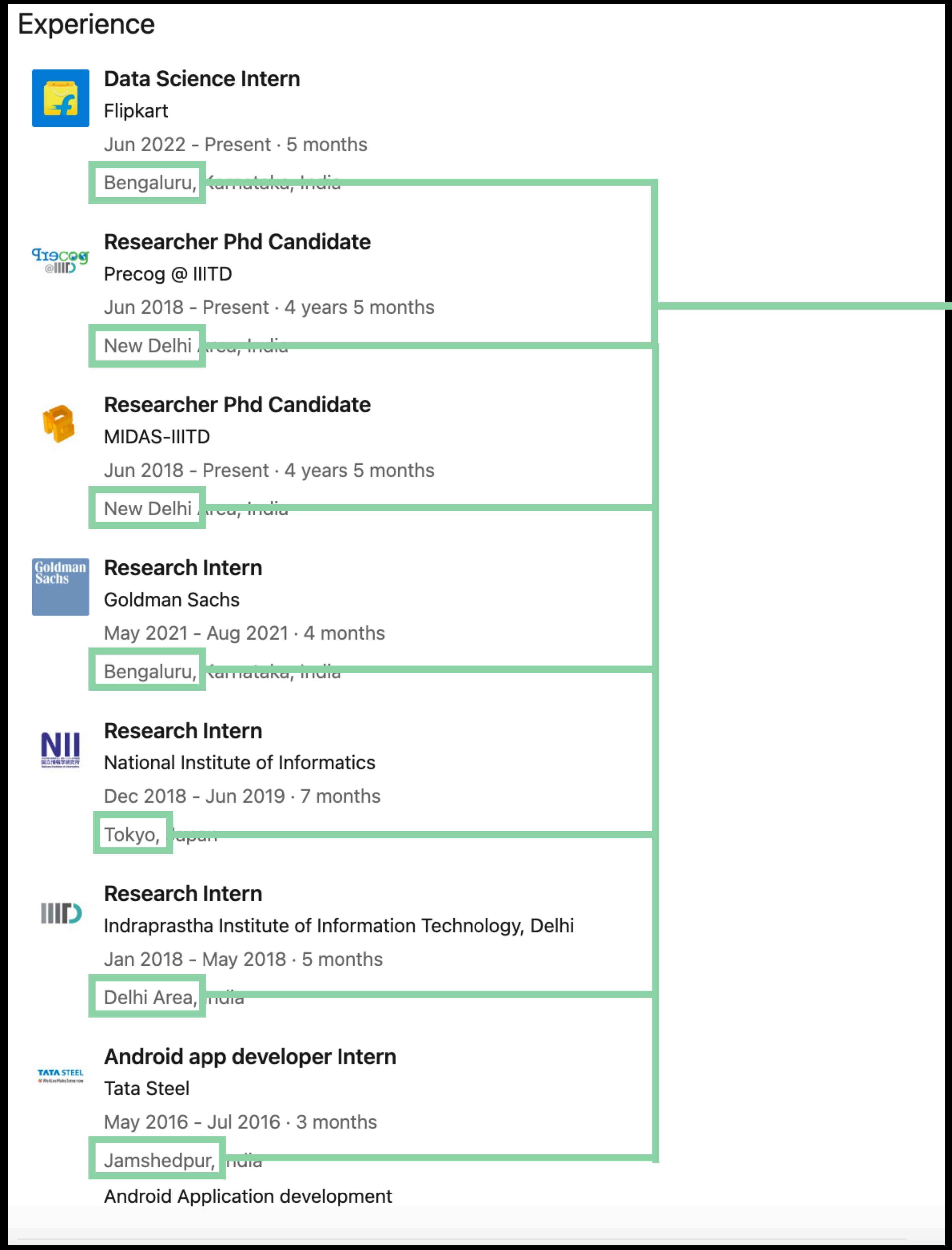
{Bengaluru, New Delhi, Tokyo, Jamshedpur}

{Bengaluru, Delhi, Rewari, Jamshedpur}



FK user

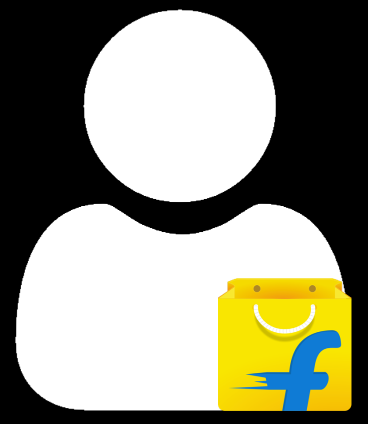
Re-Identification Validation



{Bengaluru, New Delhi, Tokyo, Jamshedpur}

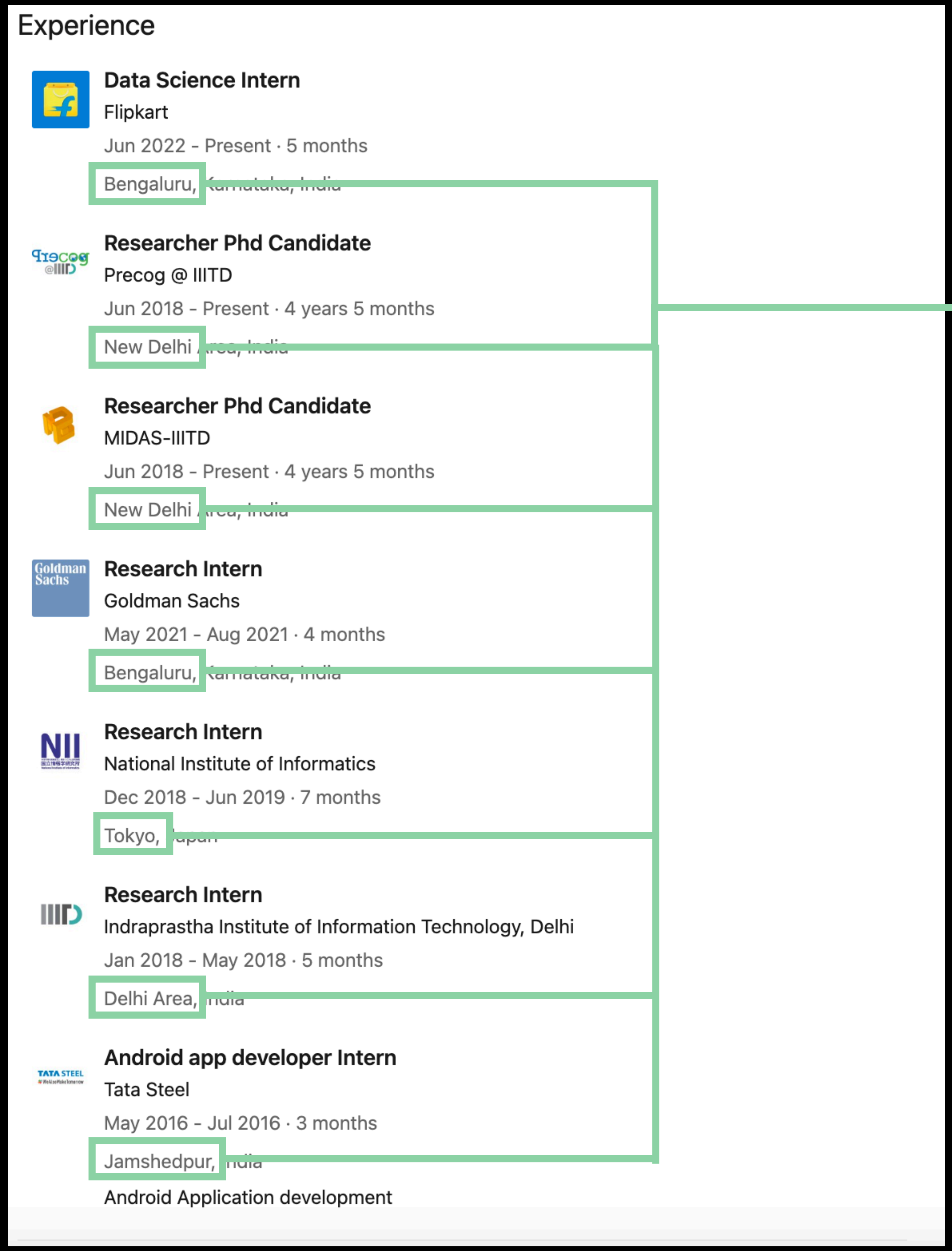
∩

{Bengaluru, Delhi, Rewari, Jamshedpur}



FK user

Re-Identification Validation


Experience

Data Science Intern
Flipkart
Jun 2022 - Present · 5 months
Bengaluru, Karnataka, India

Researcher Phd Candidate
Precog @ IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Researcher Phd Candidate
MIDAS-IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Research Intern
Goldman Sachs
May 2021 - Aug 2021 · 4 months
Bengaluru, Karnataka, India

Research Intern
National Institute of Informatics
Dec 2018 - Jun 2019 · 7 months
Tokyo, Japan

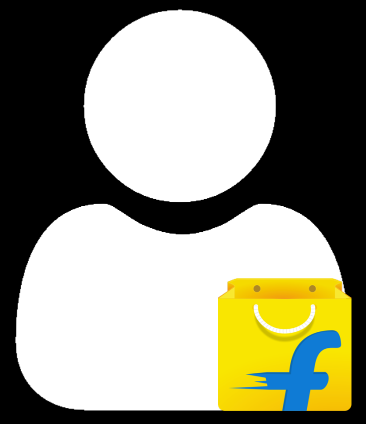
Research Intern
Indraprastha Institute of Information Technology, Delhi
Jan 2018 - May 2018 · 5 months
Delhi Area, India

Android app developer Intern
Tata Steel
May 2016 - Jul 2016 · 3 months
Jamshedpur, India
Android Application development

{Bengaluru, New Delhi, Tokyo, Jamshedpur}

$\cap$

{Bengaluru, Delhi, Rewari, Jamshedpur}

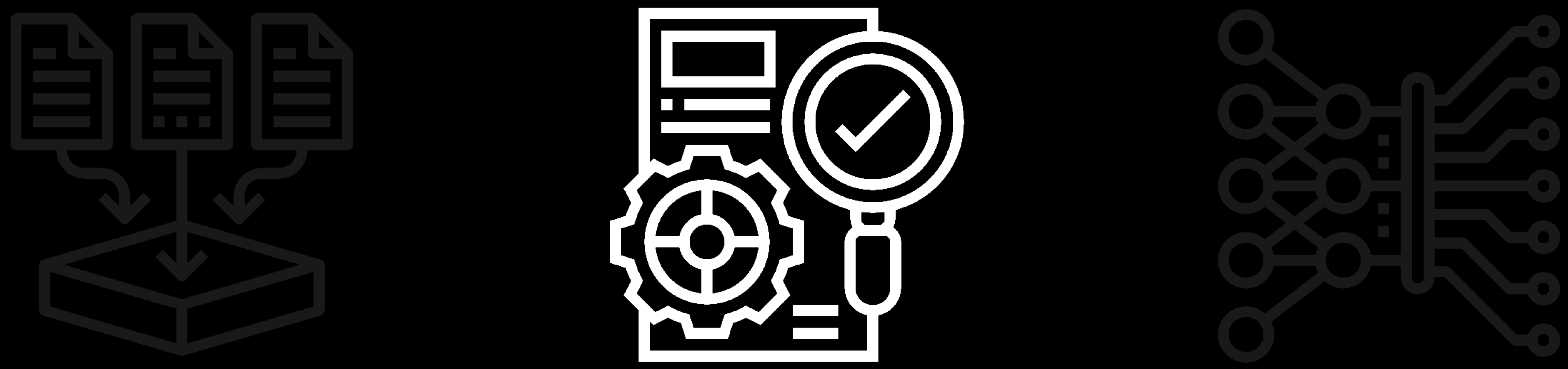

FK user

{Bengaluru, Delhi, Jamshedpur}

Re-Identification Validation

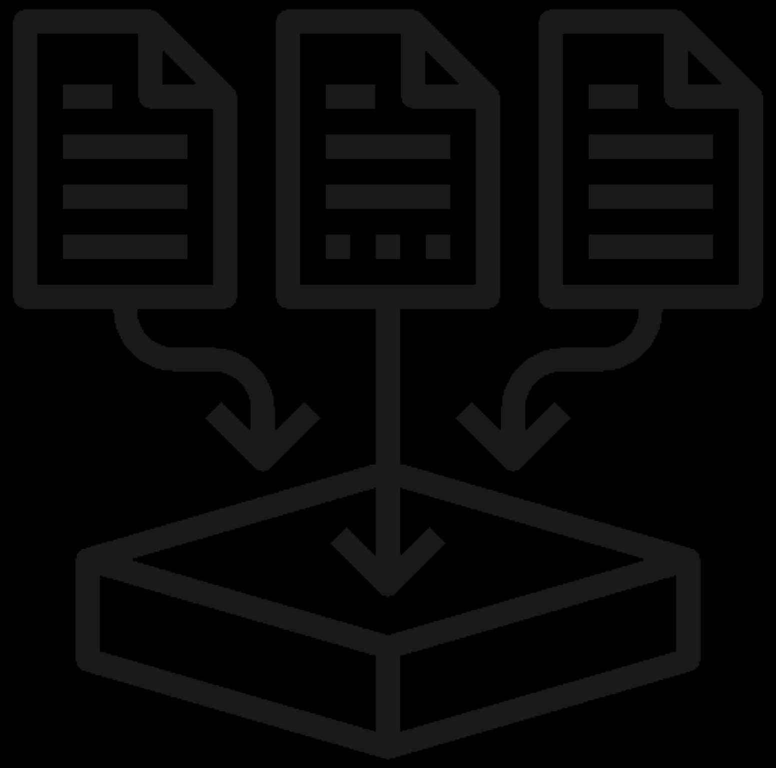
Match Threshold θ	Exact Match	Multiple Match	No Match
0.1	81.49	18.51	0.00
0.2	81.31	18.51	0.18
0.3	79.58	18.51	1.91
0.4	76.21	18.41	5.38
0.5	71.01	18.41	10.57
0.6	68.92	17.68	13.40
0.7	65.91	17.50	16.59
0.8	64.63	17.41	17.96
0.9	64.36	17.41	18.23
1.0	42.57	17.41	40.02

Proposed Flow

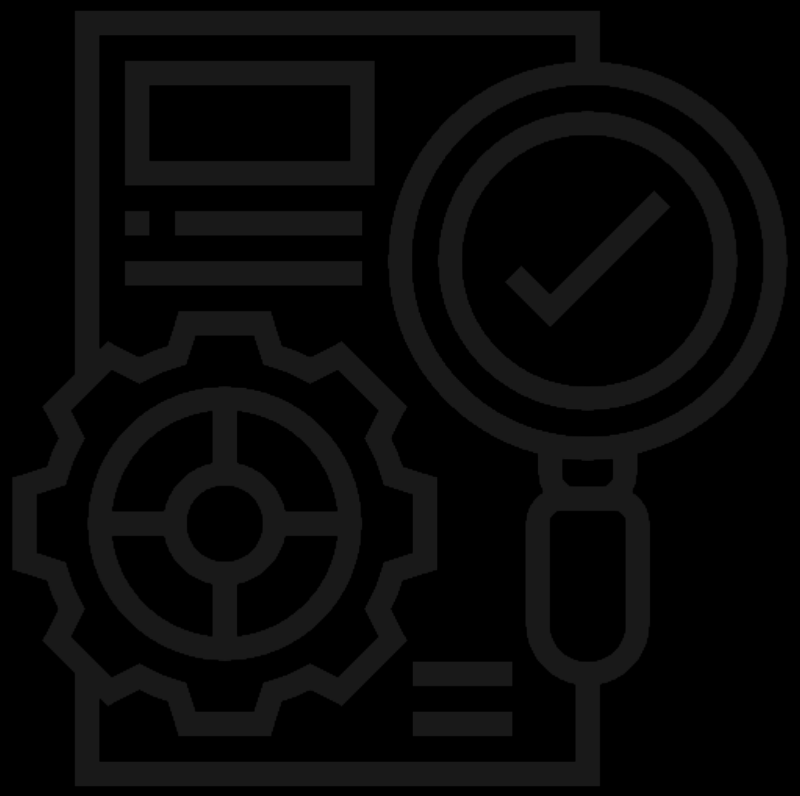


Data collection Data validation Modeling

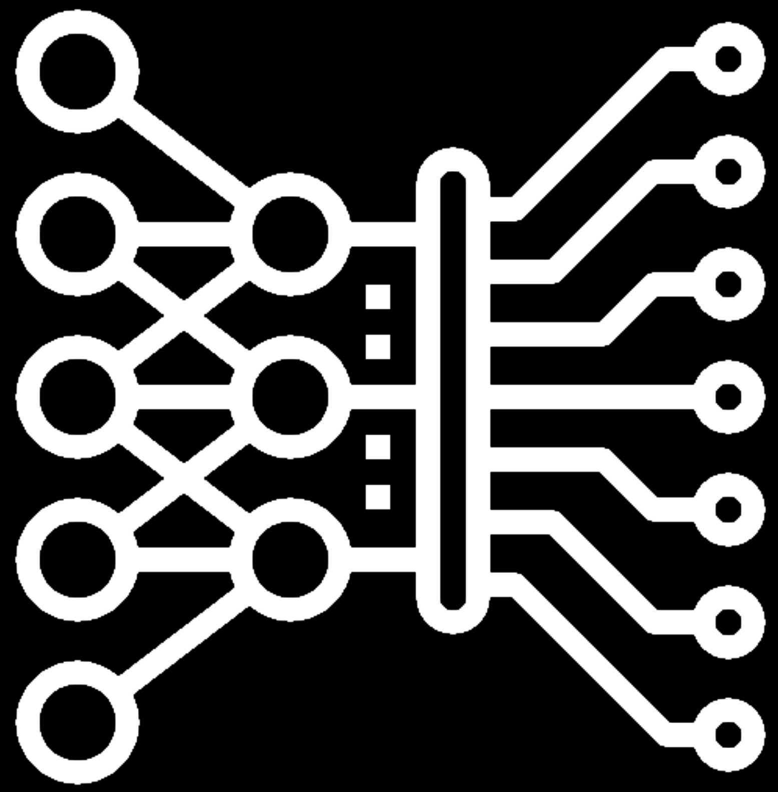
Proposed Flow



Data collection



Data validation



Modeling

Features - Internal

Features - Internal

📈 Past trends of 3 months and 1 year

➗ Ratio of RTO/Total orders

👤 User

📦 Product

🏭 Seller

🗃️ Product category

📍 Pincode

🛣️ Street name

🌆 City

⏰ Hour of the day

📆 Day of the week

Features - Social

Features - Social

Experience

Data Science Intern
Flipkart
Jun 2022 - Present · 5 months
Bengaluru, Karnataka, India

Researcher Phd Candidate
Precog @ IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Researcher Phd Candidate
MIDAS-IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Research Intern
Goldman Sachs
May 2021 - Aug 2021 · 4 months
Bengaluru, Karnataka, India

Research Intern
National Institute of Informatics
Dec 2018 - Jun 2019 · 7 months
Tokyo, Japan

Research Intern
Indraprastha Institute of Information Technology, Delhi
Jan 2018 - May 2018 · 5 months
Delhi Area, India

Android app developer Intern
Tata Steel
May 2016 - Jul 2016 · 3 months
Jamshedpur, India
Android Application development

Education

Indraprastha Institute of Information Technology, Delhi
Doctor of Philosophy - PhD · Multimodal social media computing · 9.67 CGPA
2018 - 2023
Activities and Societies: ACM Student Chapter

BML Munjal University
Bachelor of Technology - BTech · Computer Science
2014 - 2018

Jain Public School
High School
2005 - 2014

Skills

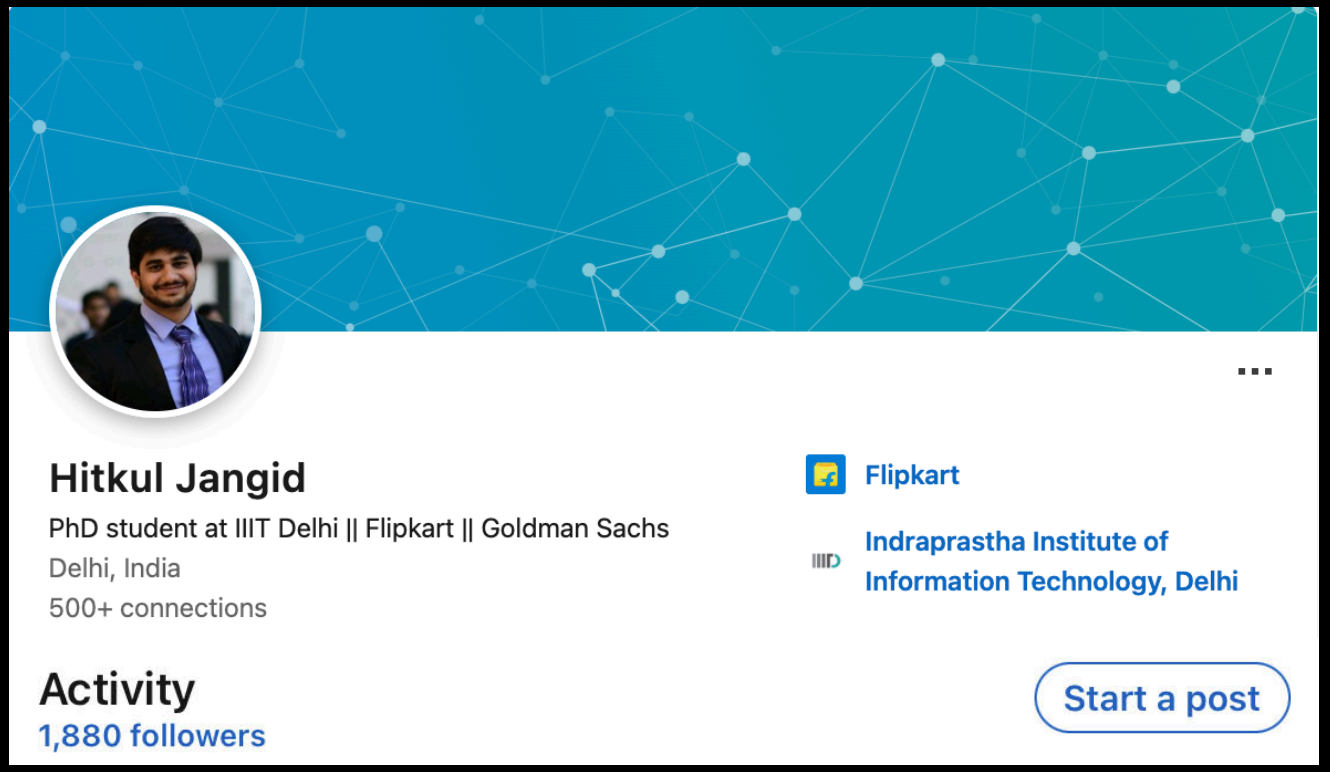
Take skill quiz

Data Science
3 endorsements

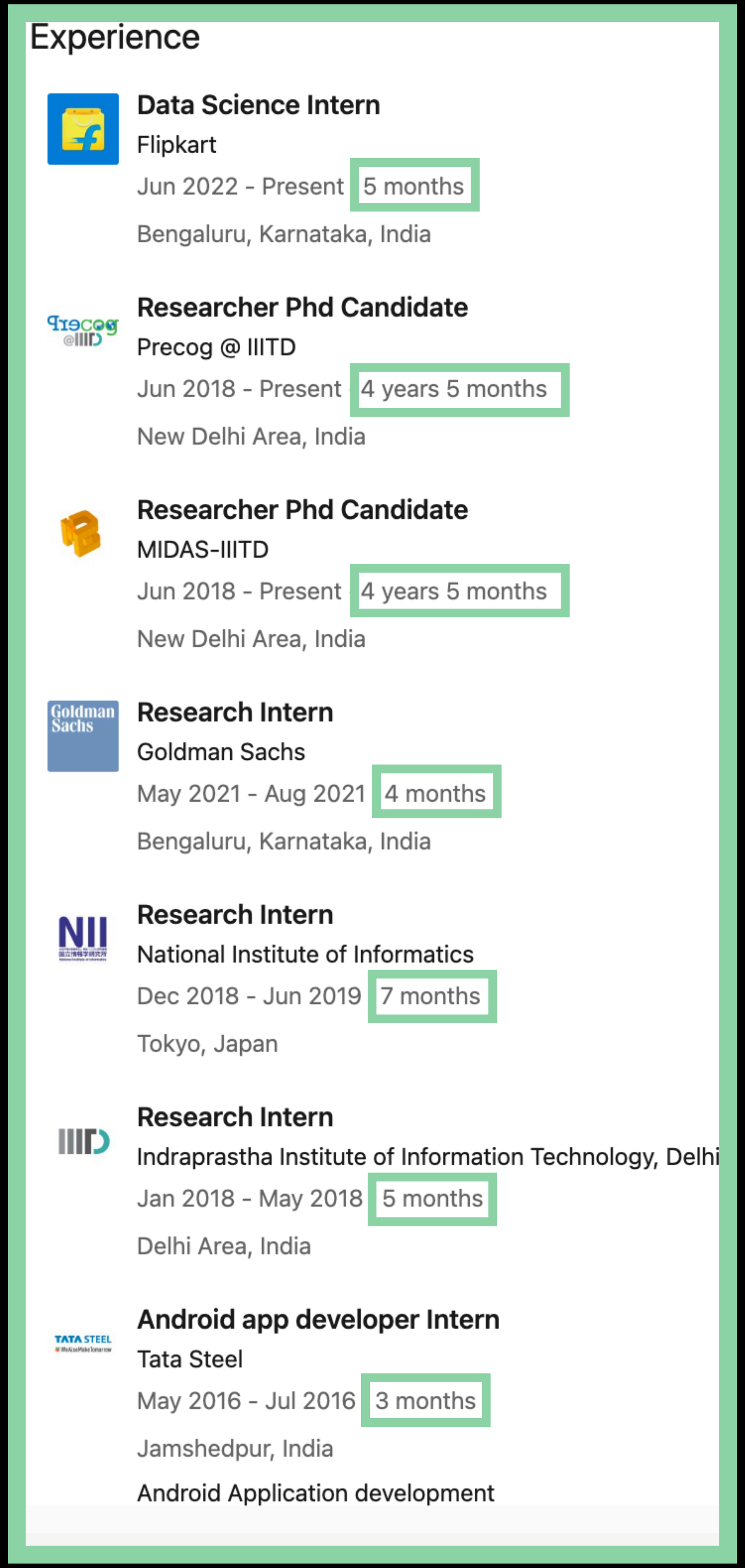
Machine Learning
Endorsed by **Karan Gupta** who is highly skilled at this
5 endorsements

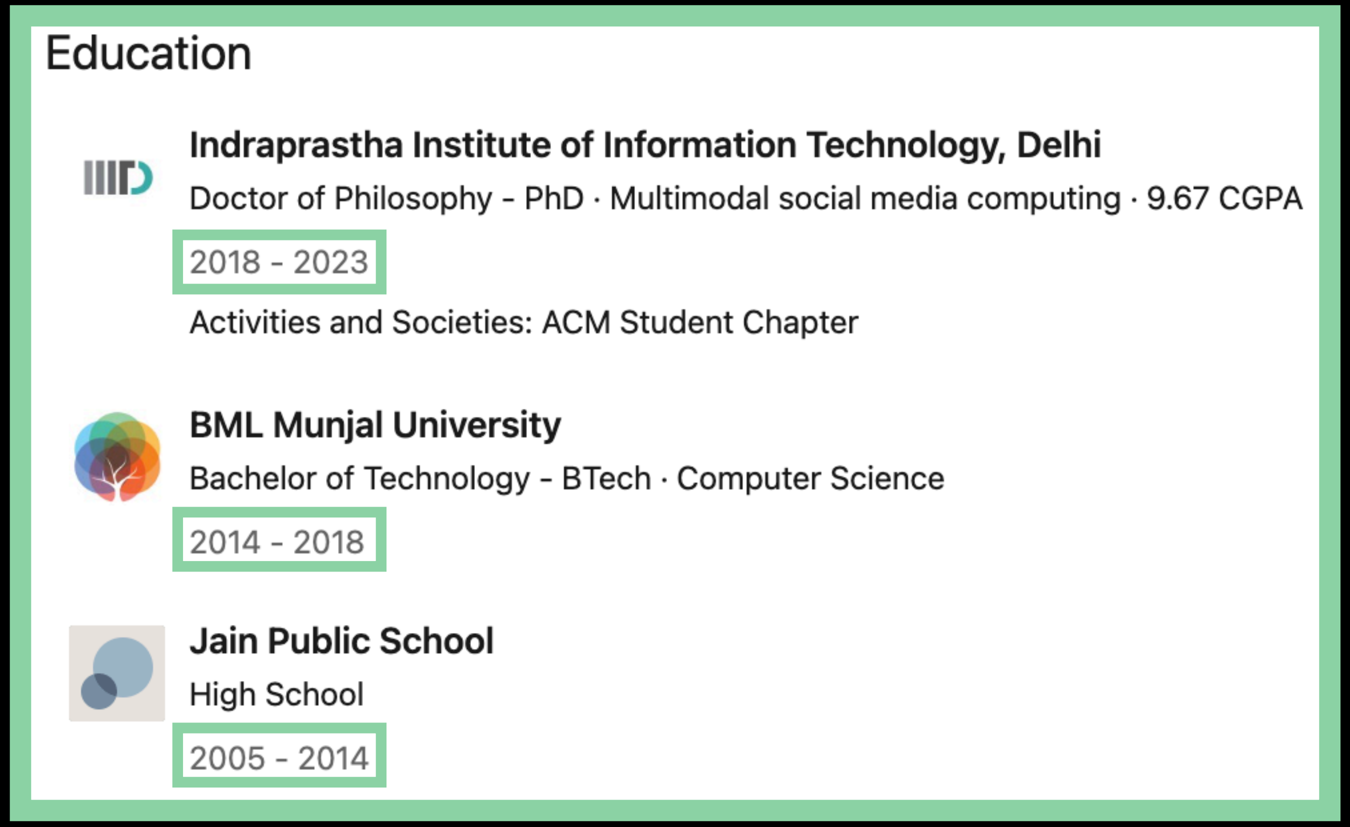
Keras
1 endorsement

Show all 20 skills →



Features - Social



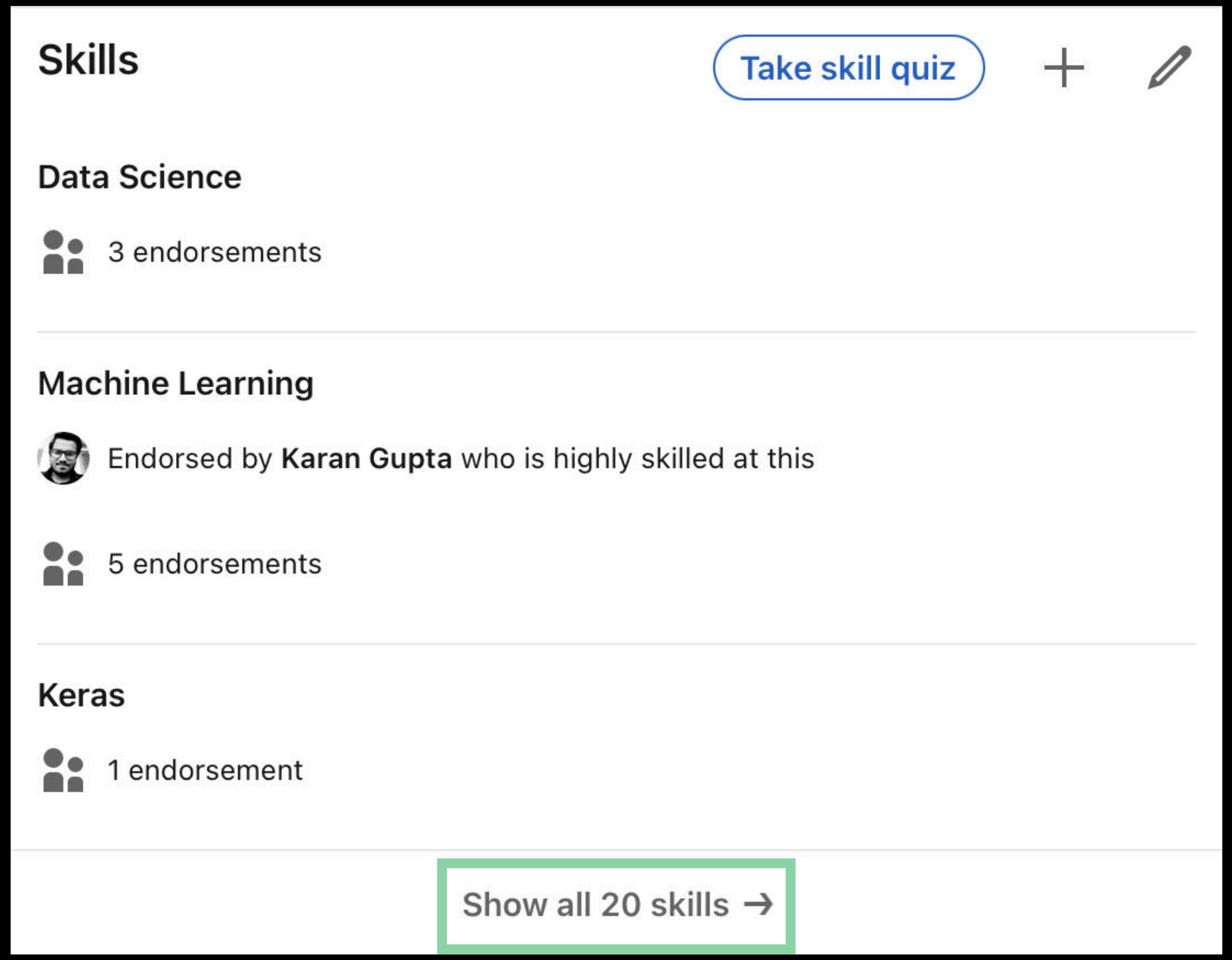


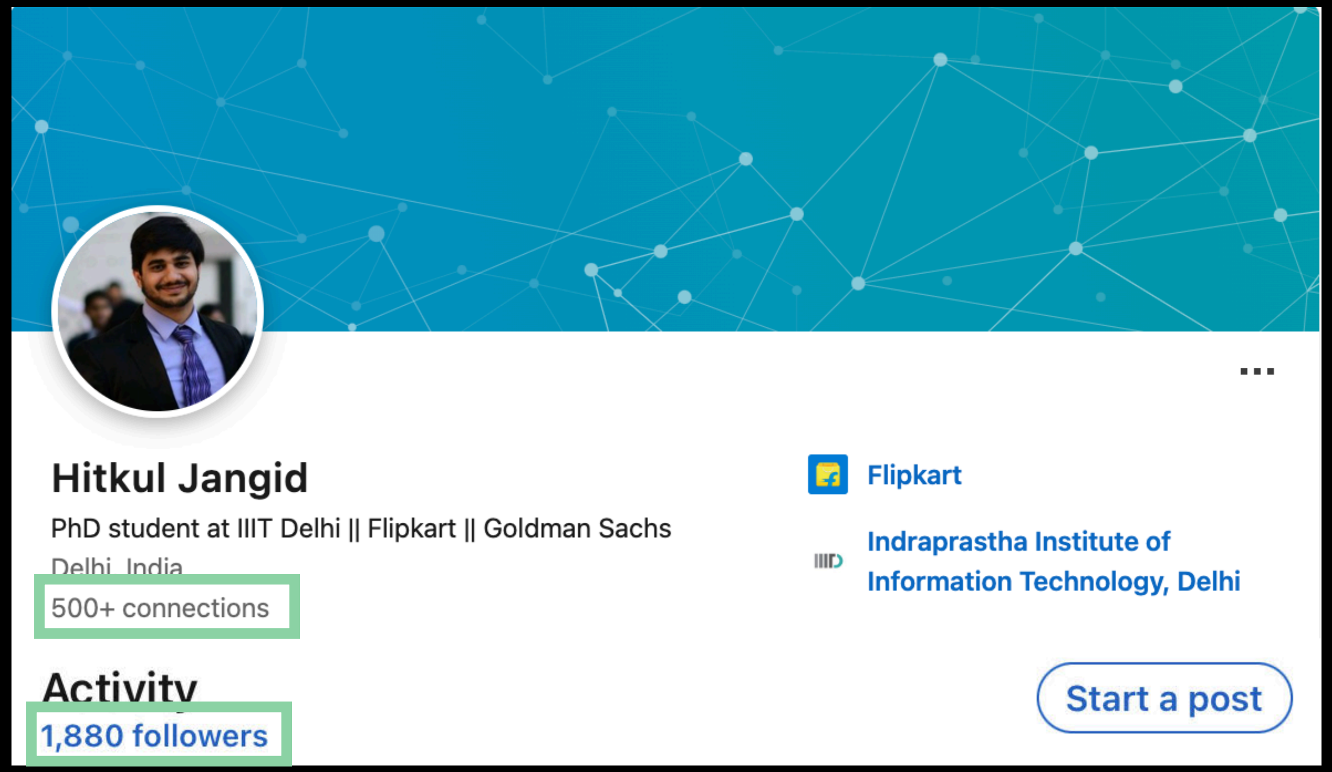
Is student
Is direct link available

# of jobs
# of educations

Years of jobs
Years of education

# of skills
# of connections
# of followers





Features - Language

Experience

Data Science Intern
Flipkart
Jun 2022 - Present · 5 months
Bengaluru, Karnataka, India

Researcher Phd Candidate
Precog @ IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Researcher Phd Candidate
MIDAS-IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Research Intern
Goldman Sachs
May 2021 - Aug 2021 · 4 months
Bengaluru, Karnataka, India

Research Intern
National Institute of Informatics
Dec 2018 - Jun 2019 · 7 months
Tokyo, Japan

Research Intern
Indraprastha Institute of Information Technology, Delhi
Jan 2018 - May 2018 · 5 months
Delhi Area, India

Android app developer Intern
Tata Steel
May 2016 - Jul 2016 · 3 months
Jamshedpur, India
Android Application development

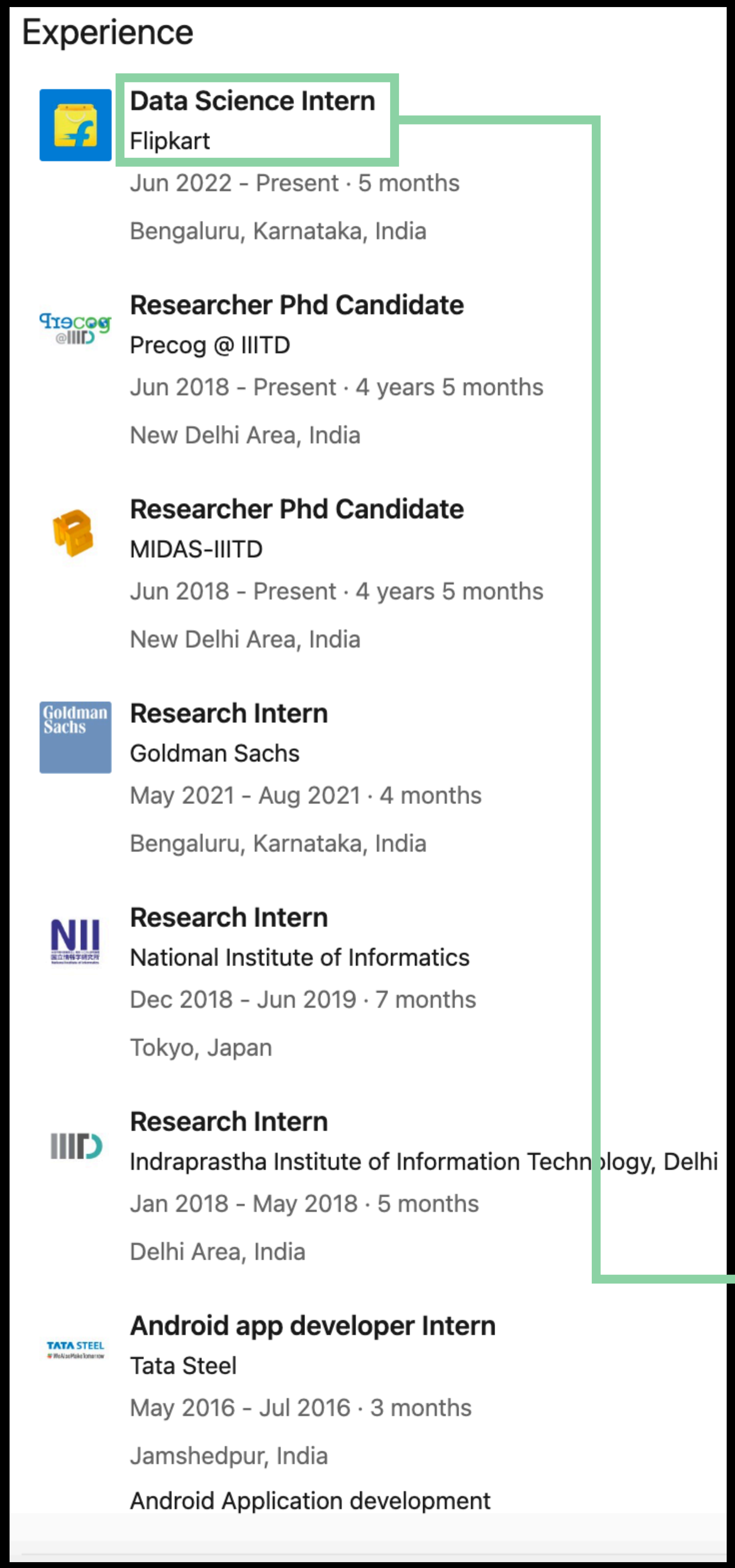
Education

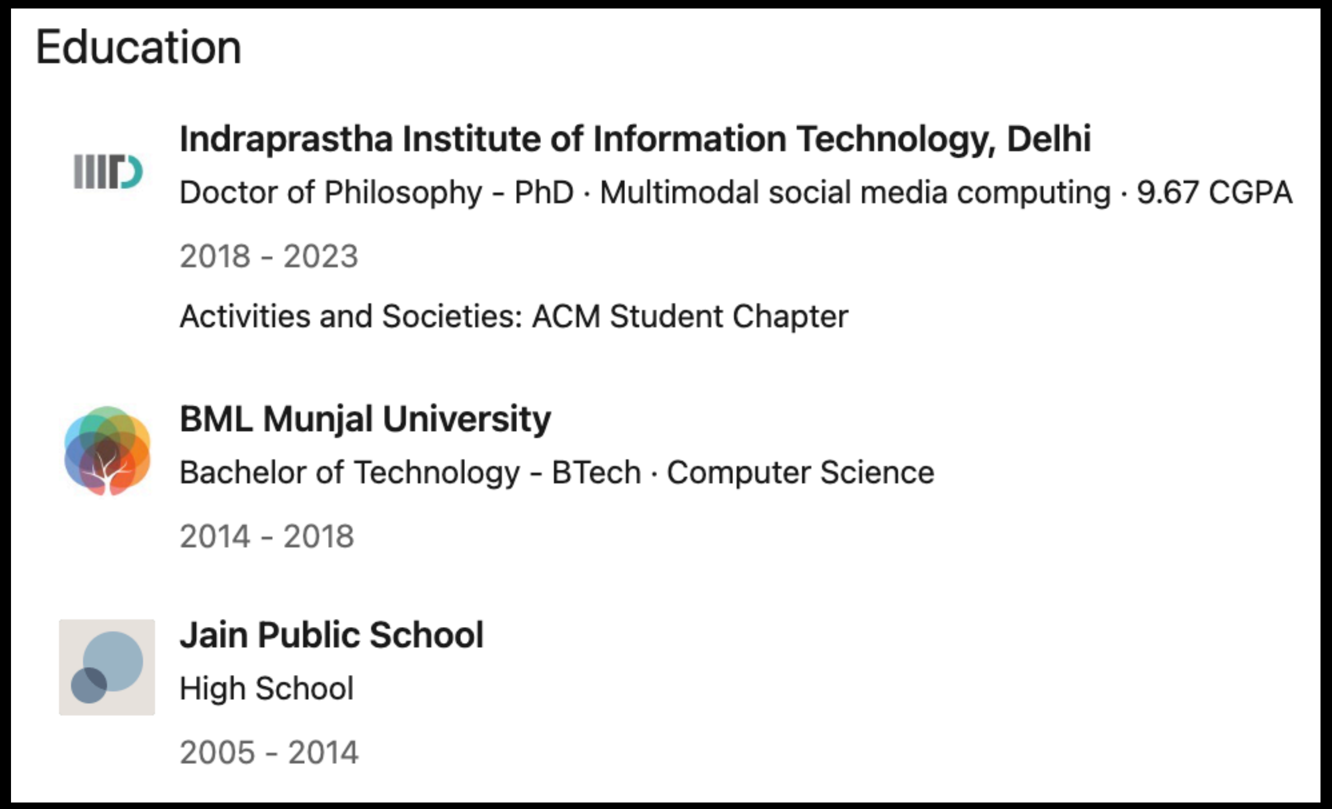
Indraprastha Institute of Information Technology, Delhi
Doctor of Philosophy - PhD · Multimodal social media computing · 9.67 CGPA
2018 - 2023
Activities and Societies: ACM Student Chapter

BML Munjal University
Bachelor of Technology - BTech · Computer Science
2014 - 2018

Jain Public School
High School
2005 - 2014

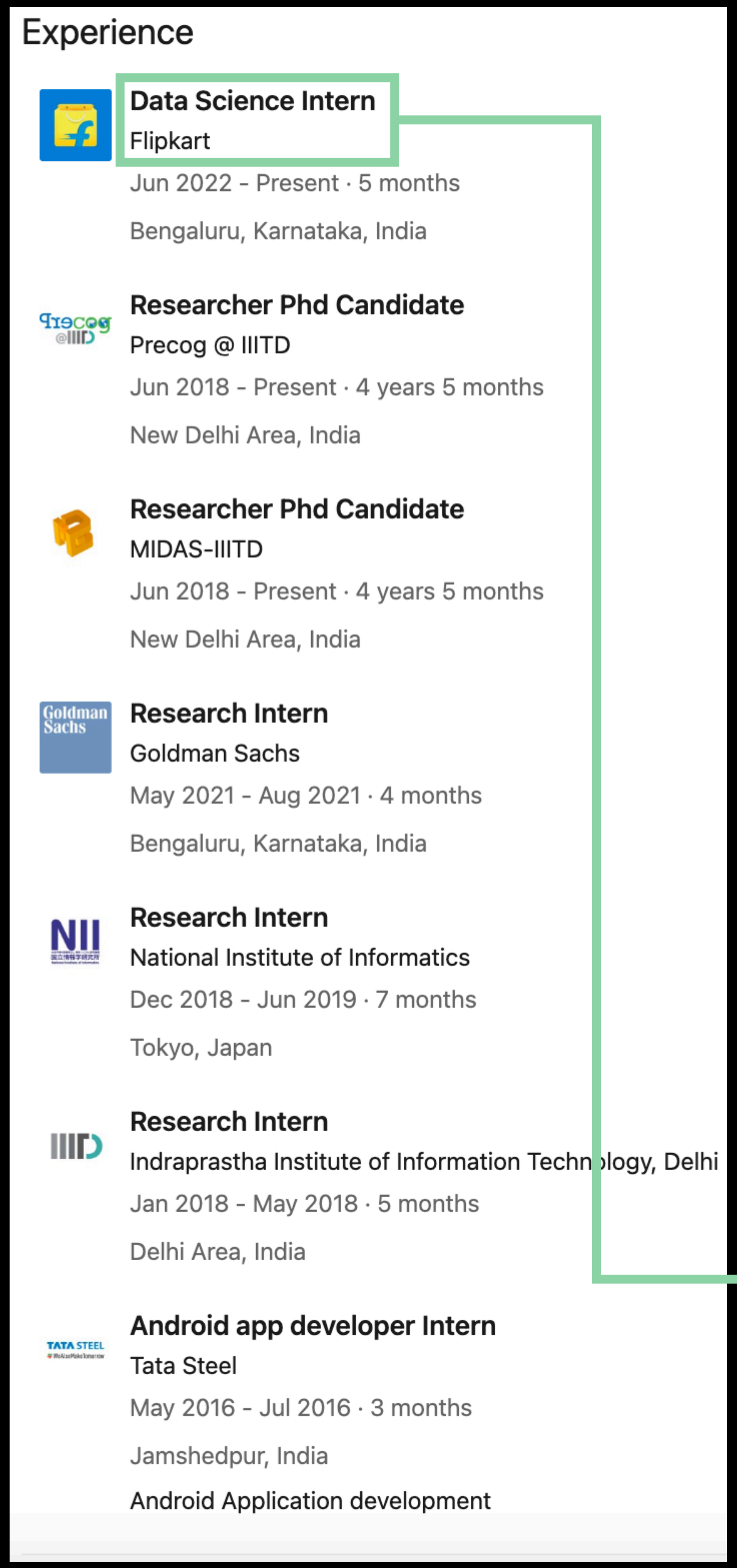
Features - Language





Data Science Intern Flipkart

Features - Language



Experience

Data Science Intern
Flipkart
Jun 2022 - Present · 5 months
Bengaluru, Karnataka, India

Researcher Phd Candidate
Precog @ IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

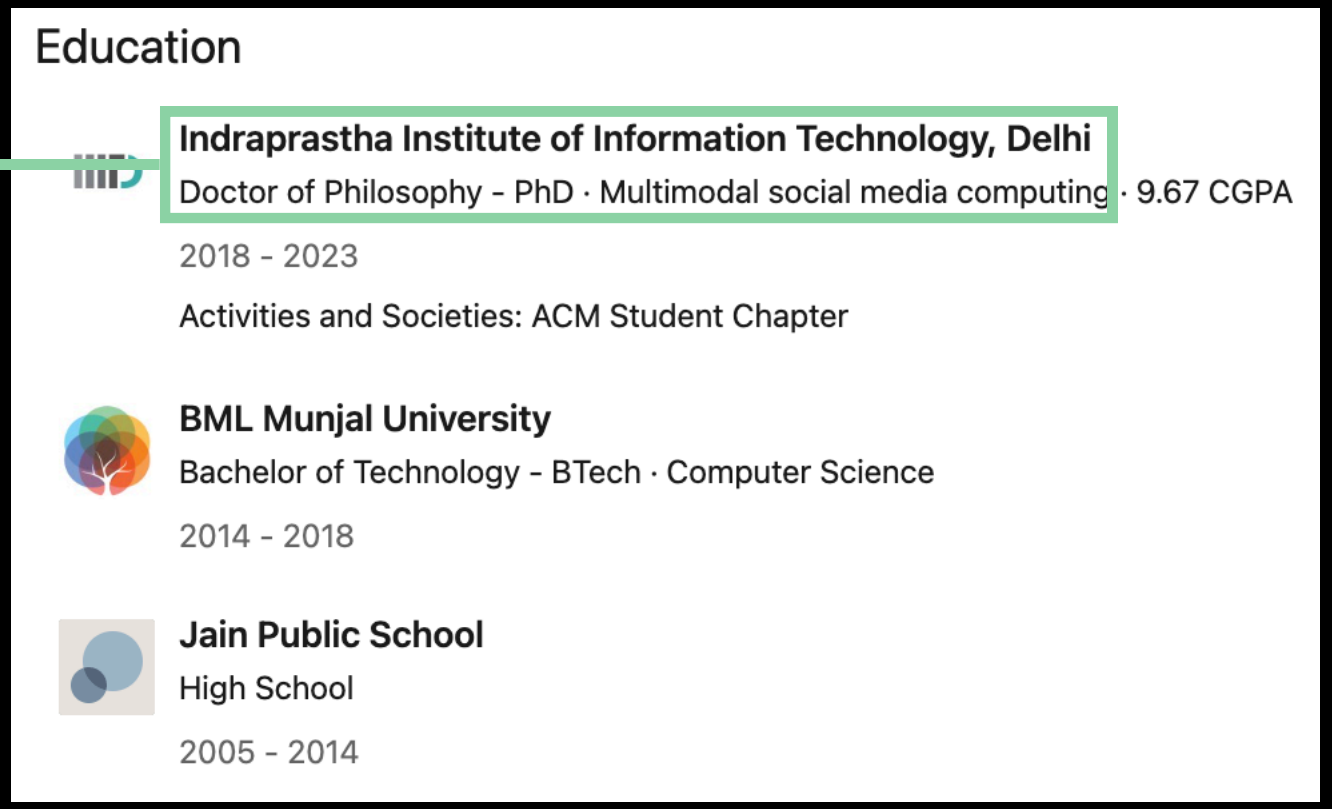
Researcher Phd Candidate
MIDAS-IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Research Intern
Goldman Sachs
May 2021 - Aug 2021 · 4 months
Bengaluru, Karnataka, India

Research Intern
National Institute of Informatics
Dec 2018 - Jun 2019 · 7 months
Tokyo, Japan

Research Intern
Indraprastha Institute of Information Technology, Delhi
Jan 2018 - May 2018 · 5 months
Delhi Area, India

Android app developer Intern
Tata Steel
May 2016 - Jul 2016 · 3 months
Jamshedpur, India
Android Application development



Education

Indraprastha Institute of Information Technology, Delhi
Doctor of Philosophy - PhD · Multimodal social media computing · 9.67 CGPA
2018 - 2023
Activities and Societies: ACM Student Chapter

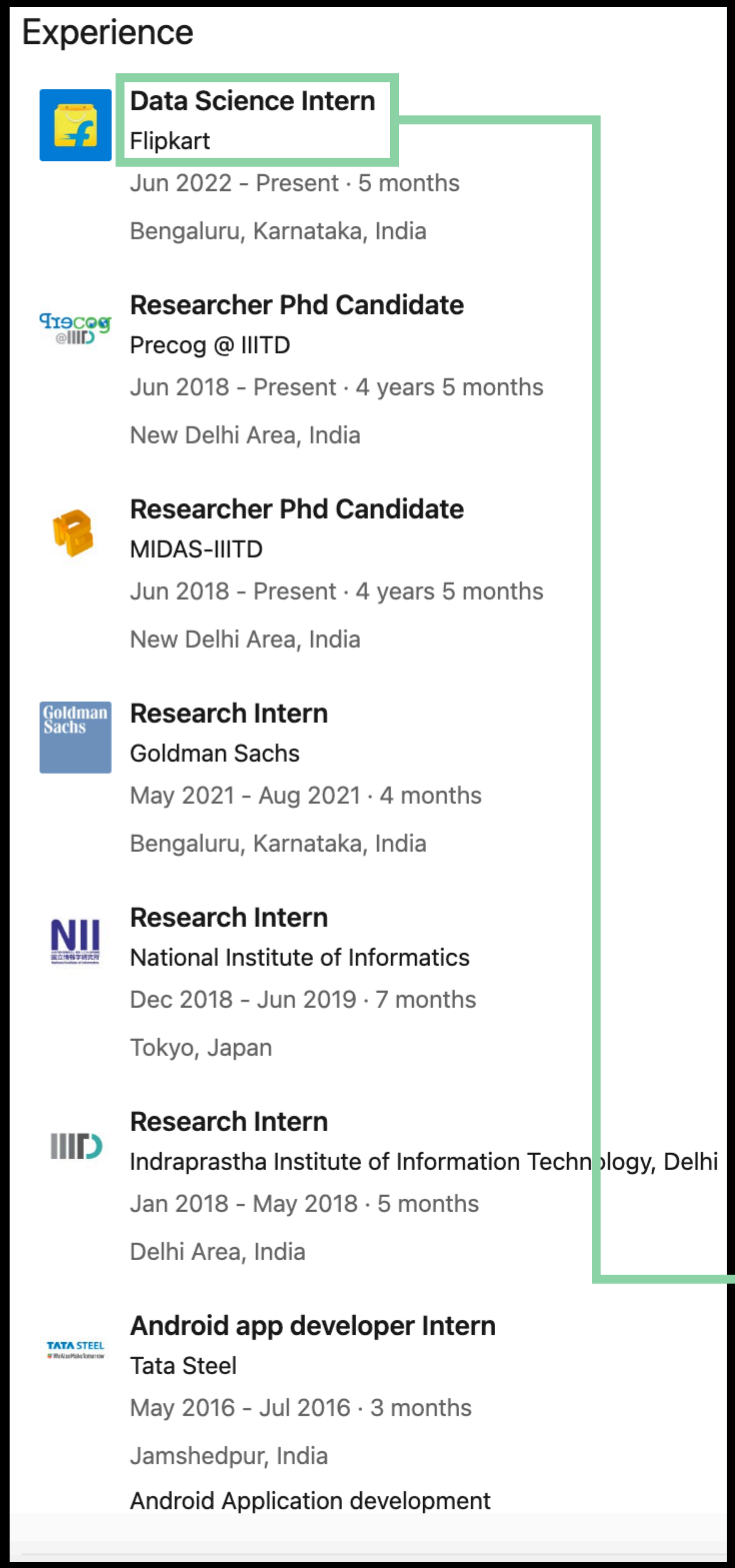
BML Munjal University
Bachelor of Technology - BTech · Computer Science
2014 - 2018

Jain Public School
High School
2005 - 2014

Doctor of Philosophy Indraprastha Institute of Information Technology

Data Science Intern Flipkart

Features - Language


Experience

Data Science Intern
Flipkart
Jun 2022 - Present · 5 months
Bengaluru, Karnataka, India

Researcher Phd Candidate
Precog @ IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

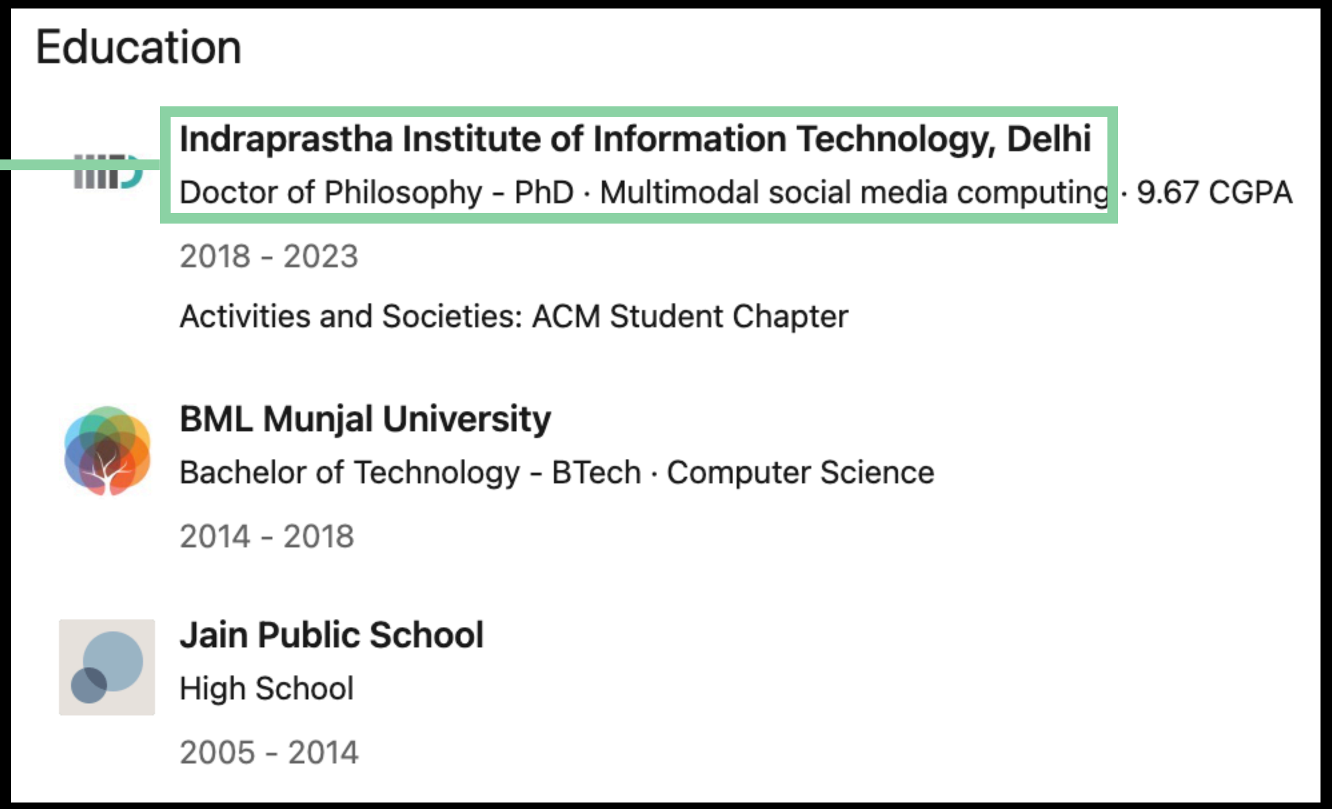
Researcher Phd Candidate
MIDAS-IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Research Intern
Goldman Sachs
May 2021 - Aug 2021 · 4 months
Bengaluru, Karnataka, India

Research Intern
National Institute of Informatics
Dec 2018 - Jun 2019 · 7 months
Tokyo, Japan

Research Intern
Indraprastha Institute of Information Technology, Delhi
Jan 2018 - May 2018 · 5 months
Delhi Area, India

Android app developer Intern
Tata Steel
May 2016 - Jul 2016 · 3 months
Jamshedpur, India
Android Application development


Education

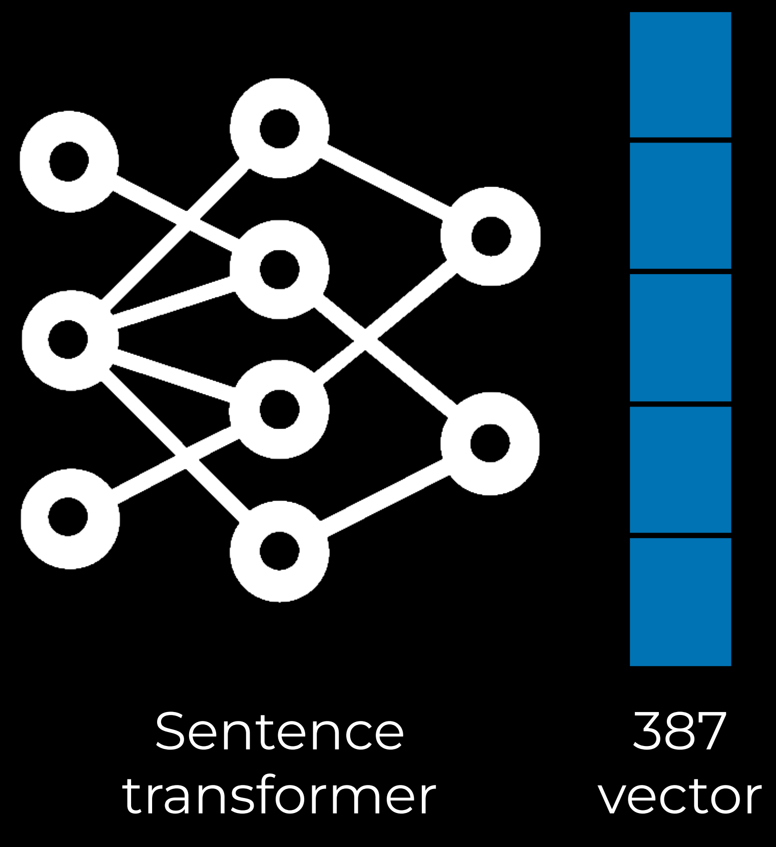
Indraprastha Institute of Information Technology, Delhi
Doctor of Philosophy - PhD · Multimodal social media computing · 9.67 CGPA
2018 - 2023
Activities and Societies: ACM Student Chapter

BML Munjal University
Bachelor of Technology - BTech · Computer Science
2014 - 2018

Jain Public School
High School
2005 - 2014

Doctor of Philosophy Indraprastha Institute of Information Technology

Data Science Intern Flipkart



Features - Language

Experience

Data Science Intern
Flipkart
Jun 2022 - Present · 5 months
Bengaluru, Karnataka, India

Researcher Phd Candidate
Precog @ IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Researcher Phd Candidate
MIDAS-IIITD
Jun 2018 - Present · 4 years 5 months
New Delhi Area, India

Research Intern
Goldman Sachs
May 2021 - Aug 2021 · 4 months
Bengaluru, Karnataka, India

Research Intern
National Institute of Informatics
Dec 2018 - Jun 2019 · 7 months
Tokyo, Japan

Research Intern
Indraprastha Institute of Information Technology, Delhi
Jan 2018 - May 2018 · 5 months
Delhi Area, India

Android app developer Intern
Tata Steel
May 2016 - Jul 2016 · 3 months
Jamshedpur, India
Android Application development

Education

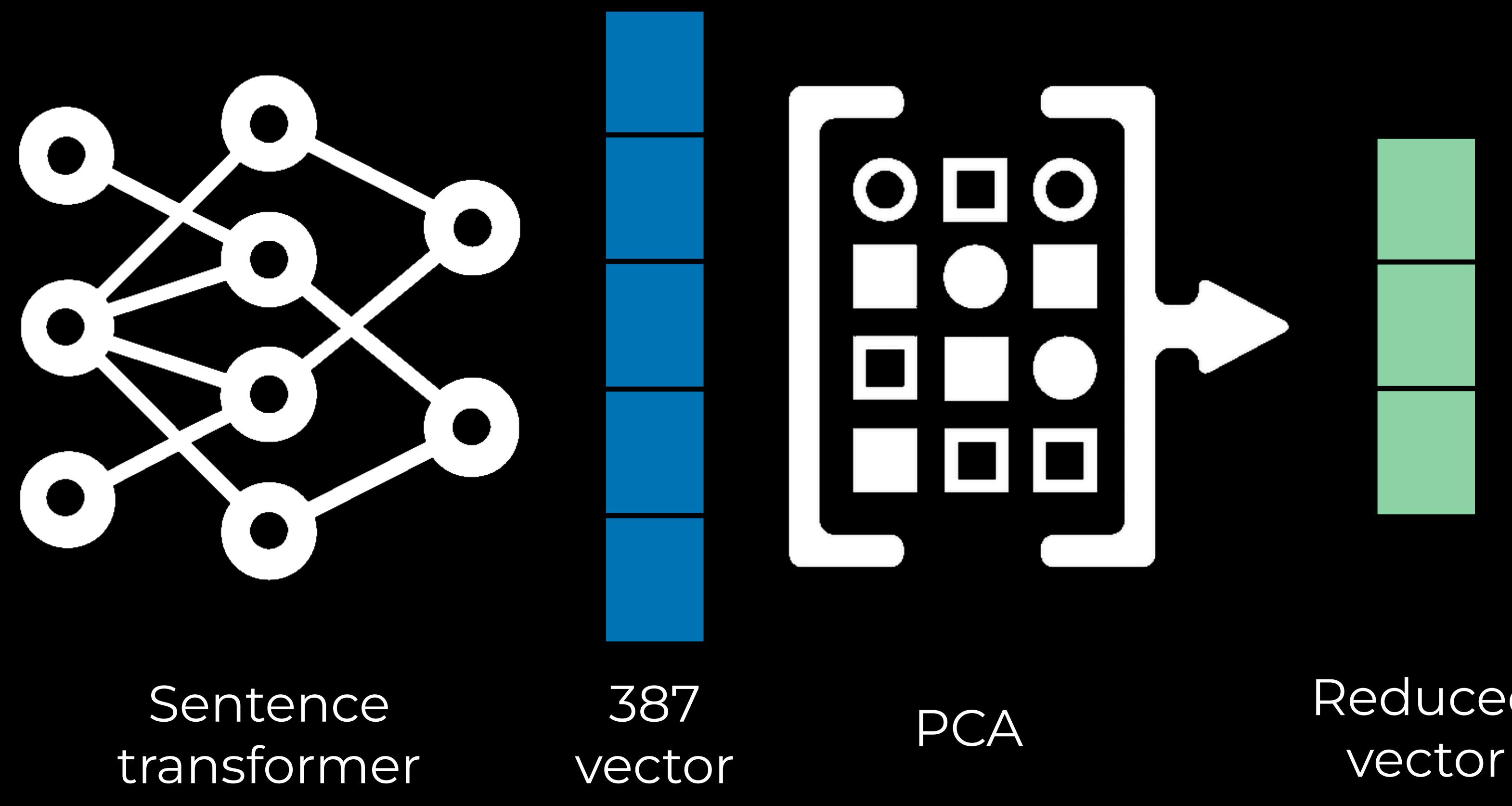
Indraprastha Institute of Information Technology, Delhi
Doctor of Philosophy - PhD · Multimodal social media computing · 9.67 CGPA
2018 - 2023
Activities and Societies: ACM Student Chapter

BML Munjal University
Bachelor of Technology - BTech · Computer Science
2014 - 2018

Jain Public School
High School
2005 - 2014

Doctor of Philosophy Indraprastha Institute of Information Technology

Data Science Intern Flipkart



Evaluation

$$Goodness = \left(\frac{|P|}{|P| + |N|} - \frac{|P| - |P_{Pred\ and\ True}|}{|P| + |N| - |P_{Pred}|} \right) \times 10^4$$

$$FPR = \frac{|P_{Pred}| - |P_{Pred\ and\ True}|}{|P_{Pred}|}$$

Results

Model	Features	Precision (%)	Recall (%)	Goodness (bps)
Random Forest	Past Trends	85.7	40.3	1,005.7
	Past Trends + Social quantitative	85.7	50.4	1,305.6
	Past Trends + Social quantitative + Social abstractive	**88.8**	**60.2**	**1,633.7**
XGBoost	Past Trends	80.0	33.6	809.3
	Past Trends + Social quantitative	82.2	39.7	994.1
	Past Trends + Social quantitative + Social abstractive	86.8	44.5	1,129.4
TabNet	Past Trends	82.4	39.4	977.0
	Past Trends + Social quantitative	78.2	30.2	716.2
	Past Trends + Social quantitative + Social abstractive	64.2	15.1	320.0

Results

Model	Features	Precision (%)	Recall (%)	Goodness (bps)
Random Forest	Past Trends	85.7	40.3	1,005.7
	Past Trends + Social quantitative	85.7	50.4	1,305.6
	Past Trends + Social quantitative + Social abstractive	**88.8**	**60.2**	**1,633.7**
XGBoost	Past Trends	80.0	33.6	809.3
	Past Trends + Social quantitative	82.2	39.7	994.1
	Past Trends + Social quantitative + Social abstractive	86.8	44.5	1,129.4
TabNet	Past Trends	82.4	39.4	977.0
	Past Trends + Social quantitative	78.2	30.2	716.2
	Past Trends + Social quantitative + Social abstractive	64.2	15.1	320.0

Results

Model	Features	Precision (%)	Recall (%)	Goodness (bps)
Random Forest	Past Trends	85.7	40.3	1,005.7
	Past Trends + Social quantitative	85.7	50.4	1,305.6
	Past Trends + Social quantitative + Social abstractive	**88.8**	**60.2**	**1,633.7**
XGBoost	Past Trends	80.0	33.6	809.3
	Past Trends + Social quantitative	82.2	39.7	994.1
	Past Trends + Social quantitative + Social abstractive	86.8	44.5	1,129.4
TabNet	Past Trends	82.4	39.4	977.0
	Past Trends + Social quantitative	78.2	30.2	716.2
	Past Trends + Social quantitative + Social abstractive	64.2	15.1	320.0

Results

Well-tuned Simple Nets Excel on Tabular Datasets

Arlind Kadra
Department of Computer Science
University of Freiburg
kadraa@cs.uni-freiburg.de

Marius Lindauer
Institute for Information Processing
Leibniz University Hannover
lindauer@tnt.uni-hannover.de

Frank Hutter
Department of Computer Science
University of Freiburg &
Bosch Center for Artificial Intelligence
fh@cs.uni-freiburg.de

Josif Grabocka
Department of Computer Science
University of Freiburg
grabocka@informatik.uni-freiburg.de

Tabular Data: Deep Learning is Not All You Need

Ravid Shwartz-Ziv RAVID.ZIV@INTEL.COM
IT AI Group, Intel
Amitai Armon AMITAI.ARMON@INTEL.COM
IT AI Group, Intel

This is definitely not all you need

DEEP LEARNING MACHINE LEARNING

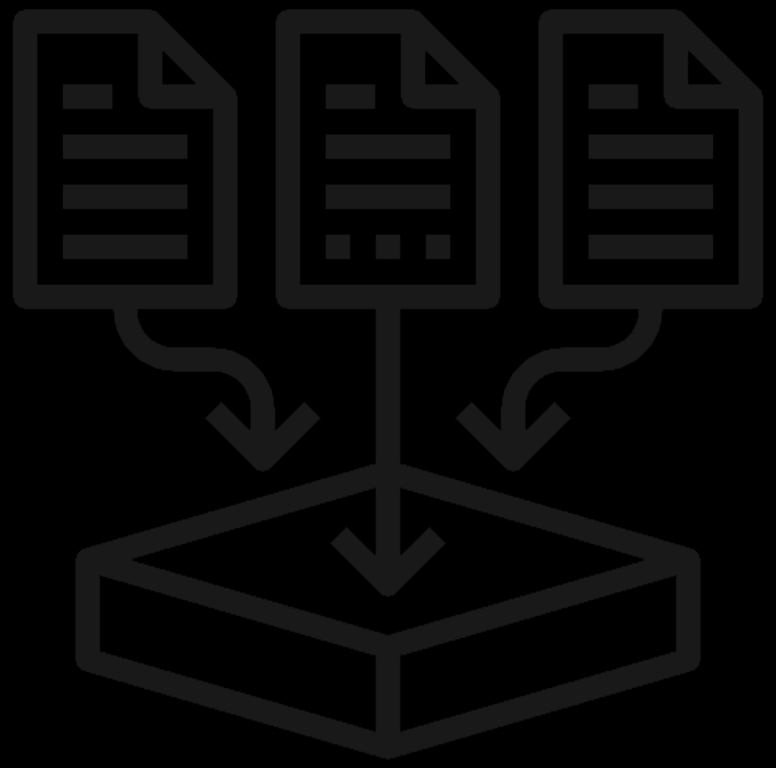
A summary of findings regarding deep learning for tabular data.

Revisiting Deep Learning Models for Tabular Data

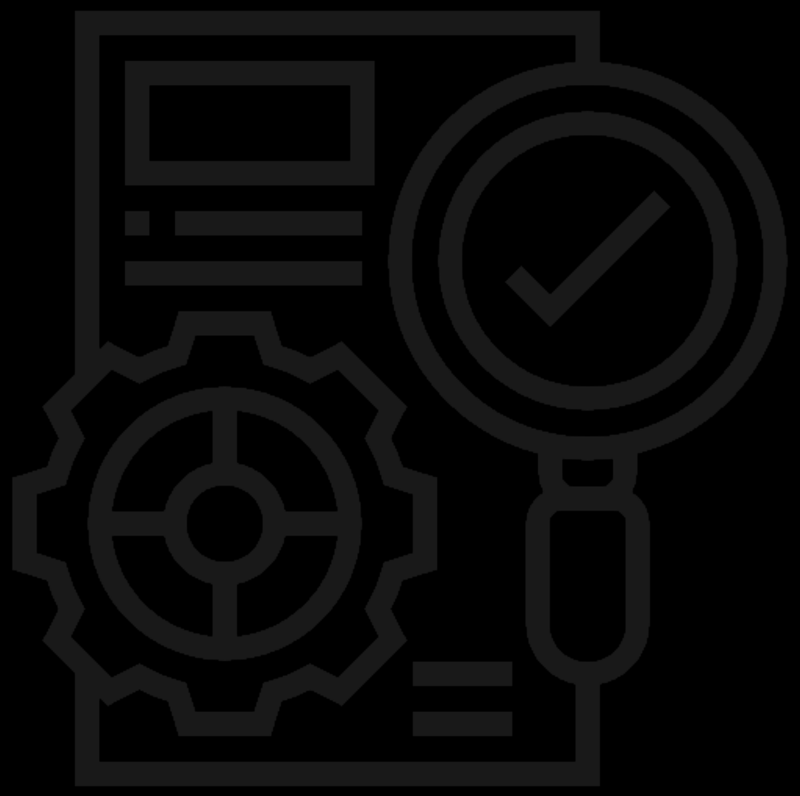
Yury Gorishniy[*†‡] **Ivan Rubachev**[†♣] **Valentin Khrulkov**[†] **Artem Babenko**[†♣]

† Yandex, Russia
‡ Moscow Institute of Physics and Technology, Russia
♣ National Research University Higher School of Economics, Russia

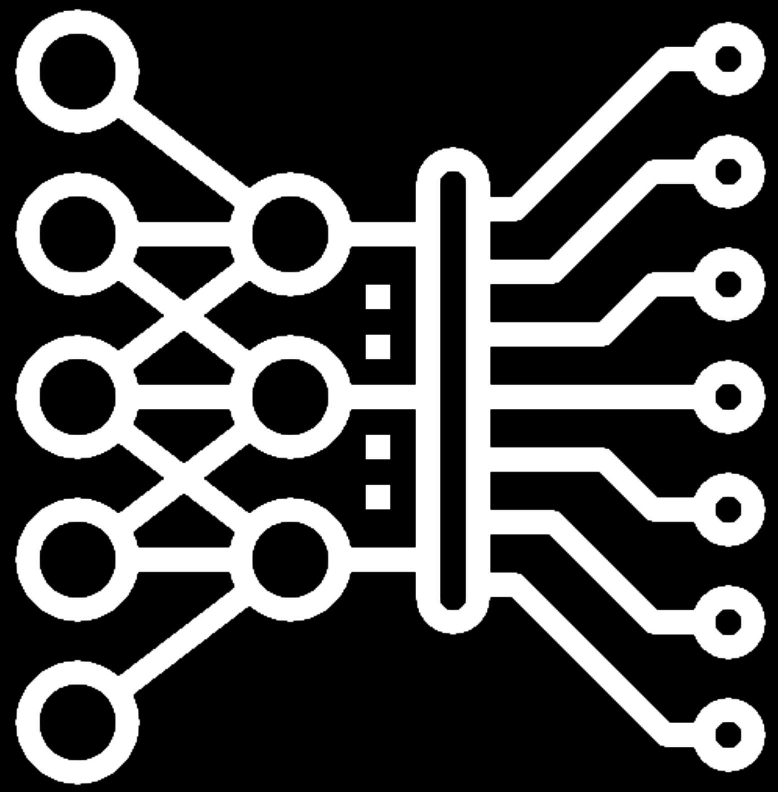
Proposed Flow



Data collection

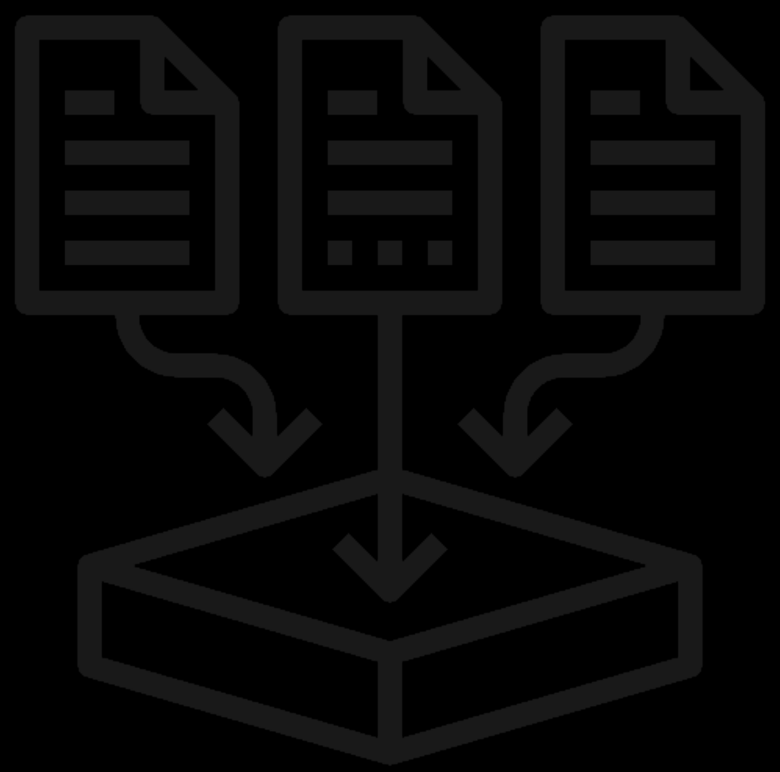


Data validation



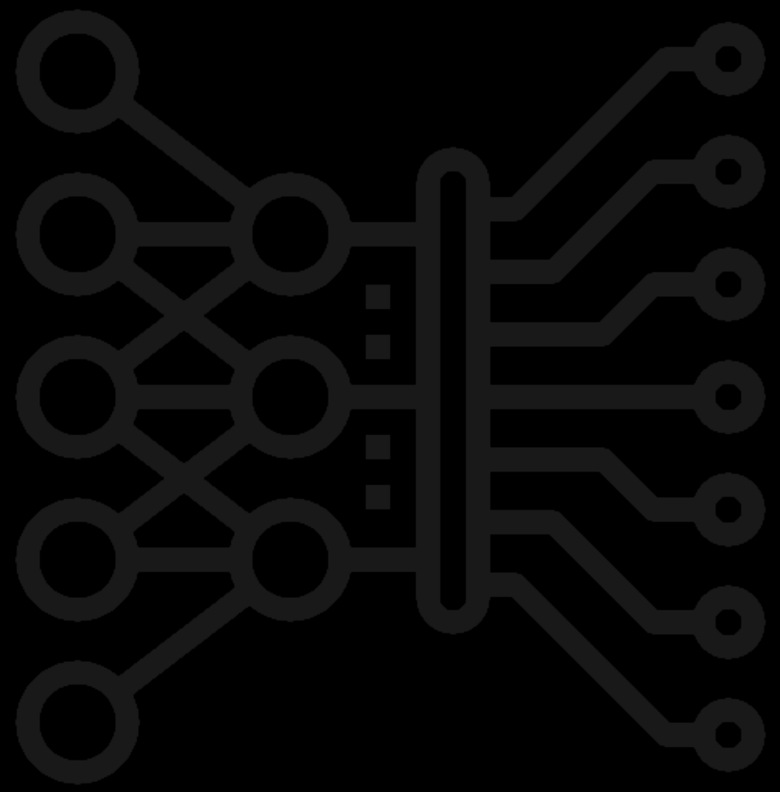
Modeling

Challenges & Future Scope



Data collection

Data validation



Modeling

Challenges & Future Scope

Remove review intersection dependency

Scale crawling setup (potentially via VMs)

Collect data from platforms apart from LinkedIn

Data collection

Data validation

Modeling

Challenges & Future Scope

Remove review intersection dependency

Scale crawling setup (potentially via VMs)

Collect data from platforms apart from LinkedIn

Data collection

Heuristics for edge cases e.g. Chennai vs Madras

Add other modalities of validation
- Language
- Images
- Network behaviour

Data validation

Modeling

Challenges & Future Scope

Remove review intersection dependency

Scale crawling setup (potentially via VMs)

Collect data from platforms apart from LinkedIn

Heuristics for edge cases e.g. Chennai vs Madras

Add other modalities of validation
- Language
- Images
- Network behaviour

More features

Finetuned language features

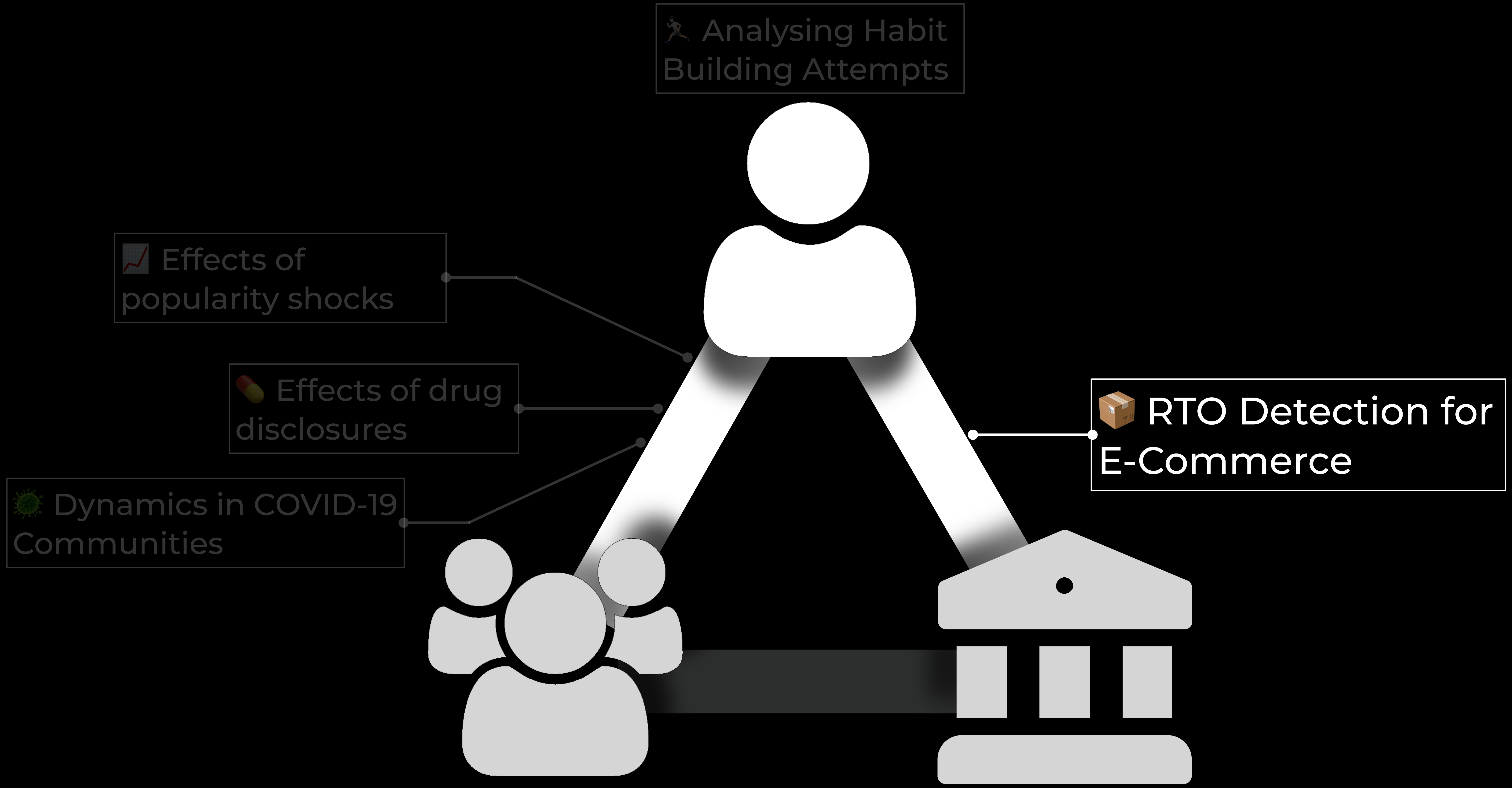
Nonlinear dimensionality reduction (UMAP)

Categorical embeddings

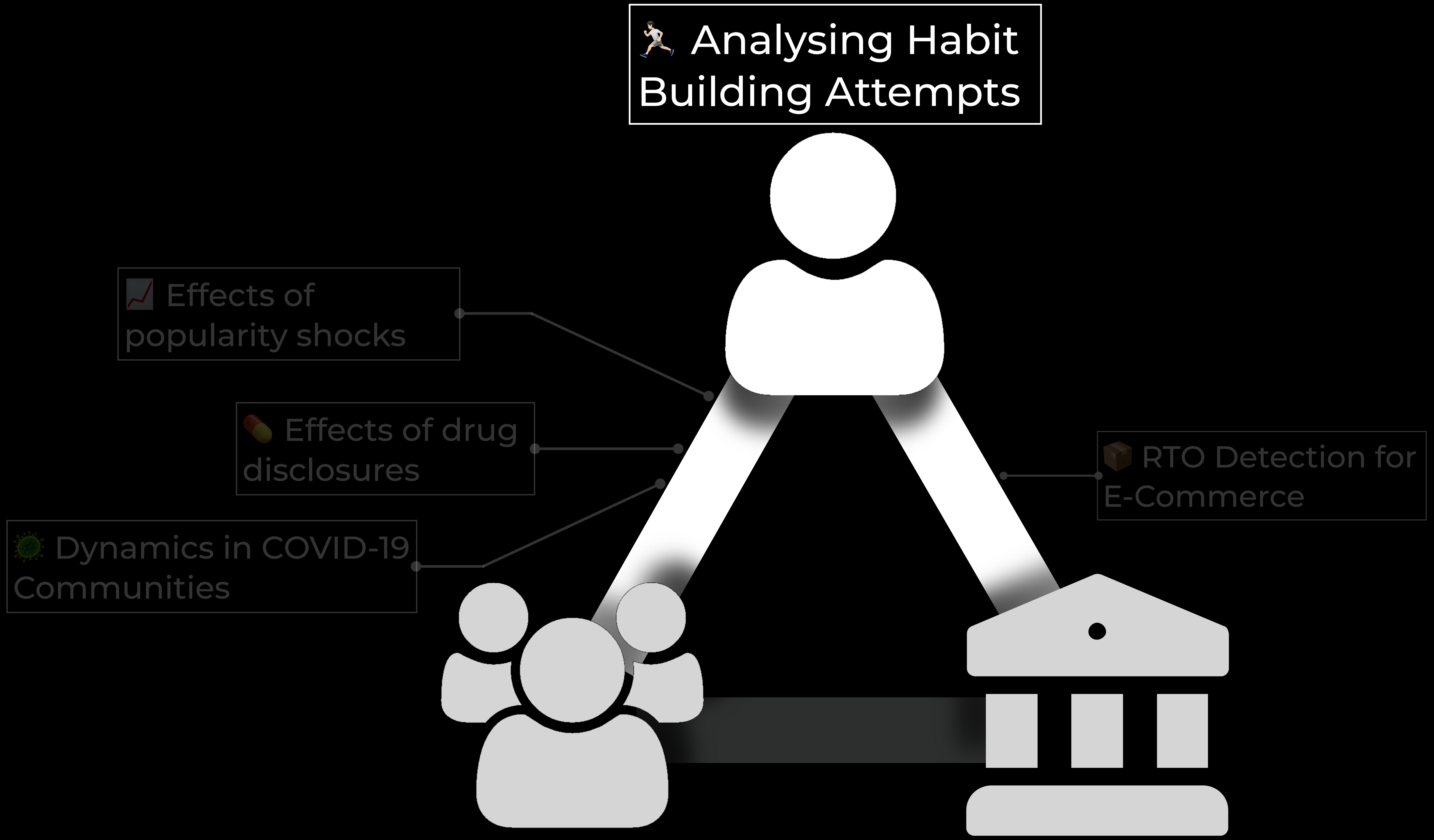
Pre-trained tasks and vectors

Transfer learning to other tasks

Our Focus


🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

Our Focus



🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

Put Your Money Where Your Mouth Is: Dataset and Analysis of Real World Habit Building Attempts

ICWSM' 24 (Dataset)

Hitkul, Rajiv Ratn Shah, Ponnurangam Kumaraguru

Rise of Self-Help

Rise of Self-Help

Self-Help titles have

3X

Between 2013 to 2019

Rise of Self-Help

Self-Help titles have

3X

Between 2013 to 2019

r/loseit

r/stopsmoking

habitica

HabitShare
Habit tracking just went social.

stickK

Strides

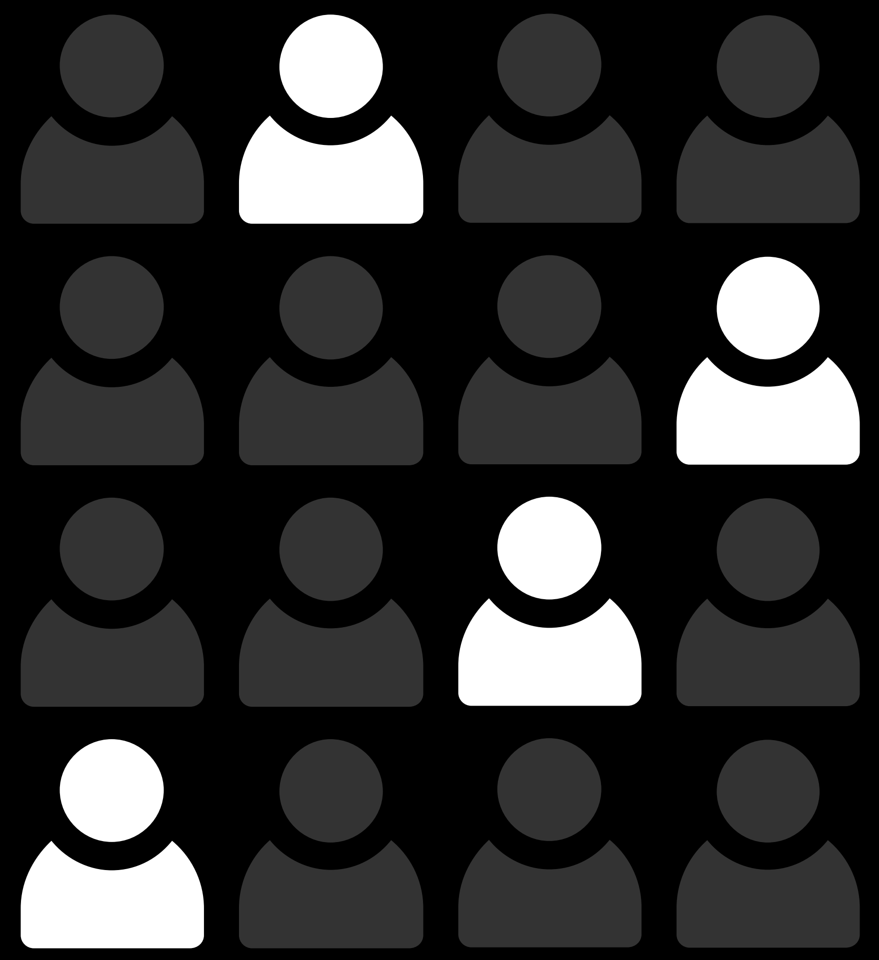
Related Work

Related Work

Homogeneous

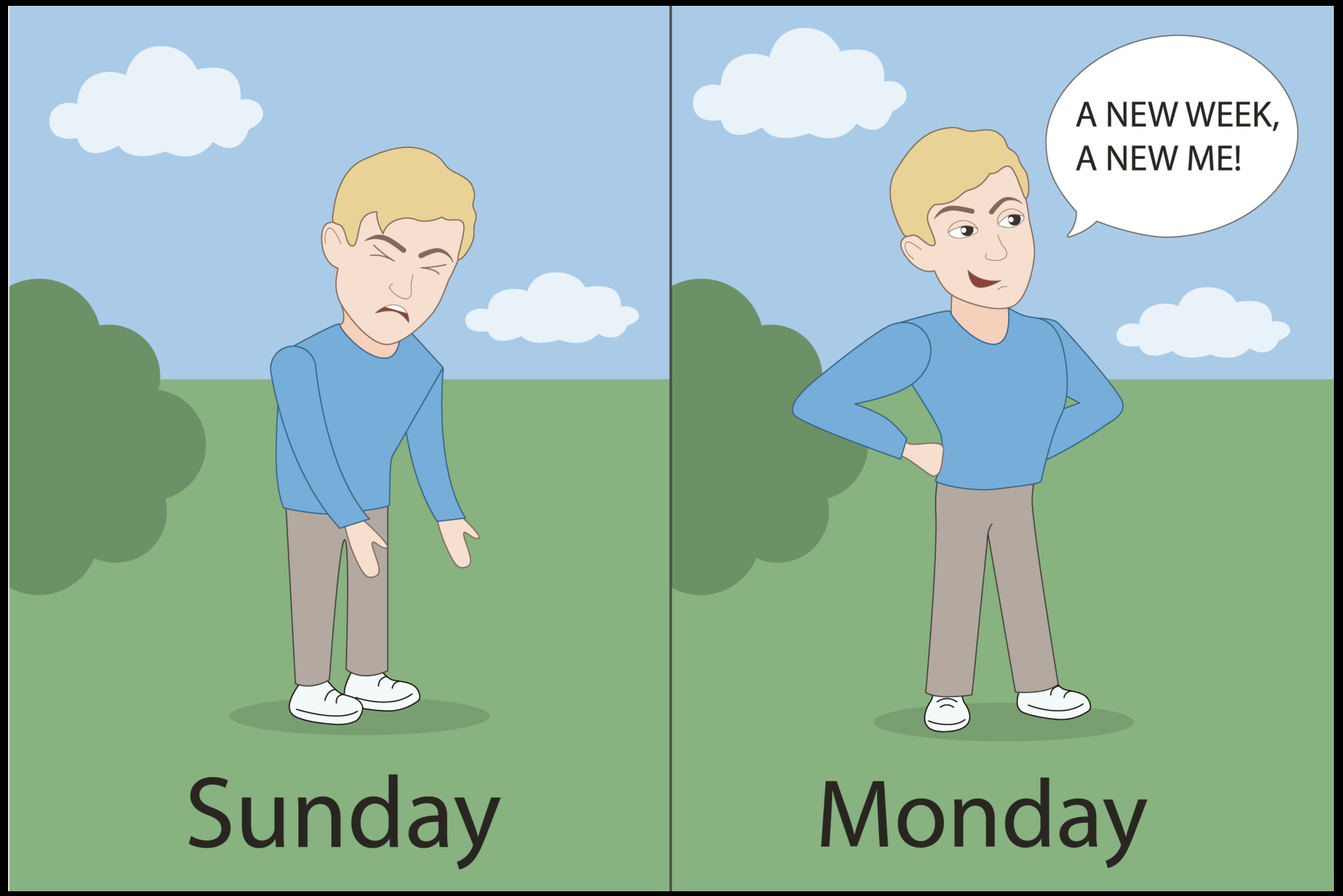
Related Work

Homogeneous



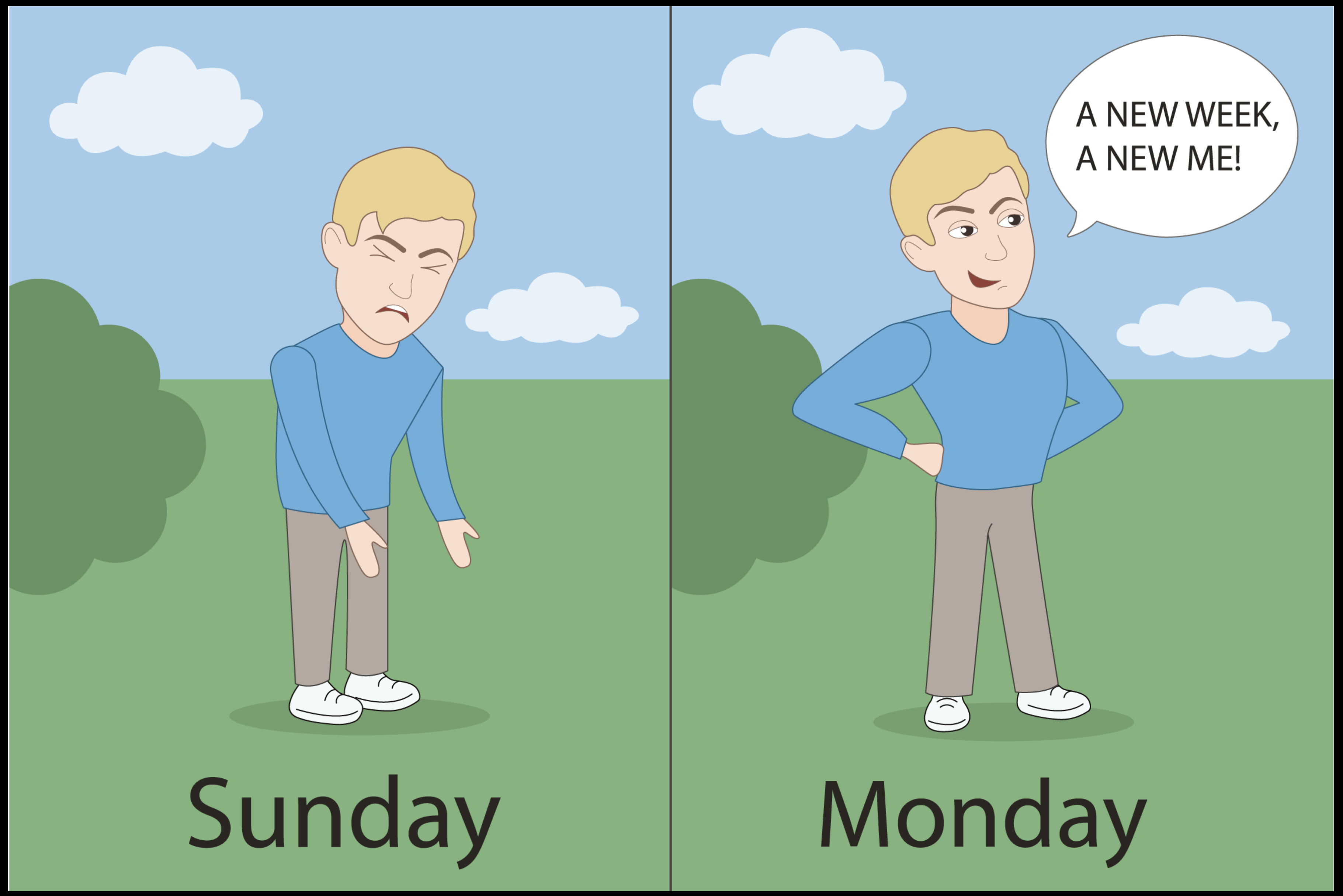
Small Sample

Research Questions



Fresh Start Effect
(Dai et al. 2014)

Research Questions



Fresh Start Effect
(Dai et al. 2014)

What is the prevalence and success rate of commitments started on such **dates**?

Research Questions

Accountability
(Rubin 2014)
(Fogg 2019)

Research Questions

Accountability
(Rubin 2014)
(Fogg 2019)

What is the extent of **accountability** methods and their effect on success rate?

Research Questions

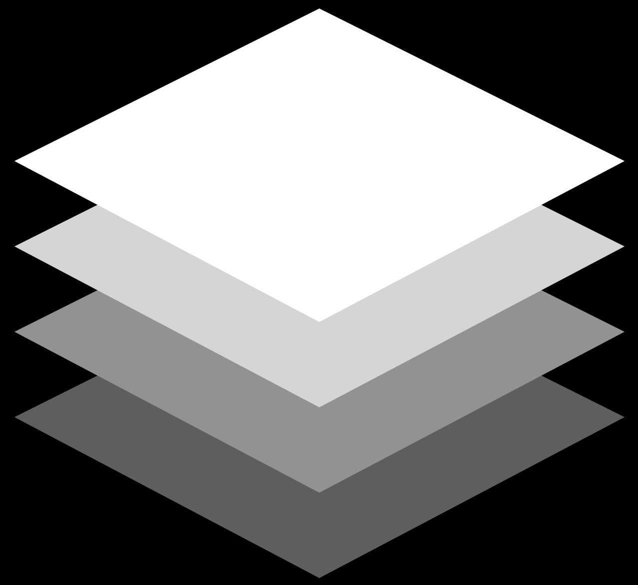
Accountability
(Kahneman 1977)

Research Questions

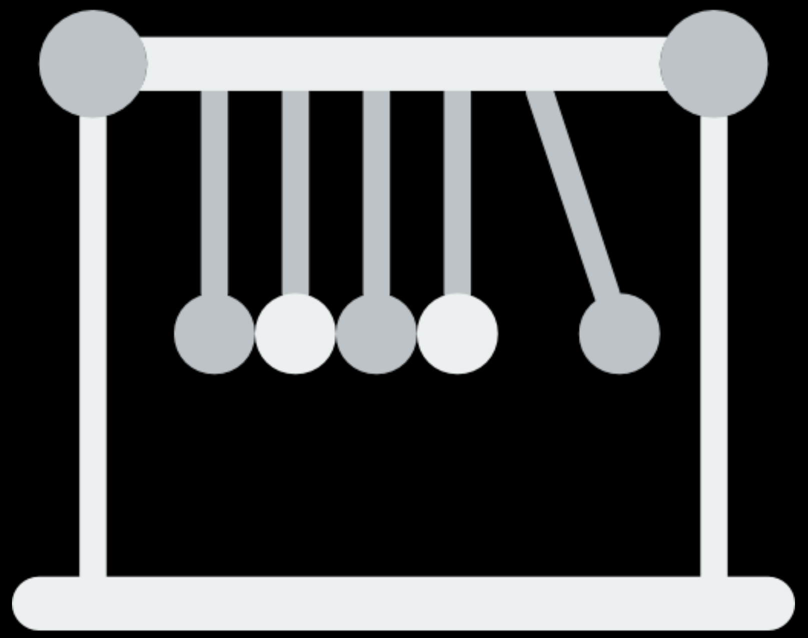
Accountability
(Kahneman 1977)

Does **monetary stake** affect the probability of success in a commitment?

Research Questions

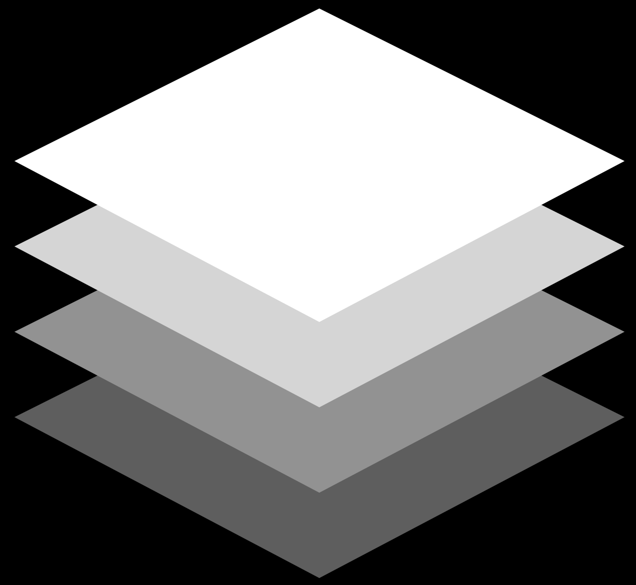


Habit Stacking
(Fogg 2019)

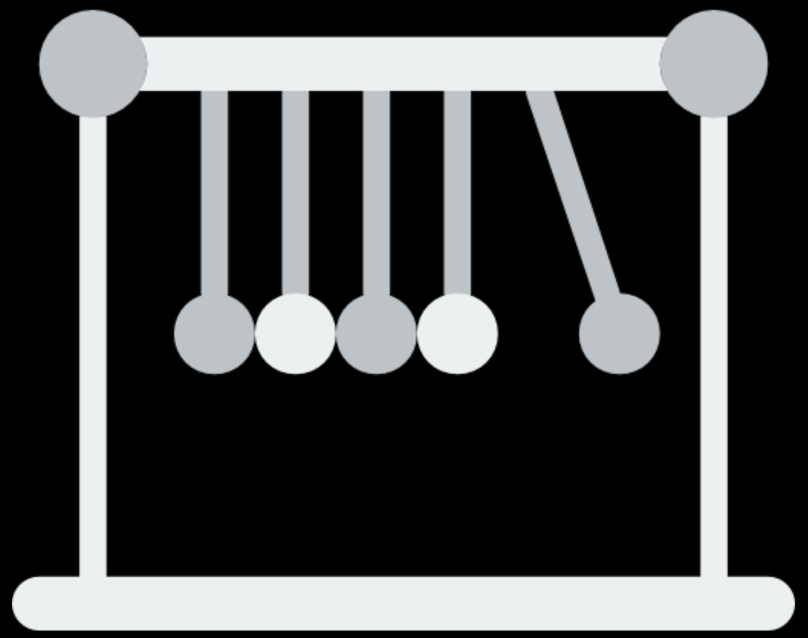


Behavioural Momentum Theory
(Dai et al. 2011)

Research Questions



Habit Stacking
(Fogg 2019)



Behavioural Momentum Theory
(Dai et al. 2011)

What is the extent of users trying to pursue **simultaneous** goals? How does it affect the success rate?

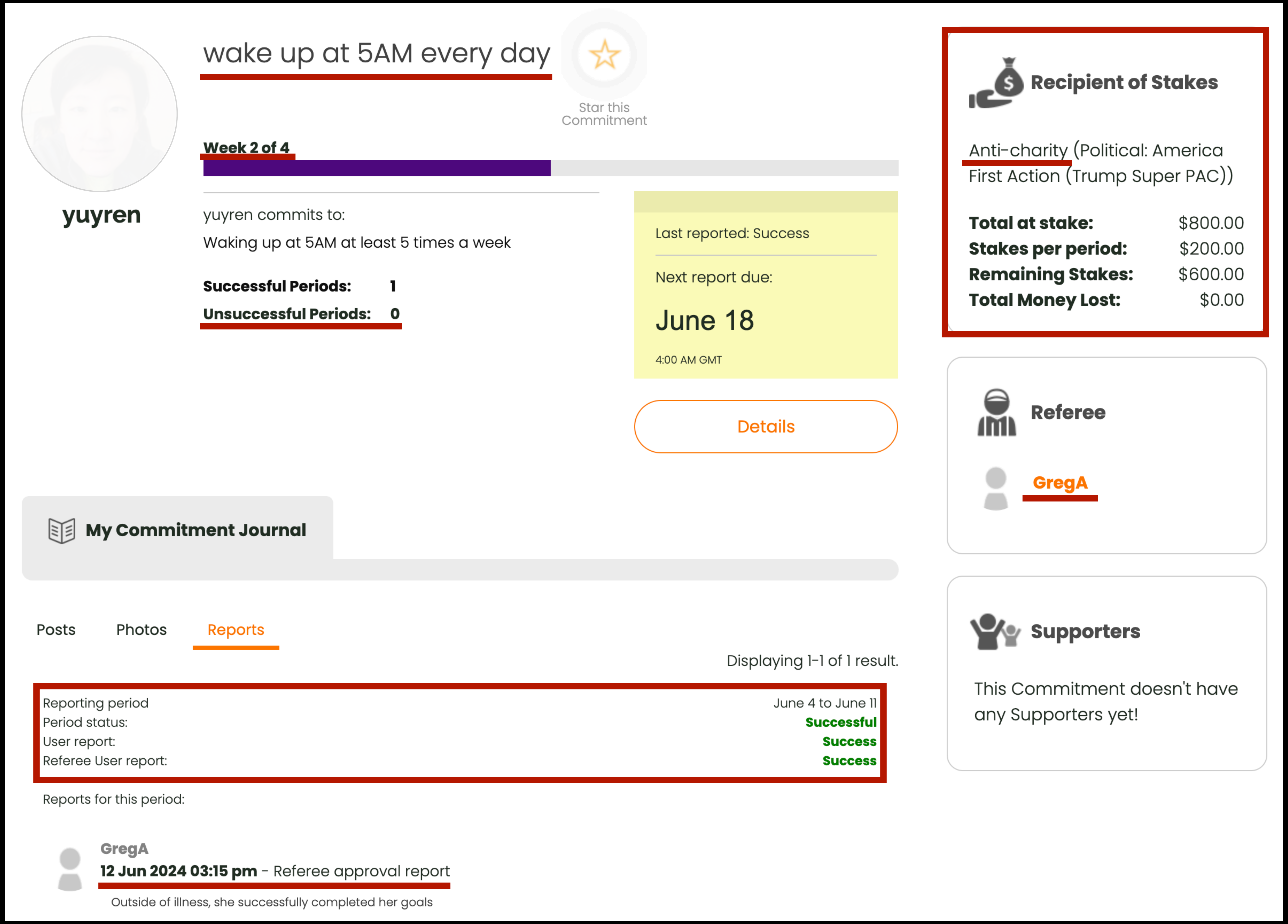
Research Questions

- What is the prevalence and success rate of commitments started on such **dates**?
- What is the extent of **accountability** methods and their effect on success rate?
- Does **monetary stake** affect the probability of success in a commitment?
- What is the extent of users trying to pursue **simultaneous** goals? How does it affect the success rate?

Data

Data

stickK


wake up at 5AM every day
Star this Commitment
yuyren
Week 2 of 4
yuyren commits to:
Waking up at 5AM at least 5 times a week
Successful Periods: 1
Unsuccessful Periods: 0
Last reported: Success
Next report due:
June 18
4:00 AM GMT
Details
Recipient of Stakes
Anti-charity (Political: America First Action (Trump Super PAC))
Total at stake: $800.00
Stakes per period: $200.00
Remaining Stakes: $600.00
Total Money Lost: $0.00
Referee
GregA
Supporters
This Commitment doesn't have any Supporters yet!
My Commitment Journal
Posts
Photos
Reports
Displaying 1-1 of 1 result.
Reporting period June 4 to June 11
Period status: Successful
User report: Success
Referee User report: Success
Reports for this period:
GregA
12 Jun 2024 03:15 pm - Referee approval report
Outside of illness, she successfully completed her goals

Data

Data

19 October 2007 to 17 August 2023

Data

742,923 Users

19 October 2007 to 17 August 2023

Data

742,923 Users

397,456 Commitments

19 October 2007 to 17 August 2023

Data

742,923 Users

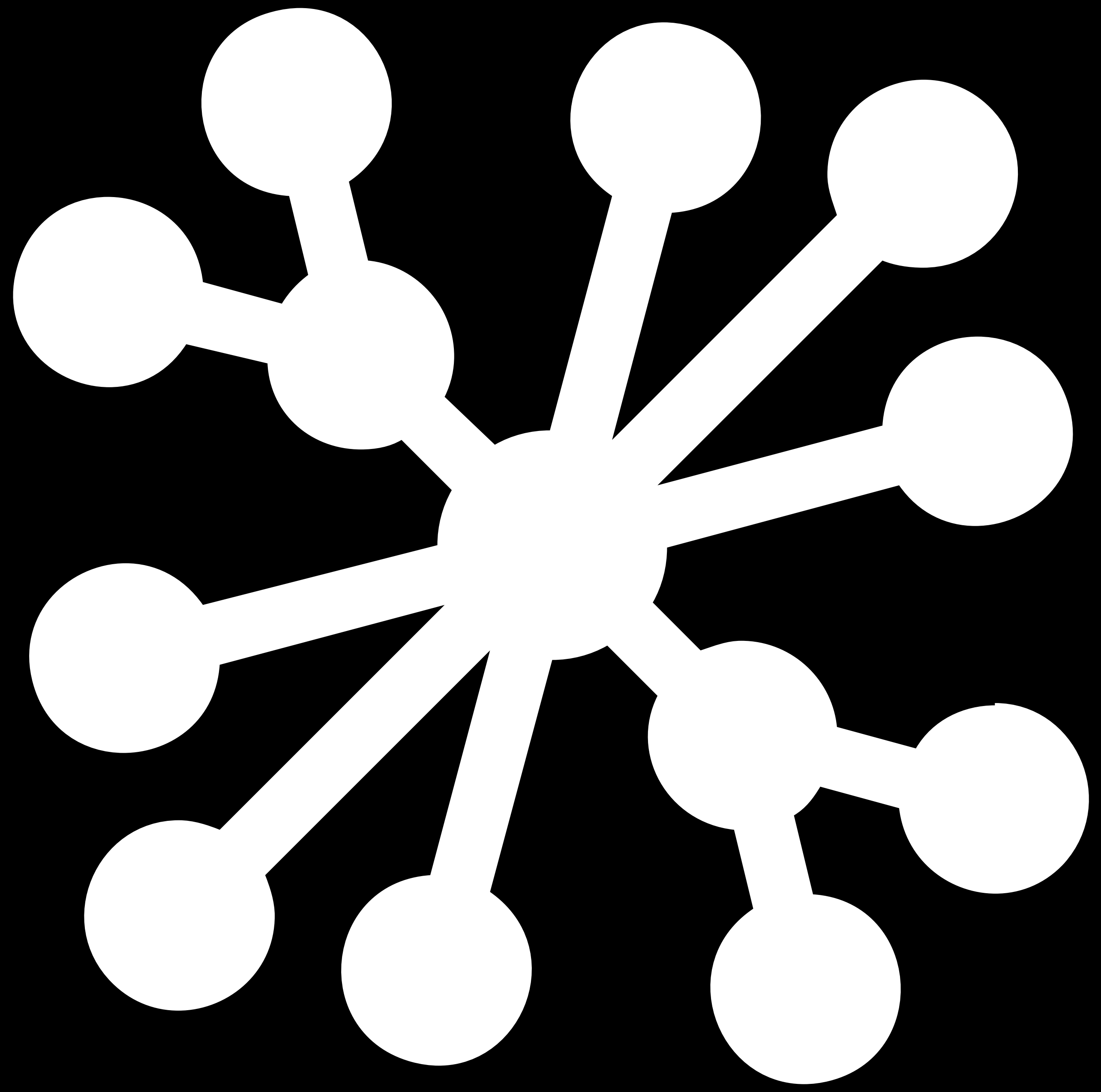
397,456
Commitments

$ 35,598,253
at Stake

19 October 2007 to 17 August 2023

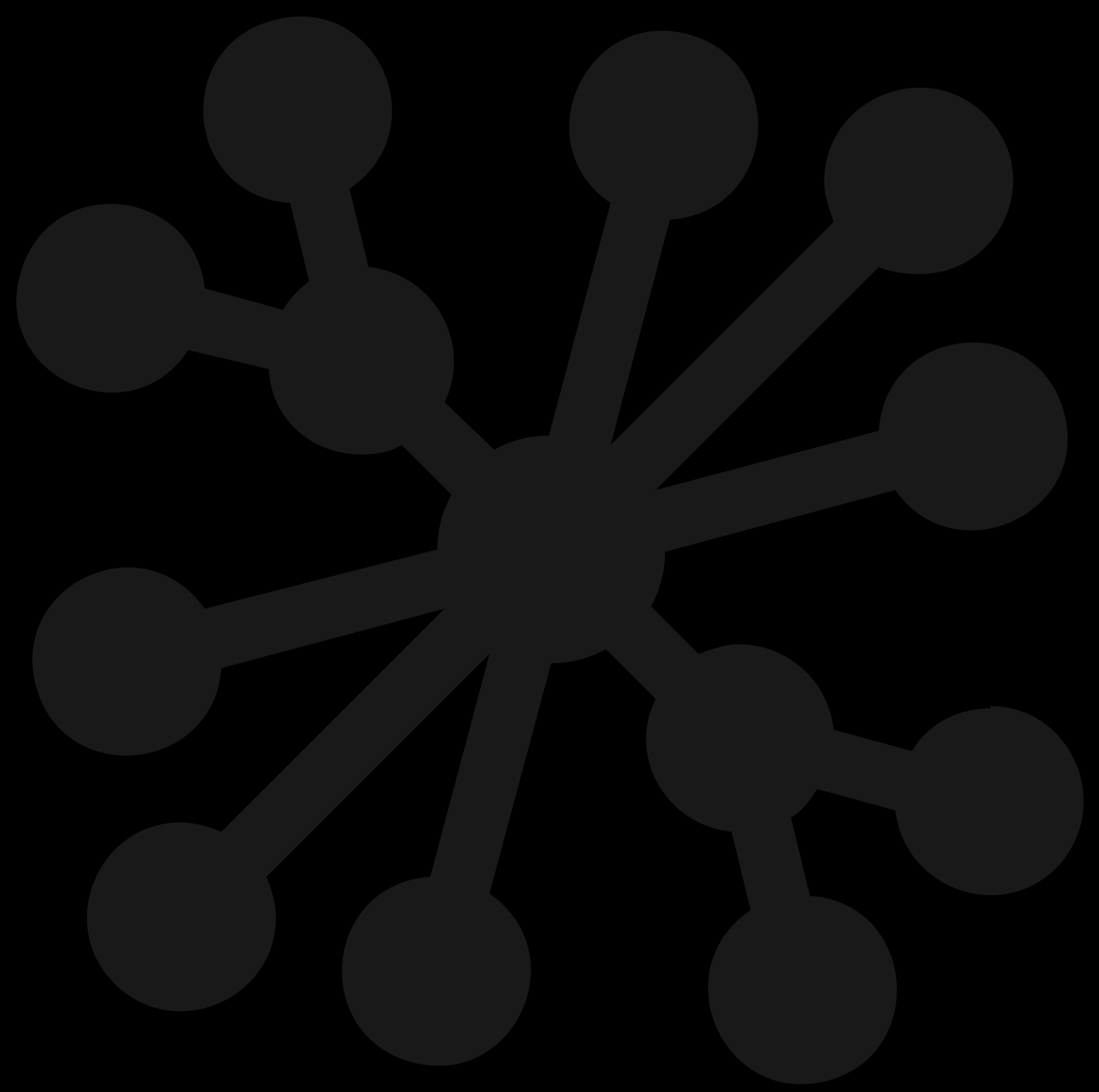
Commitment Classification

Commitment Classification



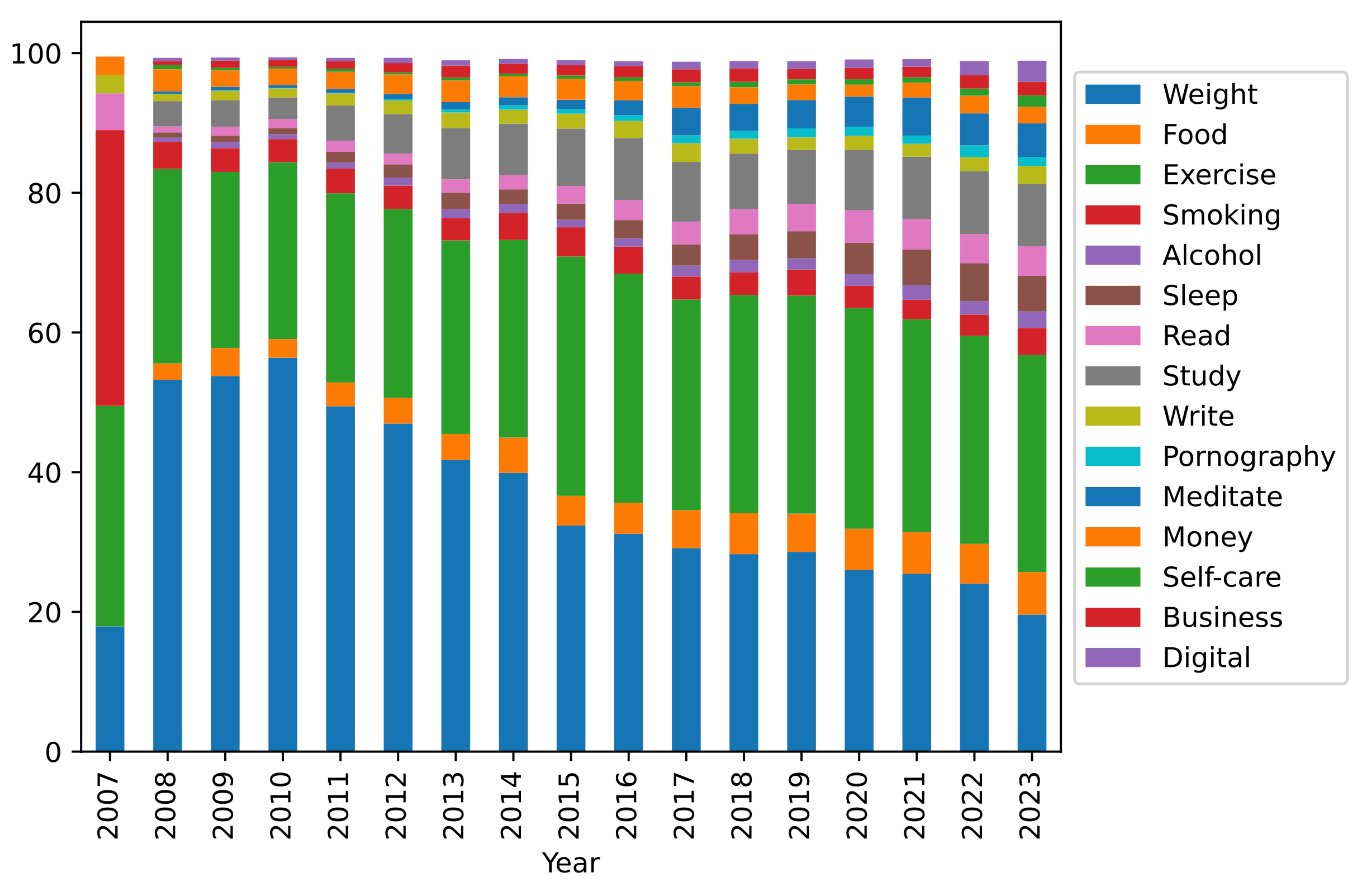
Word2Vec based relevance feedback

Commitment Classification

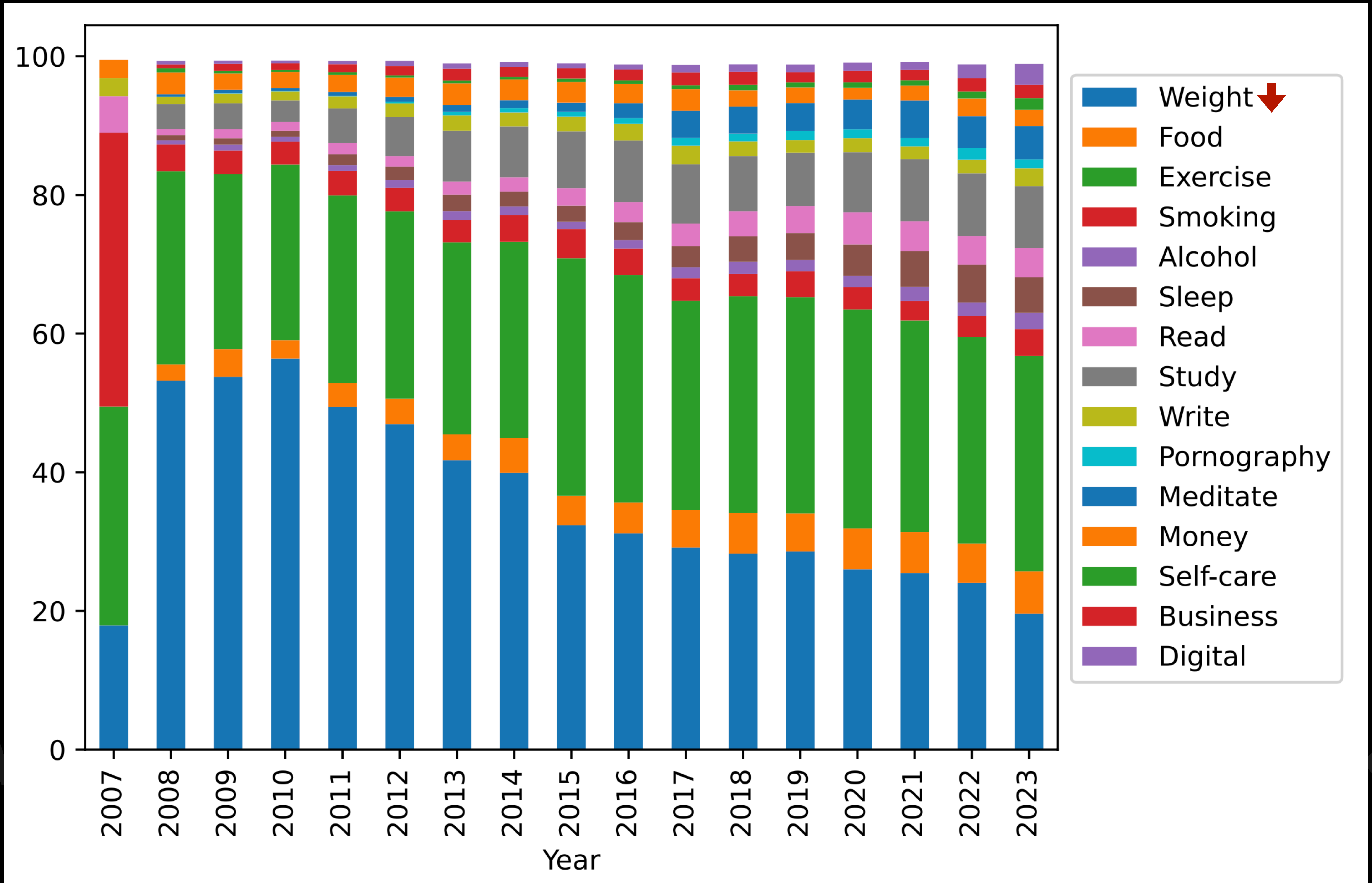


Word2Vec based relevance feedback

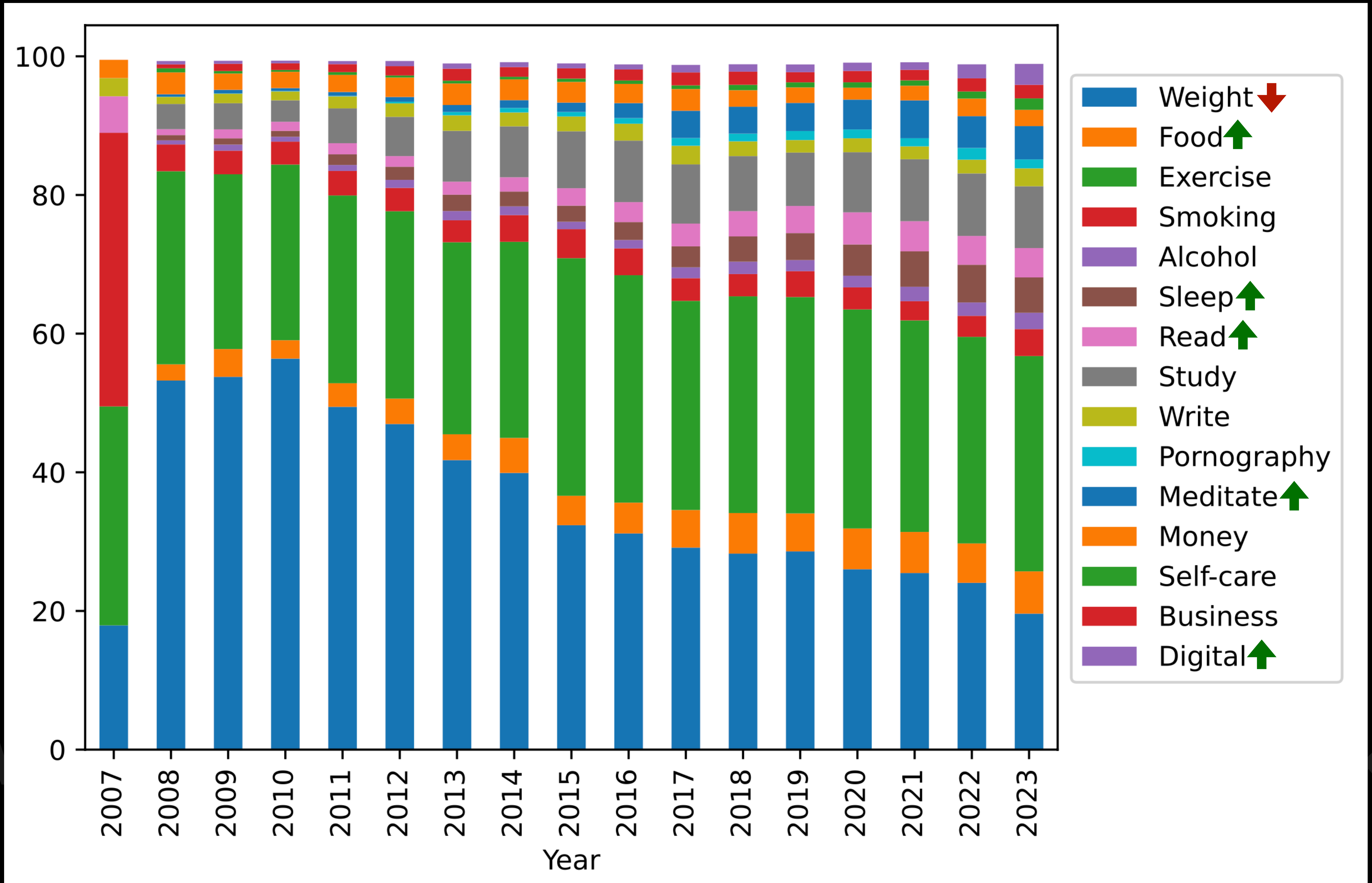
Commitment Classification



Commitment Classification

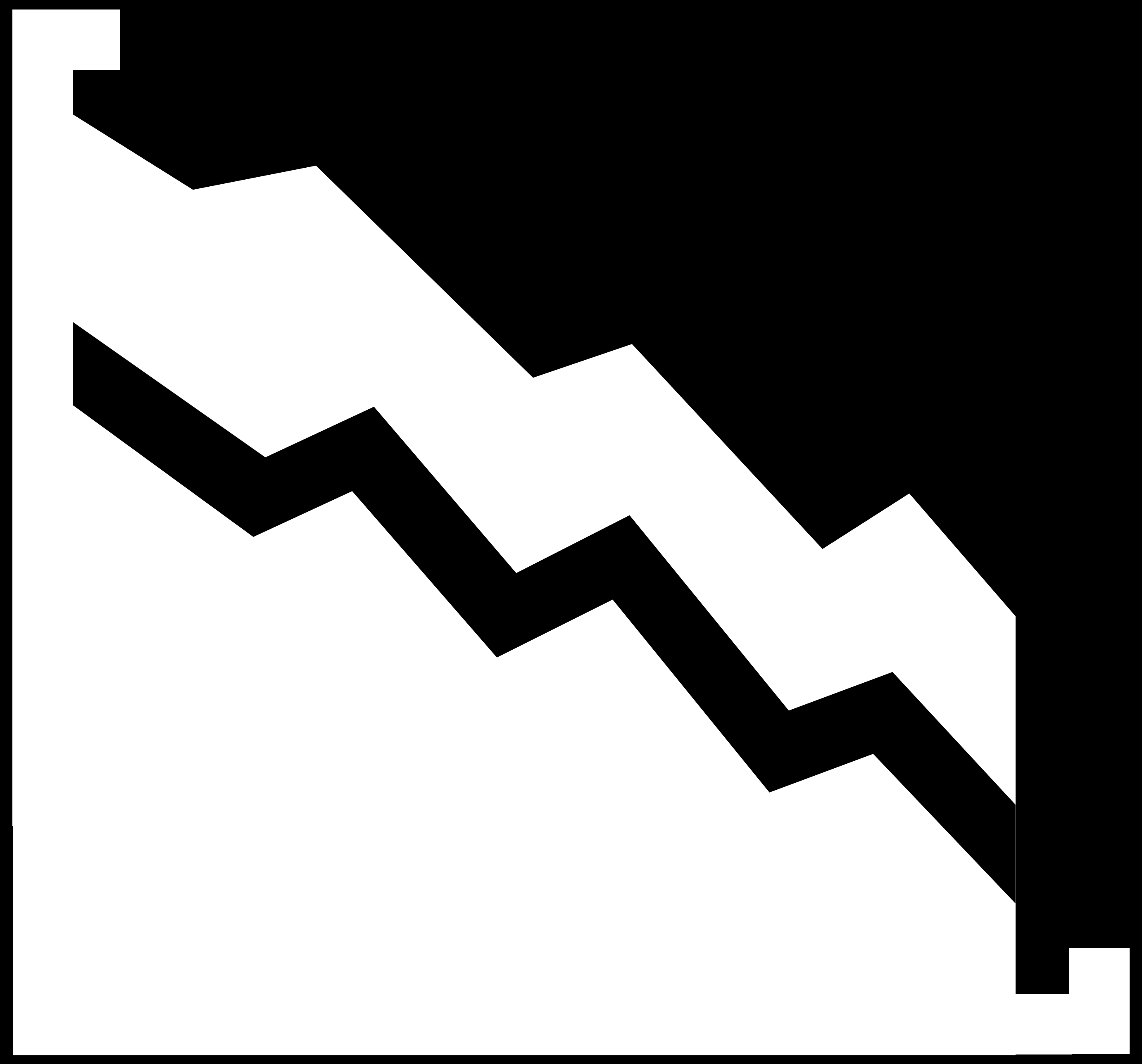


Commitment Classification



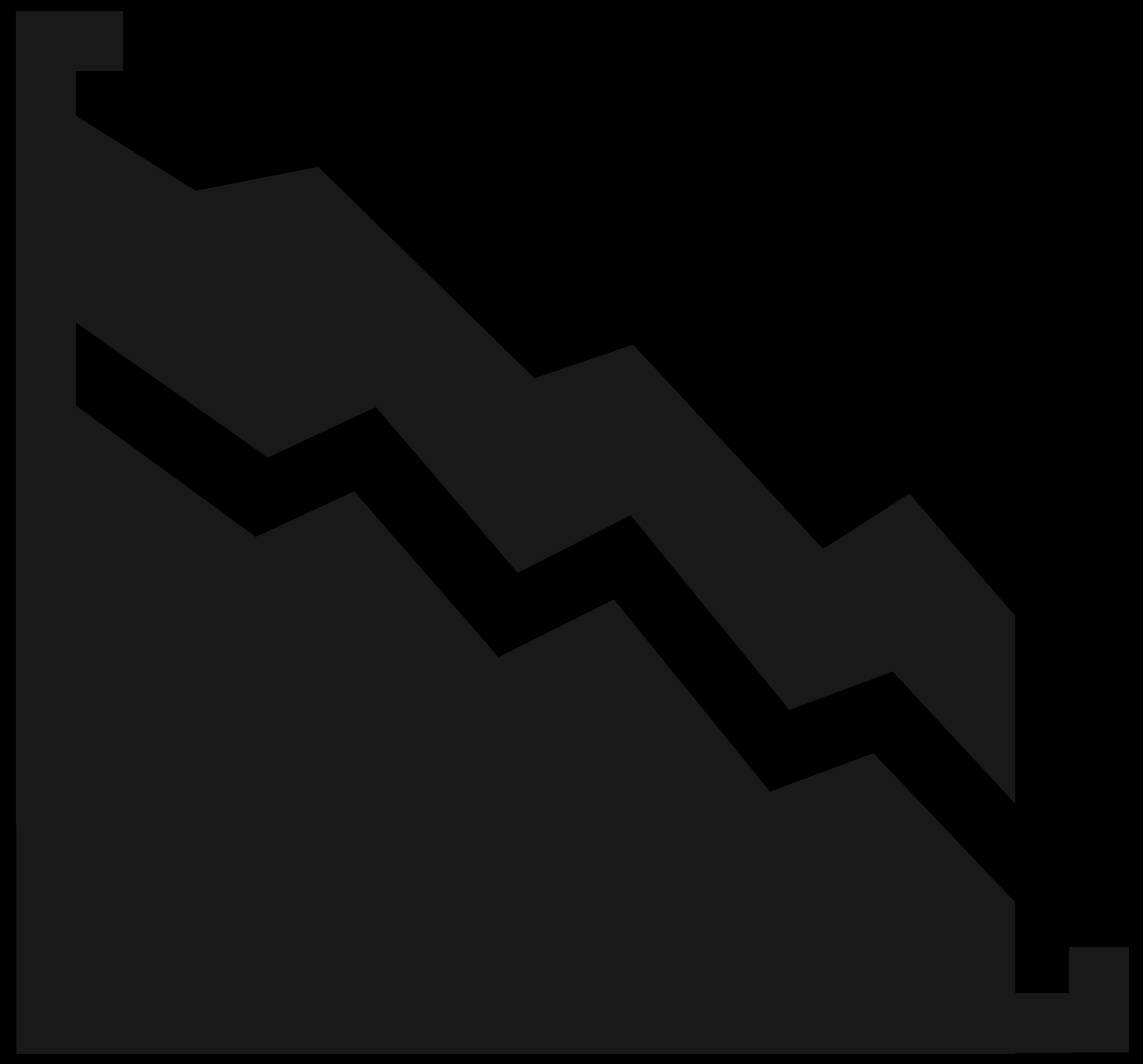
Method

Method



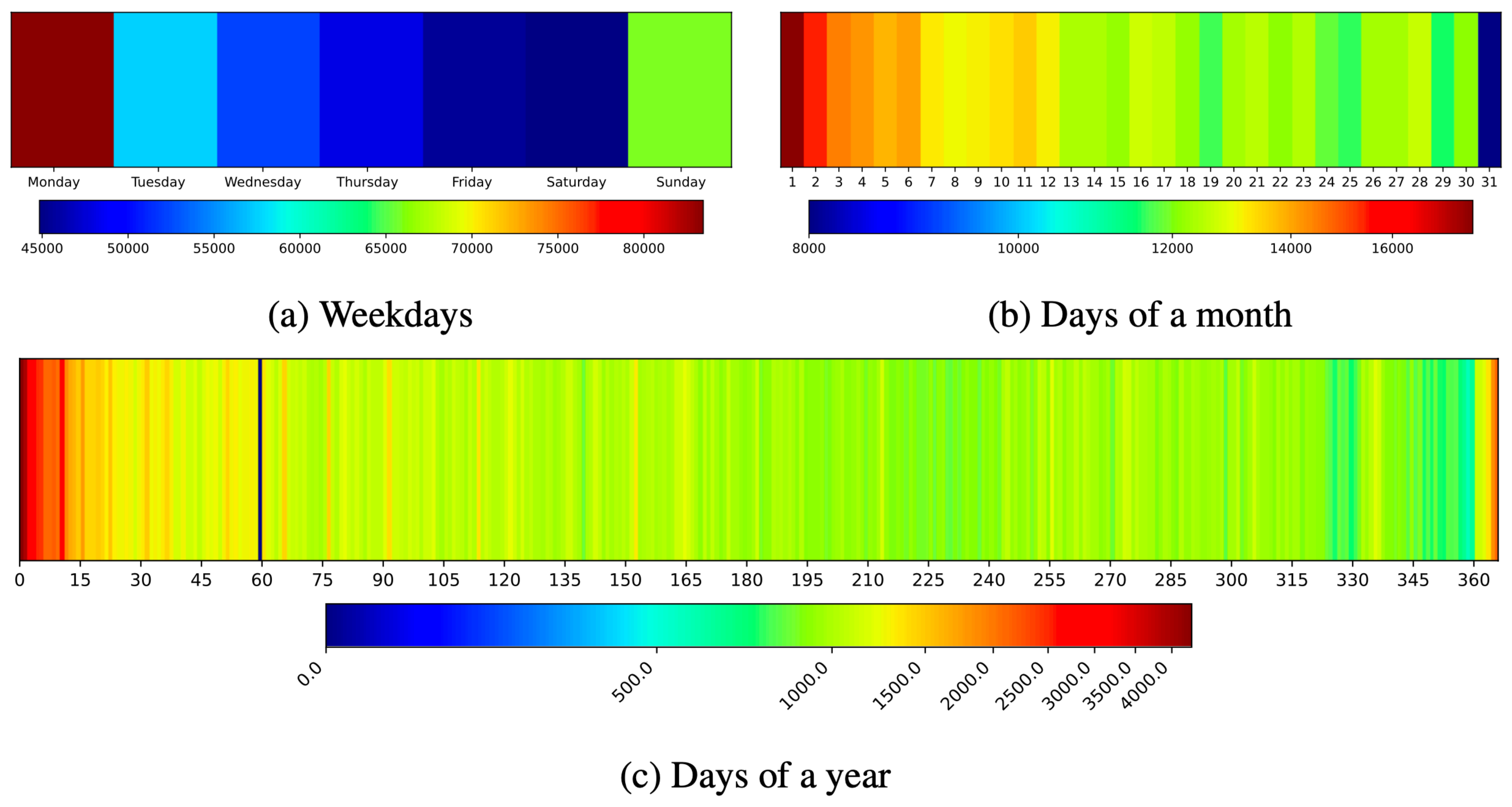
Survival Analysis / CoX Regression

Fresh Start - Extent



Survival Analysis / CoX Regression

Fresh Start - Extent

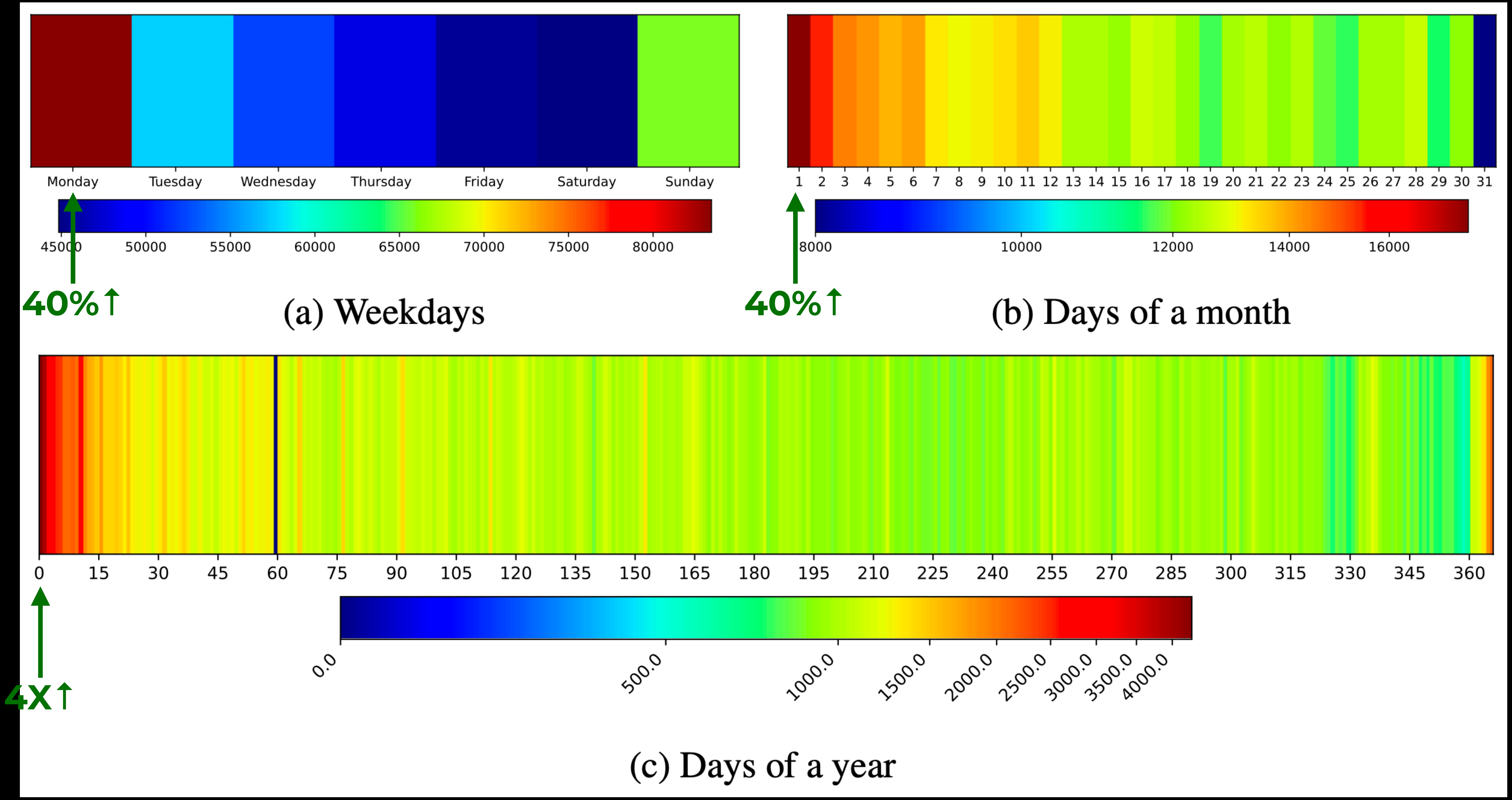


(a) Weekdays

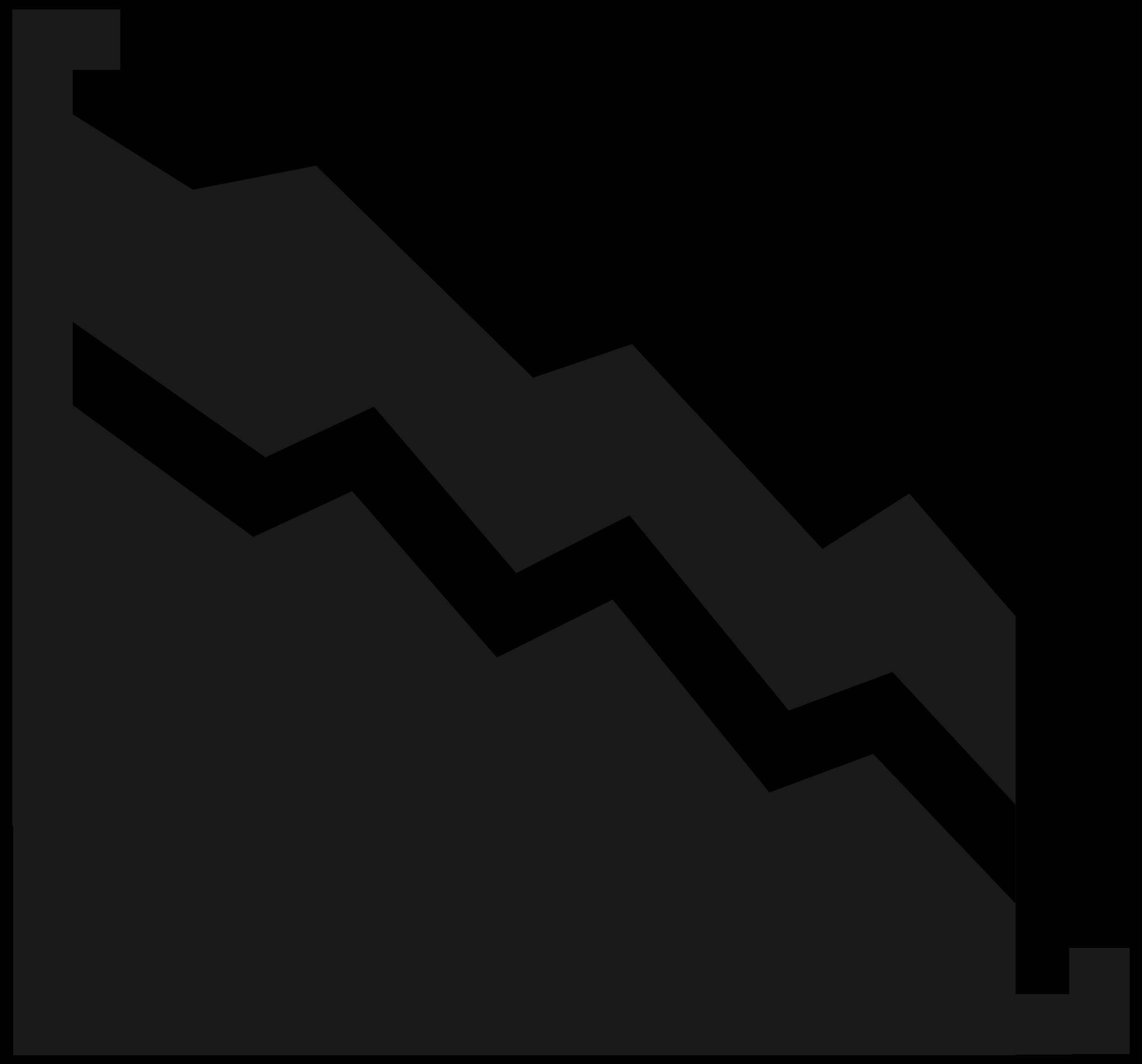
(b) Days of a month

(c) Days of a year

Fresh Start - Extent

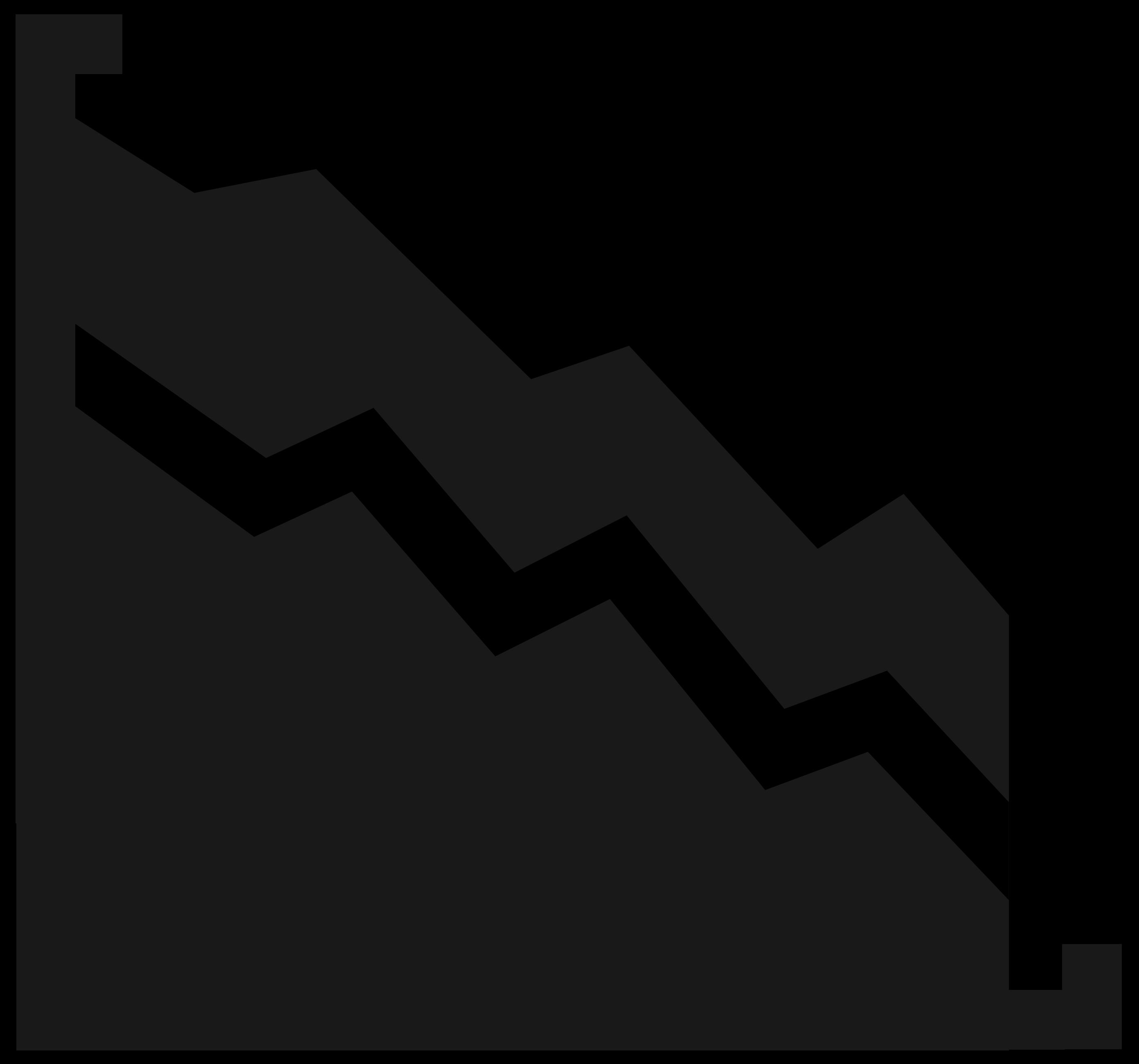


Fresh Start - Extent



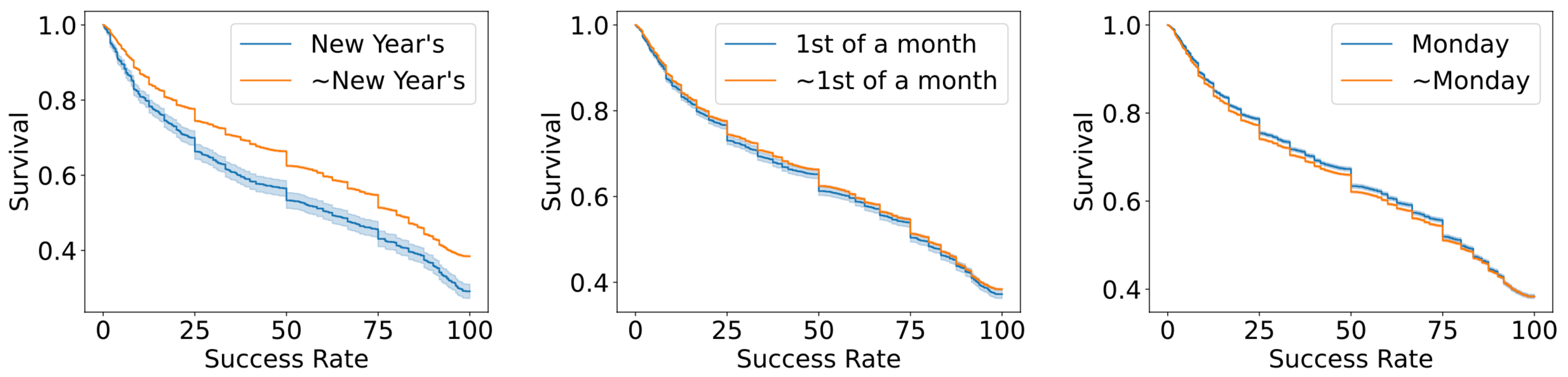
Survival Analysis / CoX Regression

Fresh Start - Extent



Survival Analysis / CoX Regression

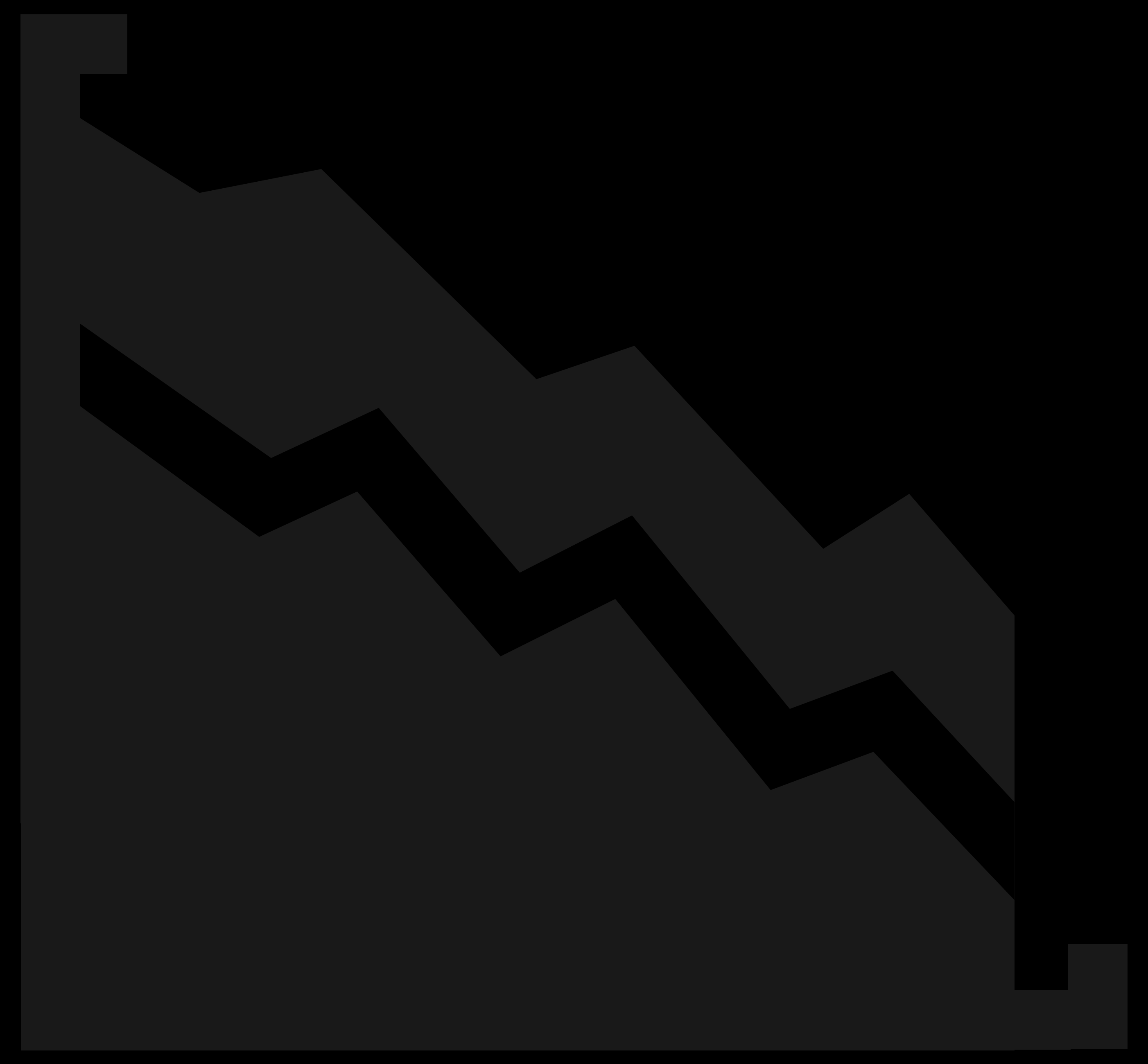
Fresh Start - Effect



(a) New Year. $p < 0.005$ (b) 1^{st} of a month. $p = 0.01$ (c) Monday. $p = 0.03$

Survival Analysis / CoX Regression

Accountability - Extent



Survival Analysis / CoX Regression

Accountability - Extent

19%

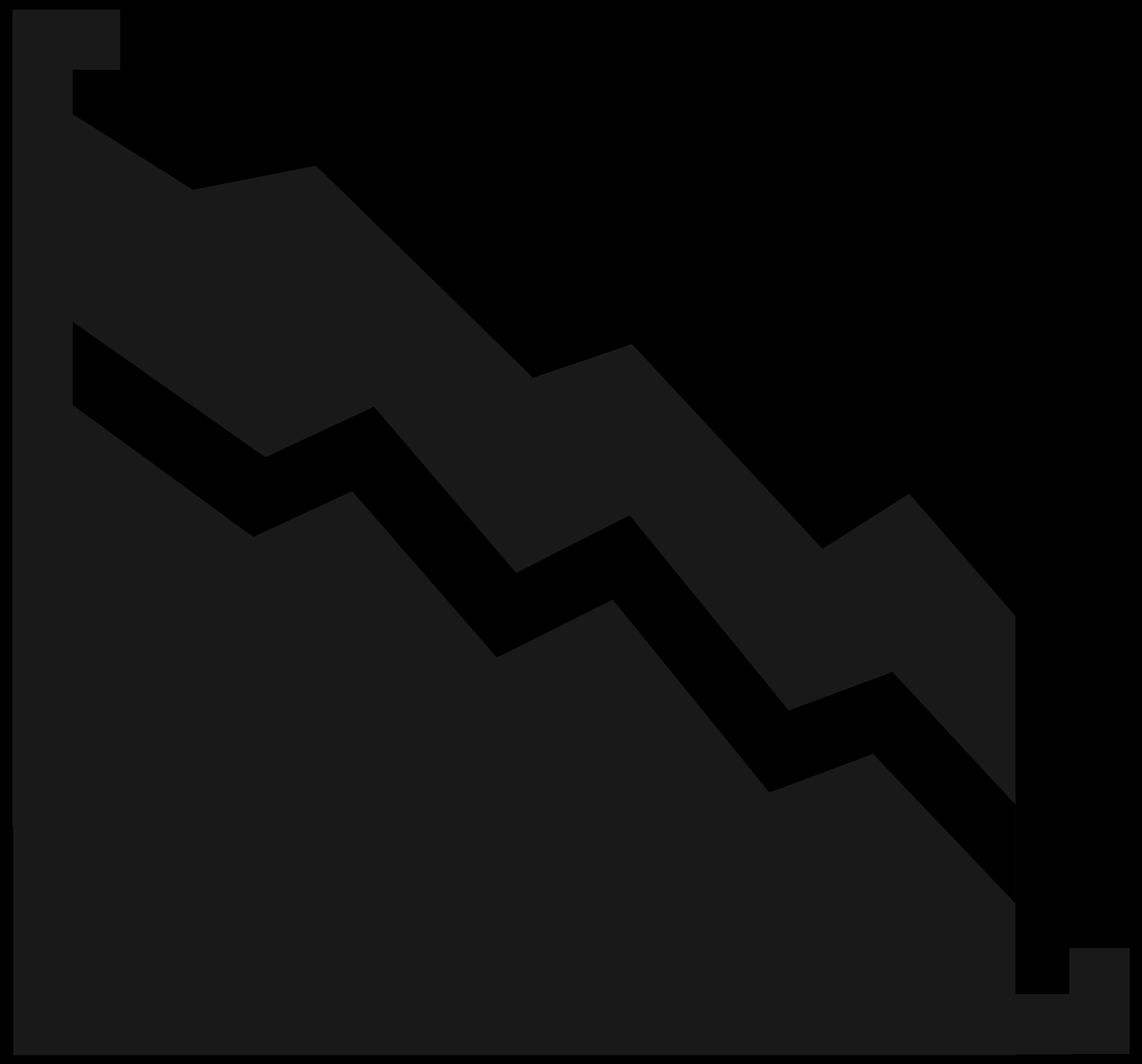
Referee
(external accountability)

8.6%

Supporters
(Social accountability)

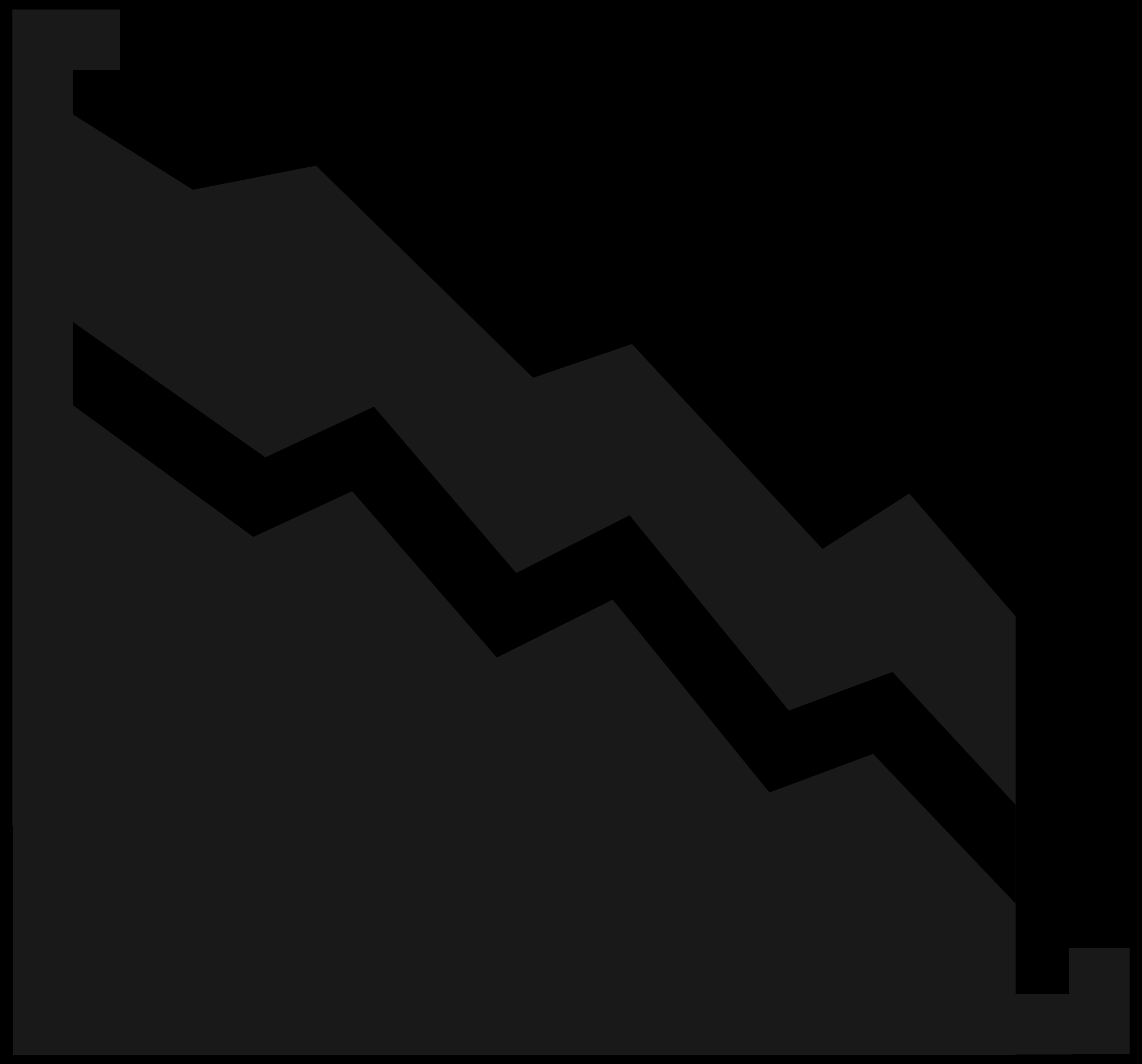
Survival Analysis / CoX Regression

Accountability - Extent



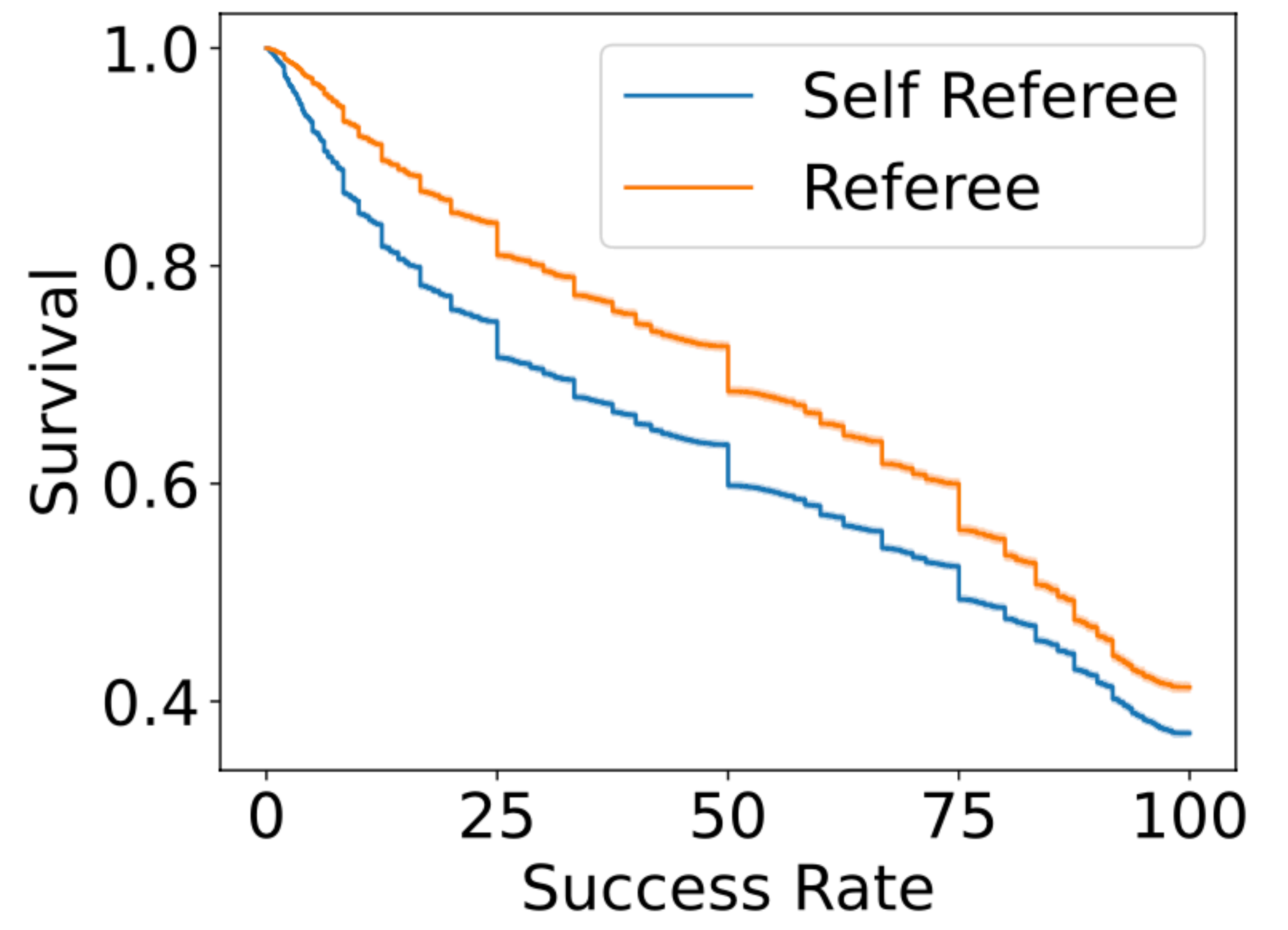
Survival Analysis / CoX Regression

Accountability - Extent



Survival Analysis / CoX Regression

Accountability - Effect



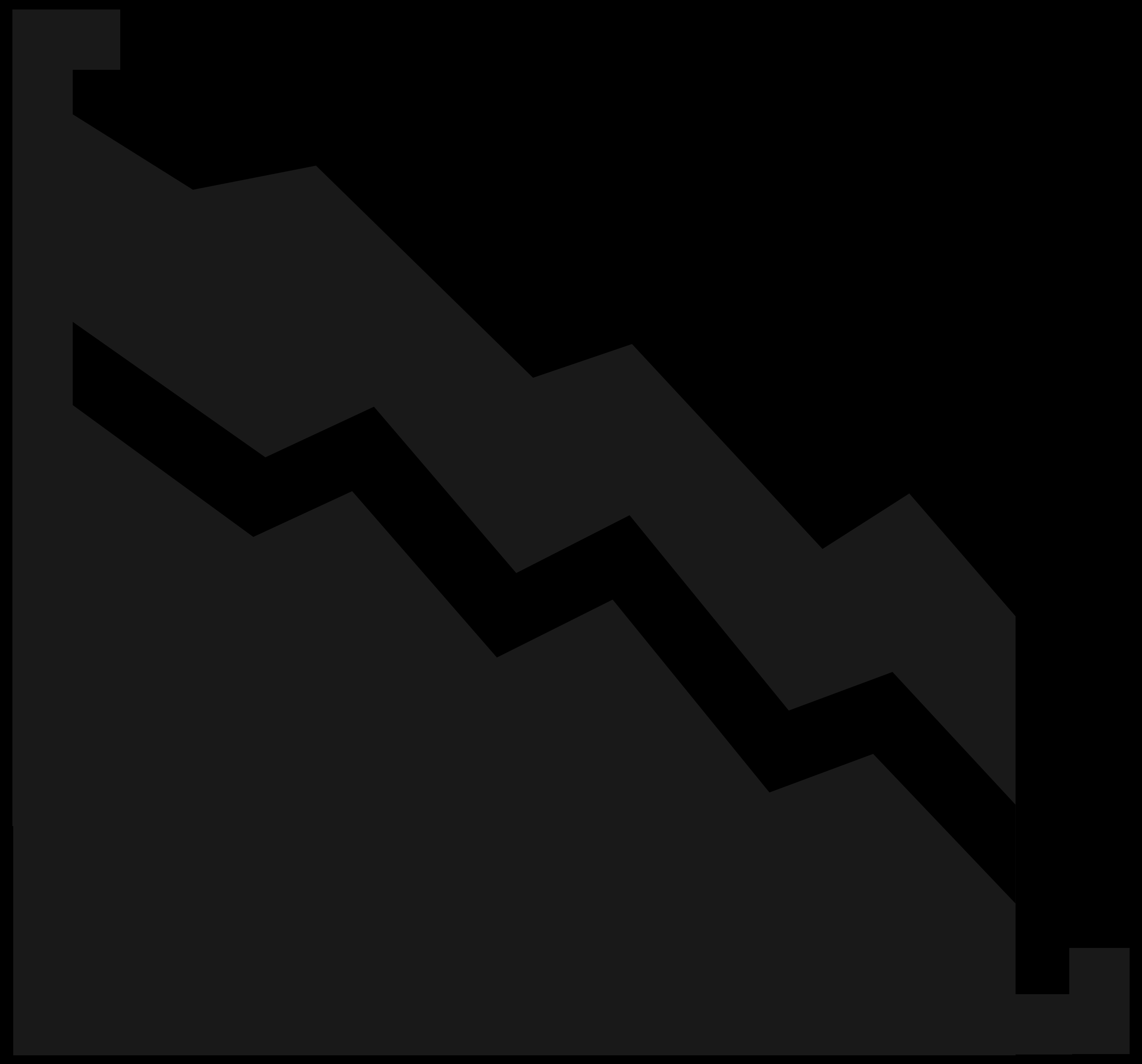
(e) Referee. $p < 0.005$

Covariate	HR	95% CI
# of reports	1.0447***	[1.0437, 1.0457]
Length of commitment	1.0045***	[1.0043, 1.0046]
# of supporters	0.9797***	[0.9725, 0.9870]
$ on stake per period	0.9474***	[0.9467, 0.9482]

$^{***}p <= 0.005$

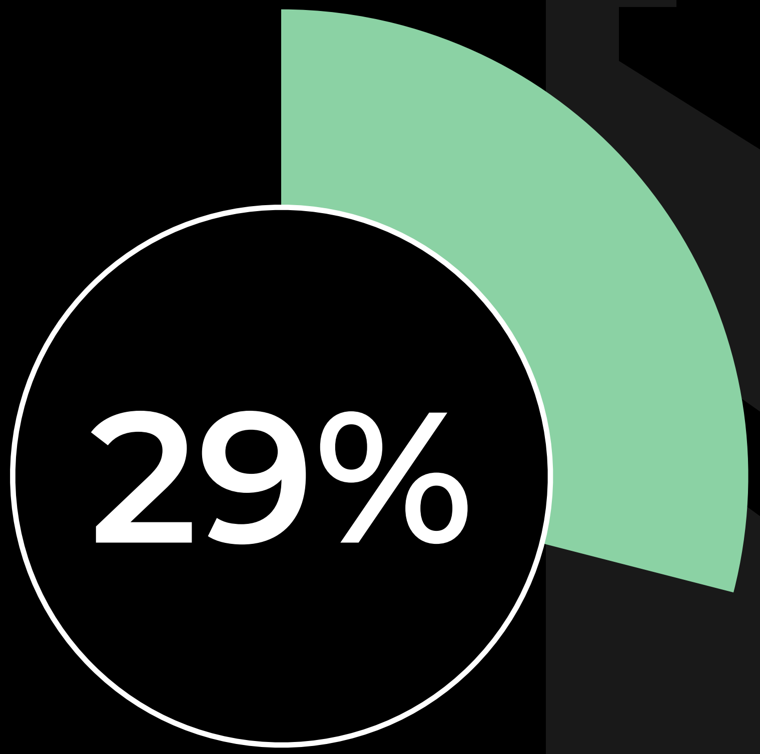
ysis / CoX Regression

Monetary - Extent



Survival Analysis / CoX Regression

Monetary - Extent

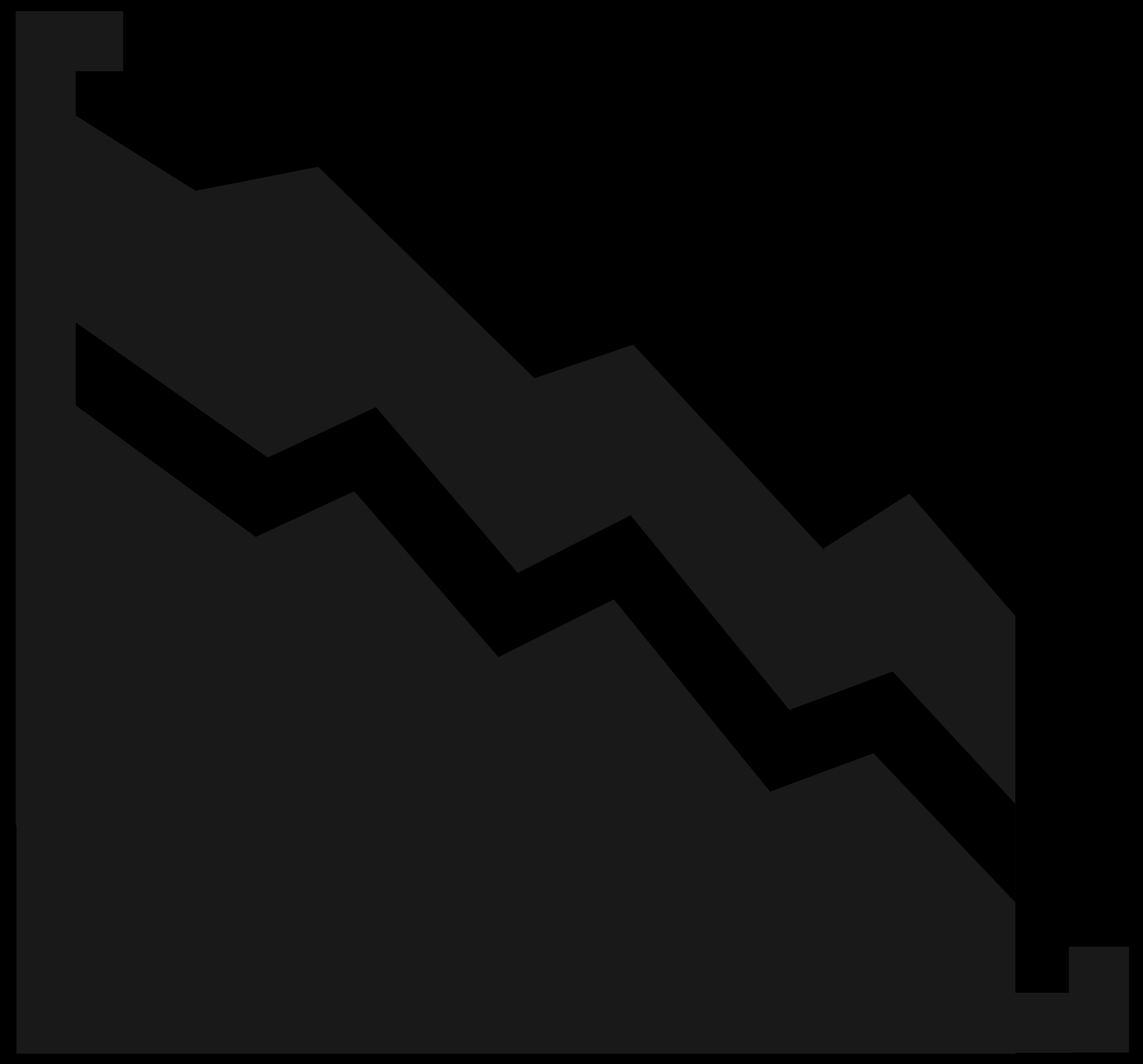


Monetary Stake

Type of stake	Frequency	Percentage
No stake	282,366	71.0%
Anti-charity	60,368	15.2%
Charity	27,053	6.8%
Friend	21,711	5.5%
StickK	5,958	1.5%

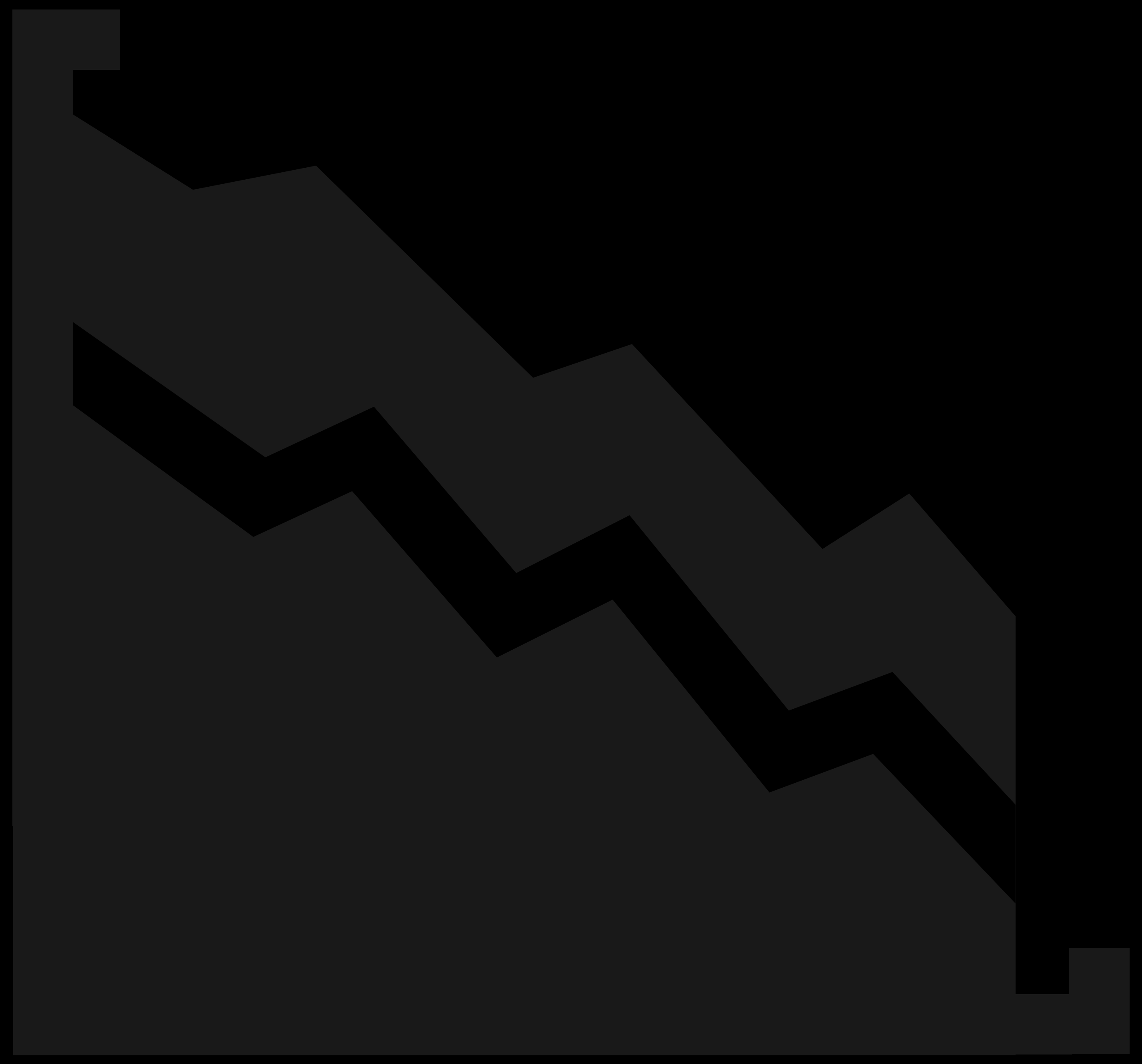
Survival Analysis / CoX Regression

Monetary - Extent



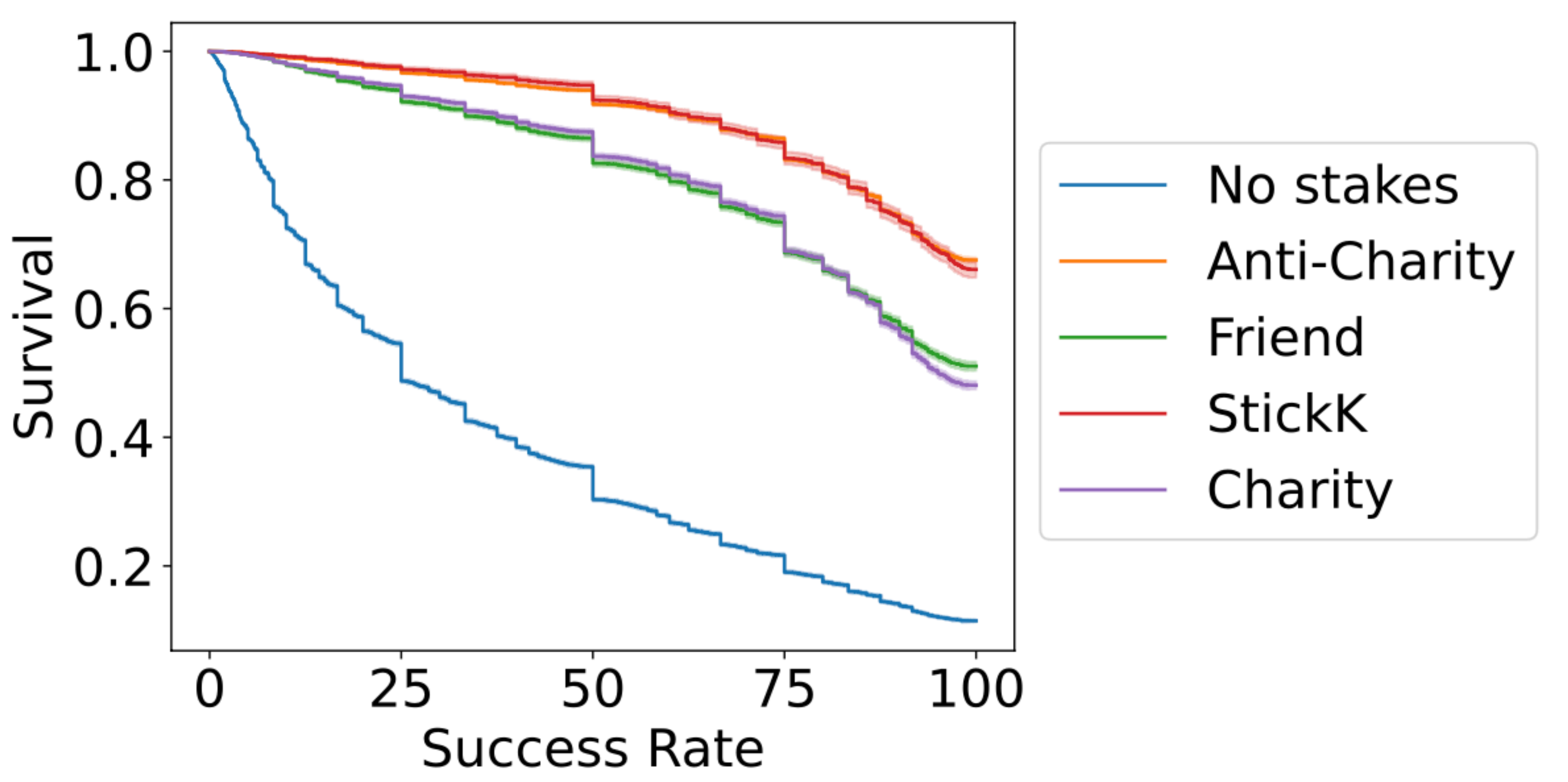
Survival Analysis / CoX Regression

Monetary - Extent



Survival Analysis / CoX Regression

Monetary - Effect

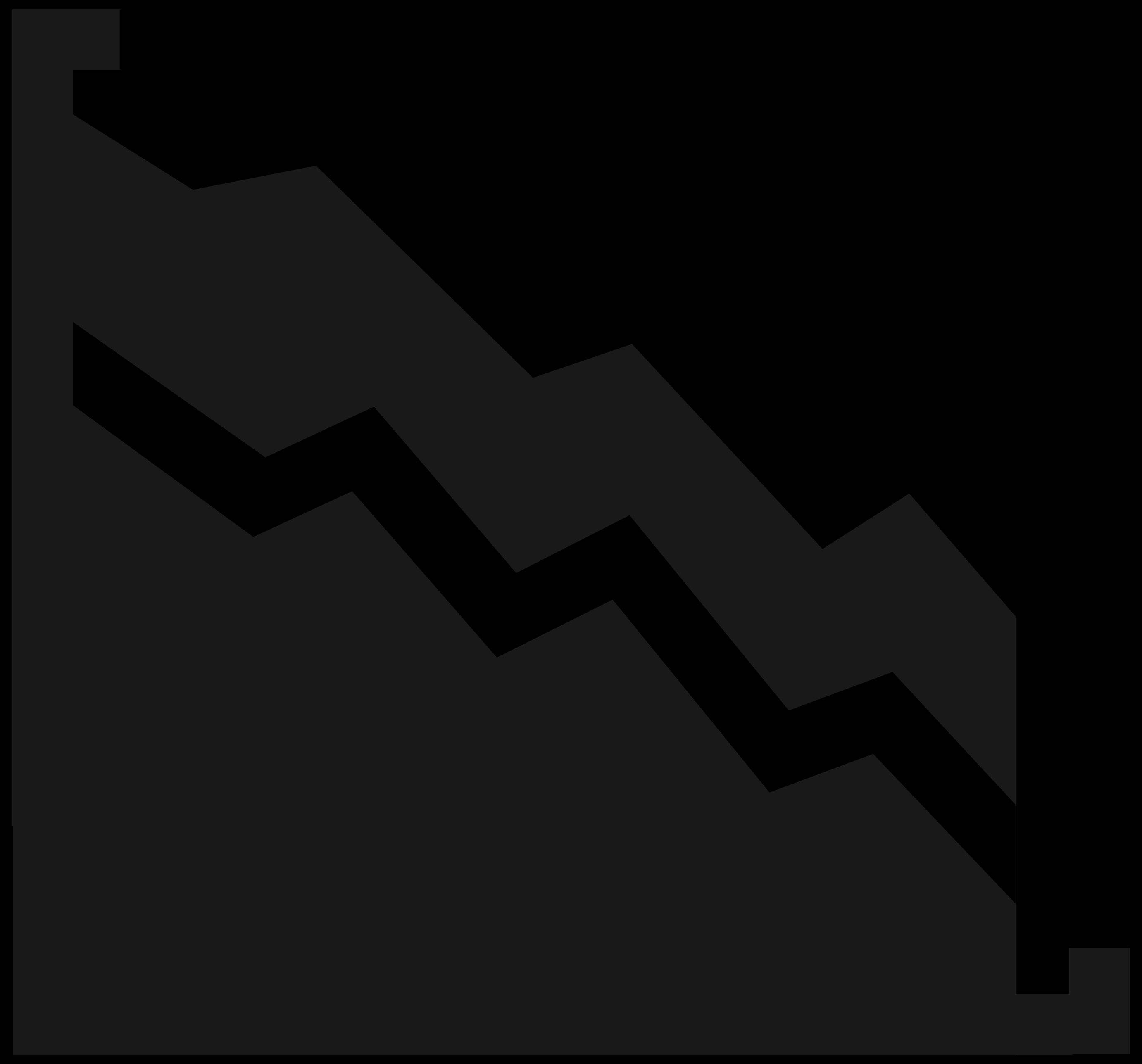


Covariate	HR	95% CI
# of reports	1.0447***	[1.0437, 1.0457]
Length of commitment	1.0045***	[1.0043, 1.0046]
# of supporters	0.9797***	[0.9725, 0.9870]
$ on stake per period	0.9474***	[0.9467, 0.9482]

*** $p <= 0.005$

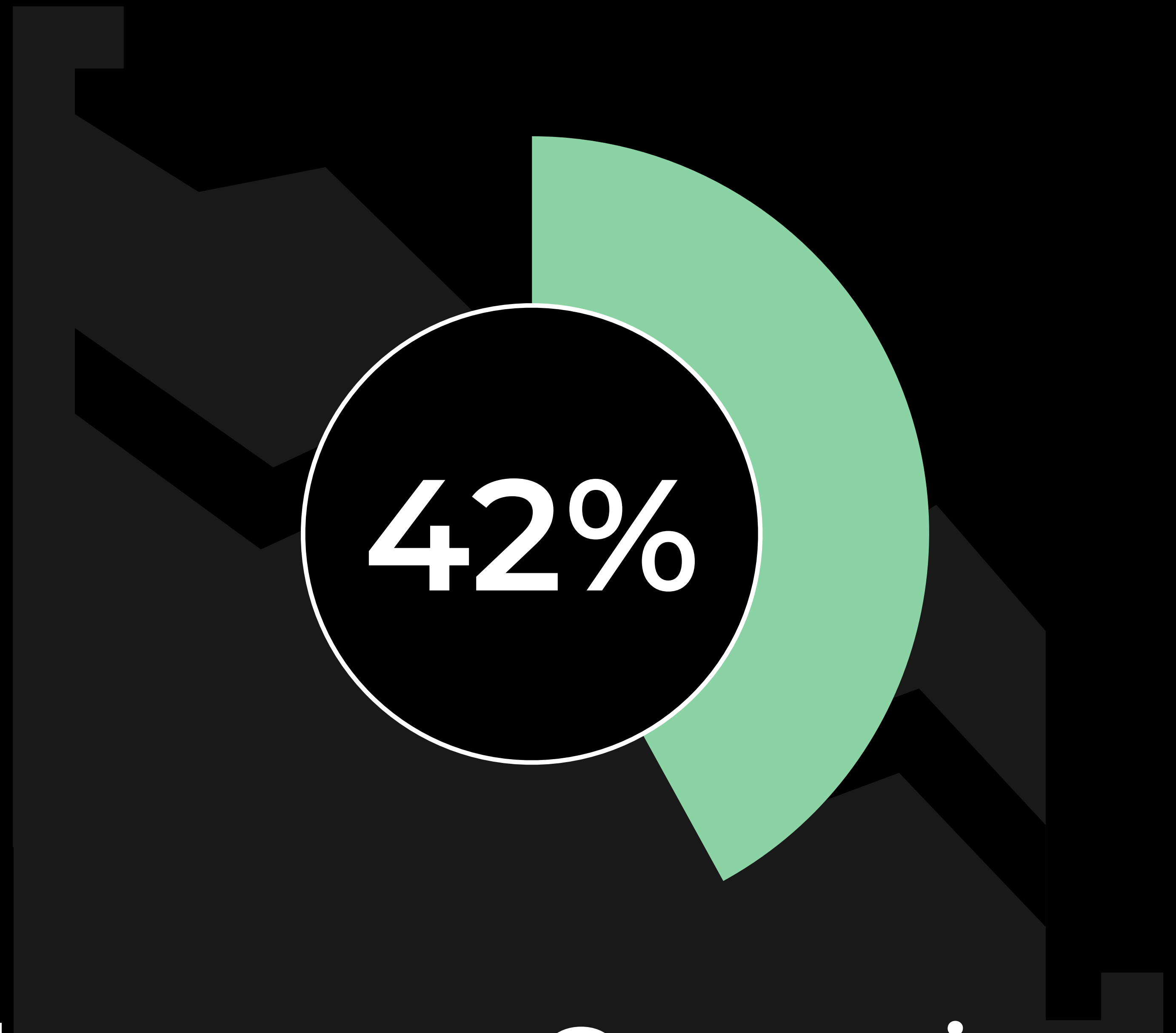
Survival Analysis / CoX Regression

Simultaneity - Extent



Survival Analysis / CoX Regression

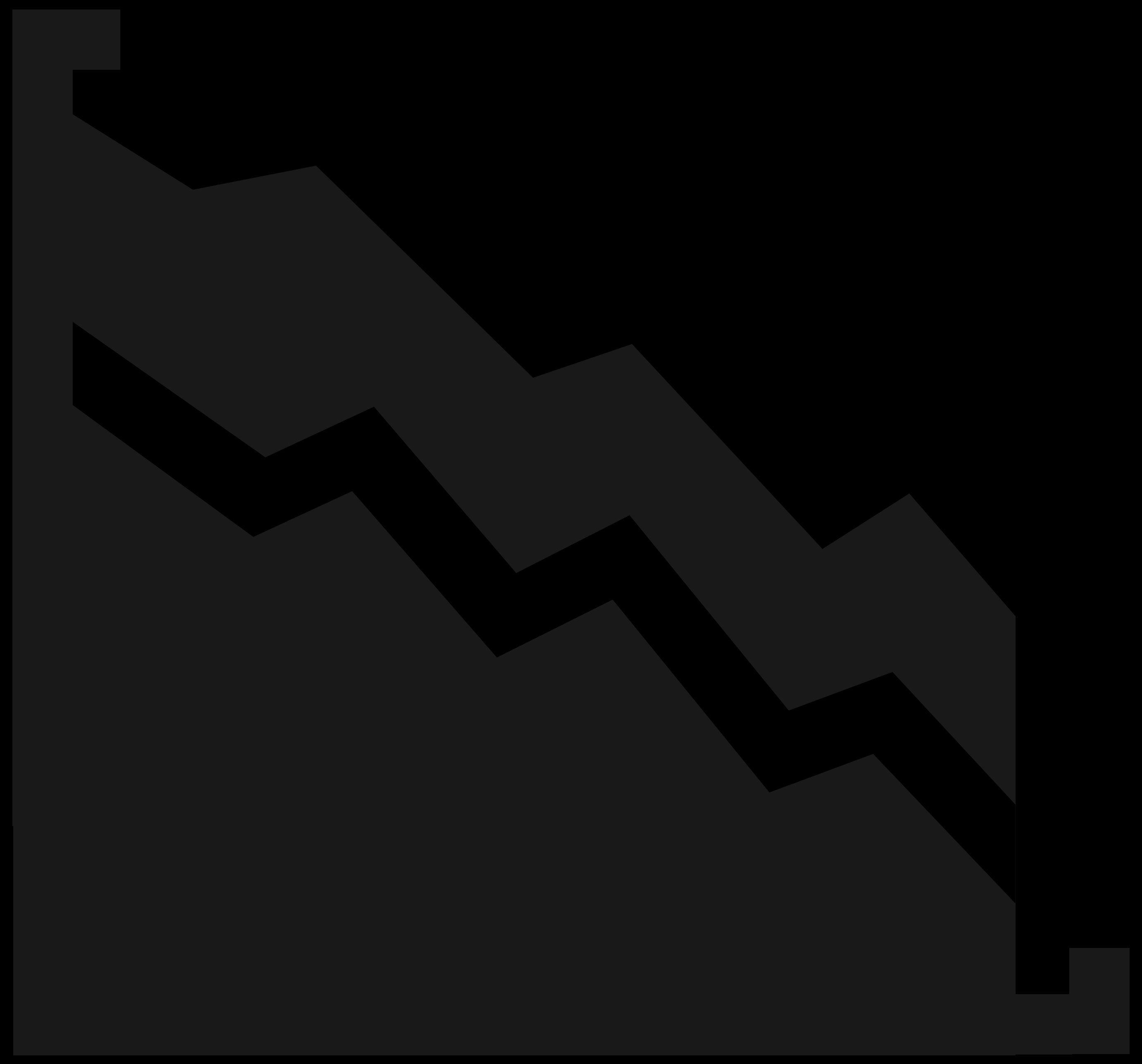
Simultaneity - Extent



Simultaneous Commitments

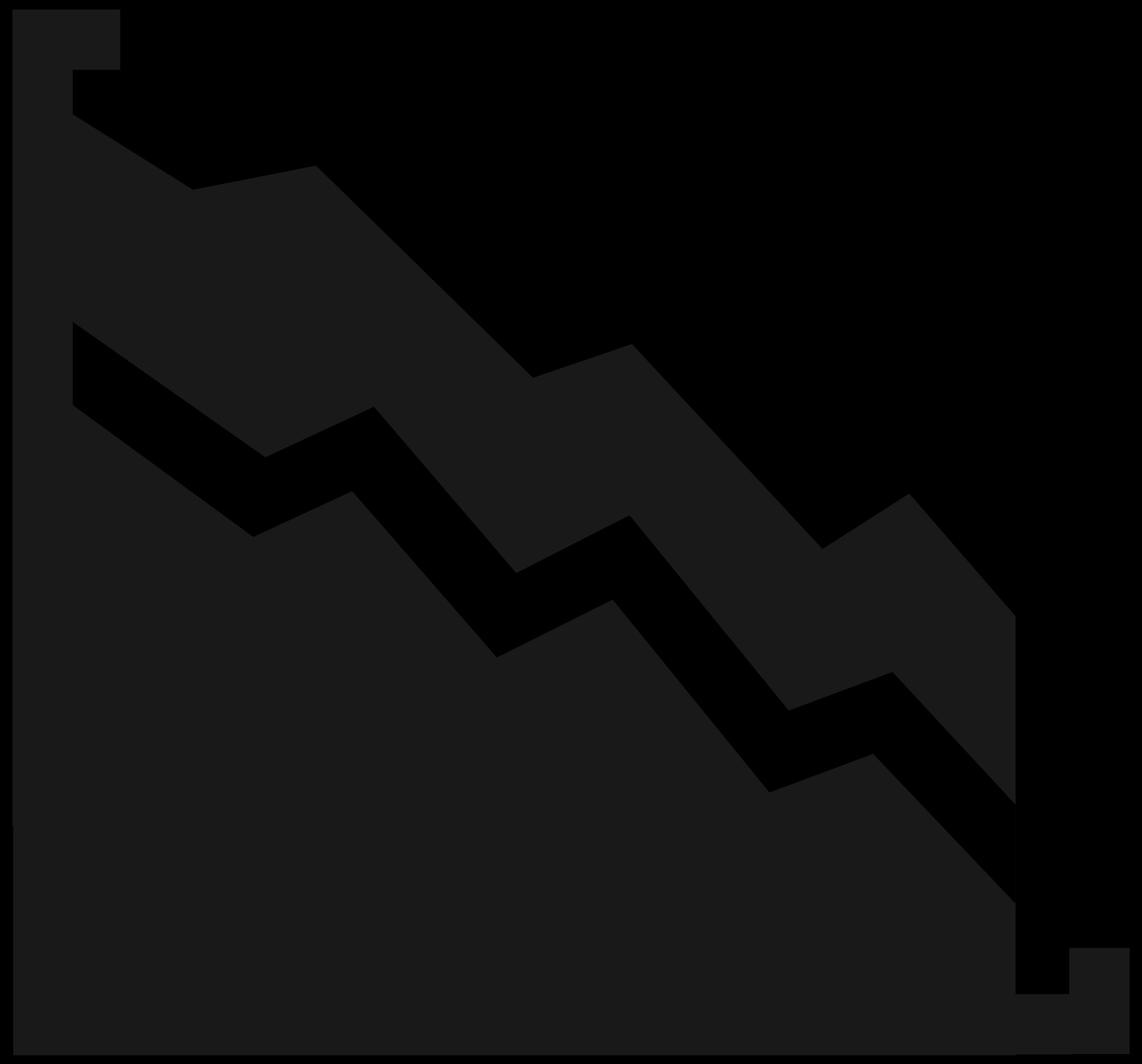
Survival Analysis / CoX Regression

Simultaneity - Extent



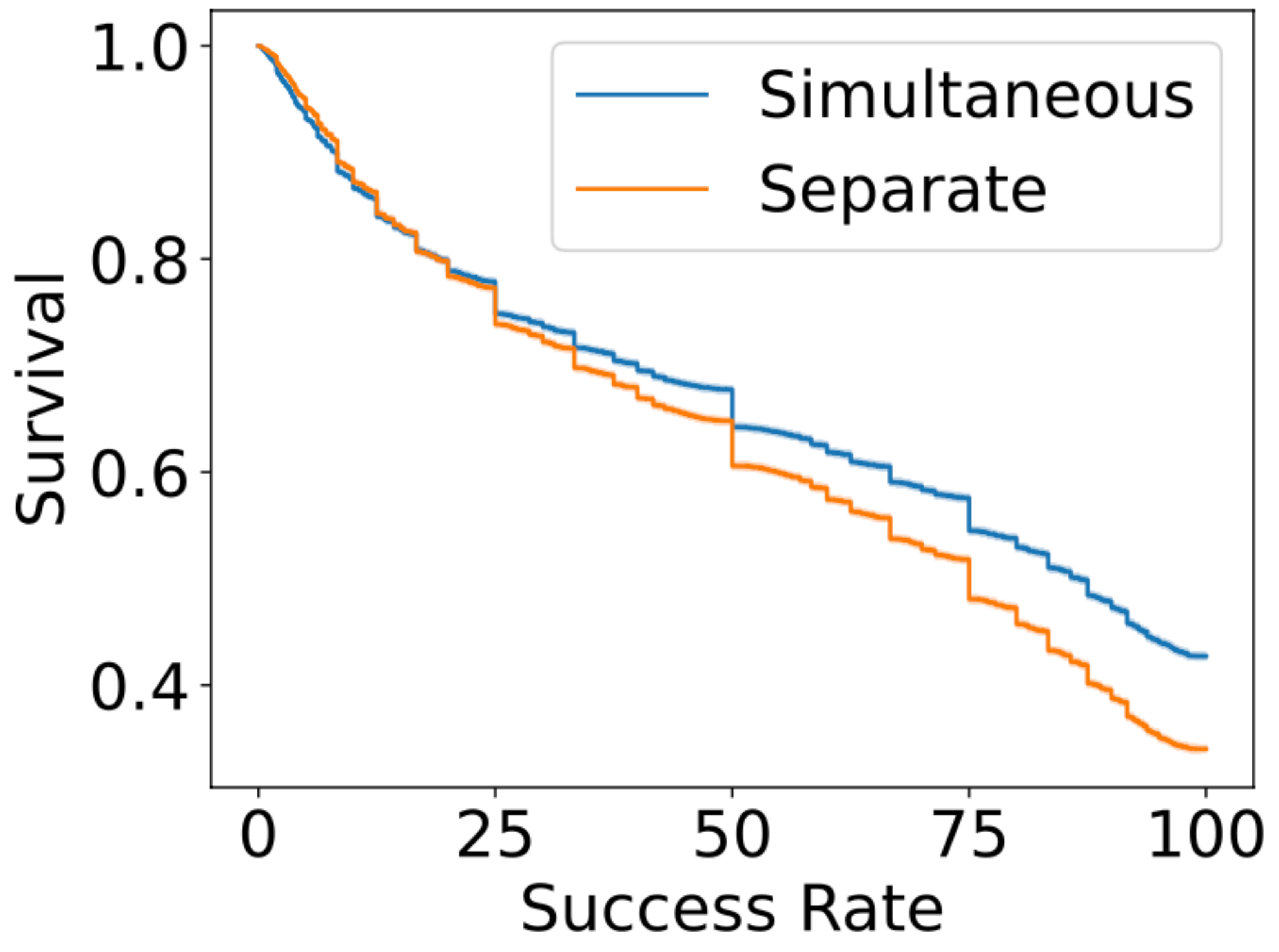
Survival Analysis / CoX Regression

Simultaneity - Extent



Survival Analysis / CoX Regression

Simultaneity - Effect



(f) Simultaneity. $p < 0.005$

Contributions

Contributions

First paper to study real world heterogeneous habit building attempts and provide insights into how to plan a goal

Contributions

- First paper to study real world heterogeneous habit building attempts and provide insights into how to plan a goal
- First large scale dataset of habit building attempts

Contributions

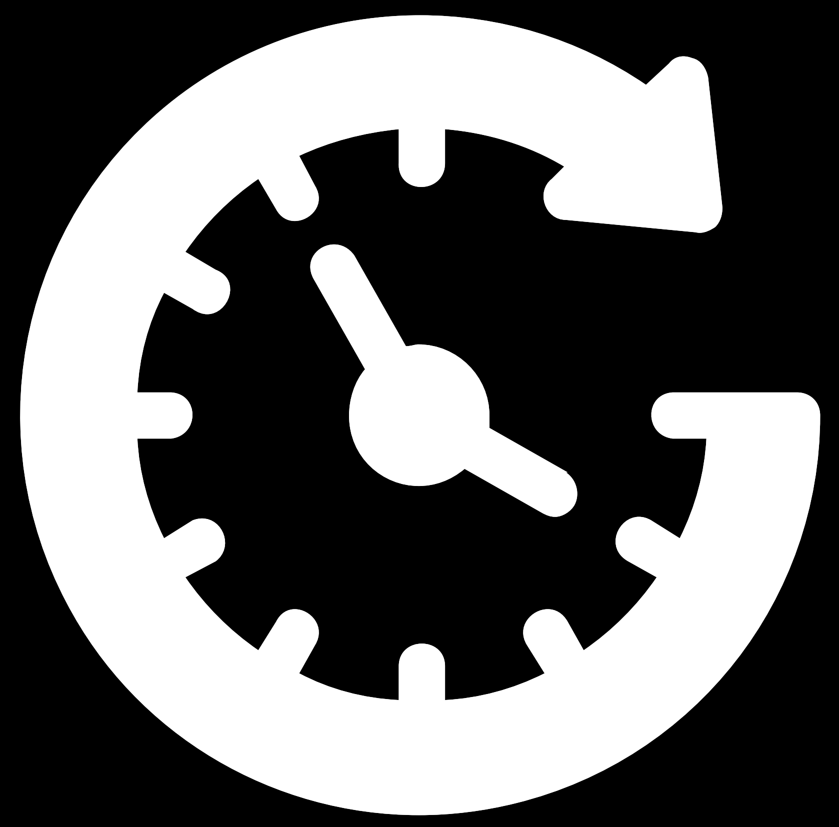
- First paper to study real world heterogeneous habit building attempts and provide insights into how to plan a goal
- First large scale dataset of habit building attempts
- Use of niche platforms to study characteristics of human behaviour

Future Work

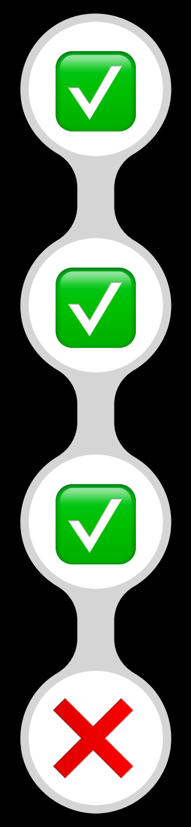
Future Work

Chronology

Future Work

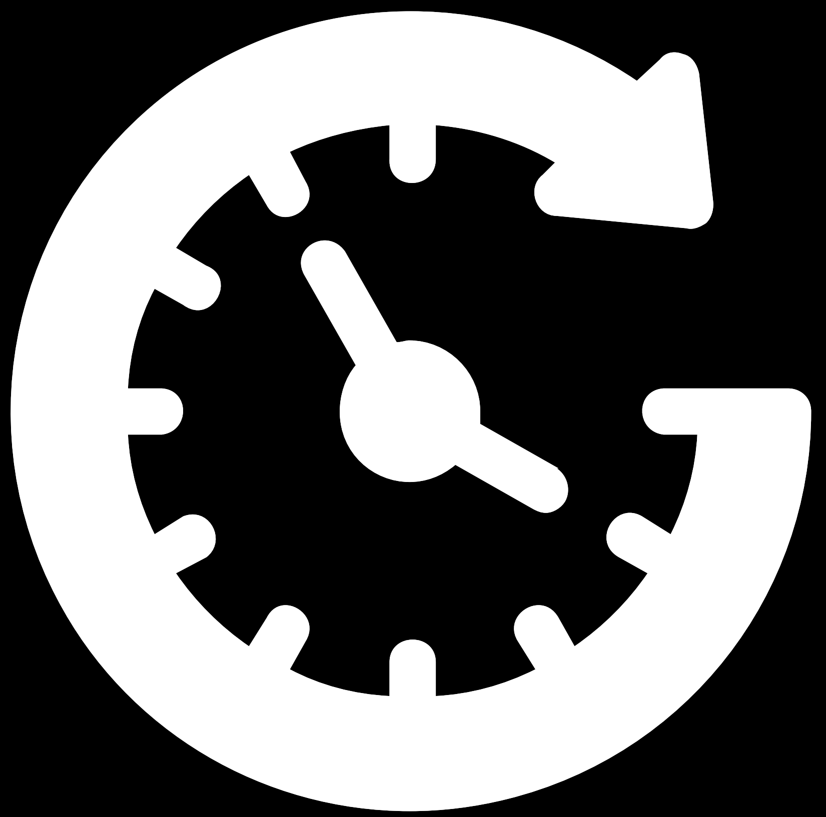


Chronology

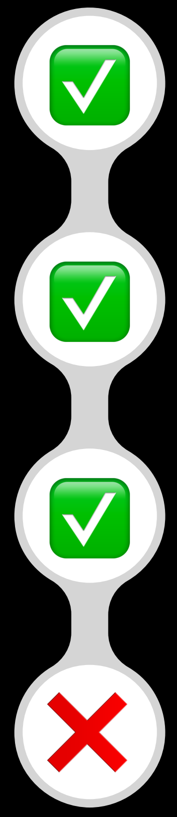


Streak

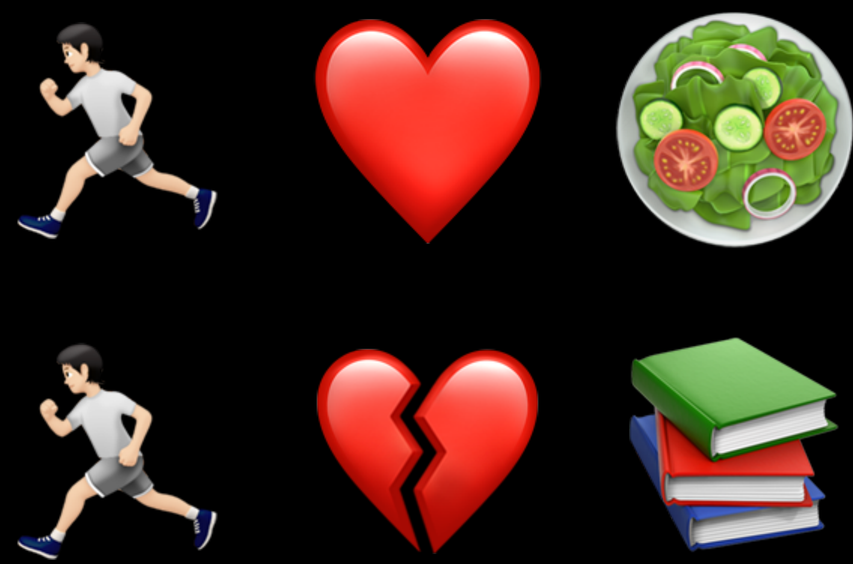
Future Work



Chronology

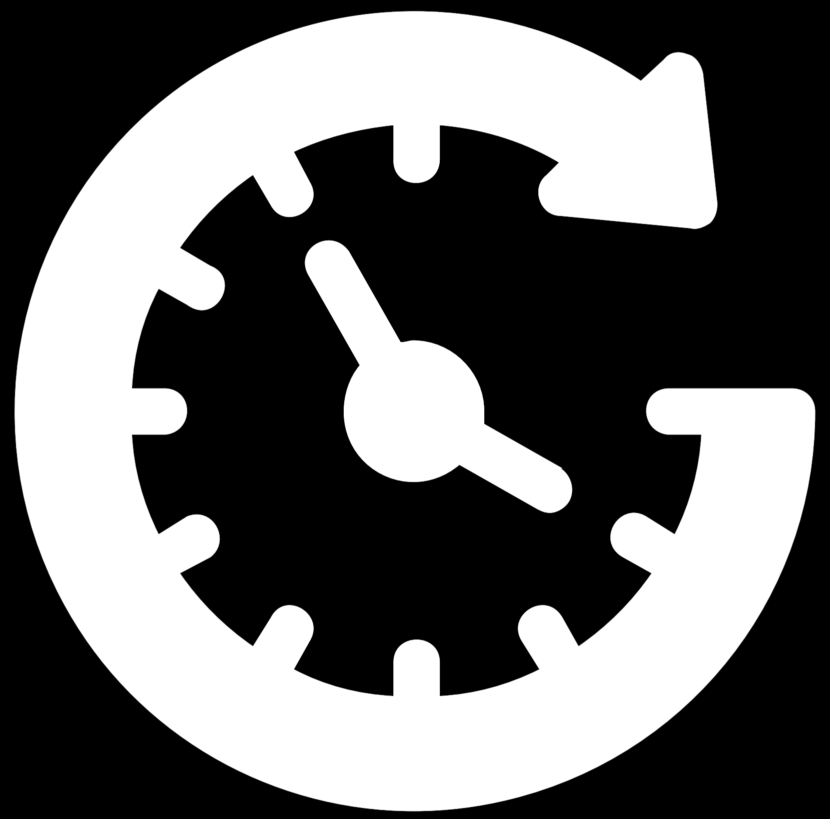


Streak

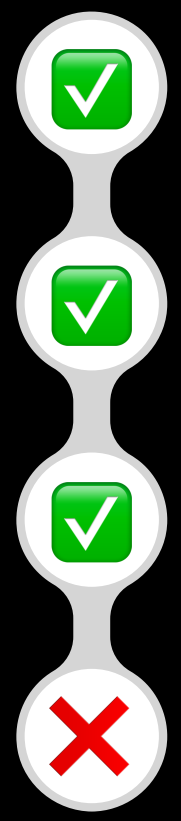


Combinations

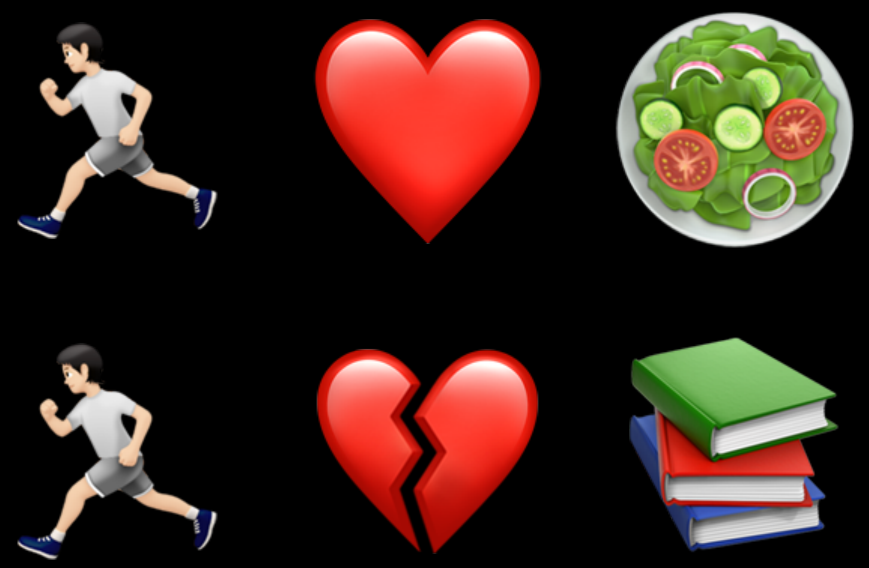
Future Work



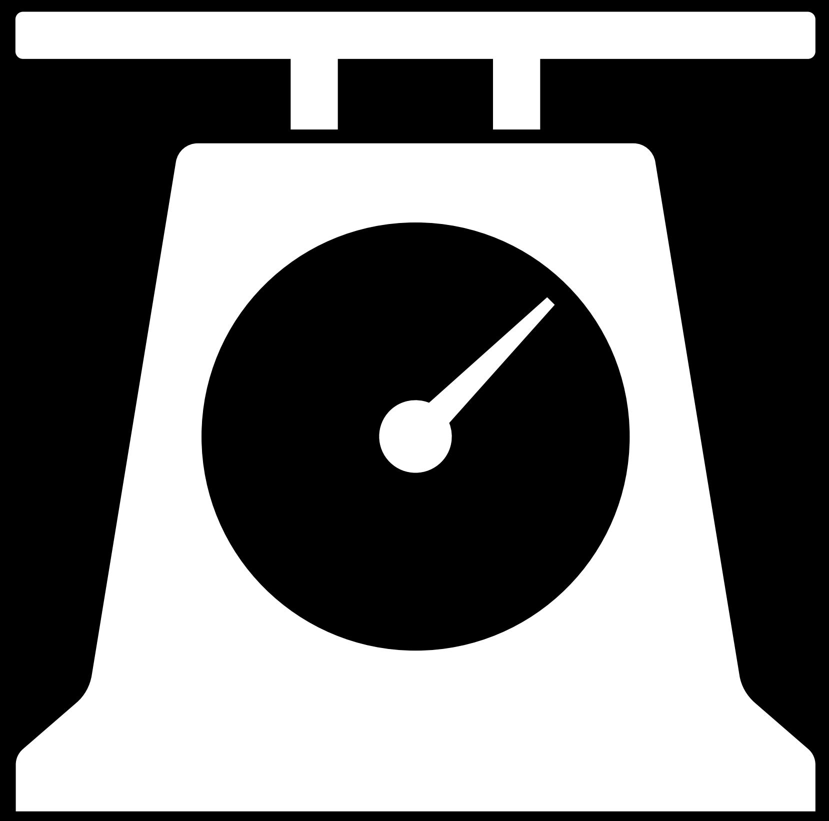
Chronology



Streak

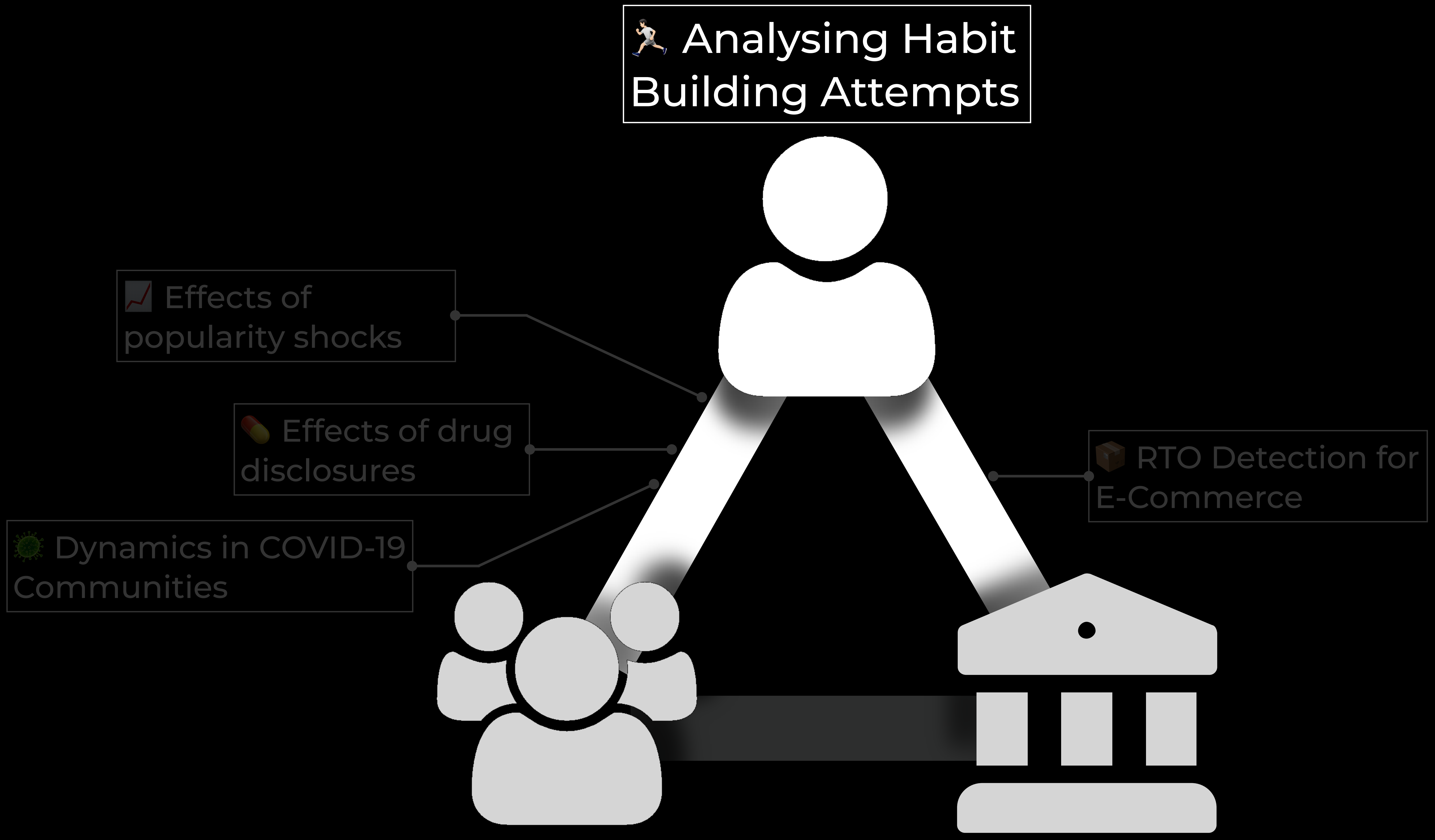


Combinations

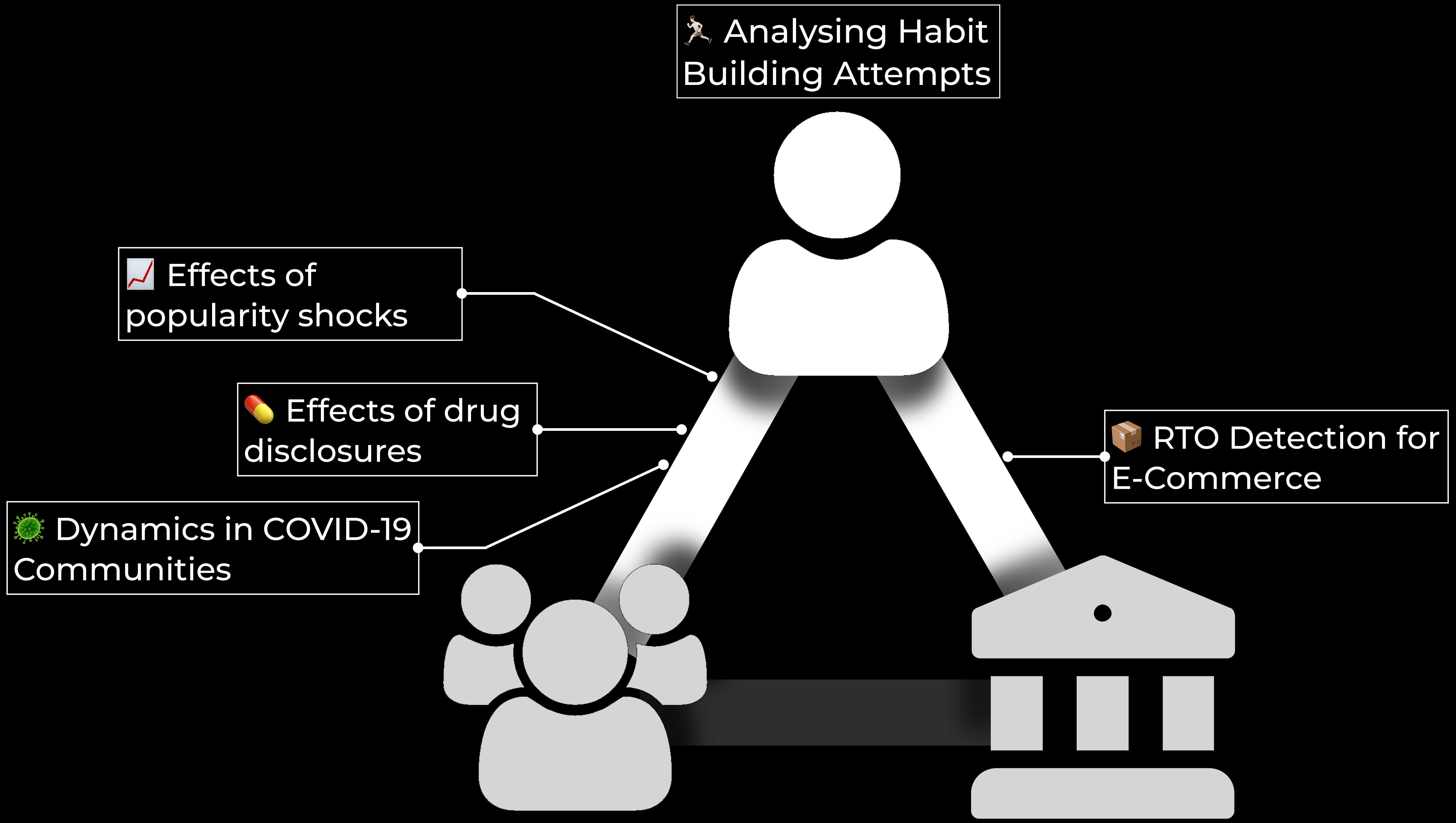


Weight Loss/Gain

Our Focus


Analysing Habit Building Attempts
Effects of popularity shocks
Effects of drug disclosures
Dynamics in COVID-19 Communities
RTO Detection for E-Commerce

Our Focus



🏃 Analysing Habit Building Attempts
📈 Effects of popularity shocks
💊 Effects of drug disclosures
🦠 Dynamics in COVID-19 Communities
📦 RTO Detection for E-Commerce

Limitations

Limitations

Generalizability

Platform Specific
Demographics
Geography

Limitations

Generalizability
Platform Specific
Demographics
Geography

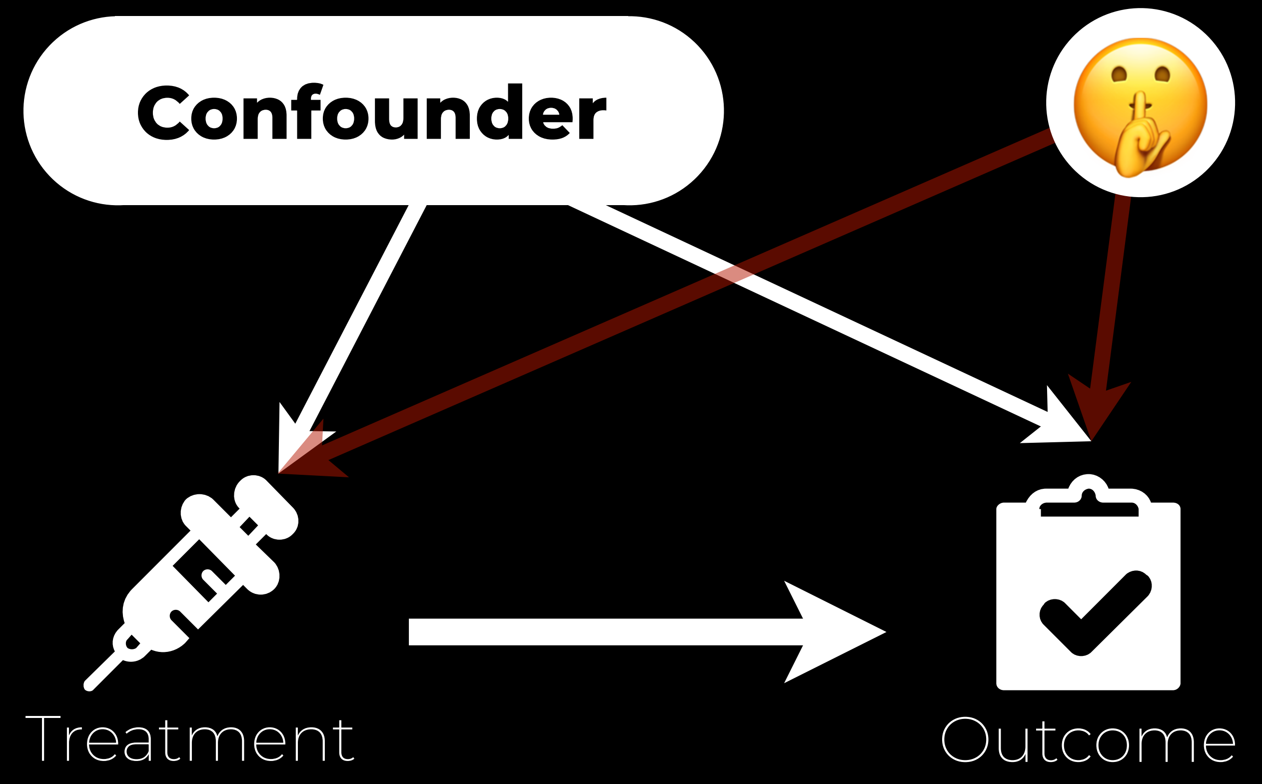
Weak Offline Proxy

Limitations

Generalizability

Platform Specific
Demographics
Geography

Weak Offline Proxy



Hidden Confounders

Future Work

Future Work

CSS Methods

Sociology

No-code Tools

Future Work

CSS Methods

Causal inference methods for unstructured social data

Resolving weak proxies
- Mixed methods
- Ensemble of weak proxies

Sociology

No-code Tools

Future Work

CSS Methods

Causal inference methods for unstructured social data

Resolving weak proxies
- Mixed methods
- Ensemble of weak proxies

Sociology

Extensive evaluation leading to complete frameworks e.g:

Operant Conditioning
- Source of Feedback
- Type of Feedback
- Dosage

No-code Tools

Future Work

CSS Methods

Causal inference methods for unstructured social data

Resolving weak proxies
- Mixed methods
- Ensemble of weak proxies

Sociology

Extensive evaluation leading to complete frameworks e.g:

Operant Conditioning
- Source of Feedback
- Type of Feedback
- Dosage

No-code Tools

Analysis tools for moderators and community owners

Publications - Thesis

ICWSM

Gurjar, O., Bansal, T., **Hitkul**, Lamba, H., and Kumaraguru, P. Effect of Popularity Shocks on User Behavior. 2022

ICWSM

Hitkul, Shah, RR., and Kumaraguru, P. Effect of Feedback on Drug Consumption Disclosures on Social Media. 2023

THE WEB CONFERENCE ACM

Hitkul, Abinaya, Saha, S., Banerjee, S., Chelliah, M., and Kumaraguru, P. Social Re- Identification Assisted RTO Detection for E-Commerce. 2023

ASONAM 2023

Hitkul, Pandey, T., Singhal, S., Kandhari, K., Tomar, K. and Kumaraguru, P. Together Apart: Decoding Support Dynamics in Online COVID-19 Communities. 2023

ICWSM

Hitkul, Shah, RR., and Kumaraguru, P. Put Your Money Where Your Mouth Is: Dataset and Analysis of Real World Habit Building Attempts. 2024

Publications - Thesis

Publications - Others

6

Peer Reviewed
(KDD, WWW, BigMM etc)

6

Technical Reports
Book Chapters

- 📺 TV News
- 👤 Face Generation
- ⚔️ Capitol Riots
- 🤡 Trolling
- 🙂🙁 Sentiment Analysis
- 😡🥰 Emotional Classification

Acknowledgment

tcs Research

Goldman Sachs

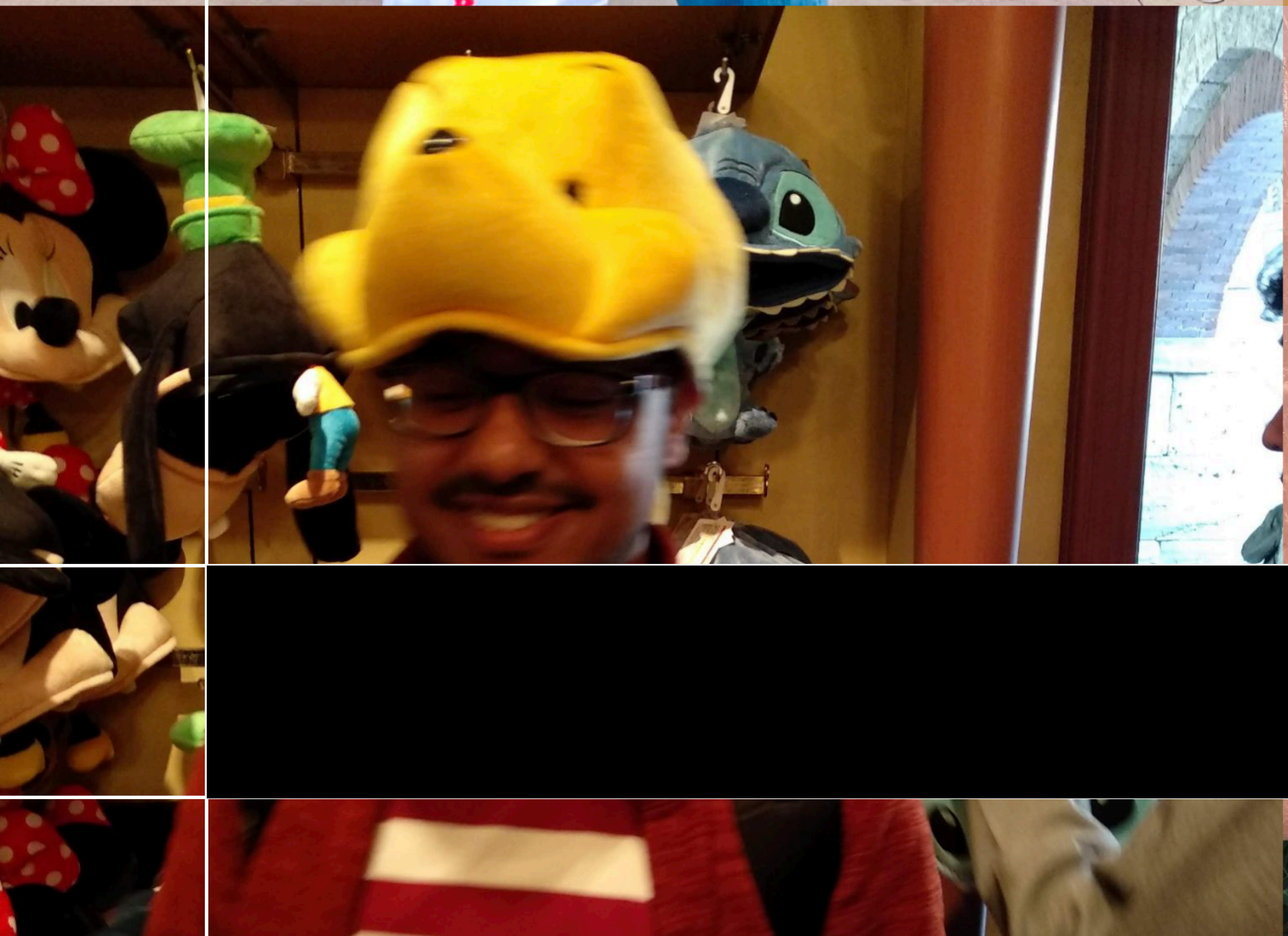
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

NII
国立情報学研究所
National Institute of Informatics

Acknowledgment

Acknowledgment

Acknowledgment



Acknowledgment

Acknowledgment

Thank You!

Do you have any questions?

hitkuli@iiitd.ac.in

hitkuljangid

@Hitkul_