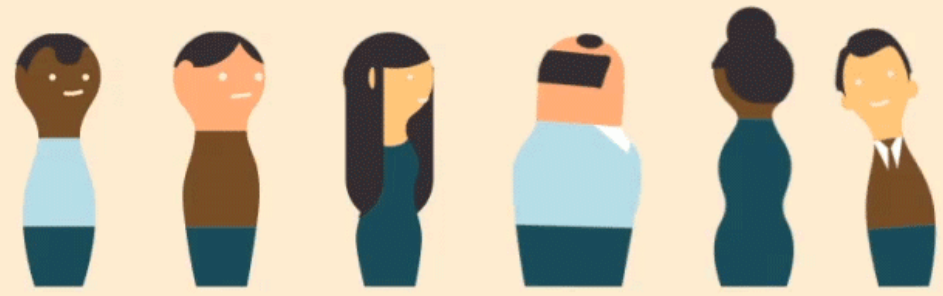


# If you Can't Beat Them, Join Them:



## A **Usability** Approach to **Interdependent Privacy** in Cloud Apps

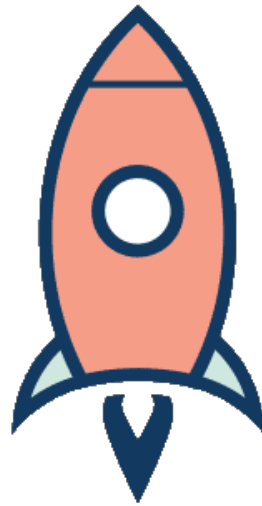
**Hamza Harkous** and Karl Aberer

**[hamzaharkous.com](https://hamzaharkous.com)**

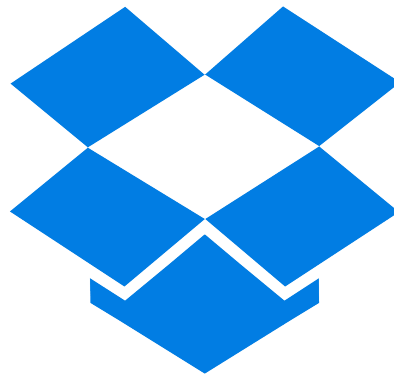
# Growing Business Adoption of Cloud Ecosystems



3<sup>rd</sup> party apps



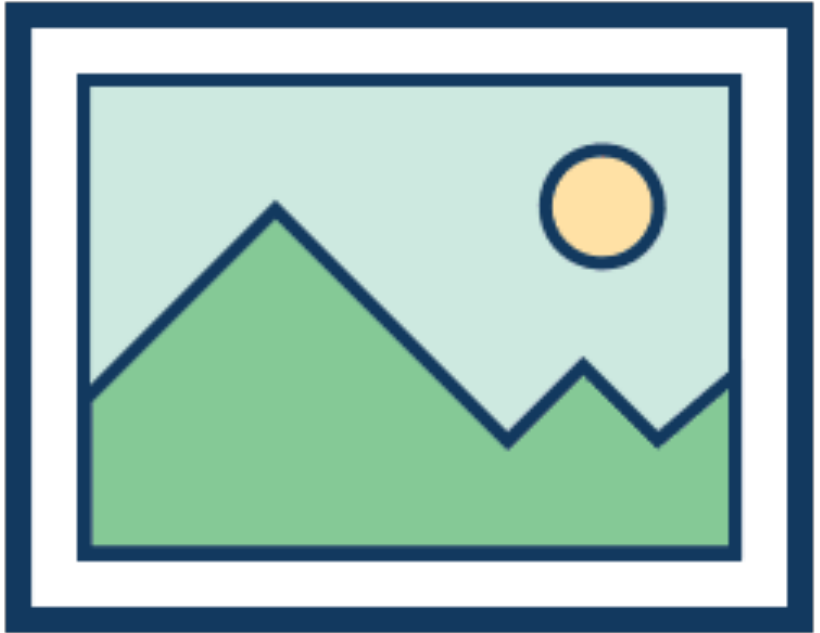
CSPs



Files



Your Files



[Register](#)[Login](#)[Menu](#) [COMPRESS IMAGE](#)[RESIZE IMAGE](#)[CROP IMAGE](#)[CONVERT TO JPG](#)[CONVERT FROM JPG](#)

## Crop IMAGE

Crop **JPG**, **PNG** or **GIF** by defining a rectangle in pixels.  
Cut your image online.

Select image



or drop image here



Organizations use **10-20 times** more  
cloud apps than their IT departments think\*.

# The Proliferation of new Adversaries:

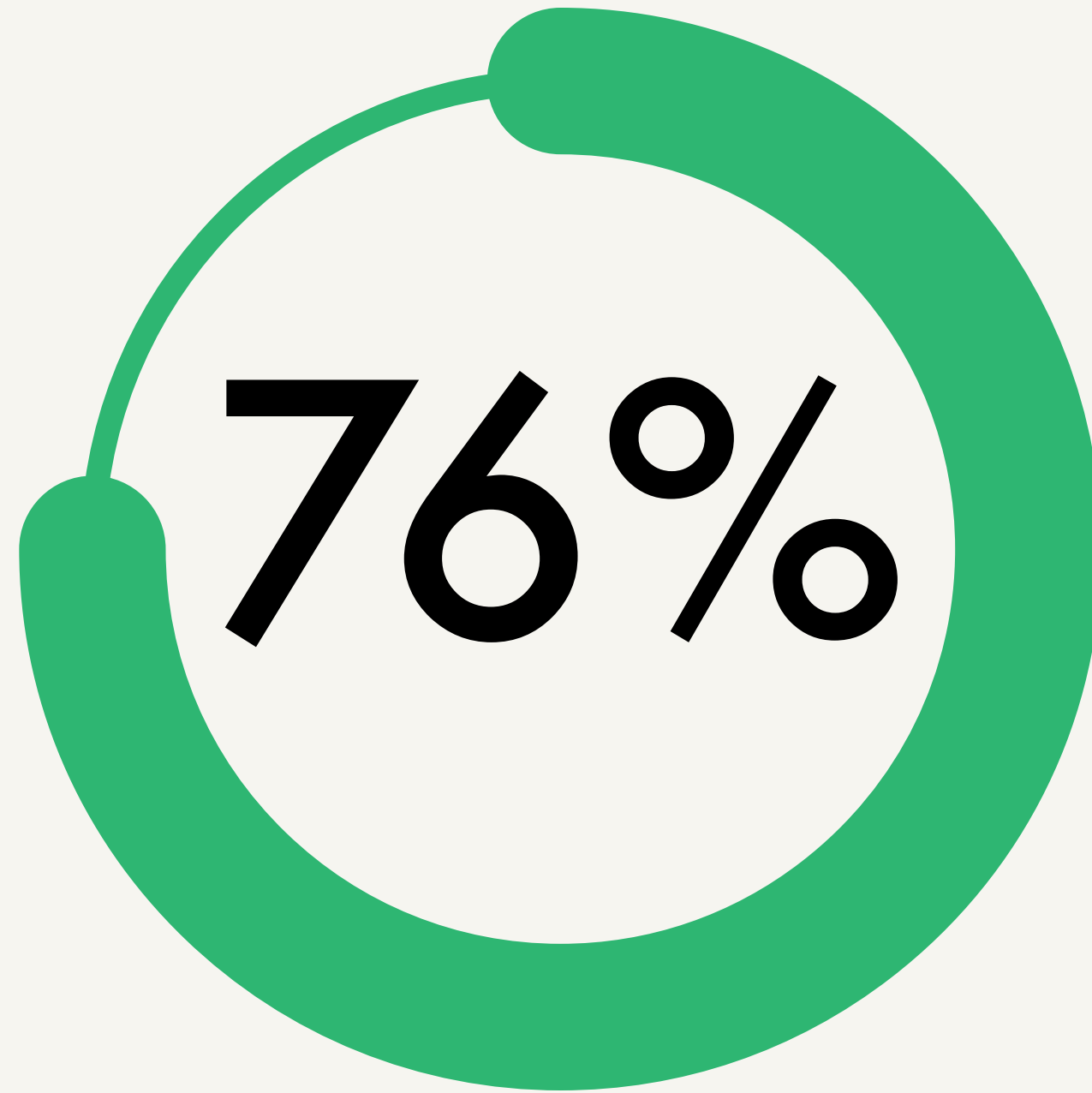
**3rd party apps**



\$13.85M

average financial impact on a company as a result of a  
cloud-storage data breach

**Elastic** Cloud Threat Labs. Q2 2015 Shadow Data Report: <https://www.elastic.net/q2-2015-shadow-data-report/>



of Google Drive apps featured on Chrome Store  
ask for **full access** to all your files\*



\*Harkous et al. The Curious Case of the PDF Converter that Likes Mozart:  
Dissecting and Mitigating the Privacy Risk of Personal Cloud Apps. (PoPETs 2016)



# 37.2%

of documents are  
shared with at least  
1 other user

\***Skyhigh Networks.** Cloud Adoption and Risk Report. 2015.

\***Elastica** Cloud Threat Labs. 1H 2016 Shadow Data Report. 2016.





37.2%

of documents are  
shared with at least  
1 other user

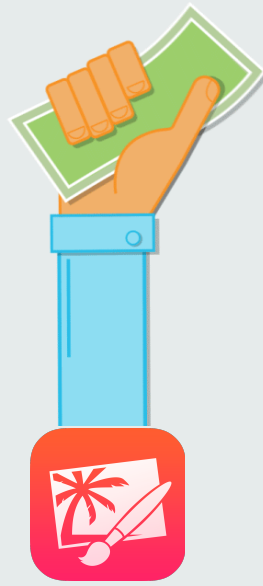


23%

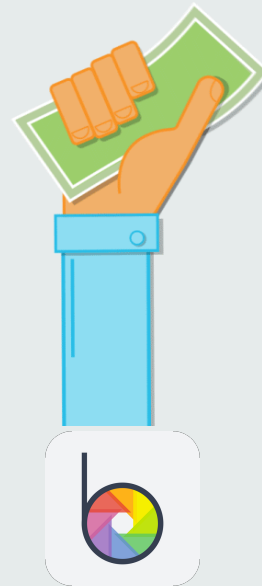
of documents are  
broadly shared  
(with all employees or  
with outsiders)

\***Skyhigh Networks.** Cloud Adoption and Risk Report. 2015.

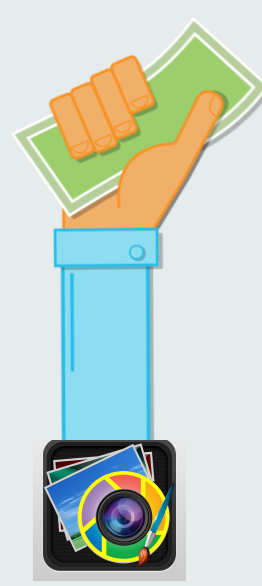
\***Elastica** Cloud Threat Labs. 1H 2016 Shadow Data Report. 2016.



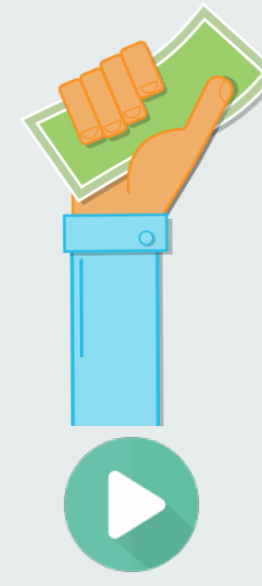
Company 1



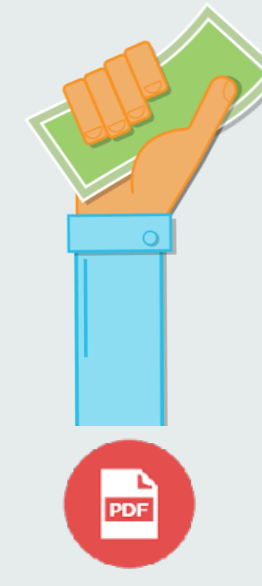
Company 2



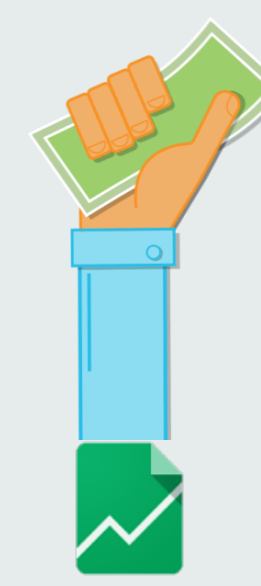
Company 3



Company 4



Company 5

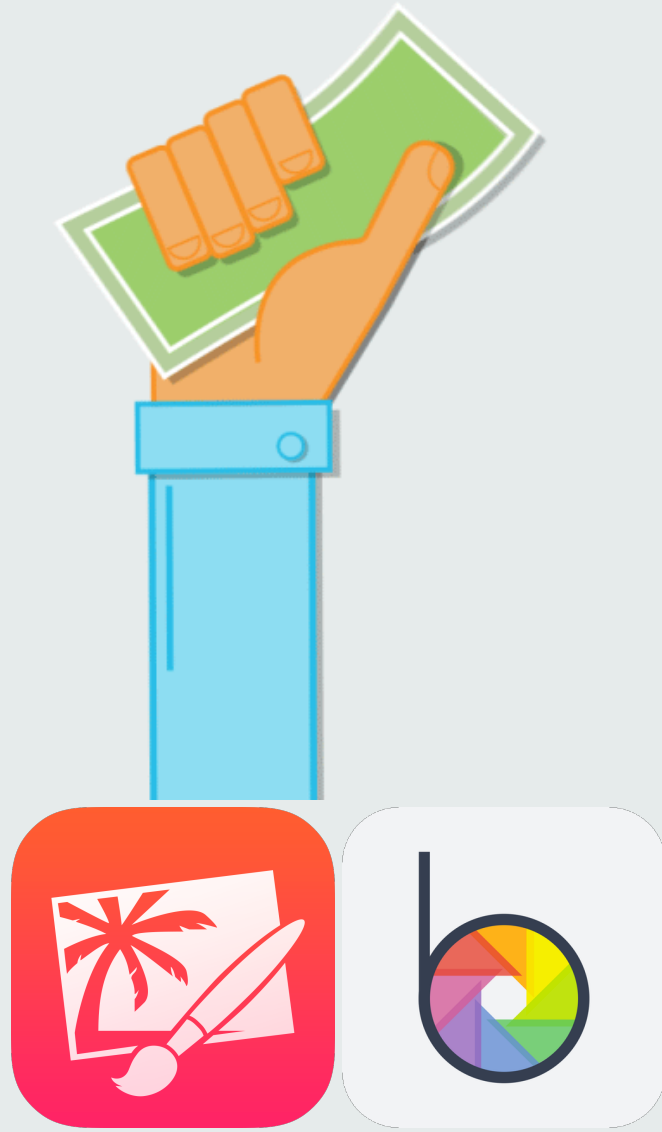


Company 6



# Too Many Shareholders

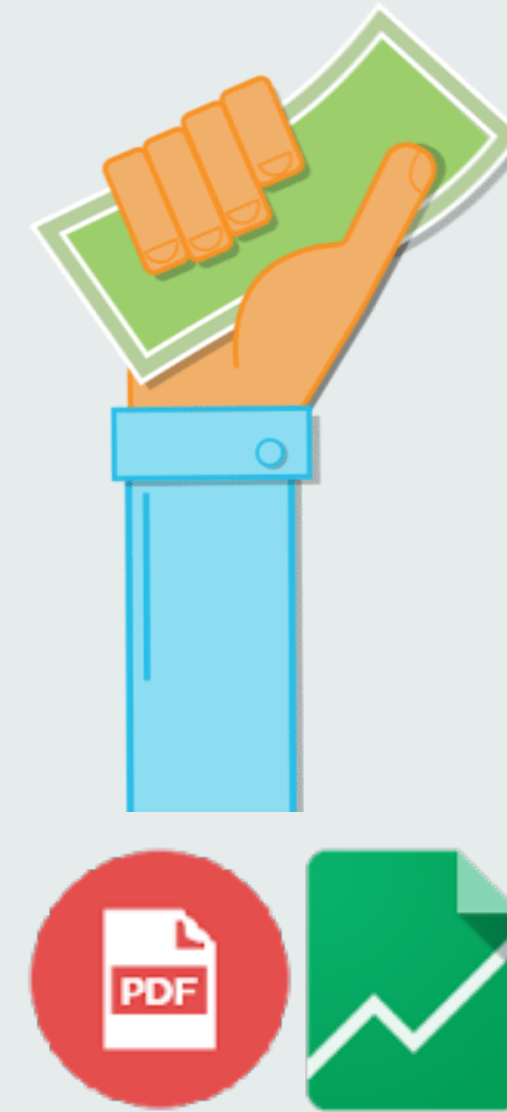
→ Larger Attack Surface



Company 1



Company 2



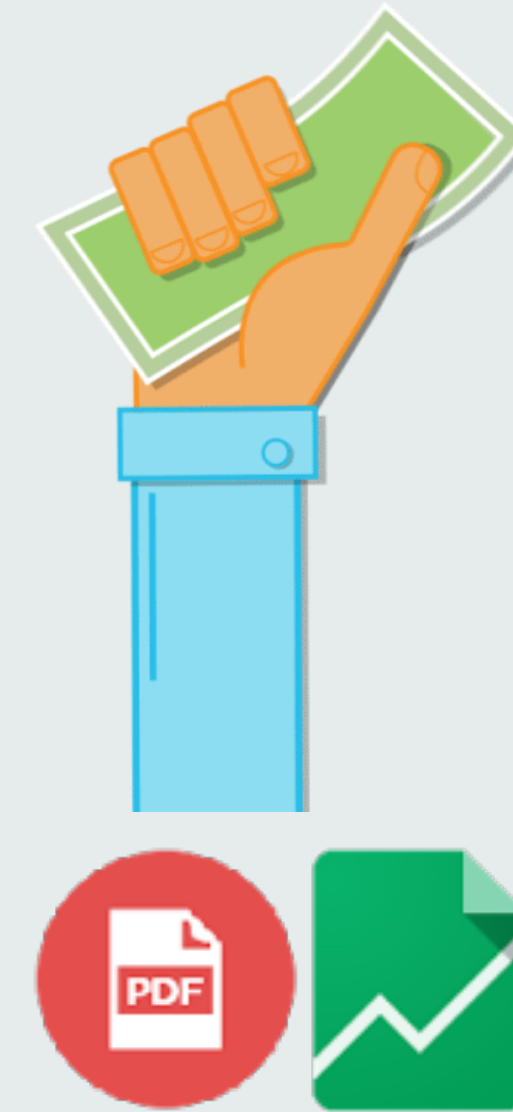
Company 3



Company 1



Company 2

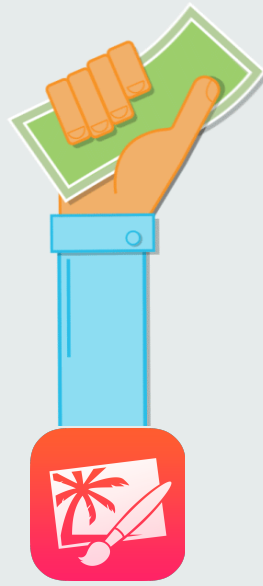


Company 3

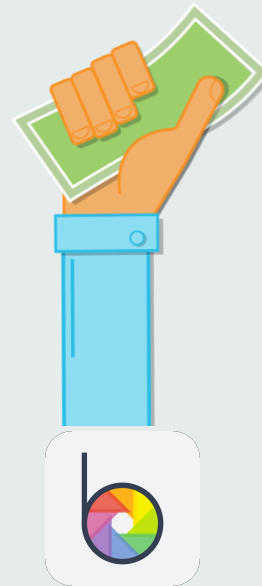
Fewer Shareholders →  
Narrower attack surface



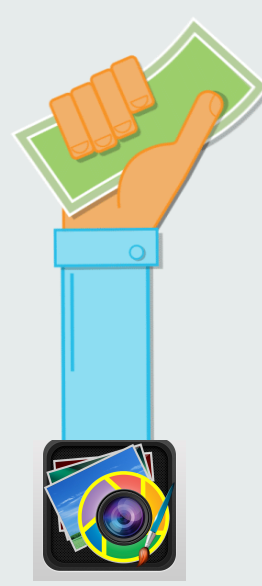
# Challenge



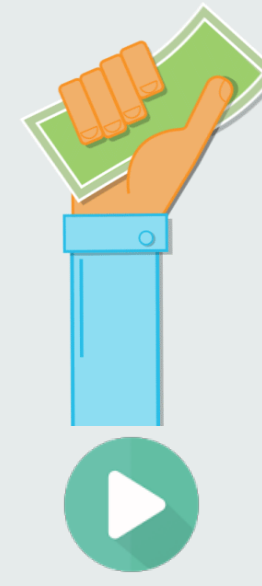
Company 1



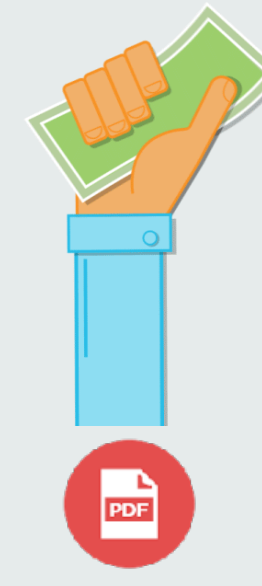
Company 2



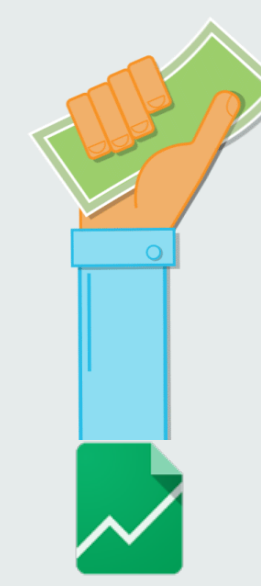
Company 3



Company 4



Company 5



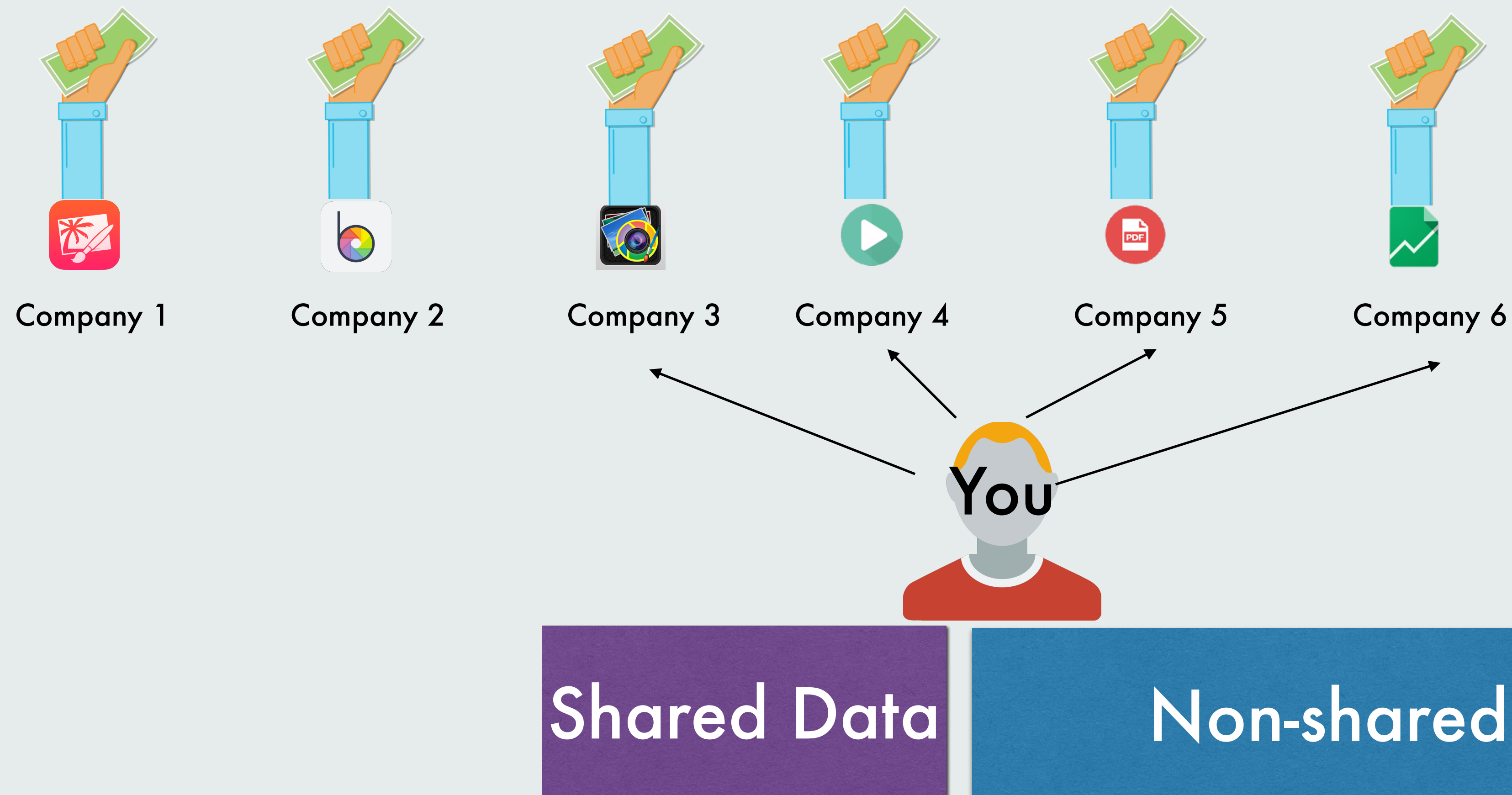
Company 6



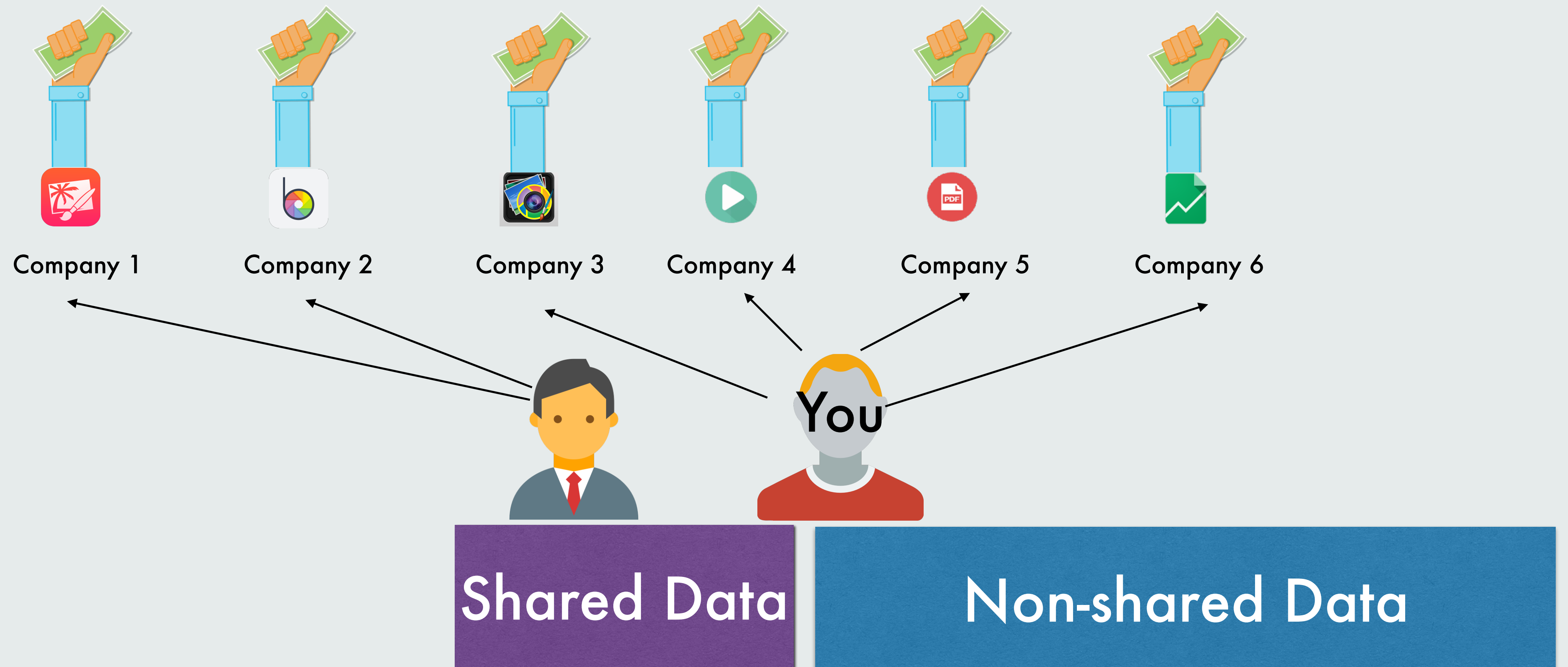
Shared Data

Non-shared Data

# Challenge



# Challenge



# Challenges



# Challenges



Cannot remember all companies



# Challenges



**Cannot remember all companies**

**Don't know what others install**

so that you don't give access to new companies

# Challenges



**Cannot remember all companies**

**Don't know what others install**  
so that you don't give access to new companies

**Cannot control what others install**

# **Interdependent Privacy**

**"The privacy of individual users is bound to be affected by the decisions of others, and could be out of their own control"\***

# Interdependent Privacy

**"The privacy of individual users is bound to be affected by the decisions of others, and could be out of their own control"\***

**Originally introduced in the context of Facebook apps**

- Biczok and Chia, Interdependent privacy: Let me share your data. In FC 2013
- Pu and Grossklags, An economic model and simulation results of app adoption decisions on networks with interdependent privacy consequences, GameSec 2014

# Interdependent Privacy

**"The privacy of individual users is bound to be affected by the decisions of others, and could be out of their own control"\***

**Originally introduced in the context of Facebook apps**

- Biczok and Chia, Interdependent privacy: Let me share your data. In FC 2013
- Pu and Grossklags, An economic model and simulation results of app adoption decisions on networks with interdependent privacy consequences, GameSec 2014

**Also used in the context of location privacy**

\* Olteanu, et al. "Quantifying interdependent privacy risks with location data." IEEE TMC (2016).





of Facebook apps get the  
**friends'** data\*



of cloud apps with full  
access get the  
**collaborators'** data



# Contributions

- **First to study the problem in the context of cloud apps.**

# Contributions

- First to study the problem in the context of cloud apps.
- **We quantify the effects of interdependent privacy in the wild.**

# Contributions

- First to study the problem in the context of cloud apps.
- We quantify the effects of interdependent privacy in the wild.
- **We propose a usable privacy solution to mitigate the issue.**

# Contributions

- First to study the problem in the context of cloud apps.
- We quantify the effects of interdependent privacy in the wild.
- We propose a usable privacy solution to mitigate the issue.
- **We showcase the network effect of privacy-aware decisions at scale.**

# Research Question-1

How significant is the impact of  
**collaborators'** app adoption  
decisions on the **users' privacy loss?**

# Dataset study:

# Google Drive+ PrivySeal




+



## PrivySeal

Your New Smart Privacy Assistant for Installing Cloud Apps

Uncover What Cloud Apps Can *Really* Know about You!


Sign in now with your  
 Google Drive! \*

\*By signing in you agree to our [privacy policy](#).



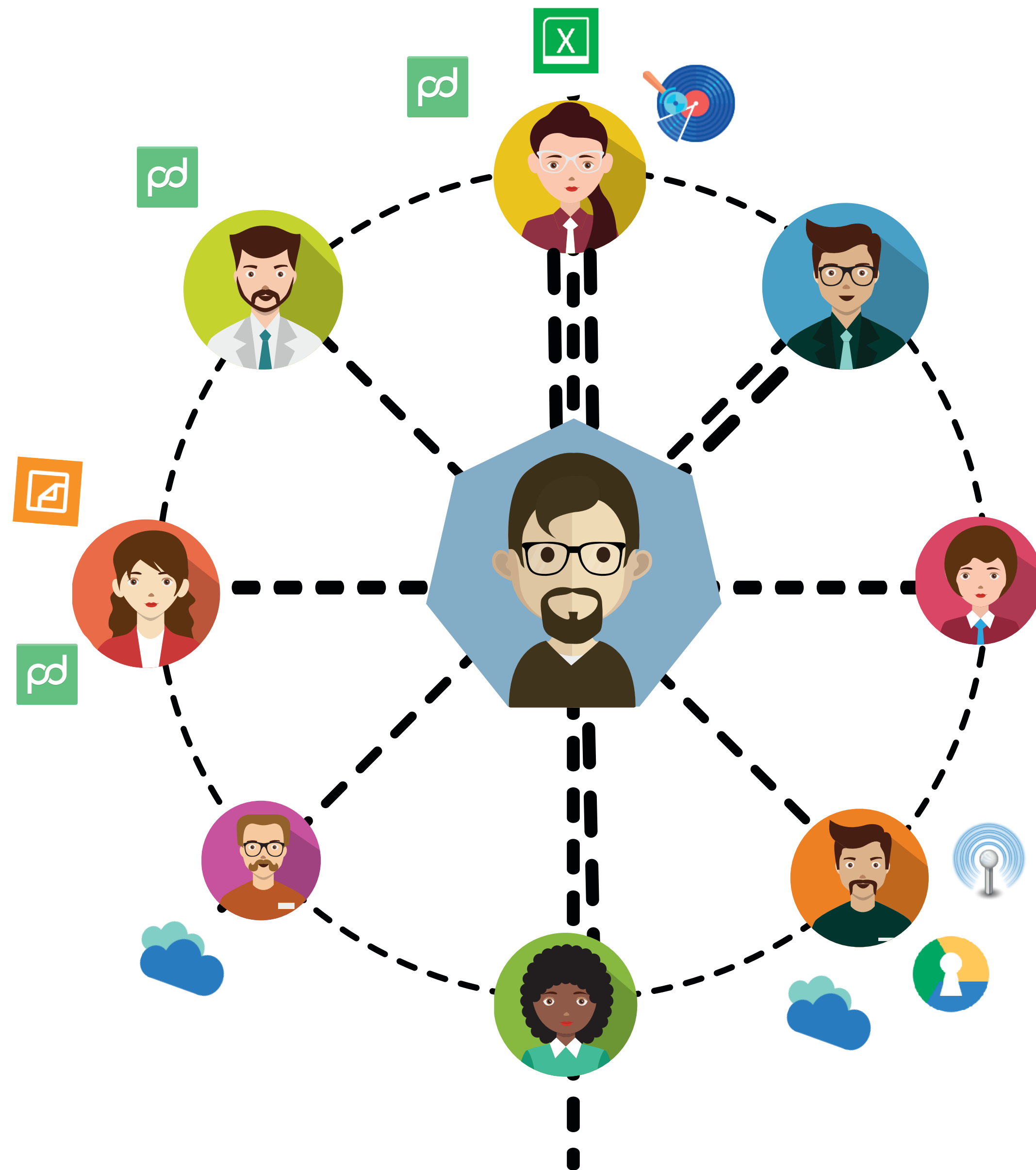
Loupe Collage wants to:

obtain permissions it doesn't need

 What do the **unneeded permissions** say about you?



# Anonymized Dataset



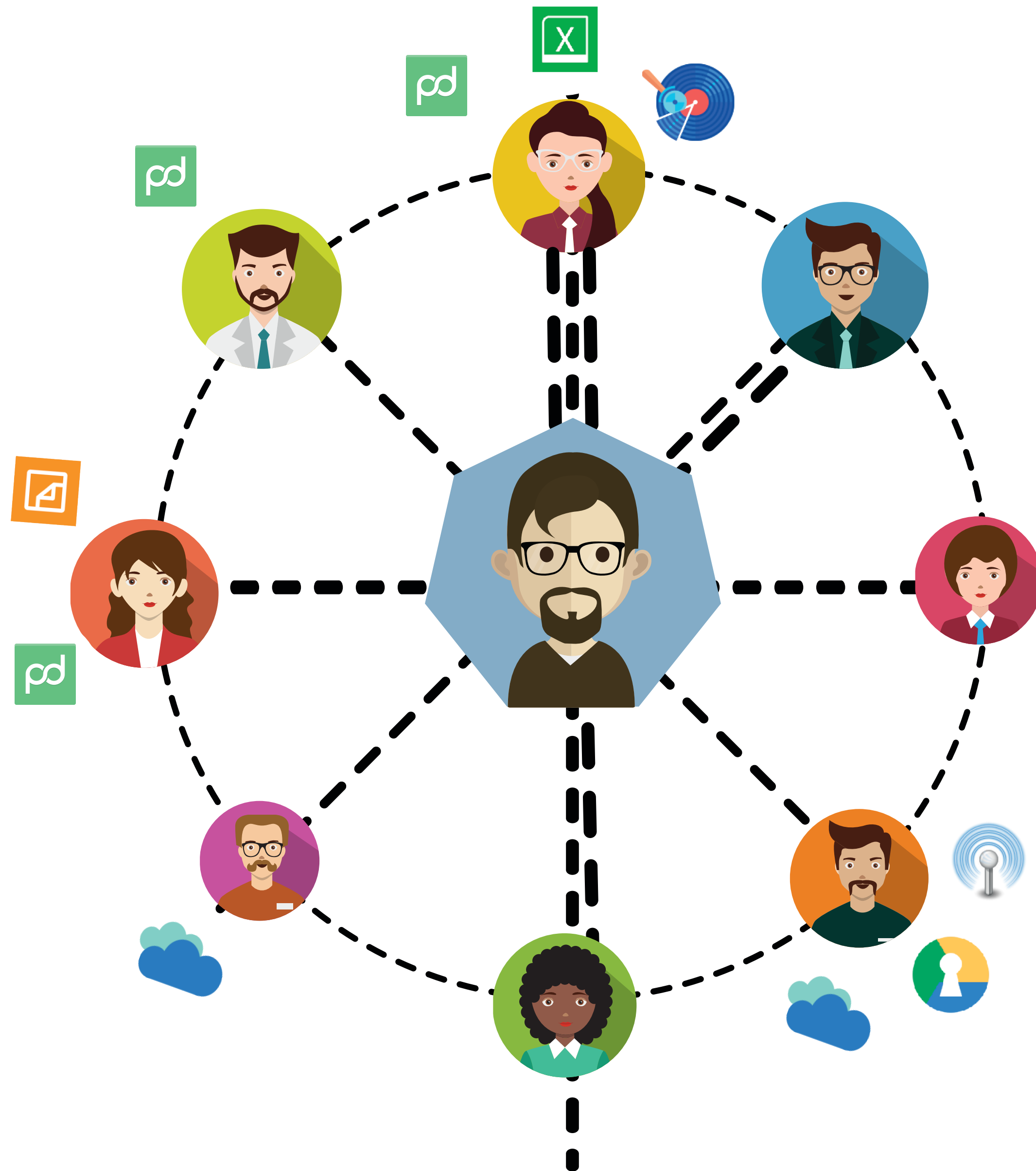
183 Google Drive users

$\geq 10$  files each

$\geq 5\%$  shared files

131 Google Drive apps

# Anonymized Dataset



183 Google Drive users

$\geq 10$  files each

$\geq 5\%$  shared files

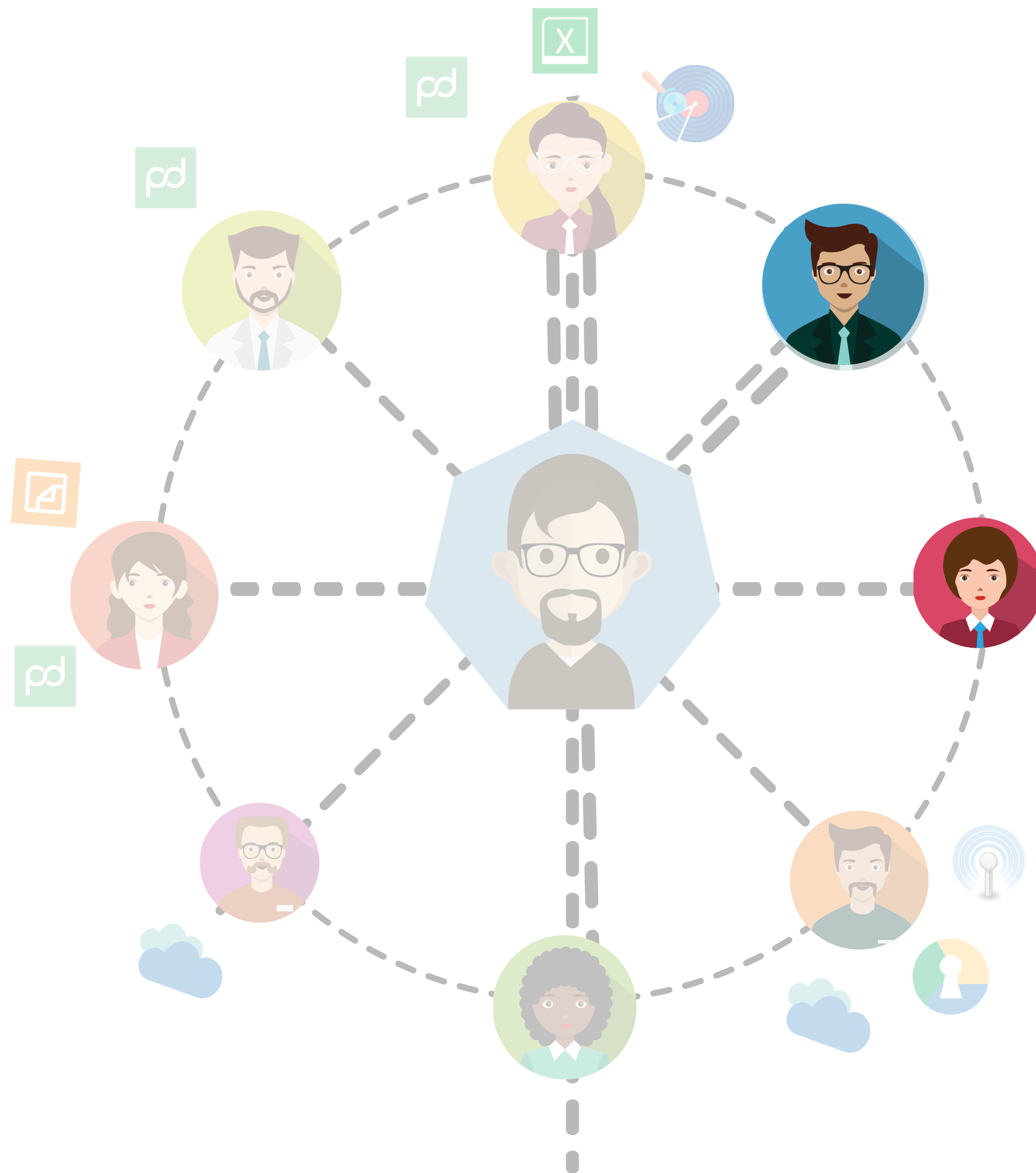
131 Google Drive apps

## Data:

- anonymized user ids
- anonymized file ids
- list of collaborators per file
- list of apps per user
- vendor for each app



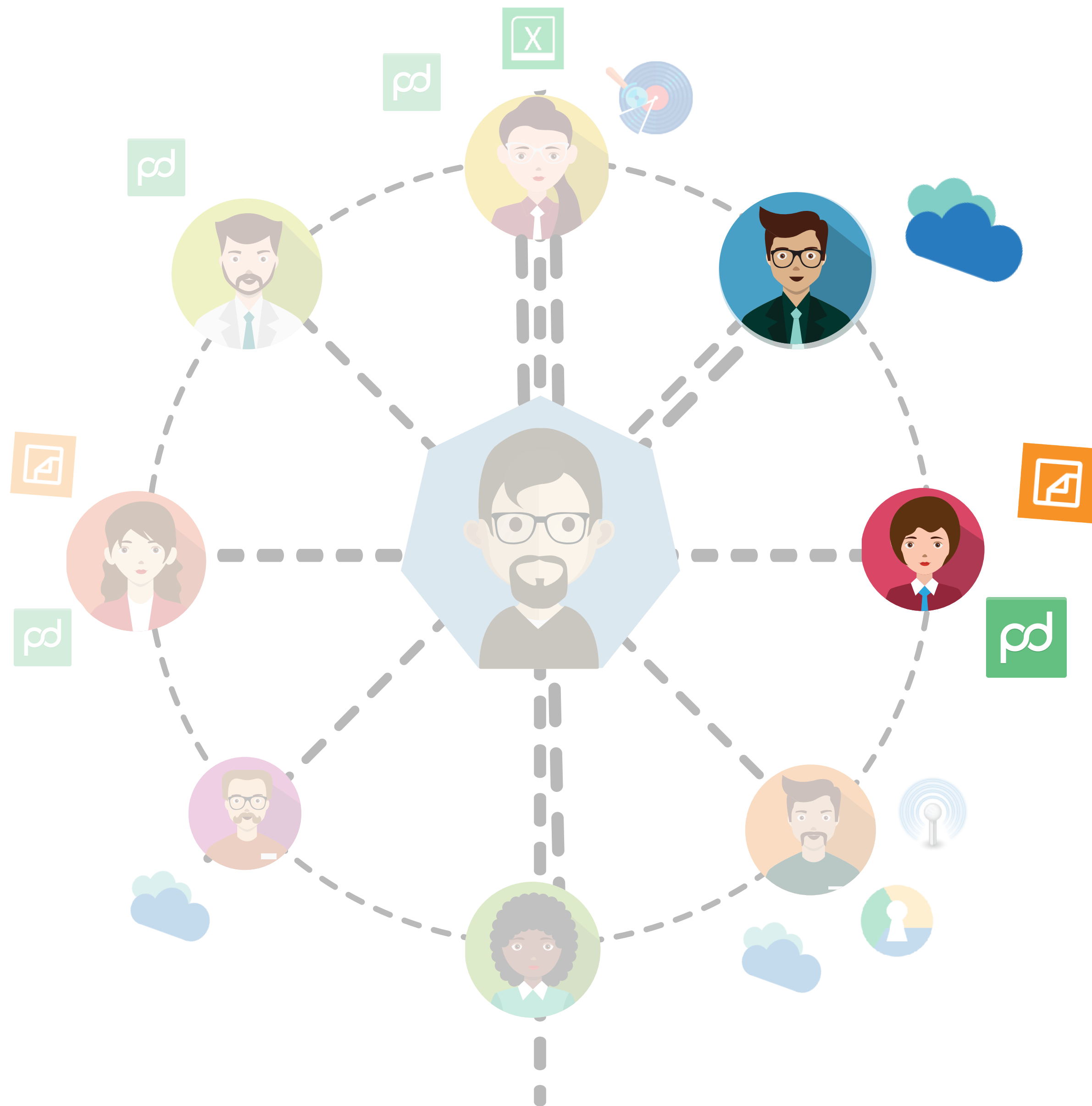
# Anonymized Dataset



**For collaborators not in the dataset, assign the apps of a random user from the dataset**



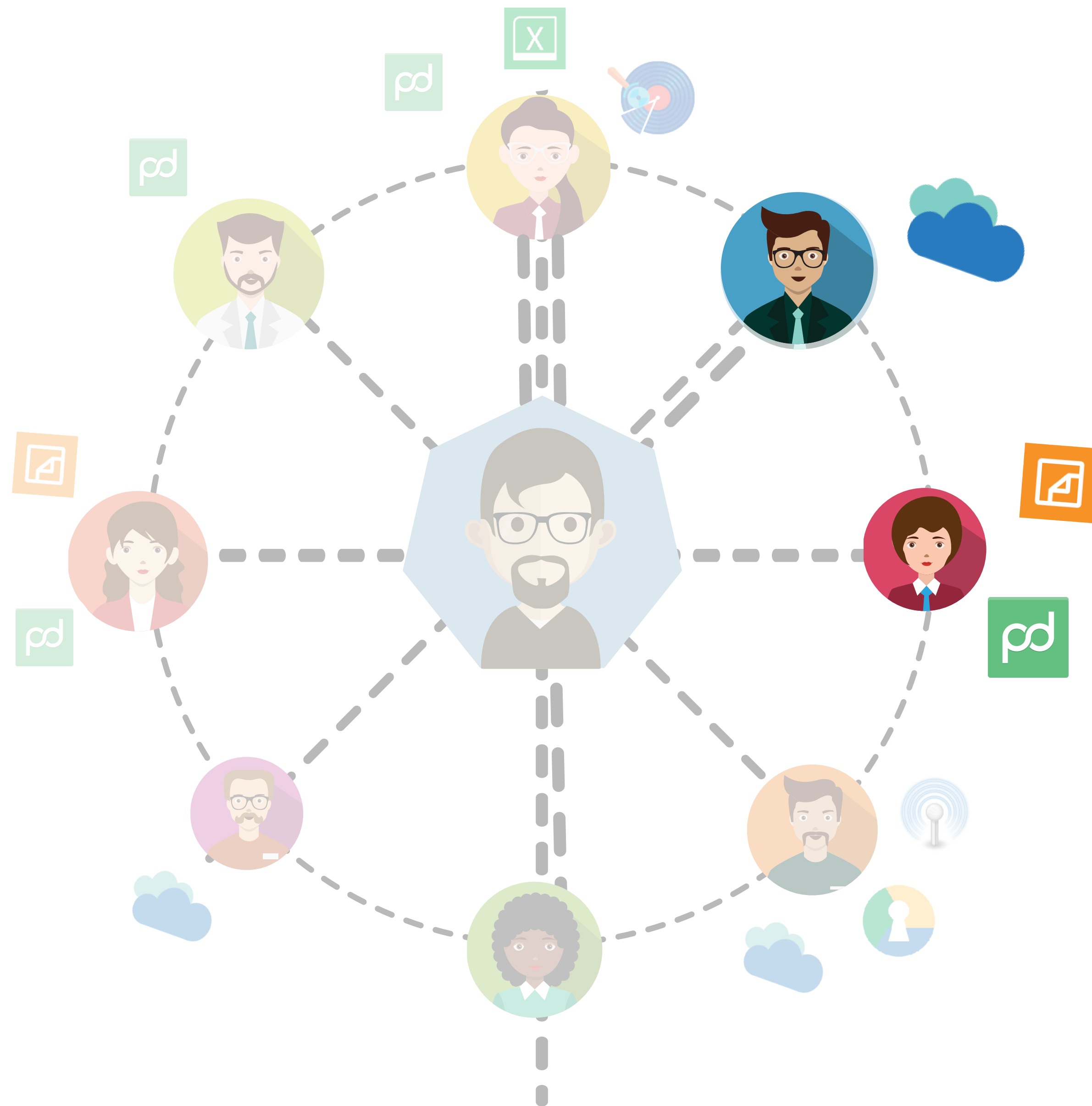
# Anonymized Dataset



For collaborators not in the dataset, assign the apps of a random user from the dataset



# Anonymized Dataset



For collaborators not in the dataset, assign the apps of a random user from the dataset

**3,422 users+collaborators**

# Metric: Vendors' File Coverage

for a user  $u$  and set  $V$  of vendors

$$VFC_u(V) = \sum_{v \in V} \frac{|F_{u,v}|}{|F_u|}$$

$|F_{u,v}|$       number of files of user  $u$  accessible by vendor  $v$

$|F_u|$       number of files of user  $u$



# Why VFC?

**Easy to relay to the user:**

**"the percentage of your files accessible by the company"**

# Why VFC?

**Easy to relay to the user:**

**"the percentage of your files accessible by the company"**

**Does not need external vendor evaluations:**

**e.g. based on reputation, number of installations, etc**

# Metric: **Vendors' File Coverage**

Part due to the **users'** decisions

$$Self-VFC_u = VFC_u(V_u)$$

$V_u$  vendors authorized by the user

# Metric: **Vendors' File Coverage**

Part due to the **users'** decisions

$$Self-VFC_u = VFC_u(V_u)$$

$V_u$  vendors authorized by the user

Part due to the **collaborators'** decisions

$$Collaborators-VFC_u = VFC_u(V_{C(u)})$$

$V_{C(u)}$  vendors authorized by the collaborators

# Metric: Vendors' File Coverage

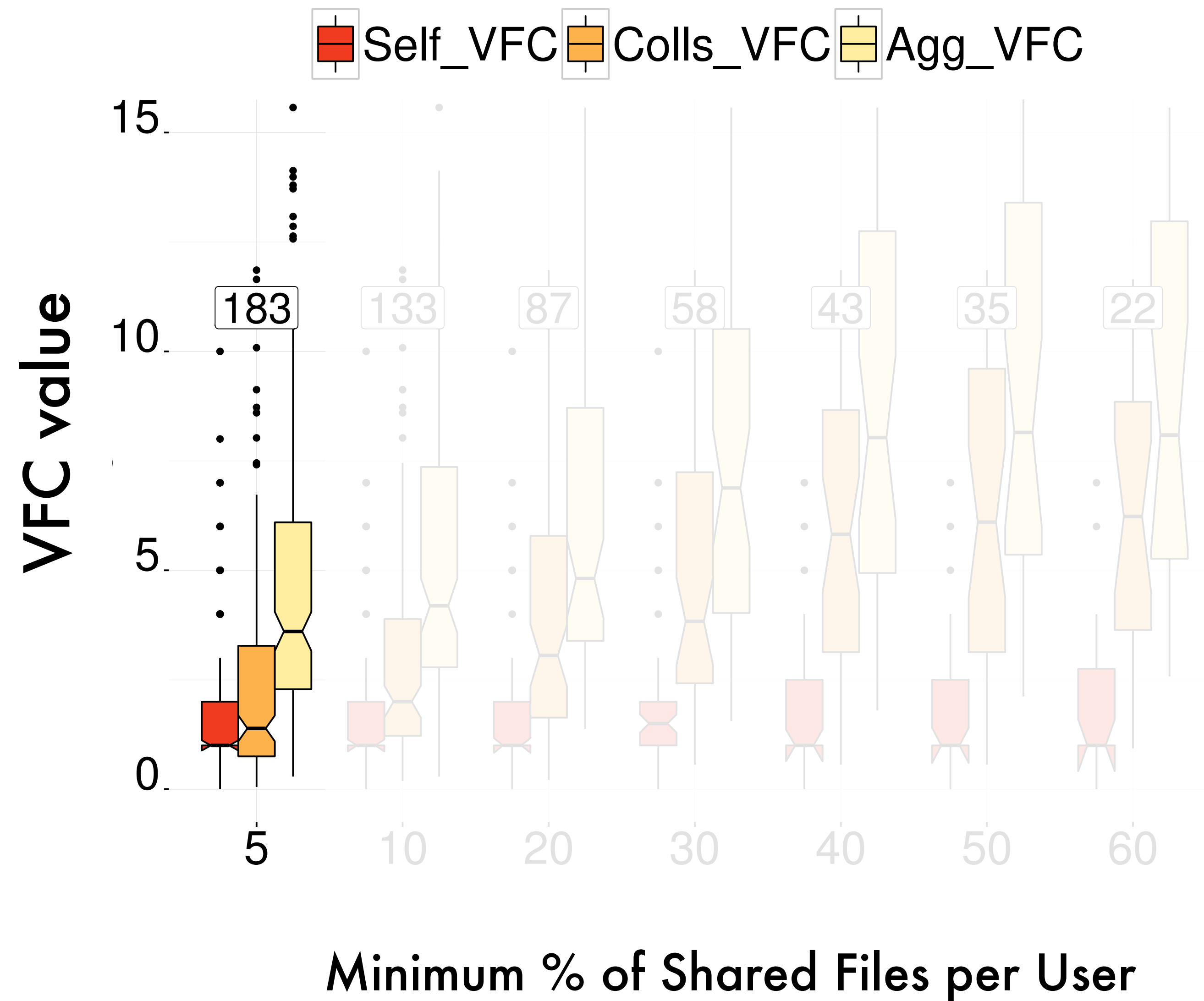
Combined metric due to the **users'** and the **collaborators**

$$\textit{Aggregate-VFC}_u = \textit{VFC}_u(V_u \cup V_{C(u)})$$

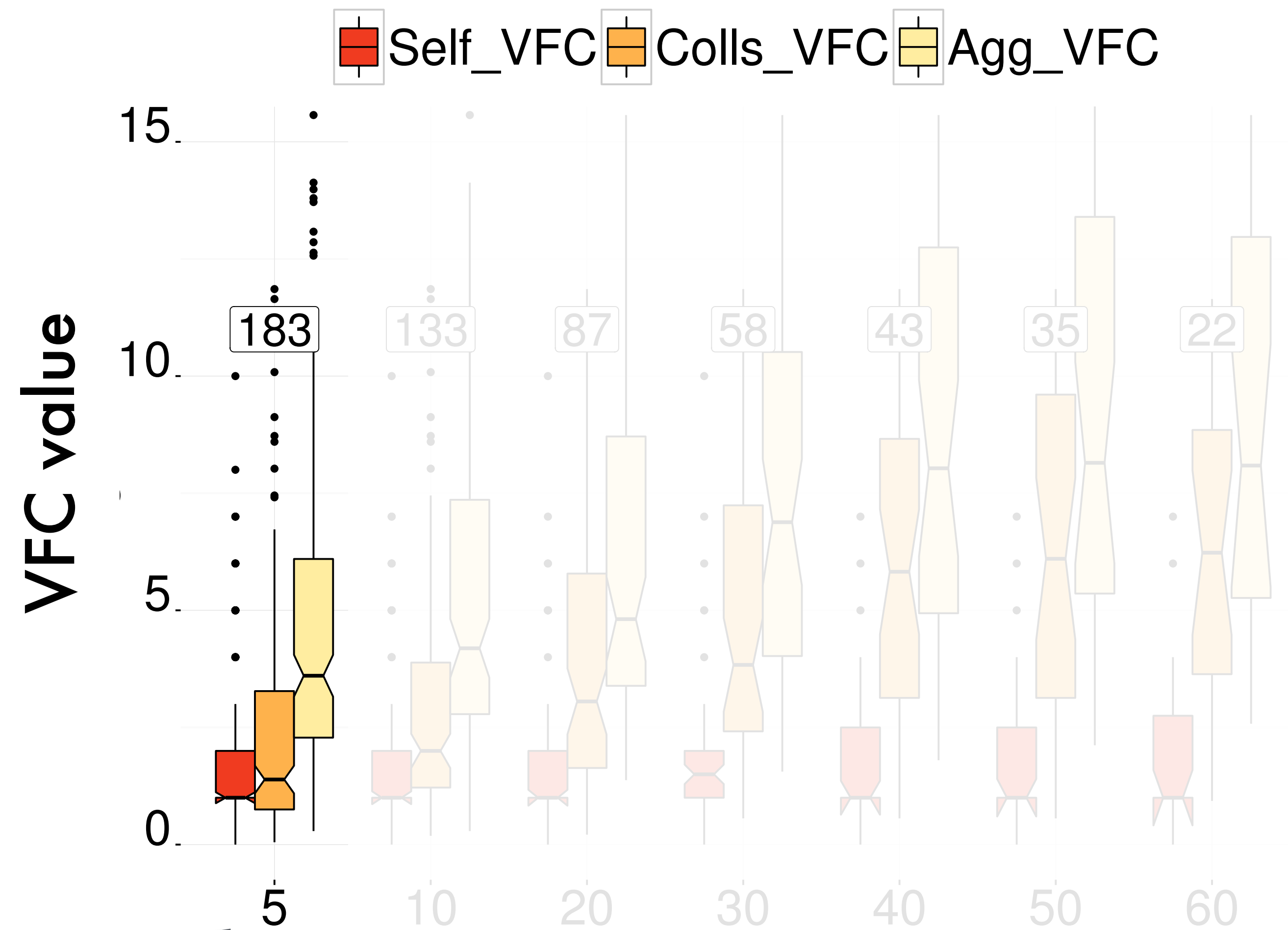
$V_u$  vendors authorized by the user

$V_{C(u)}$  vendors authorized by the collaborators

# How the *VFC* evolves with more shared data:



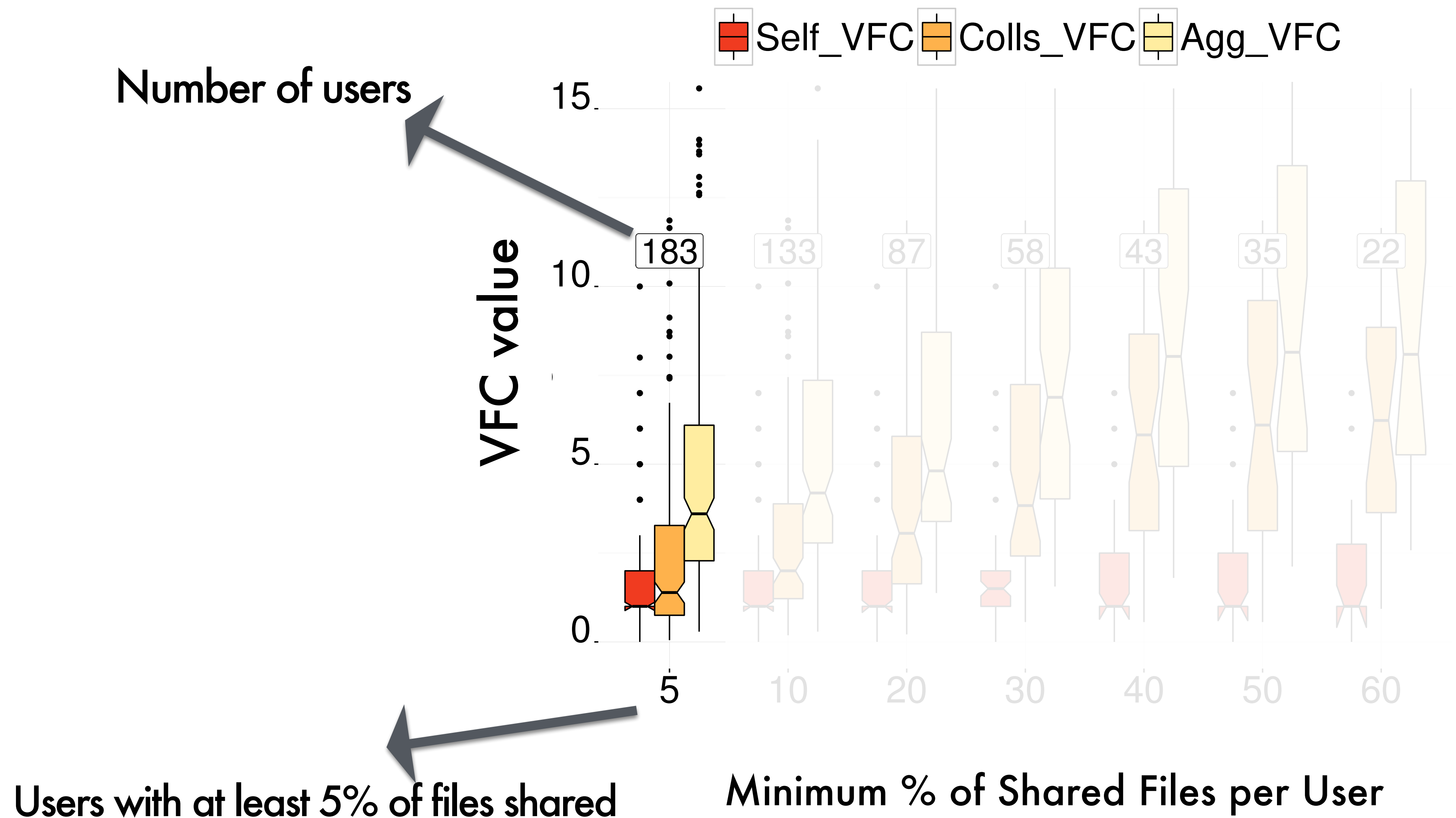
# How the *VFC* evolves with more shared data:



Users with at least 5% of files shared

Minimum % of Shared Files per User

# How the *VFC* evolves with more shared data:





How the VFC value

5

Number of users

VFC value

0

5

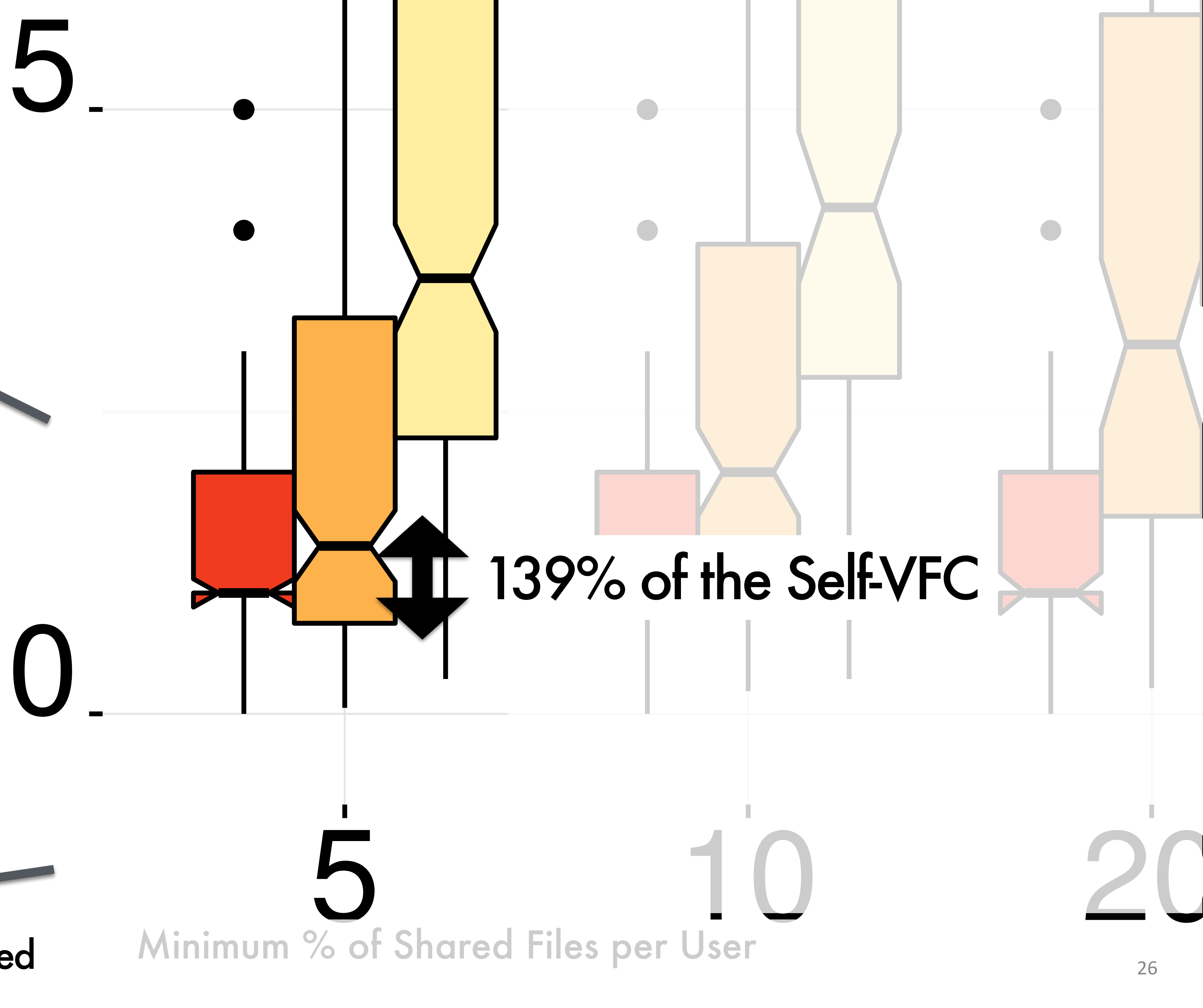
10

20

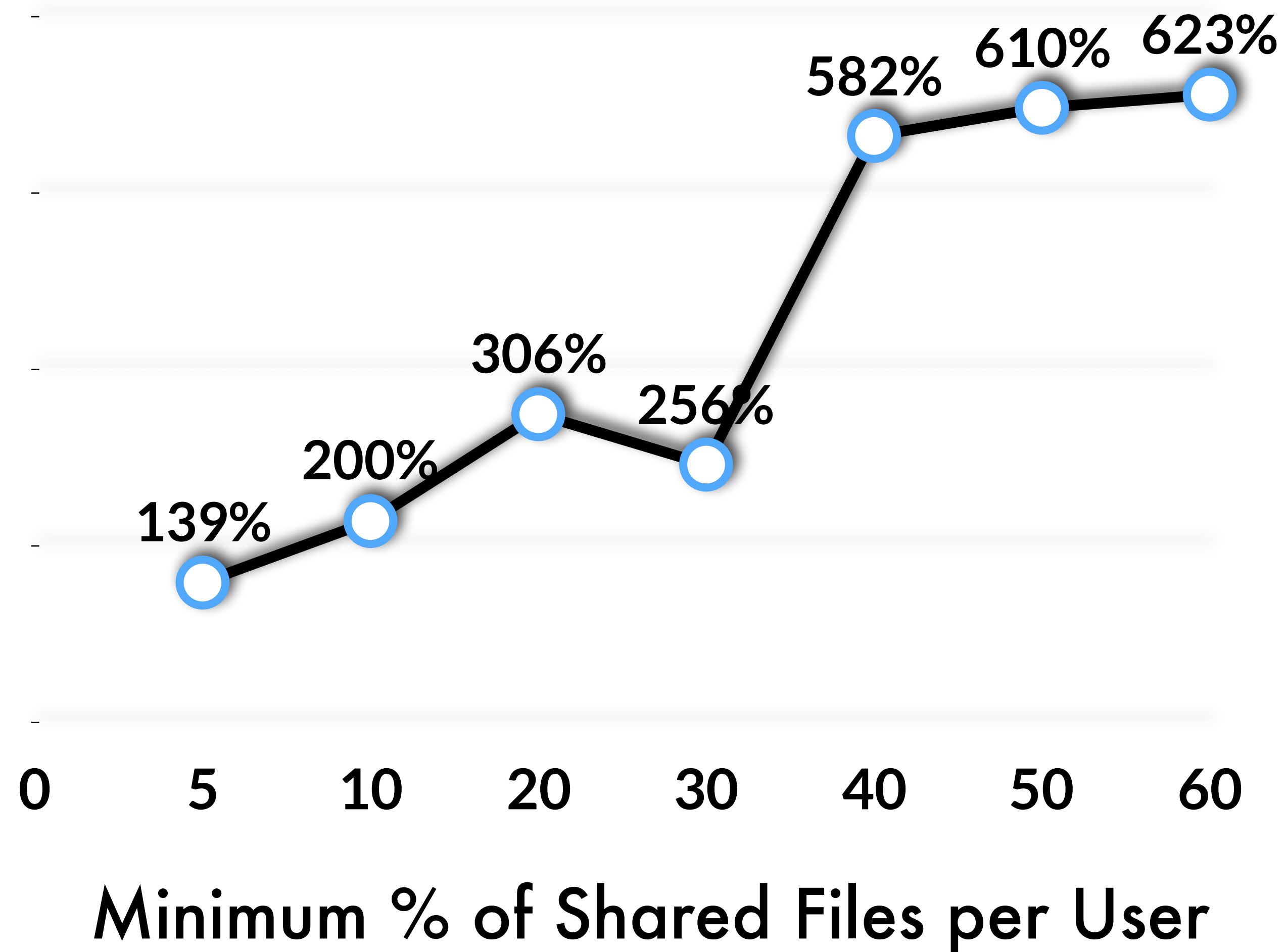
Users with at least 5% of files shared

Minimum % of Shared Files per User

139% of the Self-VFC



**% Privacy Loss** due to Collaborators relative to that due to the User



# Research Question-1

How significant is the impact of **collaborators'** app adoption decisions on the **users' privacy loss**?

Collaborators' decisions are highly significant.

They become more important as the sharing frequency increases

# Research Question-2

Do Current **Permission Models**  
help user **minimize** the VFC?

How can we **improve** them?

# Current Permission Models



## Audio Cutter

offered by [mp3cut.net](http://mp3cut.net)

★★★★★ (1313)

[Music & Radio](#)

385,203 users

VISIT WEBSITE

OVERVIEW

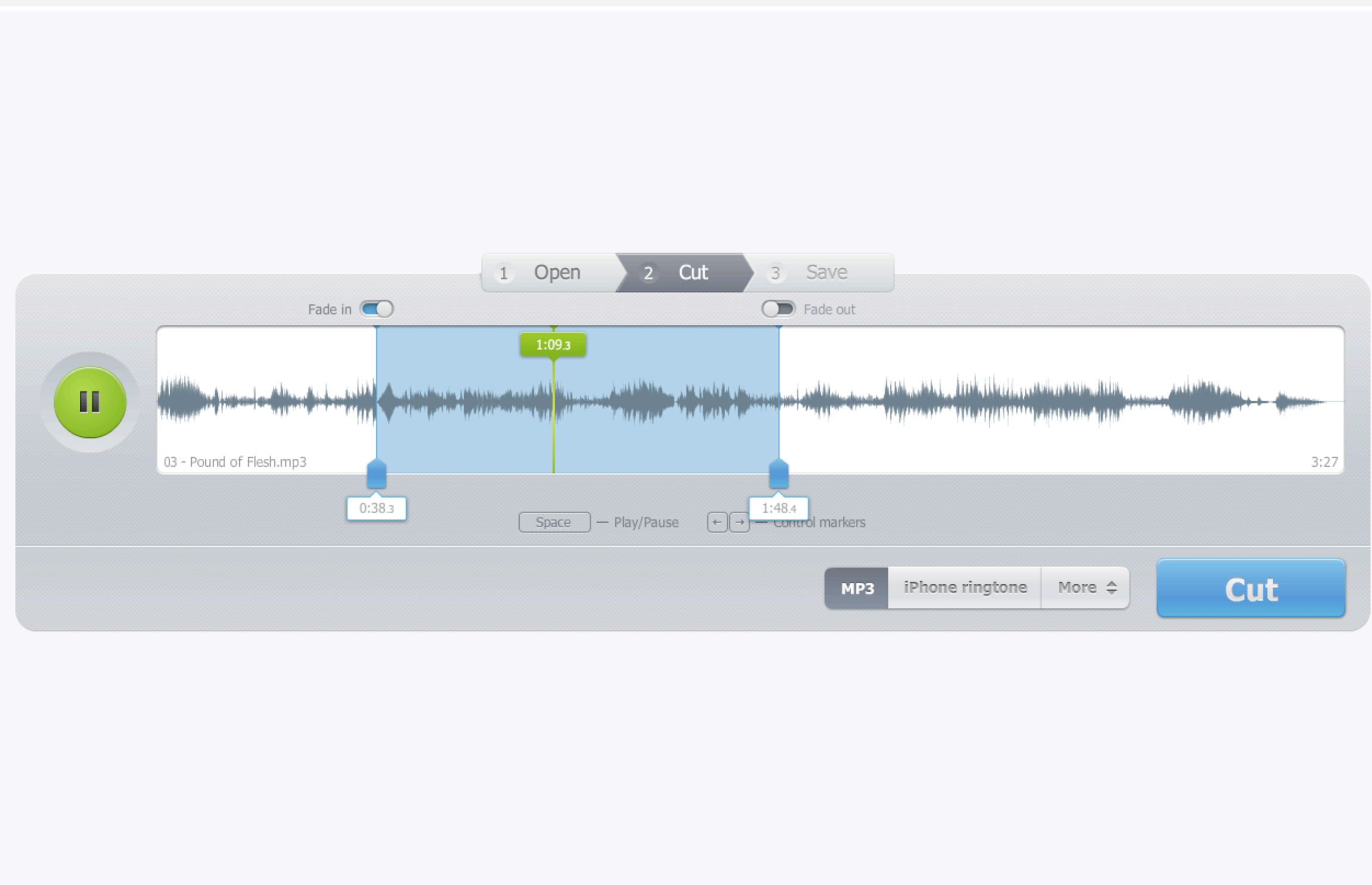
REVIEWS

SUPPORT

RELATED

G+1

863



Works with Google Drive

Compatible with your device

The easiest way to cut out a piece of music

Online Audio Cutter allows you to effortlessly cut out a desired musical fragment from an MP3 file or a file in other formats, in order, for example to set it up as a customized ringtone for your mobile phone. Our web application is free; it was designed for a single purpose, which makes it easy to use unlike complicated professional audio editors.

Features:

[Website](#)

[Report Abuse](#)

Additional Information

Version: 1.2.9

Updated: November 25, 2014

Size: 20.23KiB

Languages: See all 9

# Current Permission Models



## Audio Cutter

offered by [mp3cut.net](http://mp3cut.net)

★★★★★ (1313)

[Music & Radio](#)

385,203 users

VISIT WEBSITE

OVERVIEW

REVIEWS

SUPPORT

RELATED

G+1

863



Works with Google Drive

Compatible with your device

The easiest way to cut out a piece of music

Online Audio Cutter allows you to effortlessly cut out a desired musical fragment from an MP3 file or a file in other formats, in order, for example to set it up as a customized ringtone for your mobile phone. Our web application is free; it was designed for a single purpose, which makes it easy to use unlike complicated professional audio editors.

Features:

[Website](#)

[Report Abuse](#)

Additional Information

Version: 1.2.9

Updated: November 25, 2014

Size: 20.23KiB

Languages: See all 9



# Current Permission Models



▼ Online Audio Cutter would like to:



View and manage the files in your Google Drive



View the files in your Google Drive



View and manage Google Drive files and folders that you have opened or created with this app



Add itself to Google Drive



By clicking Allow, you allow this app and Google to use your information in accordance with their respective terms of service and privacy policies. You can change this and other [Account Permissions](#) at any time.

Deny

Allow

# **History-Based (HB) Insights Model**



# History-Based (HB) Insights Model

You are made aware of the **percentage** of your files that the **vendor** already has (i.e. the Aggregate VFC)

(due to **your** or your **collaborators'** decisions)



# Baseline **Permission Models**



Online Player wants to:



View your basic profile info.



Add itself to Google Drive.



View and manage the files in your Google Drive.



# HB Insights Permission Models



Online Player wants to:



View your basic profile info.



Add itself to Google Drive.



View and manage the files in your Google Drive.



The app's company (**driveplayer.com**) already has access to:

**70%** of your files



Your friend Lisa has already installed an app from **driveplayer.com**. So this company already has access to the files you share with Lisa.

# HB Insights **Permission Models**



Online Player wants to:



View your basic profile info.



Add itself to Google Drive.



Selecting the vendor with **maximum existing access** results in **minimizing the aggregate VFC**

(proof in the paper)

...

70% of your files



Your friend Lisa has already installed an app from **driveplayer.com**. So this company already has access to the files you share with Lisa.

# HB Insights **Permission Models**



Online Player wants to:



View your basic profile info.



Add itself to Google Drive.



Selecting the vendor with **maximum existing access** results in **minimizing the aggregate VFC**  
(proof in the paper)

...

70% of your files

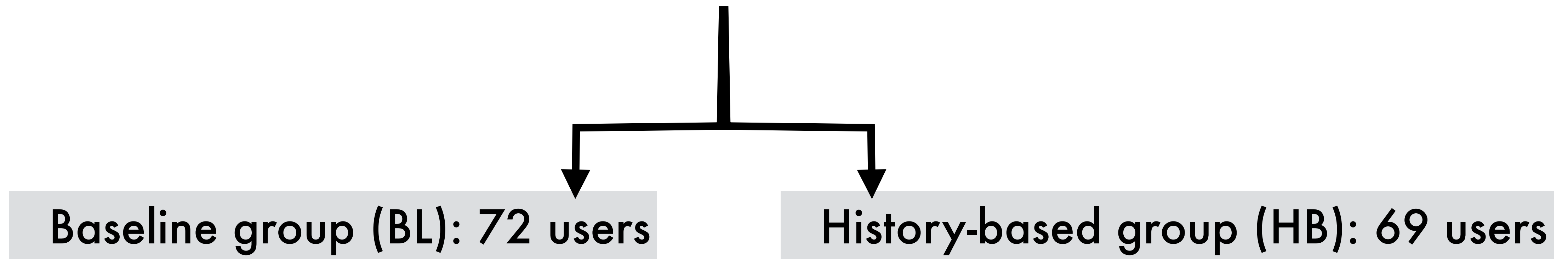


Your friend Lisa has already installed an app from **driveplayer.com**. So this company already has access to the files you share with Lisa.

# User study

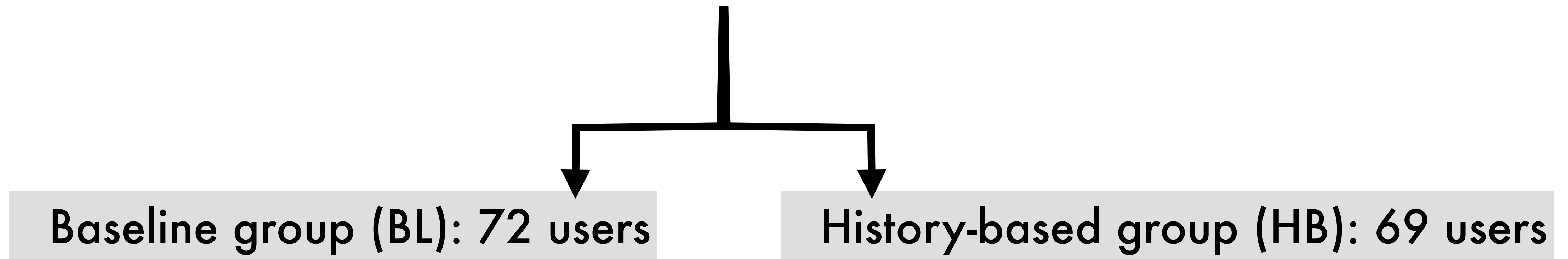
# Online Study Setup

Recruited users via **Crowdflower** (141 users)



# Online Study Setup

Recruited users via **Crowdfunder** (141 users)



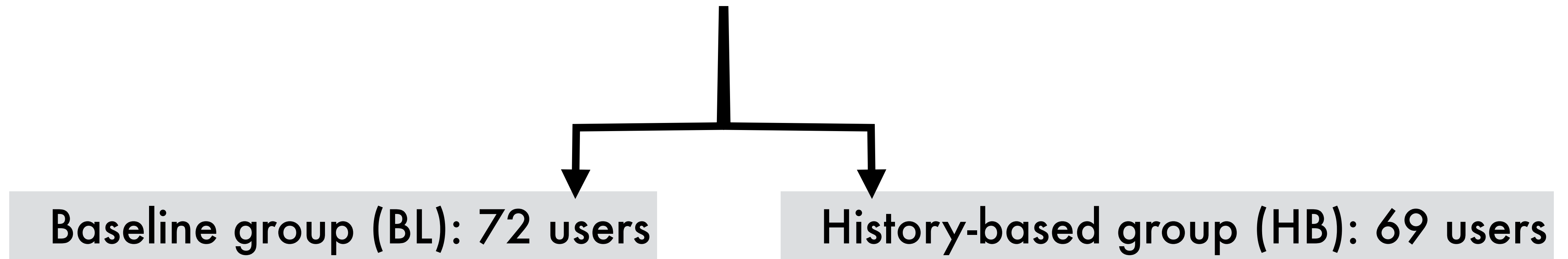
**Introductory Survey**





# Online Study Setup

Recruited users via **Crowdfunder** (141 users)



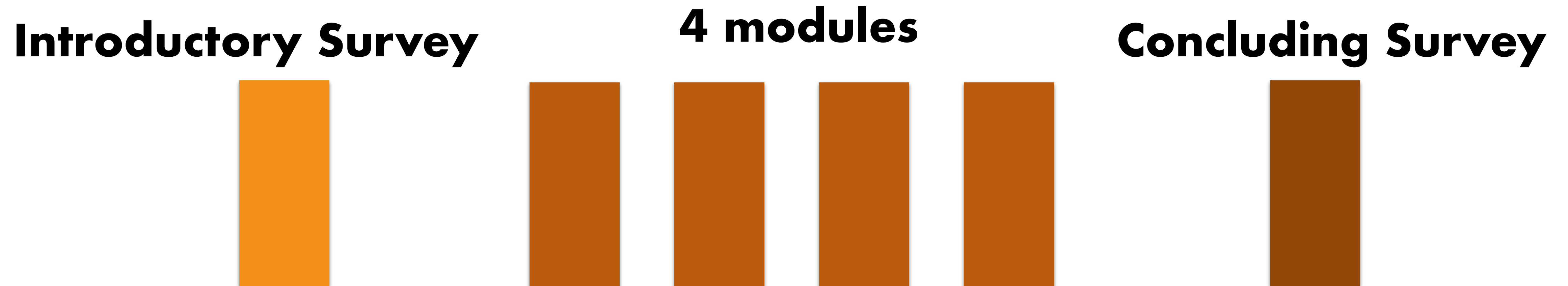
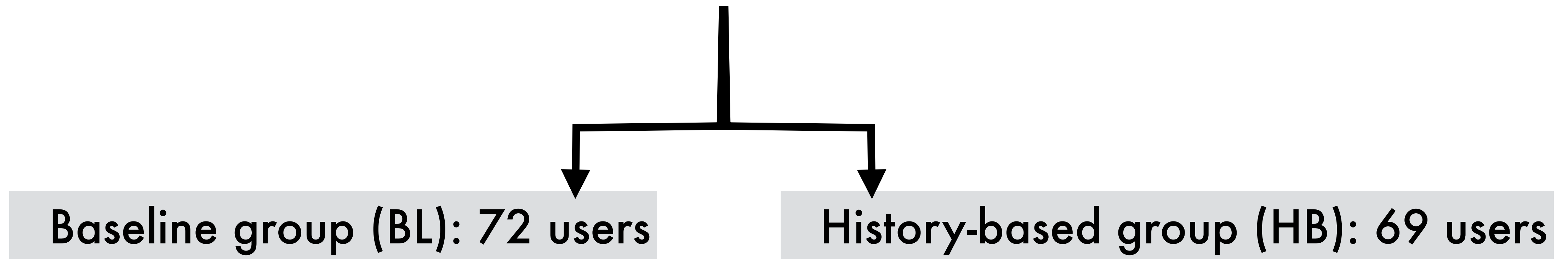
**Introductory Survey**

**4 modules**



# Online Study Setup

Recruited users via **Crowdfunder** (141 users)



# Demographics

<b>Age</b>	18-62	(median 31 years)
<b>Gender</b>	35.5%	Female
	64.5%	Male
<b>Occupation</b>	59.6%	full-time employees
	14.2%	student
	6.4%	part-time worker
	8.5%	self-employed
	5.0%	homemaker
	6.4%	Unemployed/retired
<b>IT Experience</b>	41.8%	Have worked or studied in IT
<b>Degree</b>	19.1%	High school
	7.1%	Trade/tech./vocational training
	51.1%	Associate or Bachelor's degree
	22.7%	Post Graduate Degree
<b>Countries</b>	35.0%	USA
	37.5%	IND
	7.5%	GBR
	6.9%	DEU
	6.9%	CAN
	7.4%	AUS+IRL+ NLD + PAK

# Module 1

How likely are users to select an app from the **same company** they installed from before?

You installed



by: Company 1

# Module 1

How likely are users to select an app from the **same company** they installed from before?

You installed  . Would you choose:  
by: Company 1

  
by Company 1

OR ?

  
by Company 2

# Module 1 - HB Group

## Task:

As explained, we now start from scratch. Consider that this is the first app you will install. Please install any application from the company: **thetimetube.com**. (Only one such app exists, and you can click on the app to view its info.)



### Video to GIF Converter

**Company:** thetimetube.com

**Description:** This app allows you to create animated GIFs from videos directly. You can open a video file from your Google Drive and computer.



### Online Audio Converter

**Company:** online-audio-converter.com

**Description:** Convert audio files on your Google Drive from any format to another



### NitroSafe

**Company:** nitrosafe.org

**Description:** Malware scanning for Google Drive: Searches for malware, viruses, trojans and other nasty files in your Google Drive.



### PDF Mergy

**Company:** mytools.com

**Description:** Allows to merge PDF files from your Google Drive with a simple interface.

# Module 1 - HB Group

## Task:

As explained, we now start from scratch. Consider that this is the first app you will install. Please install any application from the company: **thetimetube.com**. (Only one such app exists, and you can click on the app to view its info.)



### Video to GIF Converter

**Company:** thetimetube.com

**Description:** This app allows you to create animated GIFs from videos directly. You can open a video file from your Google Drive and computer.



### Online Audio Converter

**Company:** online-audio-converter.com

**Description:** Convert audio files on your Google Drive from any format to another



### NitroSafe

**Company:** nitrosafe.org

**Description:** Malware scanning for Google Drive: Searches for malware, viruses, trojans and other nasty files in your Google Drive.



### PDF Mergy

**Company:** mytools.com

**Description:** Allows to merge PDF files from your Google Drive with a simple interface.



# Results

Probability of installing the **more privacy preserving** option

50%

Baseline

75.4%

History-based

Fisher's exact test

p-value = 0.003



# Results

**Baseline groups participants looked for other reasons:**

(more comprehensive description, more professional logo, a better sounding name, or a more trustable URL)

**Difficulty of remembering the previous vendors**

(e.g. 2 participants justified by mentioning that the app comes from the same vendor, but actually selected a different one)

# Module 2

How likely are users to select the **same app** that their **collaborator** have used before?

# Module 2

How likely are users to select the **same app** that their **collaborator** have used before?

Your friend  
John  
installed



by: Company 1

# Module 2

How likely are users to select the **same app** that their **collaborator** have used before?

Your friend  
John  
installed



by: Company 1

. Would you choose:

OR

?



by: Company 1



by: Company 2

# Results

Probability of installing the **more privacy preserving** option

52.8%

Baseline

88.4%

History-based

Fisher's exact test

p-value < 0.001

# Results



# Results

**Quote:** “Mytools.com is the maker of PDF Mergy and since they already have ALOT of access to my files (thanks to John), might as well stick with the brand and not open up more files to another company.”



# Module 3

How likely are users to select an app from the **same vendor** that their **collaborator** have used before?

Your friend  
Lisa  
installed



Company 1



# Module 3

How likely are users to select an app from the **same vendor** that their **collaborator** have used before?

Your friend  
Lisa  
installed



Company 1

. Would you choose:



Company 1

OR

?



Company 2

# Results

Probability of installing the **more privacy preserving** option

58.3%

Baseline

82.6%

History-based

Fisher's exact test

p-value = 0.002

# Module 4

How likely are users to consider the **differences** in access levels of vendors that **collaborators** authorized?

# Module 4

Your friend  
**Lisa**  
installed



Company 1

# Module 4

Your friend  
**Lisa**  
installed



Company 1

Your friend  
**John**  
installed



Company 2

# Module 4

Your friend  
**Lisa**  
installed



Company 1

Your friend  
**John**  
installed



Company 2

Your share  
more files  
with **Lisa**

# Module 4

Your friend  
**Lisa**  
installed



Company 1

Your friend  
**John**  
installed



Company 2

Your share  
more files  
with **Lisa**

. Would you choose:



Company 1

OR

?



Company 2

# Results

Probability of installing the **more privacy preserving** option

44.4%

Baseline

82.6%

History-based

Fisher's exact test

p-value = <0.001



# Results

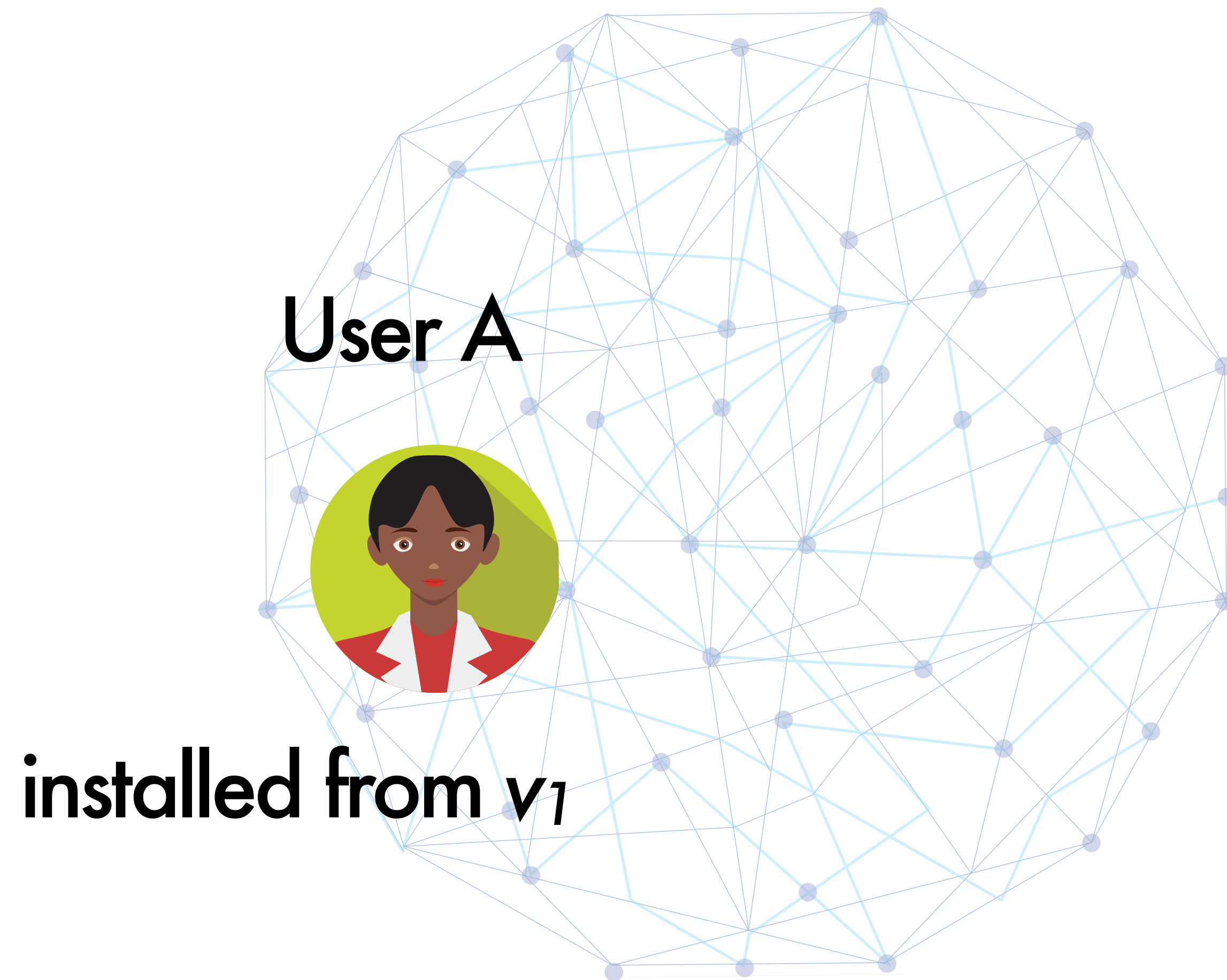
**Quote:** “This is the app that John already uses, and he has access to all of my files. The PDF Mergy app is used by Lisa, but she only has access to part of my files.”



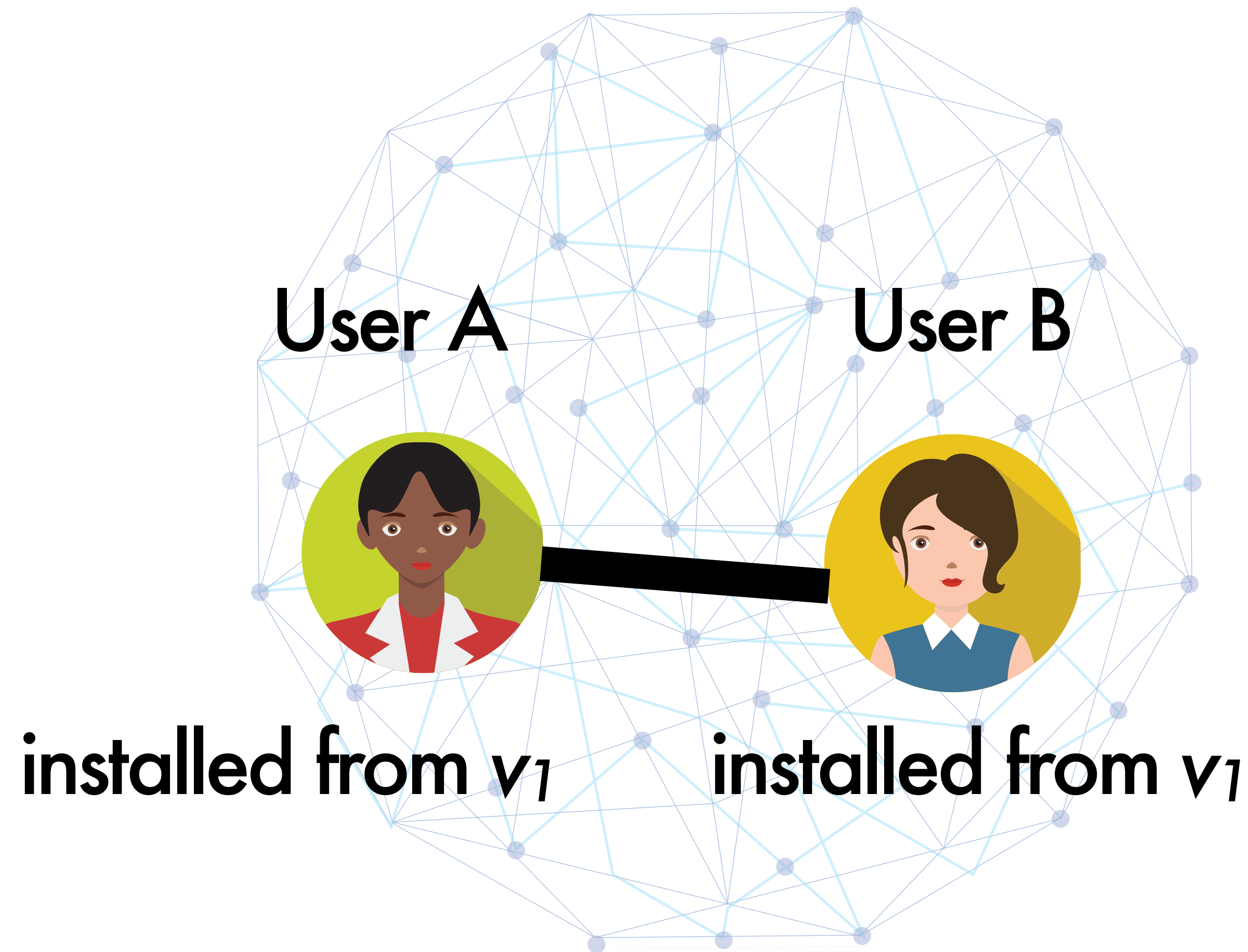
# Research Question-3

How do the History-based Decisions  
affect users' privacy at **scale**?

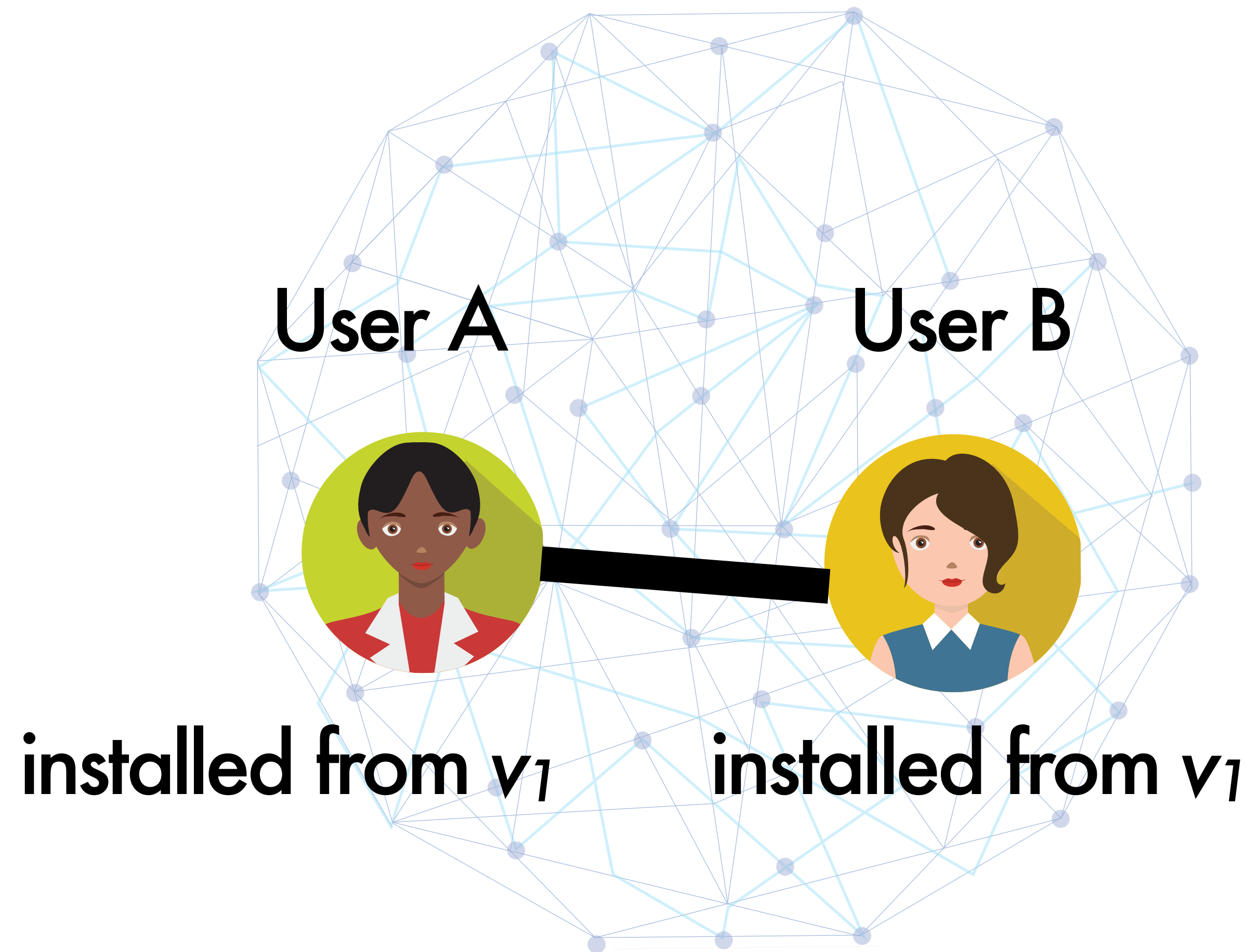
# Idea: Study the network effect of History-based decisions



# Idea: Study the network effect of History-based decisions

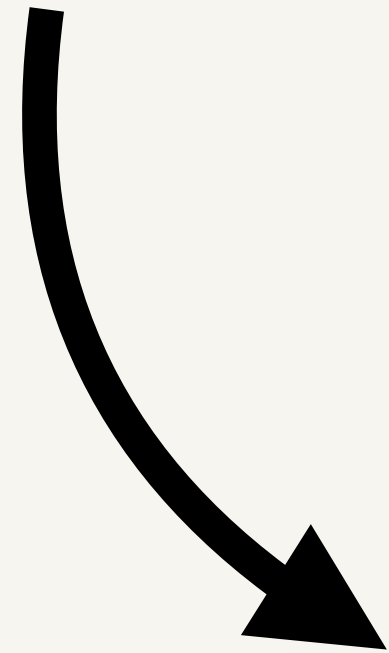


# Idea: Study the network effect of History-based decisions



VFC for User A increases by 0

**Issue:** We do not have access to large networks of cloud storage and apps' users.



**Create networks with similar properties.**

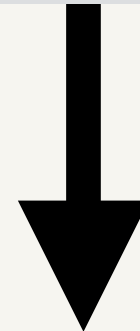
# Network 1: Inflated Google Drive Network

Start from PrivySeal users' network.



# Network 1: Inflated Google Drive Network

Start from PrivySeal users' network.

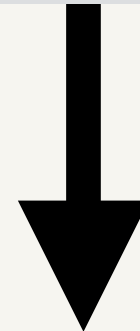


Extract the **degree distribution**.

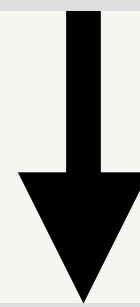


# Network 1: Inflated Google Drive Network

Start from PrivySeal users' network.



Extract the **degree distribution**.

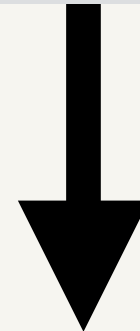


Create a large scale **connected** graph with a similar distribution using the **Configuration Model**\*.

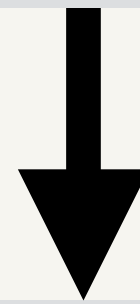
\*Newman. The structure and function of complex networks. SIAM review, 2003

# Network 1: Inflated Google Drive Network

Start from PrivySeal users' network.



Extract the **degree distribution**.



Create a large scale **connected** graph with a similar distribution using the **Configuration Model**\*.

- 18,000 users
- 138,440 edges
- average degree: 15

\*Newman. The structure and function of complex networks. SIAM review, 2003

# Network 2: Author Collaboration Network

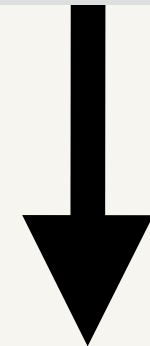


Rely on Microsoft Academic Graph  
(papers, authors, affiliations)

# Network 2: Author Collaboration Network



Rely on Microsoft Academic Graph  
(papers, authors, affiliations)

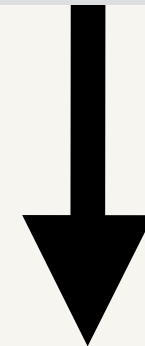


Use a snapshot of 50,000 papers to  
construct a collaboration graph

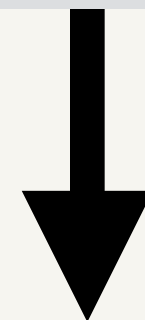
# Network 2: Author Collaboration Network



Rely on Microsoft Academic Graph  
(papers, authors, affiliations)



Use a snapshot of 50,000 papers to  
construct a collaboration graph

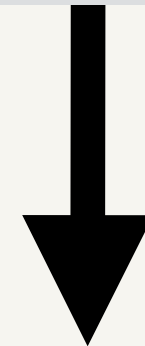


Obtain a graph

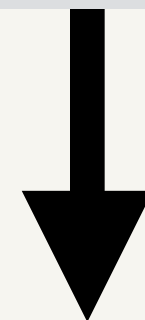
# Network 2: Author Collaboration Network



Rely on Microsoft Academic Graph  
(papers, authors, affiliations)



Use a snapshot of 50,000 papers to  
construct a collaboration graph



Obtain a graph

- 41,000 users
- 199,980 edges
- average degree: 4

# Network 3: Teams Collaboration Network

Take the previously constructed  
collaboration graph

# Network 3: Teams Collaboration Network

Take the previously constructed  
collaboration graph



Compute the Strongly Connected  
Components (SCC)

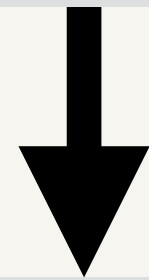


# Network 3: Teams Collaboration Network

Take the previously constructed  
collaboration graph



Compute the Strongly Connected  
Components (SCC)



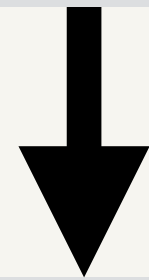
Each team is a SCC.

# Network 3: Teams Collaboration Network

Take the previously constructed collaboration graph



Compute the Strongly Connected Components (SCC)



Each team is a SCC.

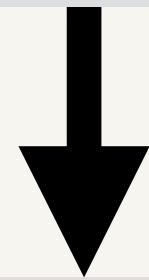
- 16,400 users
- 1700 teams

# Network 3: Teams Collaboration Network

Take the previously constructed collaboration graph



Compute the Strongly Connected Components (SCC)



Each team is a SCC.

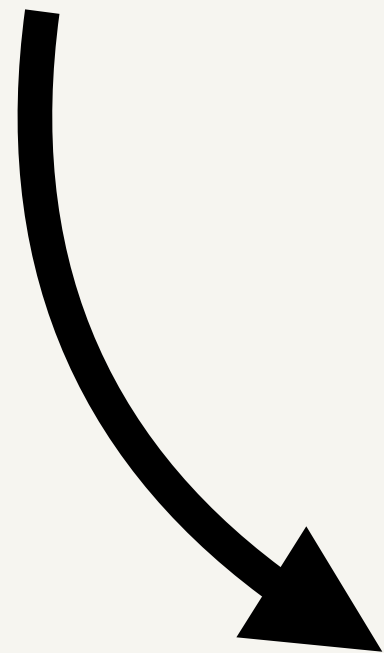
- 16,400 users
- 1700 teams

## **Assumption:**

Team members only account for other team members' decisions

# Keep simulation properties **realistic**:

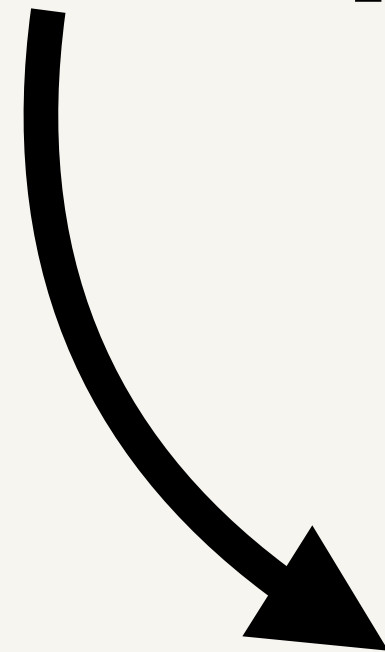
- Distribution of shared files
- Number of apps installed



Match the **PrivySeal** dataset

# Keep simulation properties **realistic**:

- Apps vendors
- App installation count



Select at random on each step among 1000  
apps from Chrome Store

# 3 User Models

Experimental Baseline Model (**EBL**)

Experimental HB Model (**EHB**)

Fully-Aware model (**FA**)  
(always selects the HB option)

# 3 User Models

## Experimental Baseline Model (**EBL**)

Users are self-interested and do not cooperate on app installation decisions.

Fully-Aware model (**FA**)  
(always selects the HB option)

# 3 User Models

Experimental Baseline Model (**EBL**)

**Users are self-interested and do not cooperate on app installation decisions.**

Fully-Aware model (**FA**)  
(always selects the HB option)



# Simulation Steps

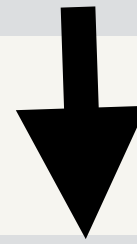
**Select a user**

(based on the installation frequency of the user)

# Simulation Steps

**Select a user**

(based on the installation frequency of the user)



**Select an app**

(based on the installation frequency of the app)

# Simulation Steps

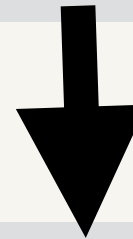
**Select a user**

(based on the installation frequency of the user)



**Select an app**

(based on the installation frequency of the app)



**Choose to install the app or one of its related apps**  
(based on the user model)

# Simulation Steps

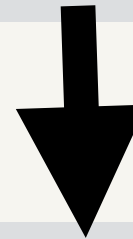
**Select a user**

(based on the installation frequency of the user)



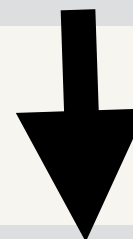
**Select an app**

(based on the installation frequency of the app)



**Choose to install the app or one of its related apps**

(based on the user model)

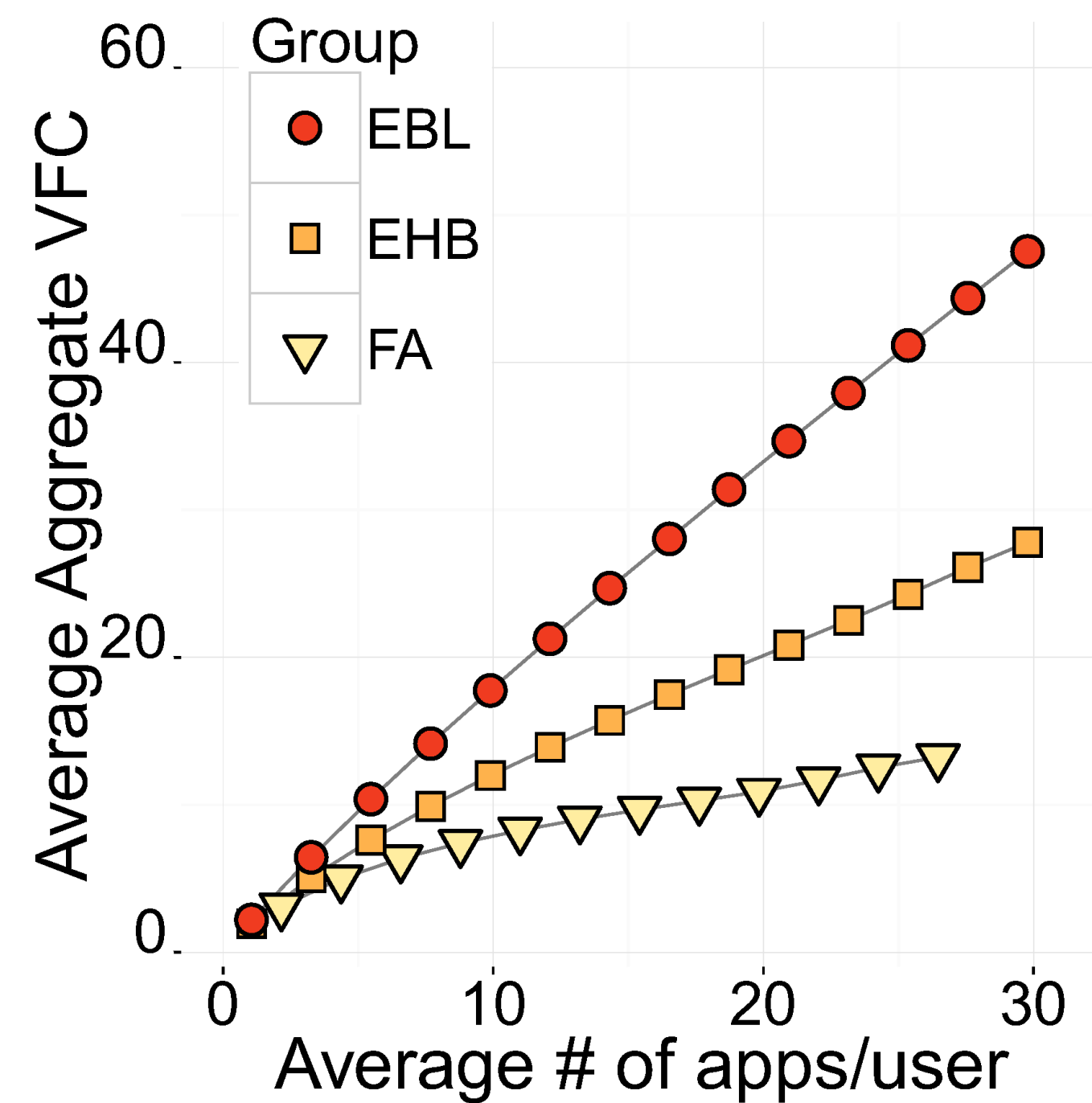


**Compute the average Aggregate VFC per user**

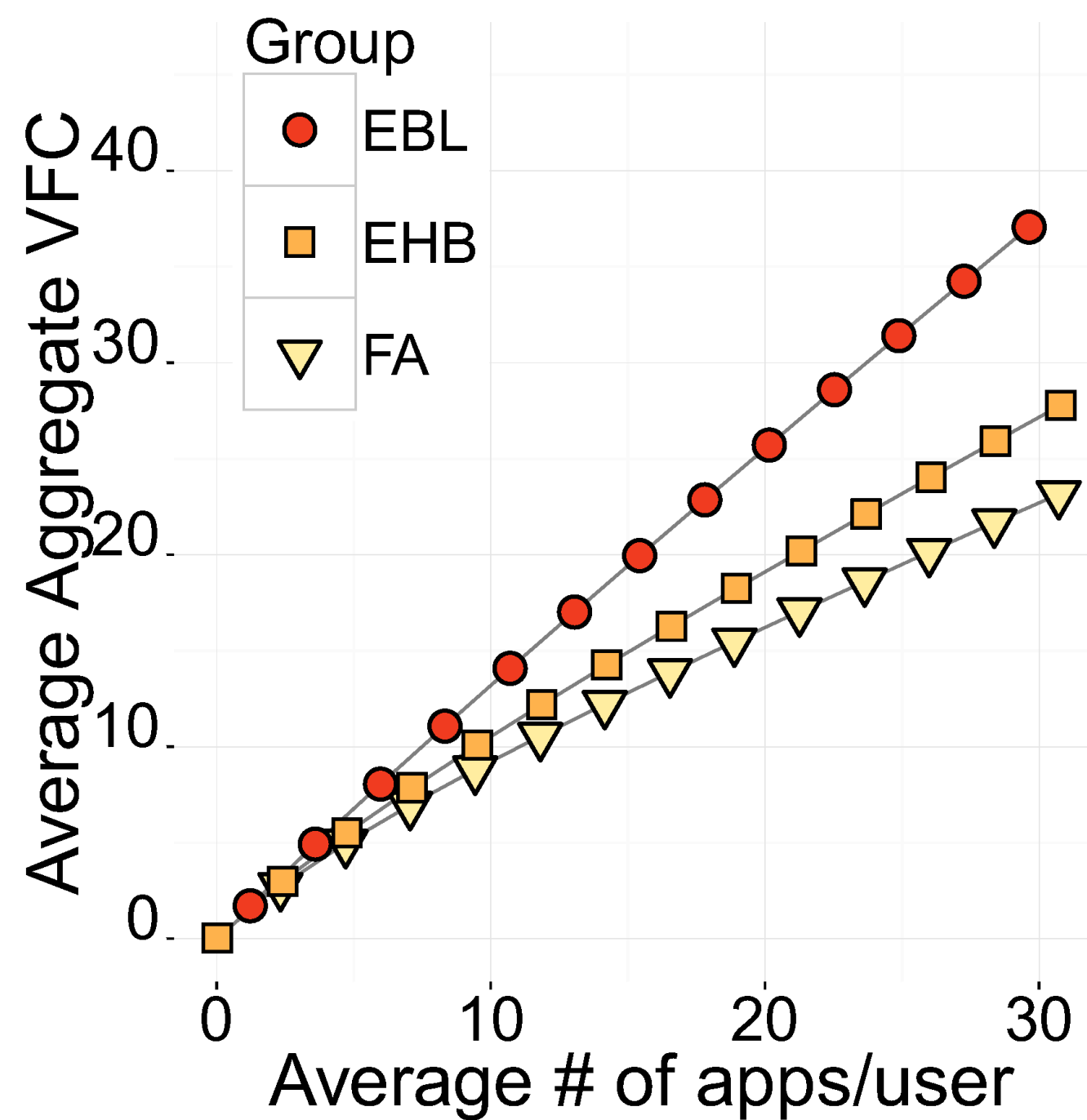
(based on the user model)

# Simulation Results

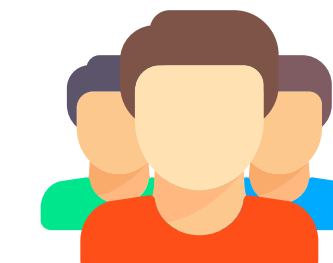
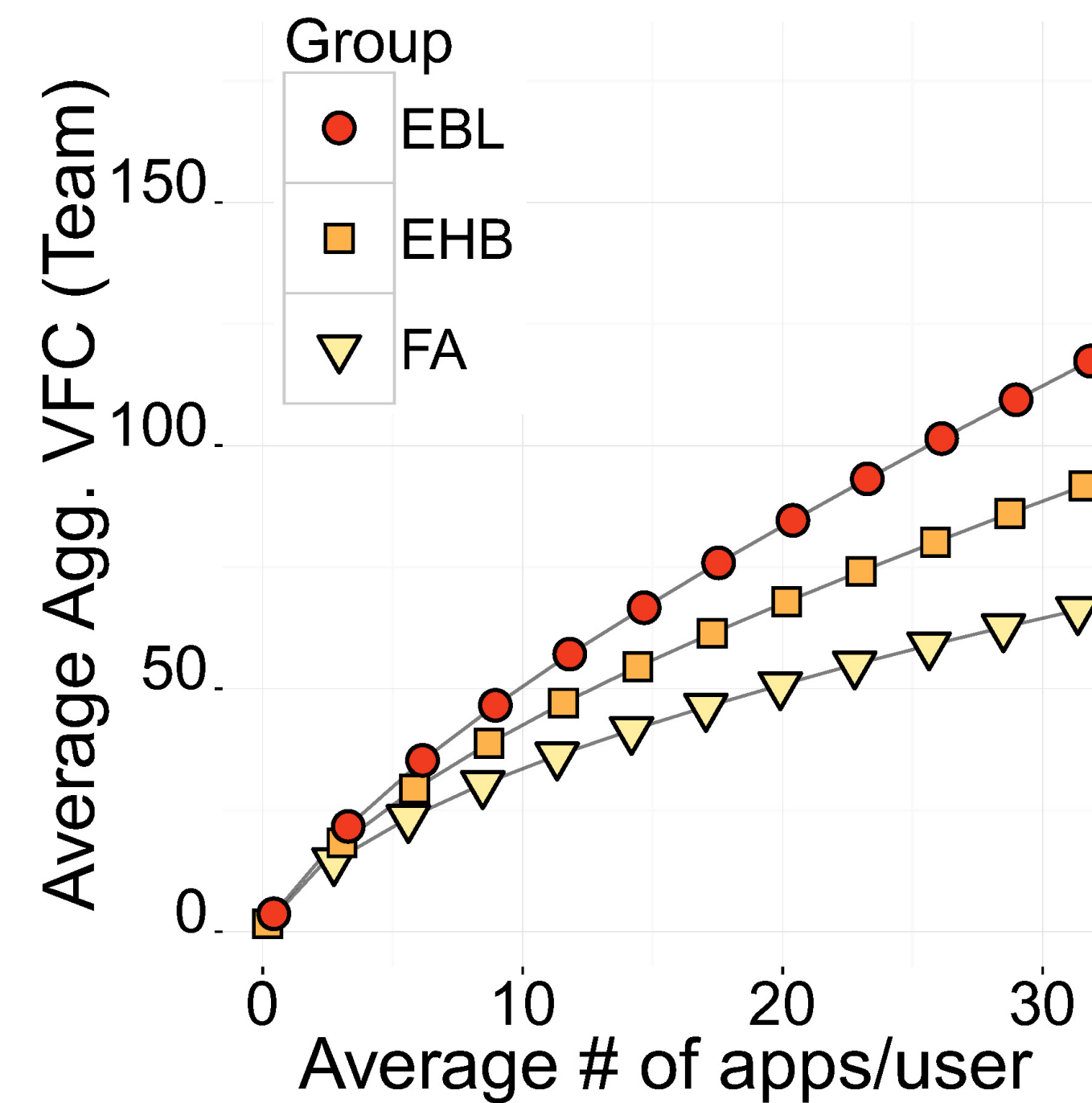
Growth of the Privacy Loss is **curtailed** as users install more apps.



Inflated Network



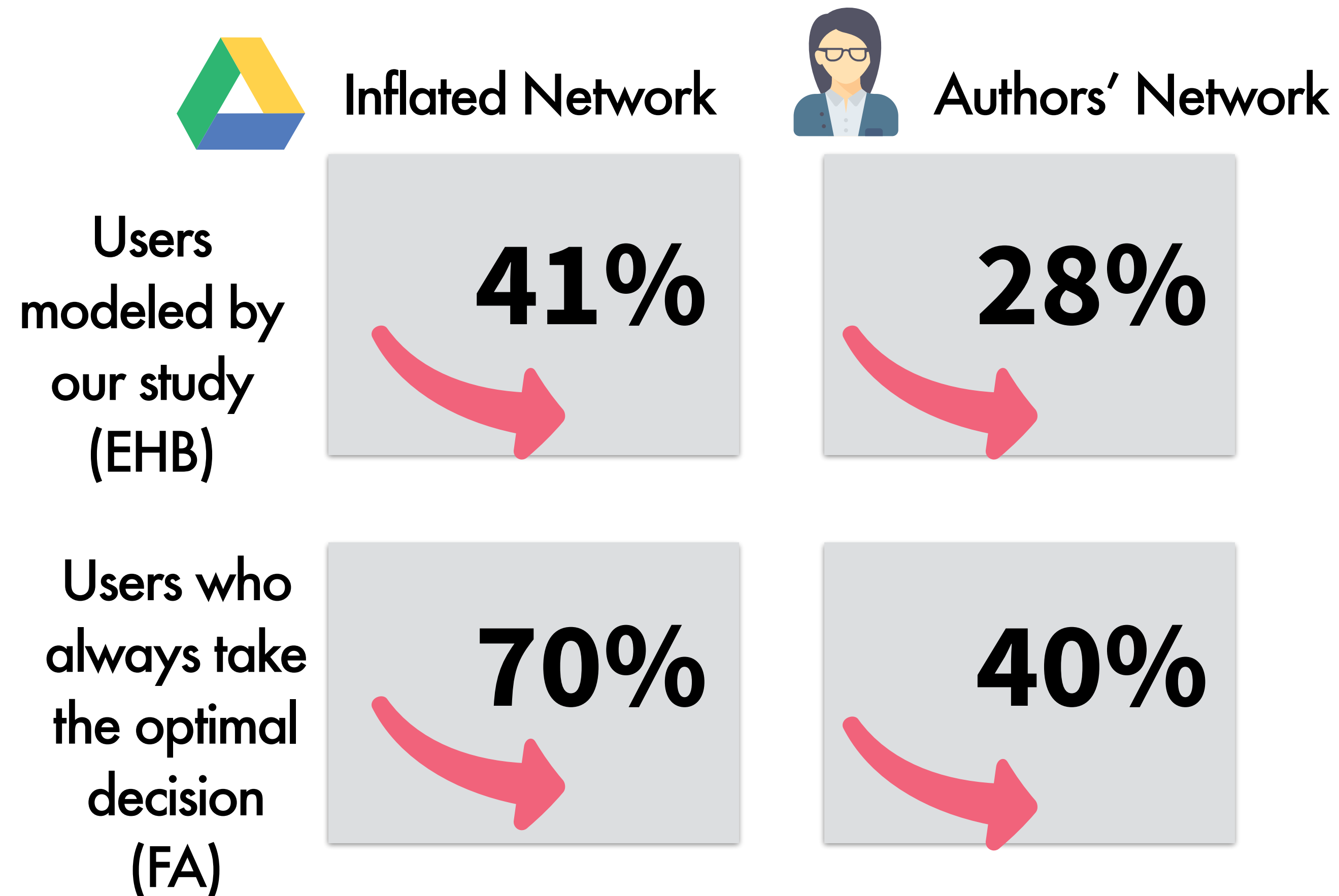
Authors' Network



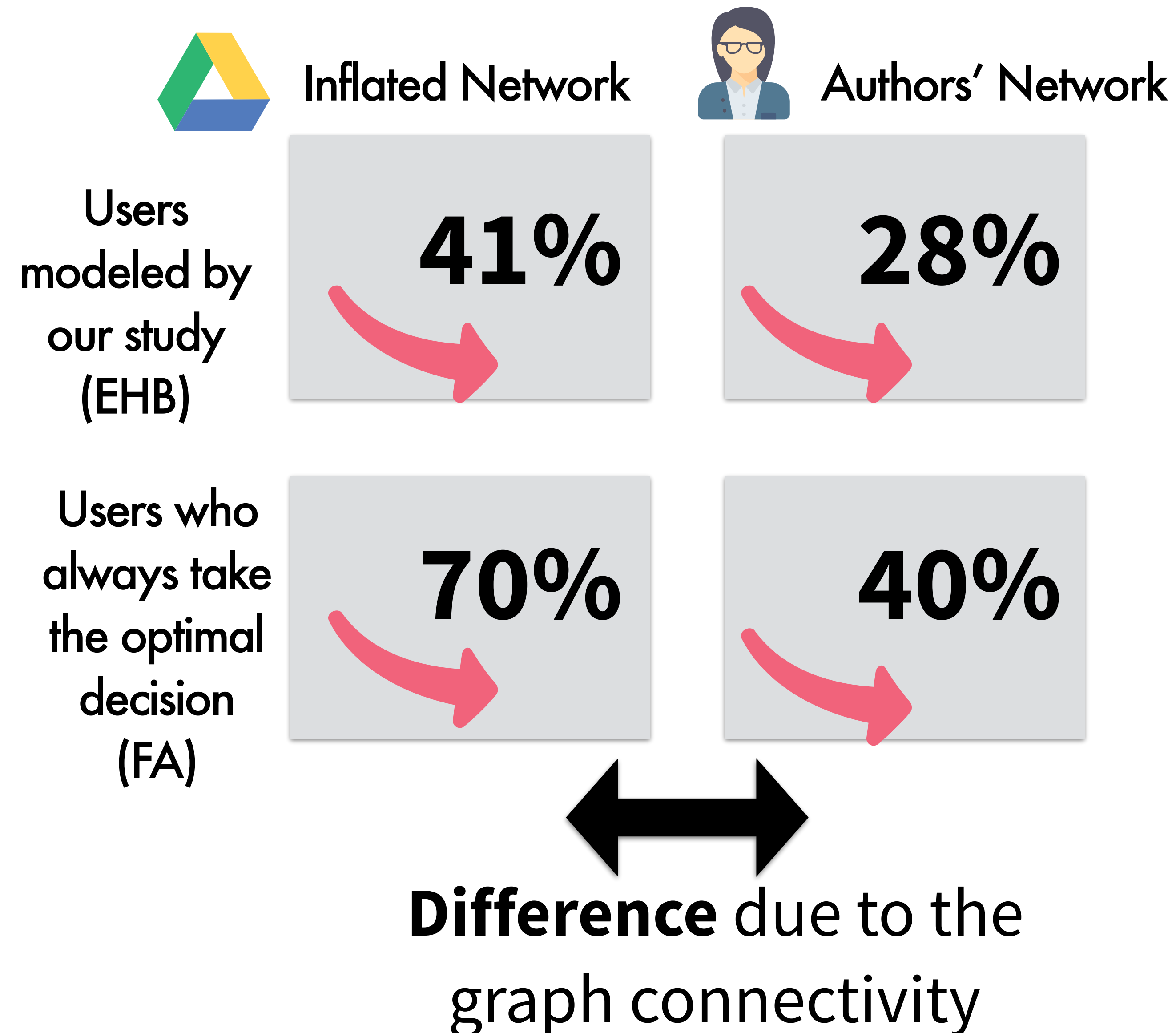
Teams' Network

Growth of the Average aggregate VFC with more apps installed by users.

How much is the **privacy loss reduction** by the **end** of the simulation  
(**w.r.t. the baseline (EBL)**)?



How much is the **privacy loss reduction** by the **end** of the simulation  
(**w.r.t. the baseline (EBL)**)?





How much is the **privacy loss reduction** by the **end** of the simulation  
(**w.r.t. the baseline (EBL)**)?



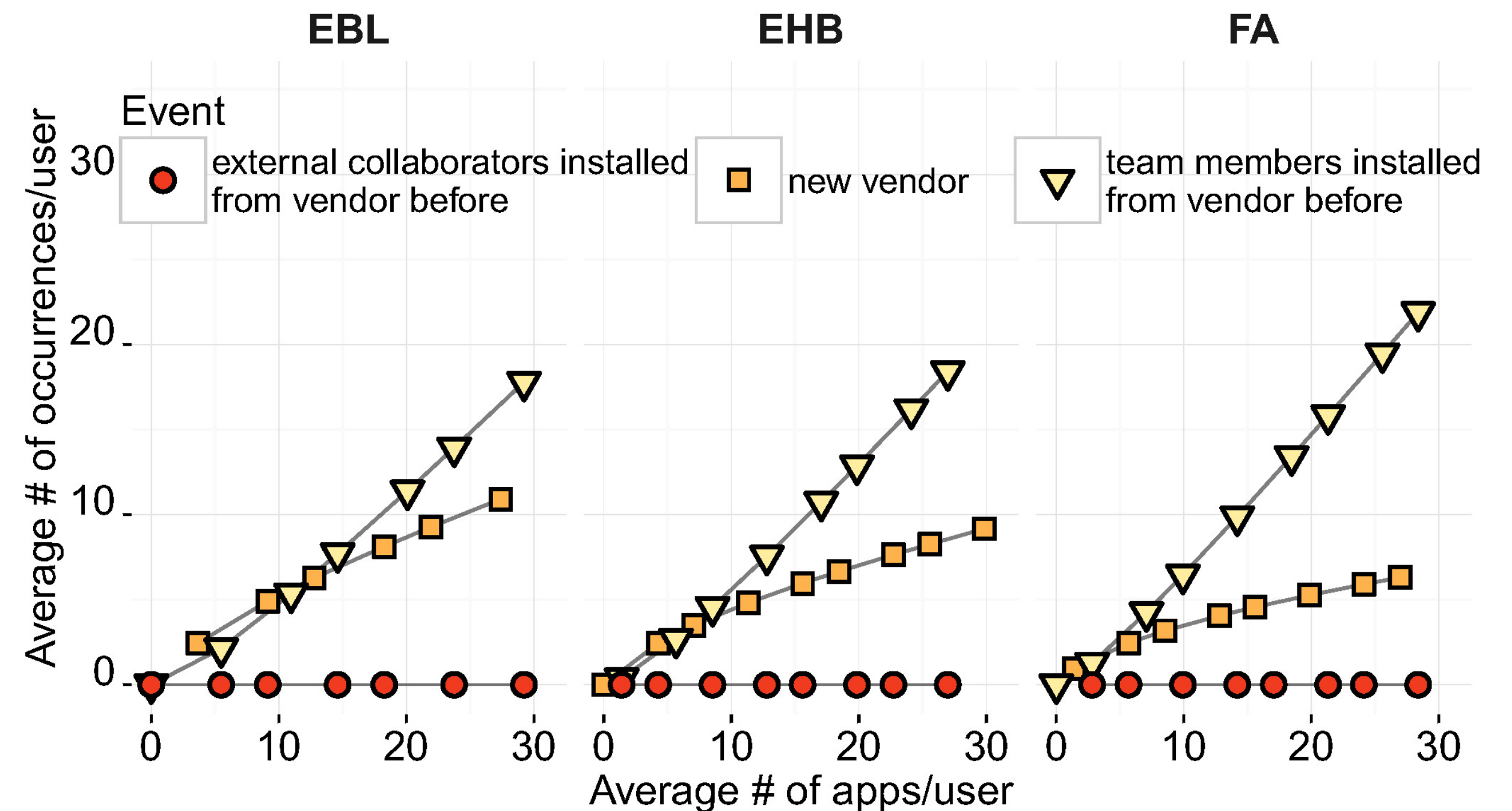
# Teams' Network

Users  
modeled by  
our study  
(EHB)

# 23%

Users who  
always take  
the optimal  
decision  
(FA)

# 45%



## Assumption:

## Team members only account for other team members' decisions

# Take-aways

The **impact of collaborators** on user's privacy significantly important.

With a **Usable privacy** solutions, we show how to mitigate this issue.

**With large networks,** the network effect of HB decisions increases.

# Future Work

History based insights are a basic building block for:  
**Data-driven, Usable privacy**

# Future Work

History based insights are a basic building block for:  
**Data-driven, Usable privacy**

Communicating risk to the users in their own language

# Future Work

History based insights are a basic building block for:  
**Data-driven, Usable privacy**

Communicating risk to the users in their own language

Context of permissions, privacy policies, etc.



[hamzaharkous.com](http://hamzaharkous.com)

[hamza.harkous@gmail.com](mailto:hamza.harkous@gmail.com)



[hamzaharkous.com](http://hamzaharkous.com)

[hamza.harkous@gmail.com](mailto:hamza.harkous@gmail.com)