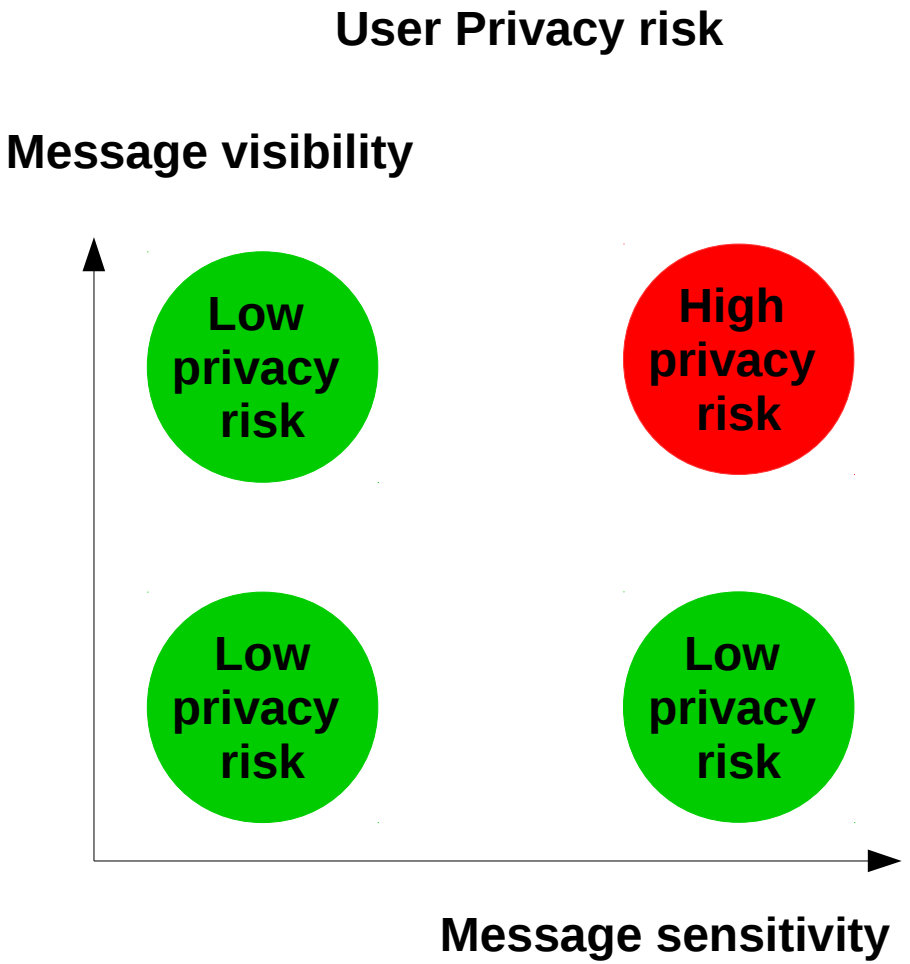
Plan of Presentation

1. Definition
2. Motivation
3. Privacy index
   - Technical
   - Social
   - Behavioral
4. Privacy setting
   - Behavioral
     - **Fuzzy c-means clustering**
     - **Item Response Theory**

1. Definition



- Privacy score is the trade-off between:
  - **Message sensitivity:**
    - Qualitative metric
  - **Message visibility:**
    - Quantitative metric

- Example:
  - Privacy score: 64%
    - 64% of my messages are visible ?
    - OK !
      - Is 64 % of my data it is sensitive ?

2. Motivation

Connection data	• SSL session • Device / log / Timezone • Cookies / Browsing history
Login data	• Email / Phone / Password
Mandatory data	• Name / birthday / gender
Extended profile data	• Education / hometown / languages • Political / religion / website / work
Application data	• Usage statistics / Scores • Permissions / Credit card
Interests	• Hobbies : Books / Music / Movies • Likes / Inspirational_people
Network data	• Family / Friends / Groups
Contextual data	• Taggable_friends / Tagged_places
Private communication Data	• **Private message** • Inbox / Outbox / Poke
Disclosed data	• Text post / Photo / Video • Check-in

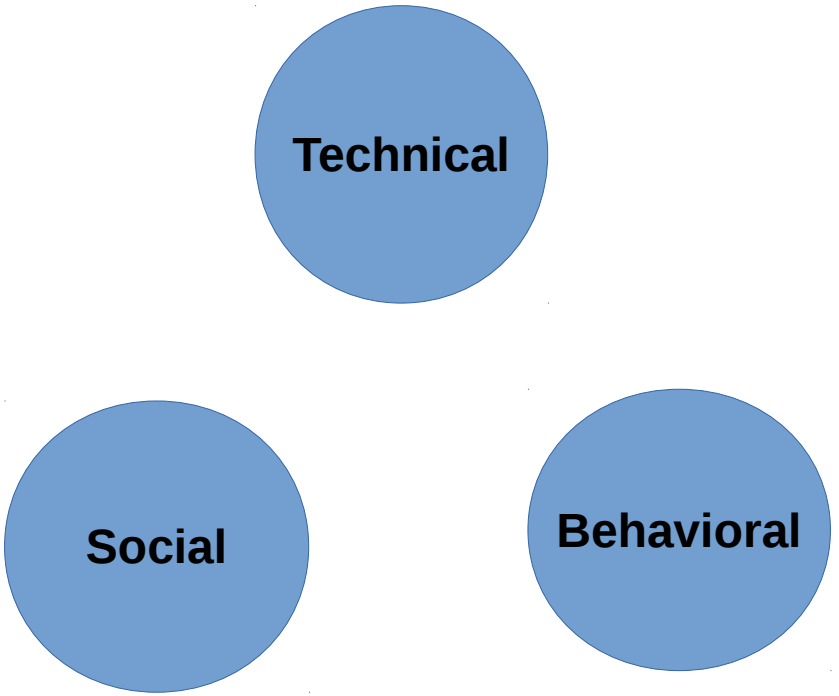
2. Motivation

Connection data	• SSL session • Device / log / Timezone • Cookies / Browsing history
Login data	• Email / Phone / Password
Mandatory data	• Name / birthday / gender
Extended profile data	• Education / hometown / languages • Political / religion / website / work
Application data	• Usage statistics / Scores • Permissions / Credit card
Interests	• Hobbies / Books / Music / Movies • Likes / Inspirational_people
Network data	• Family / Friends / Groups
Contextual data	• Taggable_friends / Tagged_places
Private communication Data	• **Private message** • Inbox / Outbox / Poke
Disclosed data	• Text post / Photo / Video • Check-in

Where are all this messages ?

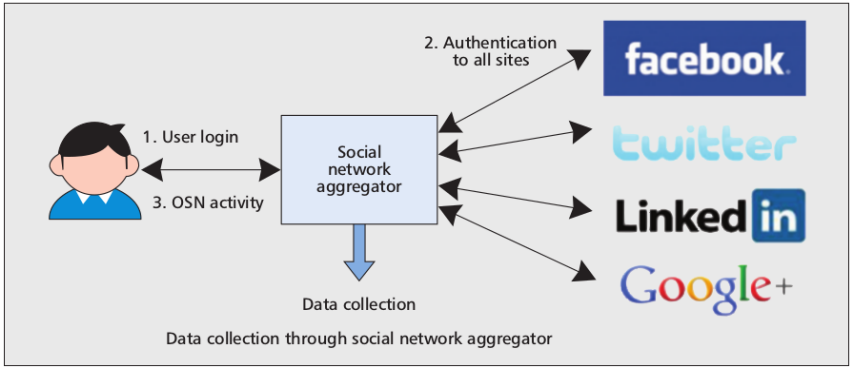
Can I measure their privacy ?

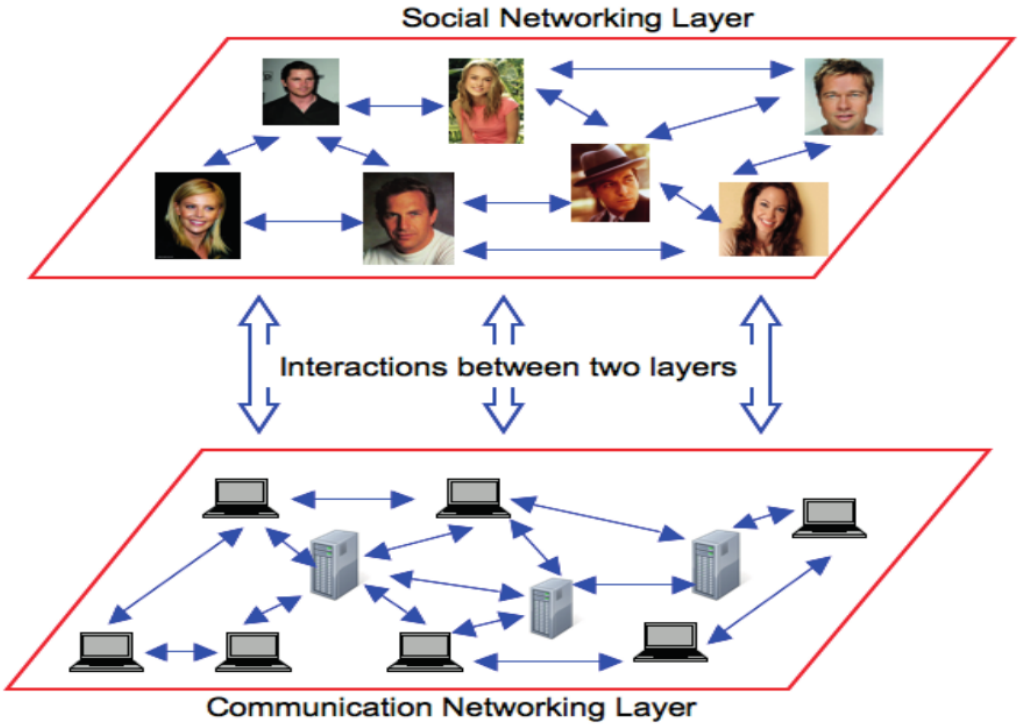
3. Privacy index



- **Behavioral**
  - Data: text, URL, Code
  - Data type: image, video
  - Nb of URL / Nb of Hash-tags
- **Social**
  - Family / Friends / Groups
  - Spammers / Fake profile
- **Technical**
  - SSL session
  - Device / log / Timezone
  - Cokies / Browsing history

3. Privacy index

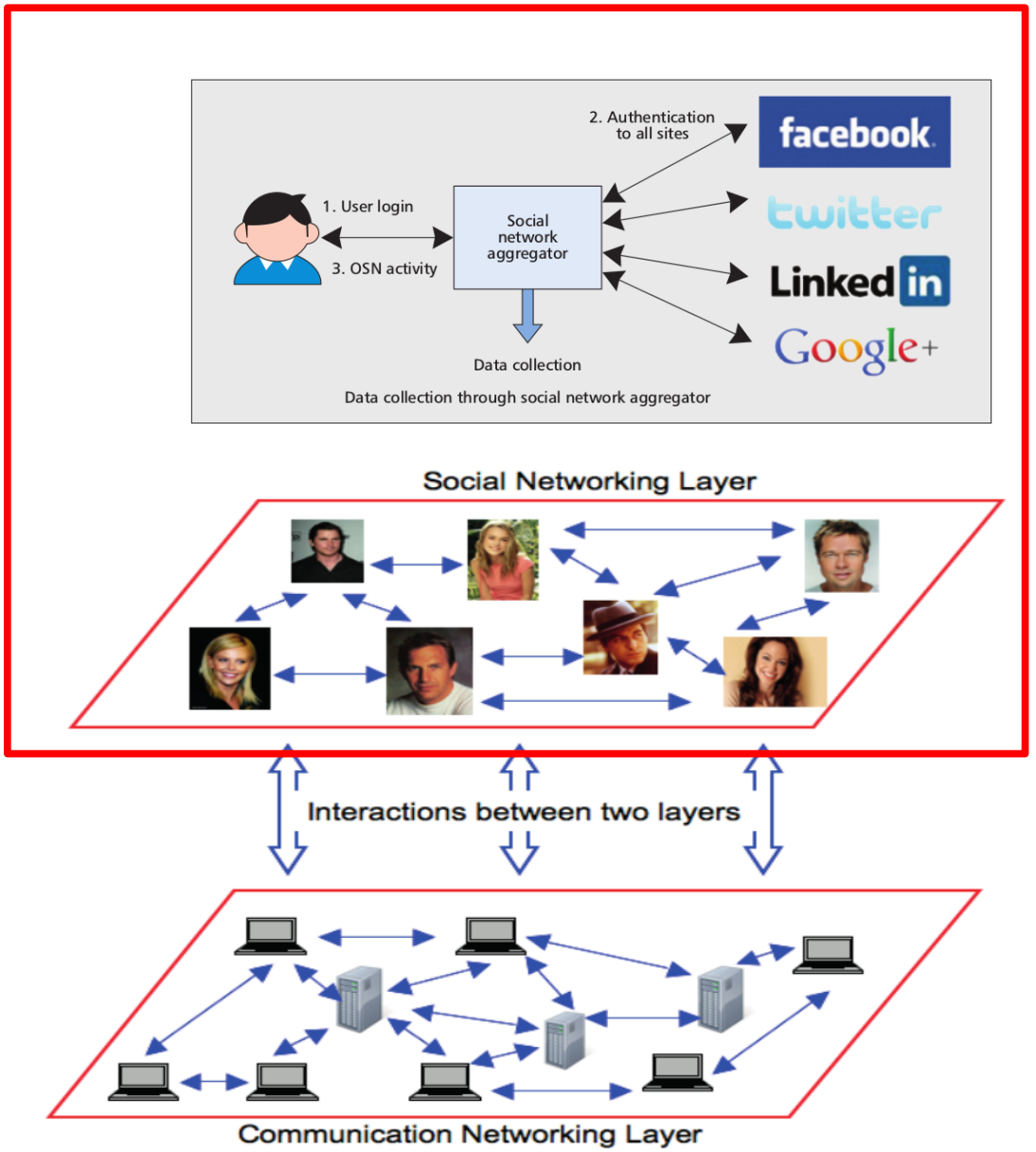




- **Behavioral**
  - Data: text, URL, Code
  - Data type: image, video
  - Nb of URL / Nb of Hash-tags
- **Social**
  - Family / Friends / Groups
  - Spammers / Fake profile
- **Technical**
  - SSL session
  - Device / log / Timezone
  - Cokies / Browsing history

3 privacy layers → 3 privacy values

3. Privacy index



- **Behavioral**
  - Data: text, URL, Code
  - Data type: image, video
  - Nb of URL / Nb of Hash-tags
- **Social**
  - Family / Friends / Groups
  - Spammers / Fake profile
- **Technical**
  - SSL session
  - Device / log / Timezone
  - Cokies / Browsing history

3 privacy layers → 3 privacy values

3. Privacy index

Controlling privacy disclosure of third party applications in online social networks (2016)

4. Privacy settings

Privacy settings Matrix

Sensitivity	β1	...	βn
User\Item	msg 1	...	msg n
User 1			
...		visibility	
User N			

- **Privacy score is the trade-off between:**
  - **message sensitivity**
  - **message visibility**

- Items visibility and sensitivity depend on:
  - **Privacy settings matrix**

- **Behavioral privacy**
  - **Examples**
    1) **Fuzzy c-means clustering**
    2) **Item Response Theory**

4. Privacy settings:
4.1 Behavioral privacy (Fuzzy c-means clustering)

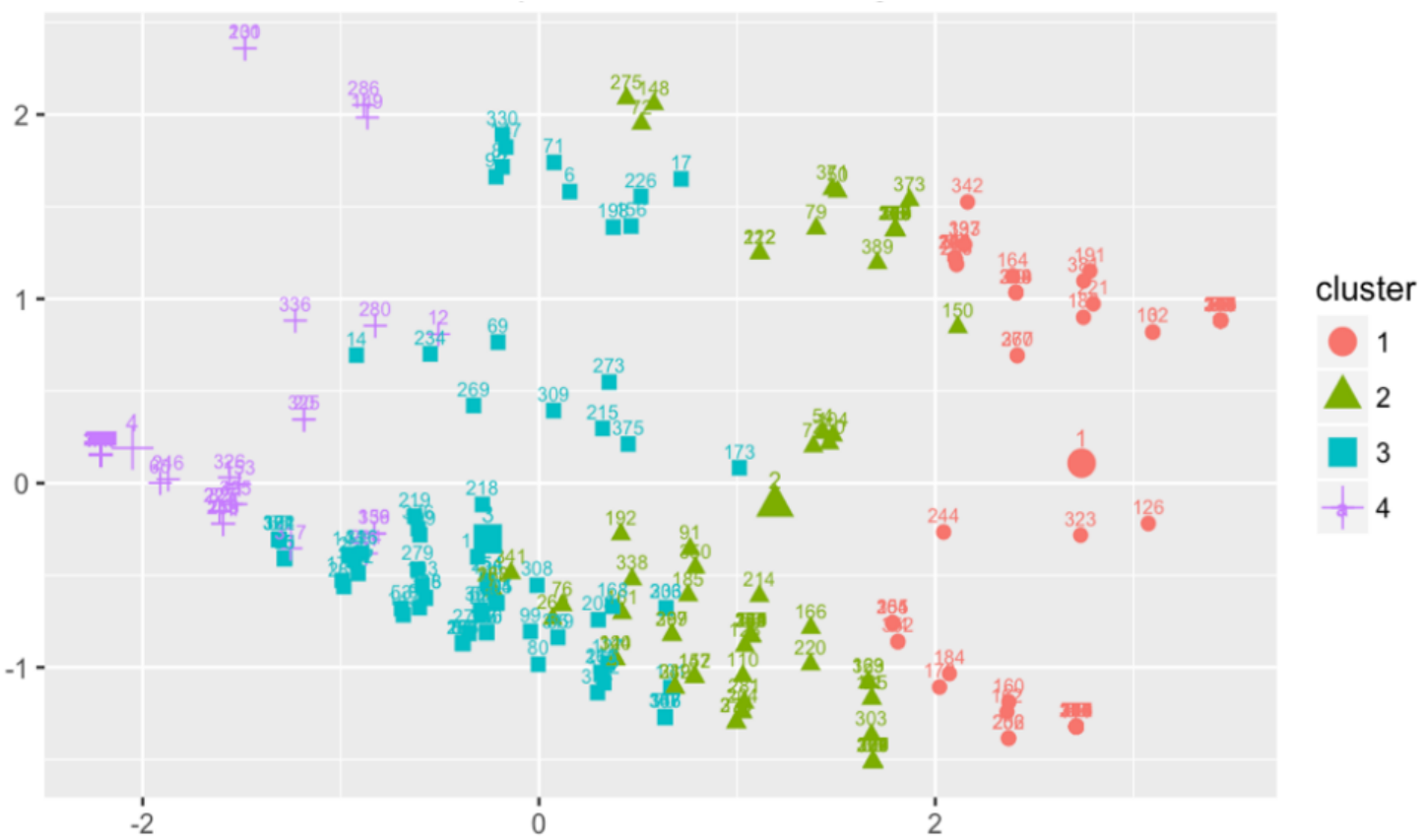
- **Input**
  - Users: $U = \{ u_1, \dots , u_N \}$
  - Privacy settings: $S = \{ s_{(1,1)}, \dots , s_{(i,v)} \}$
    - ➔ Data type $I = \{$ MyActivity, ContactMe, MyRelations, MyTopics, PersonelInfo, VoteInfo $\}$
    - ➔ Visibilities**:** $V = \{$ OnlyMe, Friends, FriendsOfFriends, Public $\}$
- **Method**
  - Fuzzy c-means clustering
- **Output**
  - **Users behavior**

User	My Activity	Contact Me	My Relations	My Topics	Personal Info	Vote Intention
1	2	3	2	3	3	2
...	...	...	...	...	...	...
N	4	4	4	2	2	1

Privacy settings matrix

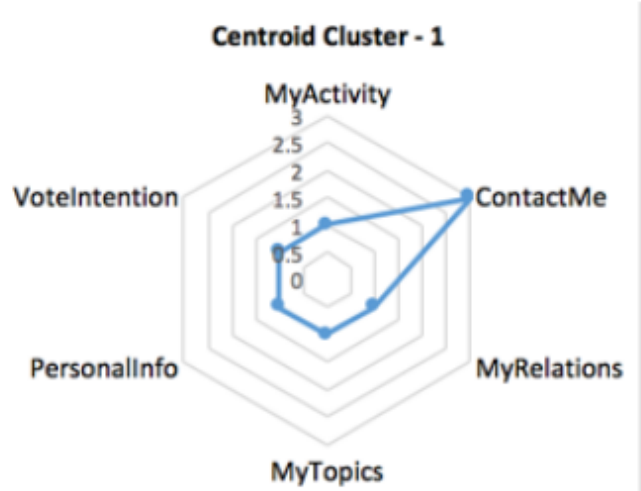
4. Privacy settings:

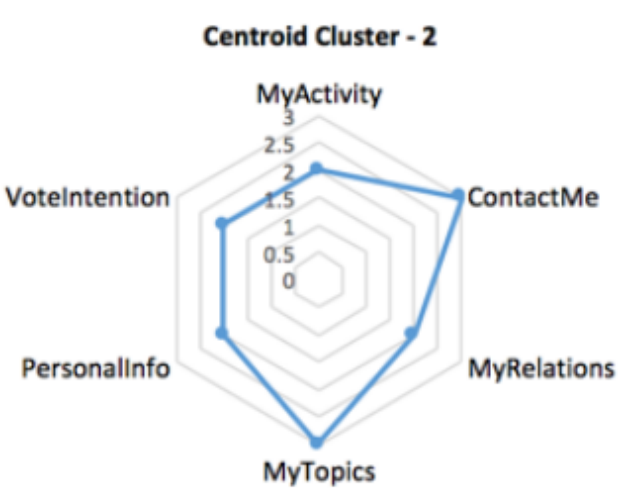
4.1 Behavioral privacy (Fuzzy c-means clustering)

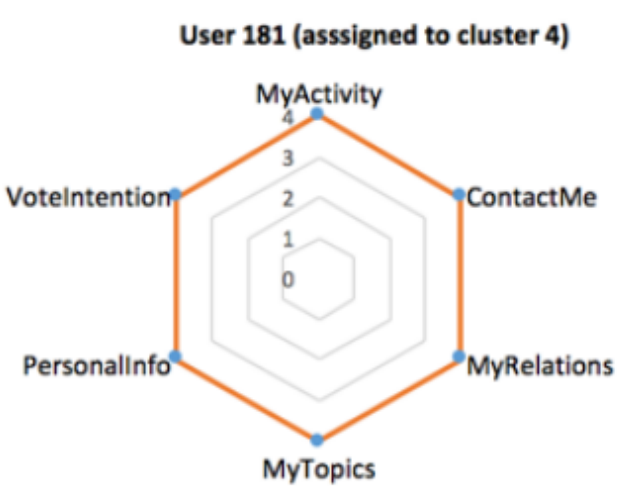


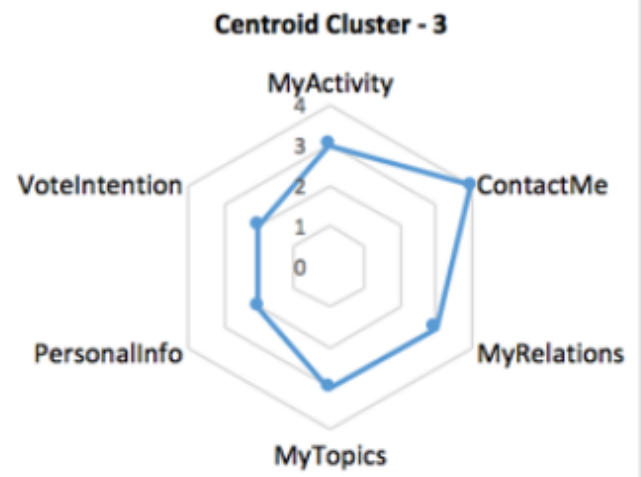
Fuzzy c-means clustering with 4 clusters

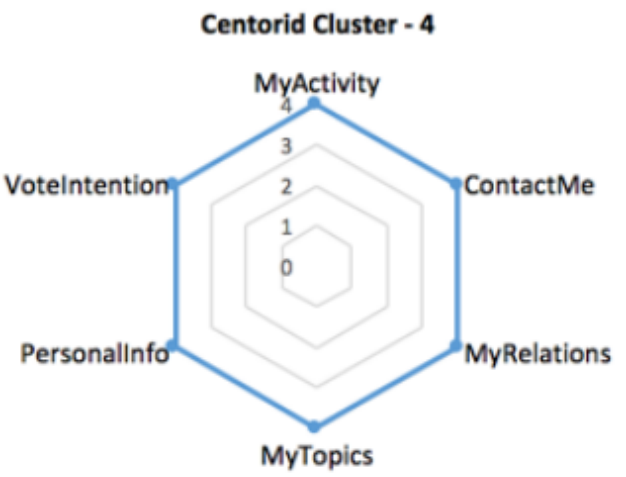
4. Privacy settings:
4.1 Behavioral privacy (Fuzzy c-means clustering)

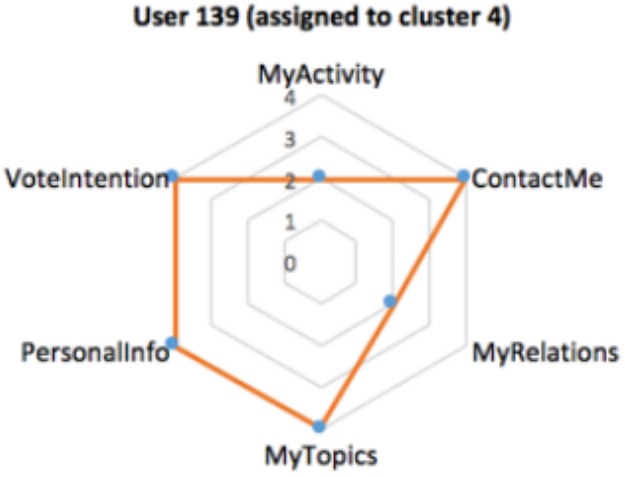






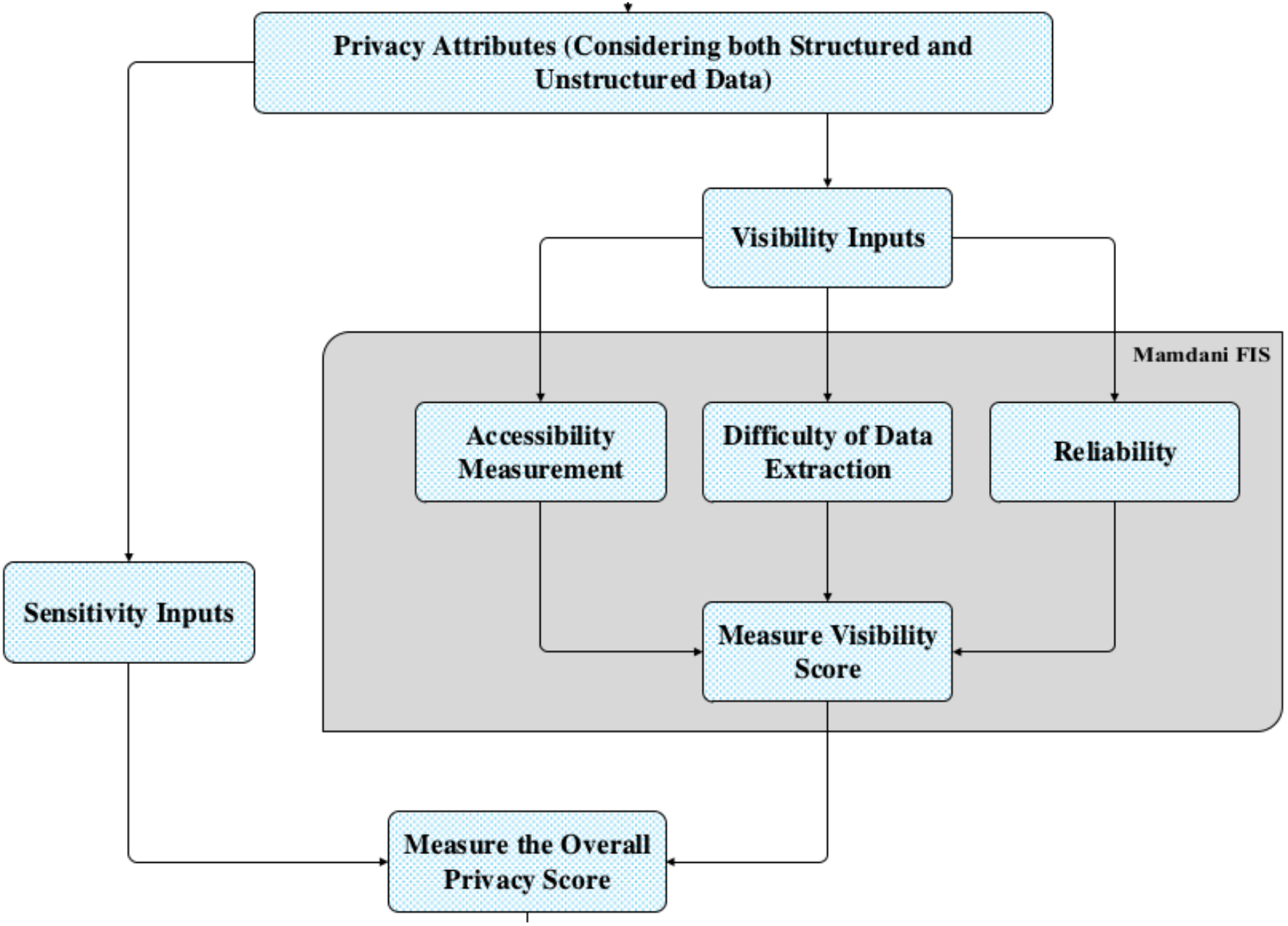






4. Privacy settings:

4.1 Behavioral privacy (Fuzzy c-means clustering)

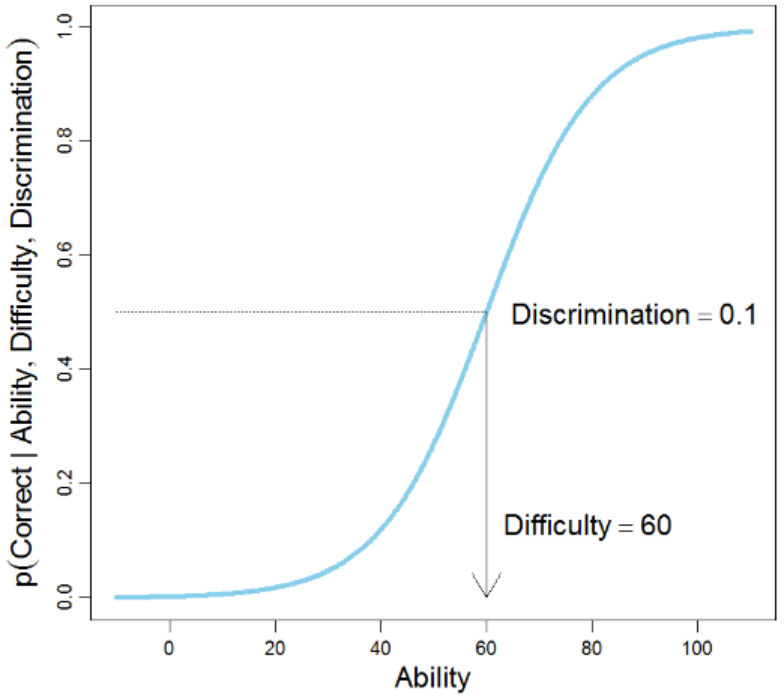


Scoring Users' Privacy Disclosure Across Multiple Online Social Networks (2017)

4. Privacy settings:
4.2 Behavioral privacy (Item Response Theory)

Privacy settings Matrix

	Sensitivity	β1	...	βn
Attitude	**User\Item**	**Item 1**	**...**	**Item n**
θ_1	**User 1**			
	...		R(i,j)	
θ_N	**User N**			



- Privacy score is the trade-off between:
  - **data sensitivity**
  - **data visibility**

- data visibility and sensitivity depend on:
  - **Privacy settings matrix**

- **data visibility depends on:**
  - Response Matrix

$$P_{ij} = Prob\{R(i,j) = k\}$$

- **Item Response Theory (IRT)**

$$P_{ij} = \frac{1}{1 + e^{\alpha_i(\theta_j - \beta_i)}}$$

4. Privacy settings:
4.2 Behavioral privacy (Item Response Theory)

		Sensitivity	β1	…	βn
Privacy	Attitude	**User\Item**	**Item 1**	**…**	**Item n**
P1	θ1	**User 1**			
		…		R(i,j)	
PN	θN	**User N**			

- $P_j = \sum_{i=1}^{n} \beta_i \cdot V_{ij}$
- $V_{ij} = P_{ij}$
- $P_{ij} = \frac{1}{1+e^{\alpha_i(\theta_j - \beta_i)}}$
- $P_{ij} = Prob\{R(i,j) = k\}$

Thank you for your attention