

# Privacy and Security in Online Social Media

---

Course on NPTEL

NOC21-CS28

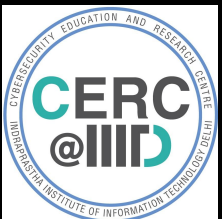
Week 8.1

Ponnurangam Kumaraguru (“PK”)

Full Professor

ACM Distinguished Speaker

fb/ponnurangam.kumaraguru, @ponguru



# What is the difficulty in matching?

---

- <https://www.facebook.com/ponnurangam.kumaraguru>
- <https://twitter.com/ponguru>
- <https://in.linkedin.com/in/ponguru>



# Ponnurangam Kumaraguru (PK)

Associate Professor @ IIIT Delhi

New Delhi Area, India | Higher Education

**500+**  
connections

Current	Data Security Council of India, IIIT-Delhi
Education	Carnegie Mellon University
Recommendations	1 person has recommended <b>Ponnurangam</b>
Websites	Personal Website

# This lecture

---

- Tracking social footprint / identities across different social network



# Other Times, Other Values: Leveraging Attribute History to Link User Profiles across Online Social Networks

Paridhi Jain  
Indraprastha Institute of  
Information Technology (IIIT-D),  
India  
paridhi@iiitd.ac.in

Ponnuram Kumaraguru  
Indraprastha Institute of  
Information Technology (IIIT-D),  
India  
pk@iiitd.ac.in

Anupam Joshi  
University of Maryland, Baltimore  
County (UMBC),  
USA  
joshi@cs.umbc.edu

## ABSTRACT

Profile linking is the ability to connect profiles of a user on different social networks. Linked profiles can help companies like Disney to build psychographics of potential customers and segment them for targeted marketing in a cost-effective way. Existing methods link profiles by observing high similarity between most recent (current) values of the attributes like name and username. However, for a section of users observed to evolve their attributes over time and choose dissimilar values across their profiles, these current values have low similarity. Existing methods then falsely conclude

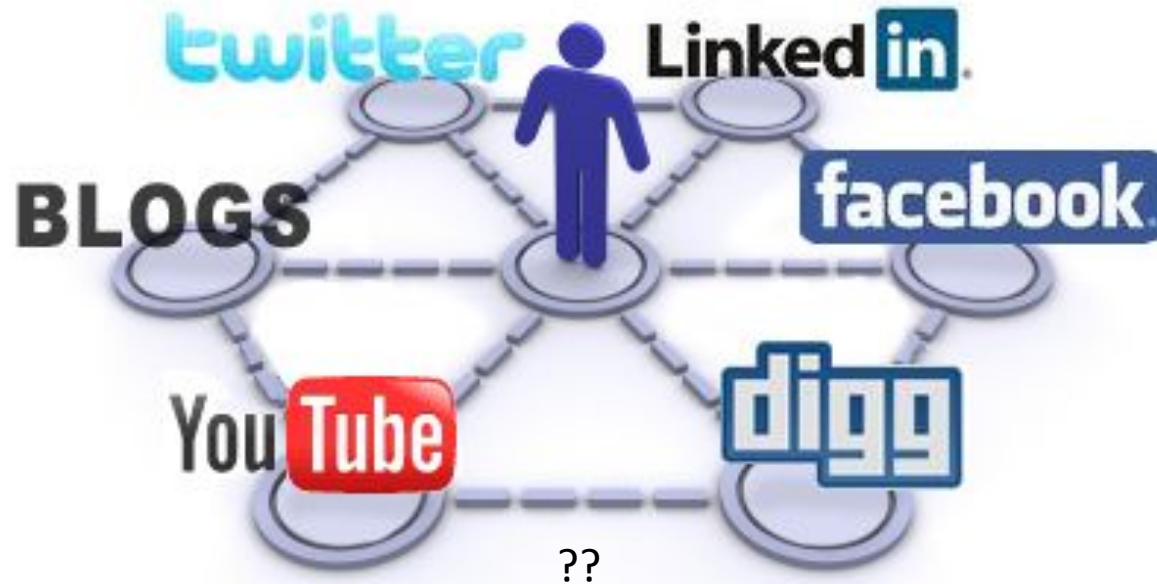
mation, lists her friends and later creates content to share with her friends. The quality, quantity and veracity of the information created and shared by her vary with the OSN, thereby resulting in dissimilar profiles of the same user, scattered on the world wide web, with no explicit links directing to one another. These disparate profiles liberate her from any privacy concerns that could emerge if the profiles were implicitly collated. However, linking these disparate unlinked profiles can benefit various stakeholders.

Companies like Disney and PepsiCo carry out psychographic segmentation based upon customers' activities, interests, opinions

Jain, P., Kumaraguru, P., and Joshi, A. Other Times, Other Values:  
Leveraging Attribute History to Link User Profiles across Online Social Networks.

# Knowing this can be useful!

---



# De-duplicating audience

---



437,632 likes



FOLLOWERS  
**153K**



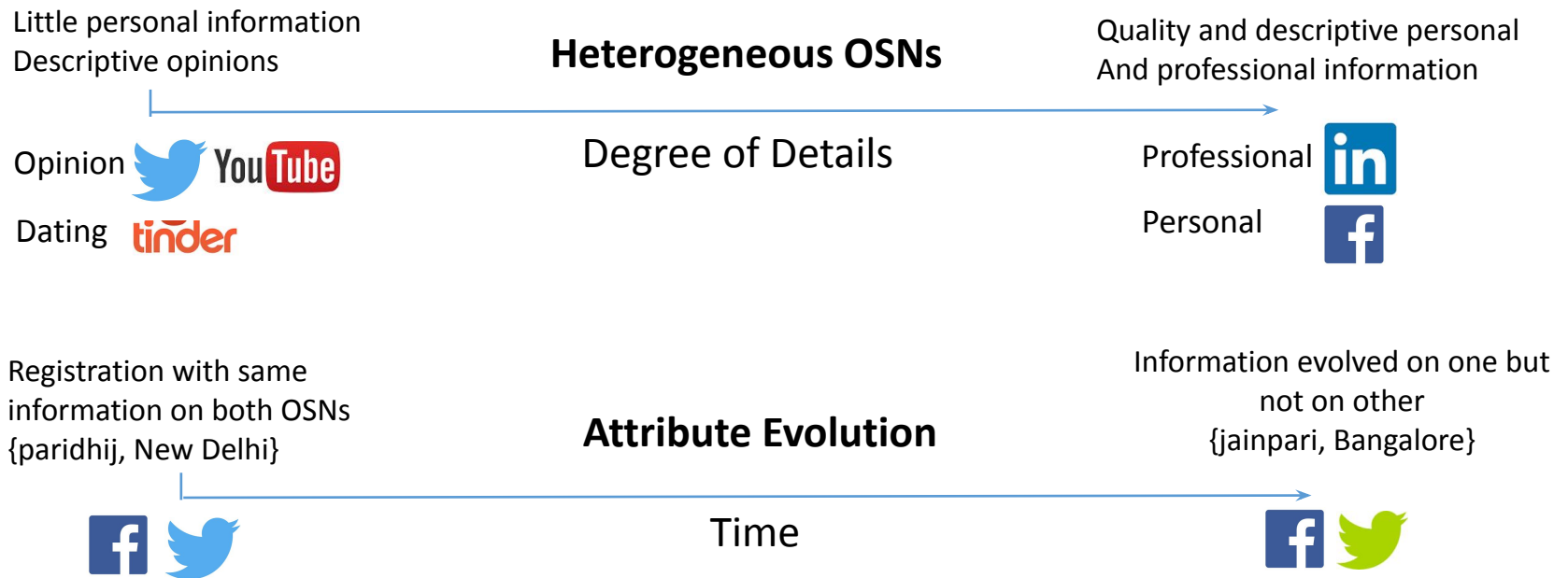
805,097 followers

Social audience = 437,632 + 153,000 + 805,097 or less??



# Challenges

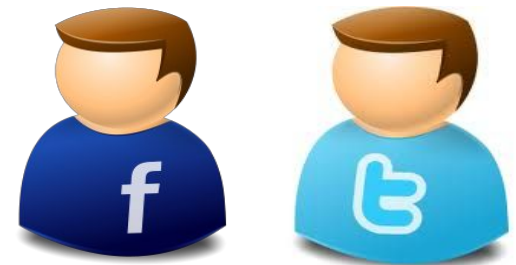
---



# Profile linking approach

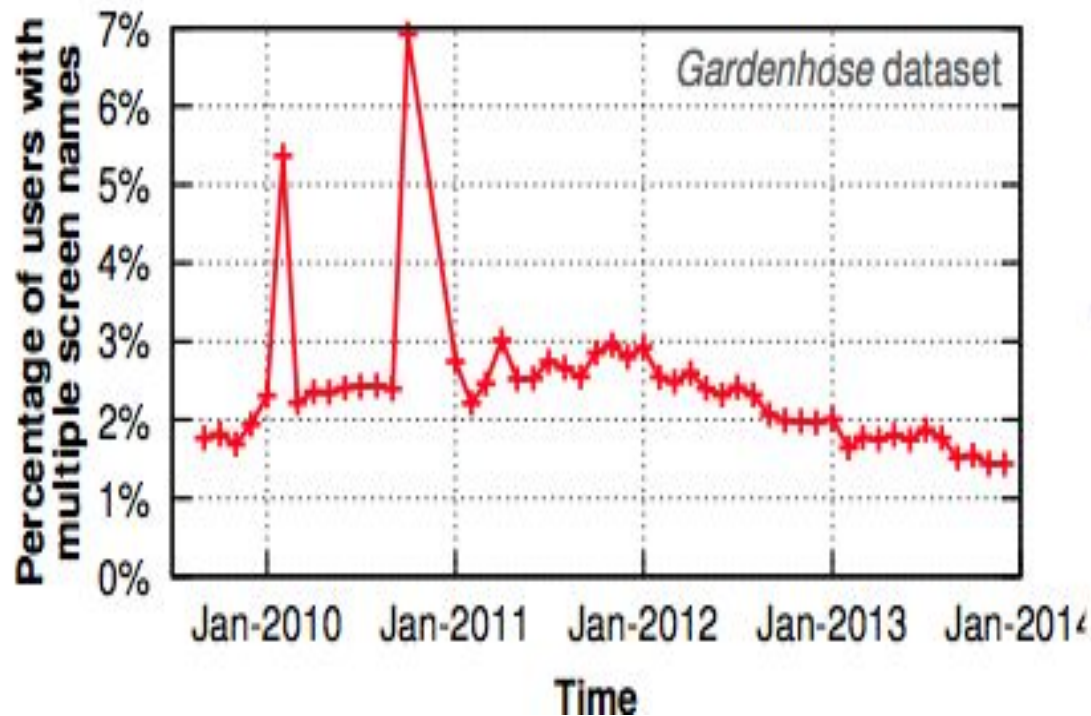
---

- List common attributes
- Compare attribute values using syntactic, semantic or graph based methods
- High similarity denote profiles refer to a single user
- Values considered here are the most recent (current) values of the attributes



# But the values change!

---

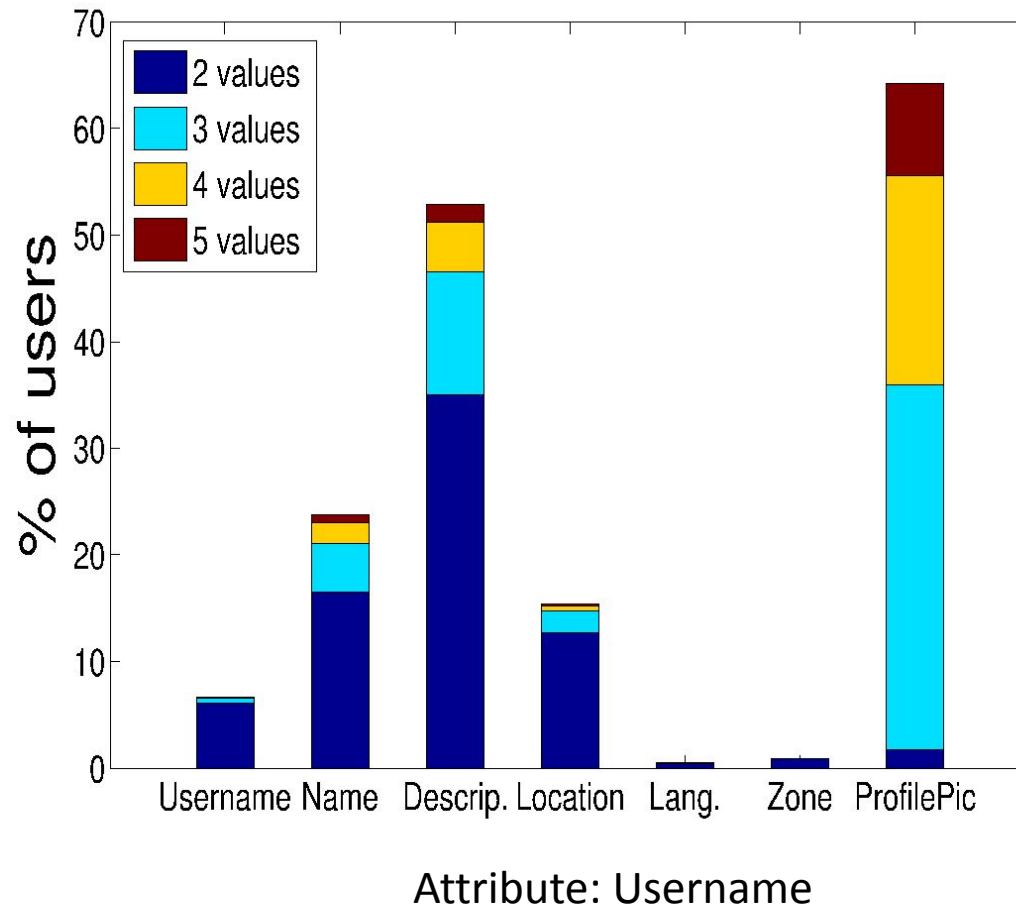


Attribute: Username

# of users tracked: 376 million [random]

Tracking period: 4 years

# Values change

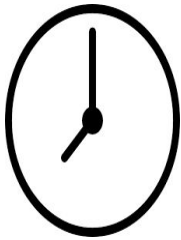


# of users tracked: 8 million [random]

Tracking period: 2 months

# Reality!

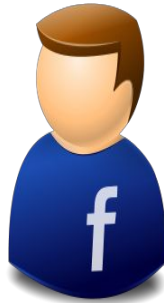
---



Registration: t1

Observation: t2

Observation: t3



@nitinsgr

@nitinsgr

@nitinsgr



@nitinsgr

@explorer\_nitin

@logicalIndian

Attribute Evolution

Unmatching values

# Problem Statement

---

*Given two user profiles and the respective **username** sets, each composed of past and current **usernames**, find if profiles refer to a single individual?*

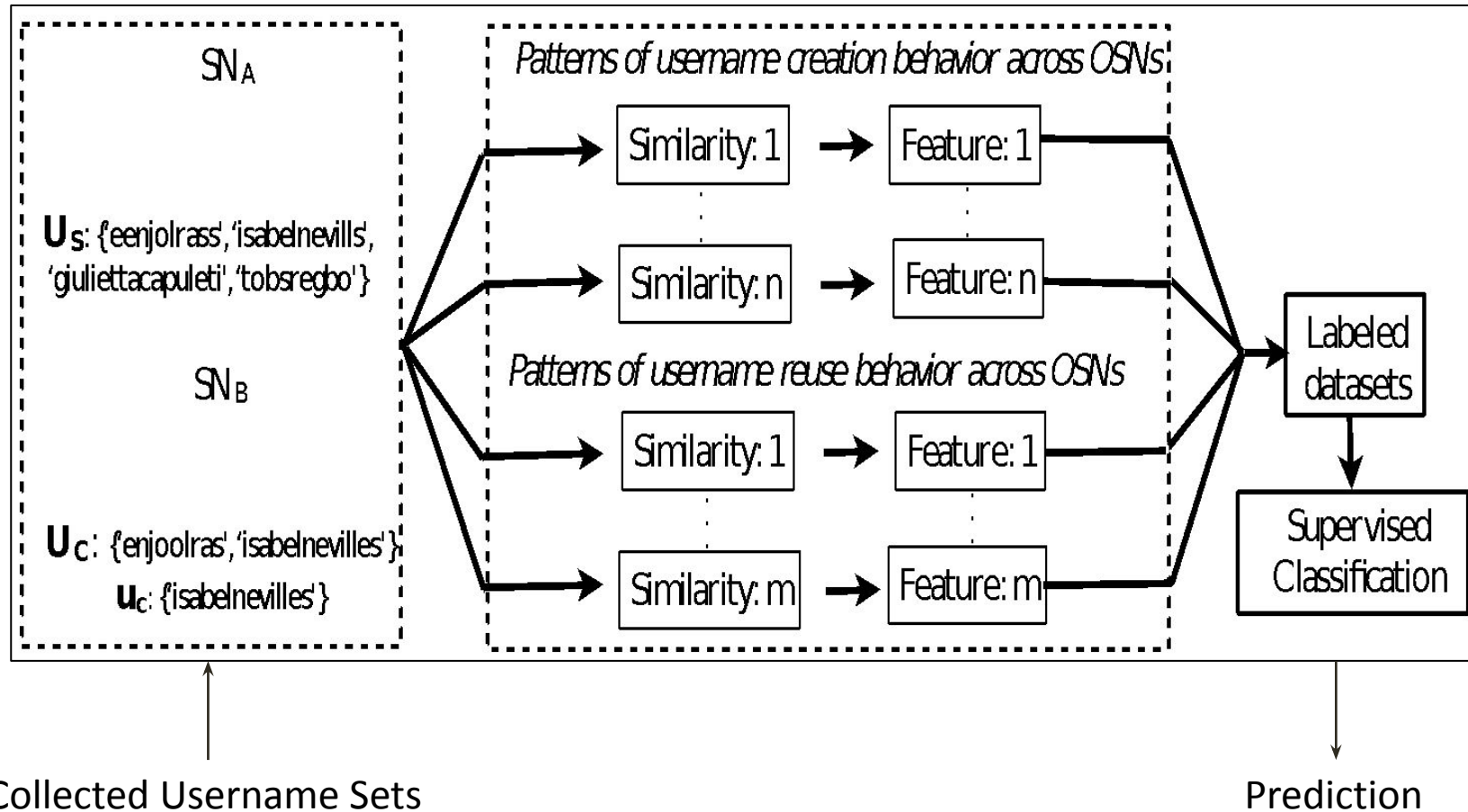


# Why only usernames?

---

- Unique attribute of a user
- Universally and publicly available attribute
- Homogenous, character and length restricted
- Easier history collection methods for username as other attributes

# Methodology



# Ground Truth Collection

- Self-identification behavior [Cross-referencing one's OSN accounts]
- Extrovert users

Twitter username →

Tumblr username on the URL →

**luzy**  
@yourgirluzy

the very growing sin a.k.a wickedpuritan

Twytter

girlonthesportingnewss.tumblr.com

Joined March 2011

Tweet to luzy

145 Photos and videos

TWEETS 22.9K FOLLOWING 246 FOLLOWERS 555 FAVORITES 153

Tweets Tweets & replies Photos & videos

luzy retweeted

the devil herself @forestfeminist · 4h

\*finds cute twitter boy\* \*finds cute twitter boy's problematic tweets\* god dammit not again

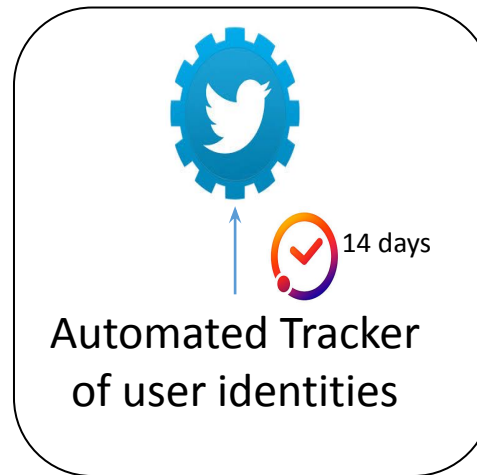
luzy @yourgirluzy · 4h

Cuando las investigaciones no están escritas en tercera persona me siento muy incómoda.

# Past Usernames Collection

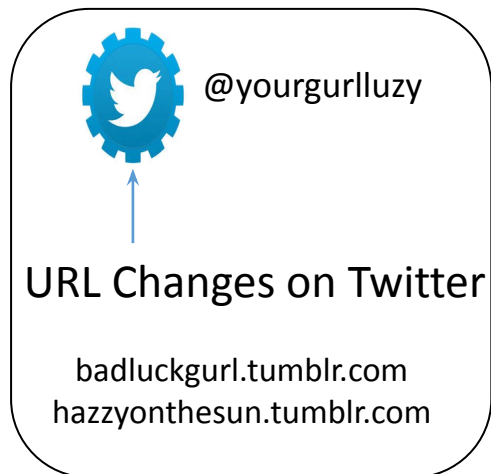
---

<http://twitter.com/yourgurlluzy>



[@badluckgirl,  
@bieblerlover]

<http://girlonthesportingnewss.tumblr.com>



[@badluckgirl,  
@hazzyonthesun]

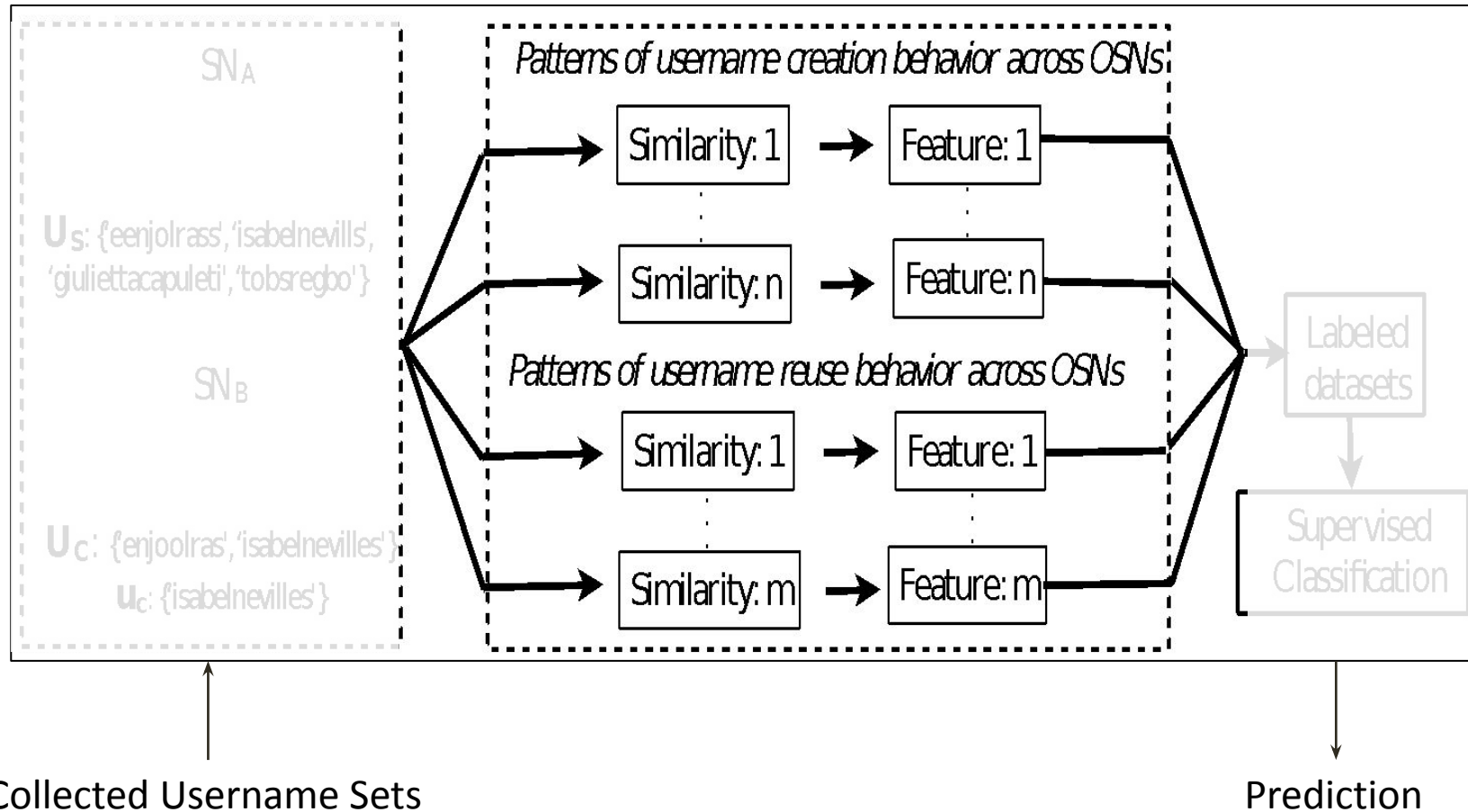
# Sample

---

- **User ID:** 595929421
- **Past usernames on Twitter:**
  - ["bigeasye\_", "reezy11\_", "epiceric\_", "soulanova", "swampson\_", "hebetheeric", "swampkidd\_"]
- **Past Usernames on Instagram:**
  - ["bigeasye\_", "epiceric17", "swampson", "hebetheeric"]}

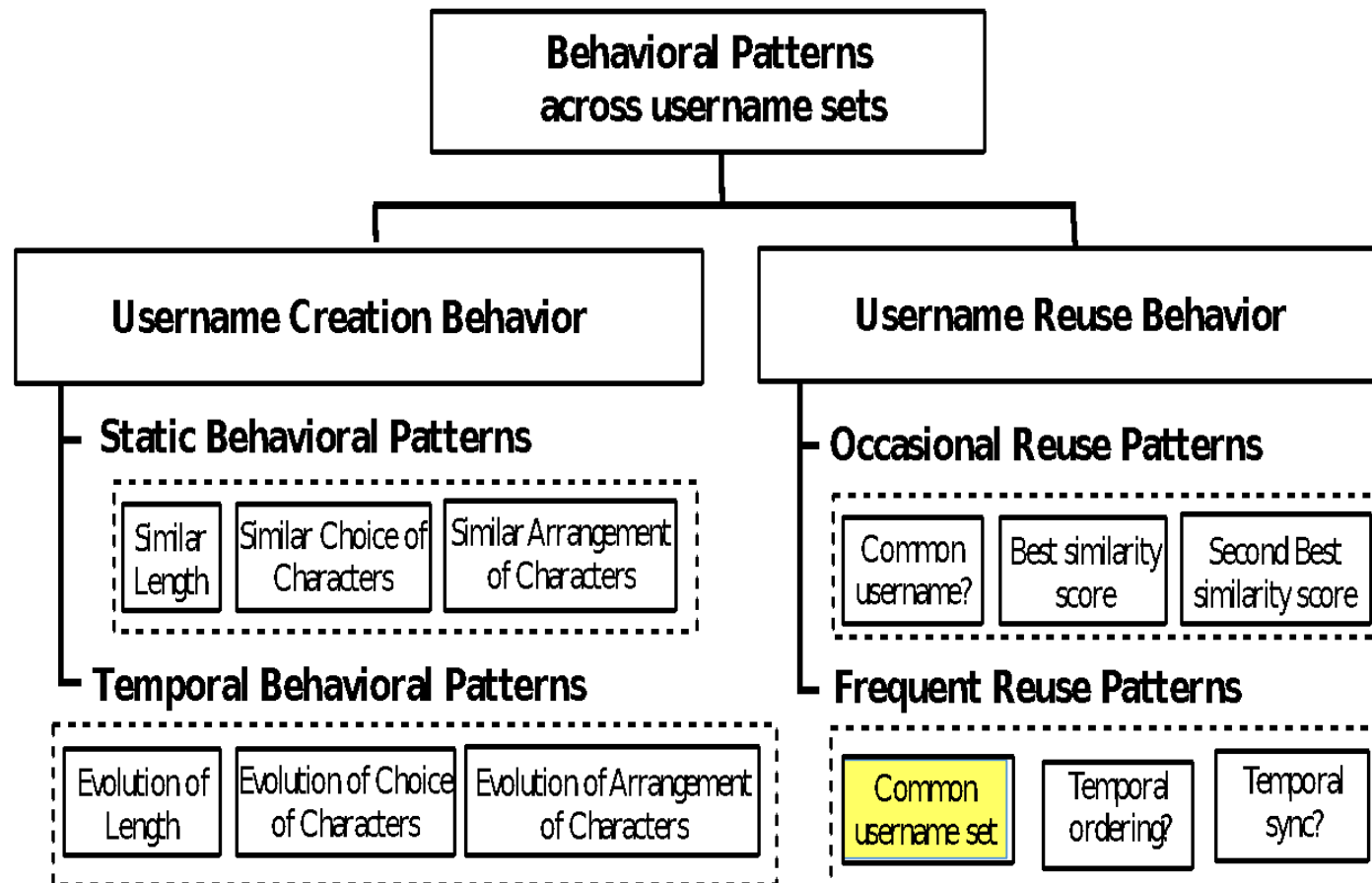
# Methodology

Assumption: Consistent user behavior within and across networks over time



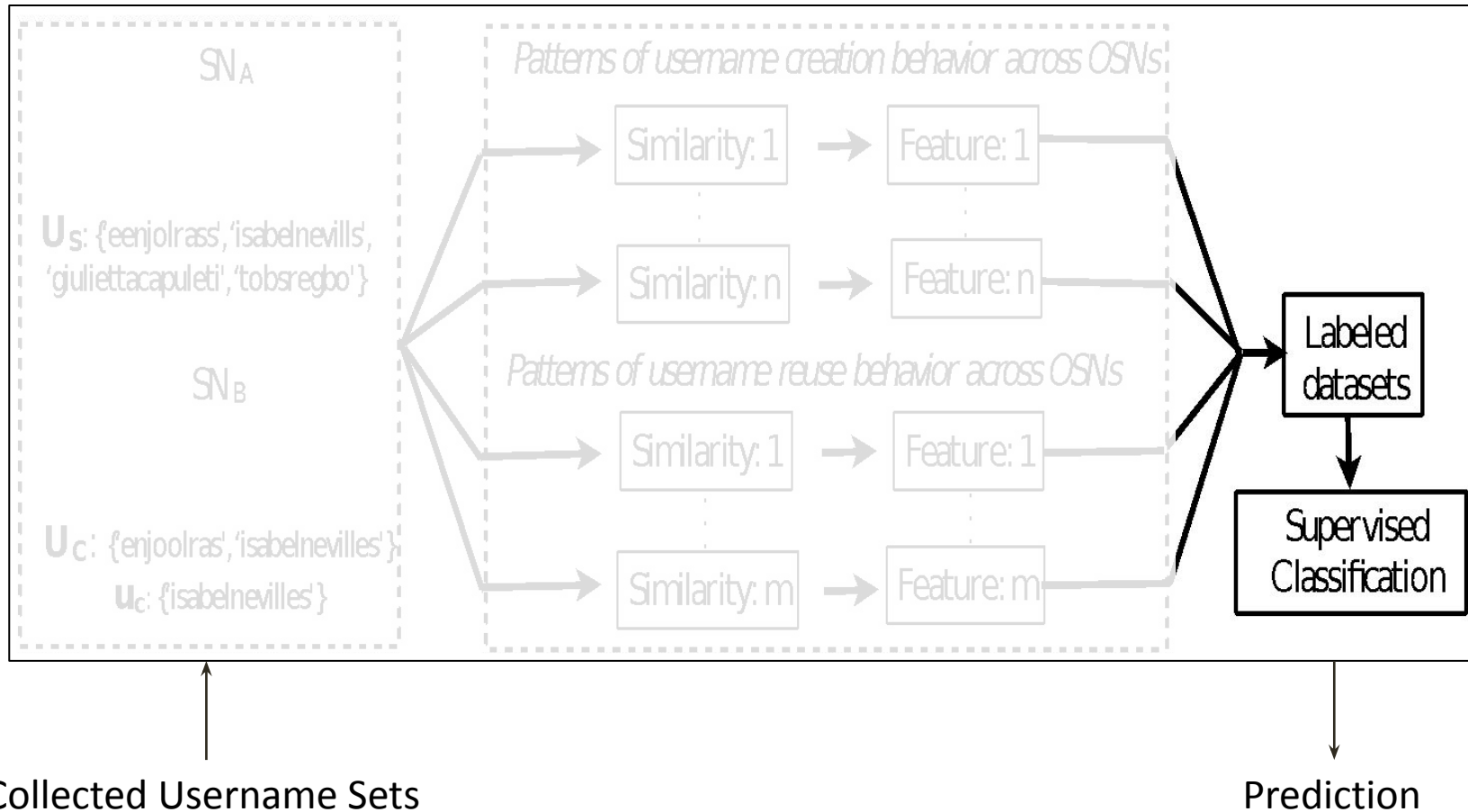


# Features



# Features: 26

# Methodology



# Datasets

---

- Linking profiles

- Twitter – Instagram
- Twitter – Tumblr
- Twitter - Facebook

- Past usernames available for both profiles:

- 21,446 positive pairs, 21,449 negative pairs

- Past usernames available only on Twitter but current username available on other profile:

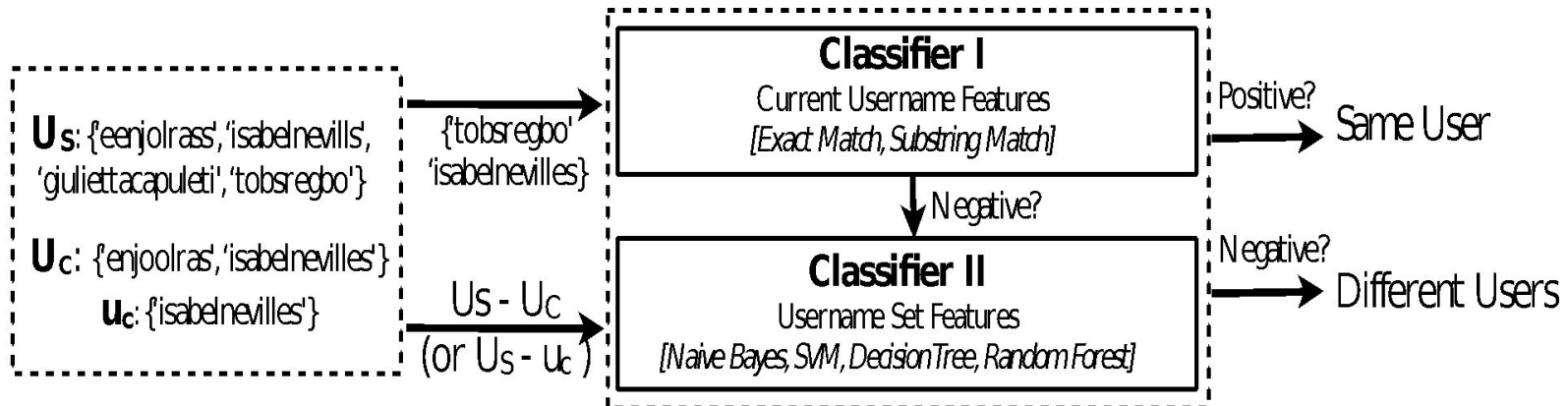
- 112,451 positive pairs, 112,451 negative pairs

# Supervised Classification

## 1. Independent Supervised Framework



## 2. Cascaded Supervised Framework



# Prediction

---

Framework Config.	Accuracy	FNR	FPR
Exact Match (b1)	<b>55.38</b>	<b>89.34</b>	<b>0.00</b>
Substring Match (b2)	60.99	78.46	0.00
Independent [Naive Bayes]	72.19	55.86	0.13
Cascaded [b1→Naive Bayes]	72.48	55.27	0.14
<b>Cascaded [b1 → SVM [Linear]]</b>	<b>76.74</b>	<b>45.16</b>	<b>1.65</b>
Cascaded [b2 → Naive Bayes]	72.51	54.97	0.17
Cascaded [b2 → SVM [Linear]]	76.84	45.16	1.25

# Prediction

---

A comparison of cascaded framework accuracy with and without Twitter-Tumblr instances

Framework Config. [History on Both or One]	Accuracy	FNR	FPR
Exact Match (b1)	55.38	89.34	0.00
<b>Cascaded [all network]</b>	<b>76.74</b>	<b>45.16</b>	<b>1.65</b>
Exact Match without Tumblr (b1)	66.17	67.51	0.00
<b>Cascaded [without Tumblr]</b>	<b>91.20</b>	<b>16.60</b>	<b>0.96</b>



# Measuring Volume of Sentiments



**Phạm Hoàng Nam** 🙄 feeling annoyed

June 27 · Hanoi, Vietnam · 🌐

I fucking **hate** Adobe Flash Player, force me updating new version nearly every two weeks. Can't leave without this shit. Fucking hate those mofos :-<

7 Likes 9 Comments



Like



Comment



Share



**Lok C. - DigitalRev TV Cameraman**

June 28, 2014 · 🌐

Apple, why? You created **#Aperture**, LR was just a copycat. 8 years, I'm happy to avoid Adobe. I always **hate** Adobe, since I was a graphic designer. Now they can charge LR for a fortune, annually. Sad day.

386 Likes 67 Comments 2 Shares



Like



Comment



Share



**Pure Sugar Studios**

May 11, 2013 · 🌐

UGH! all my **adobe** products are being temperamental and will not start! How long until the cloud is useable? Anyone else have this issue?  
[#hateadobe](#)



Like



Comment



Share



**Sarah Tolzmann** @notetosarah · 18h

when Photoshop "zones out" for a full 4 minutes and sends me into panic mode... THANKS **ADOBE**. **HATE** YOU, MEAN IT.



**Jolly Ol' Teek** @TeekTheTweeter · Aug 13

[@redlianak](#) I use Premiere Elements. I **hate** Adobe with a fiery passion, but I wouldn't recommend anything else.



[View conversation](#)



**Jeremy Miller** @jrmiller\_ · Aug 13

I **hate** that **Adobe** will lock you into a monthly subscription for a year. That's some BS.



**Michael Laing** @Michael\_Laing · Aug 12

[@IAmPlxmas](#) Both programs have advantages and disadvantages, C1 give a better final edit but LR is faster. I also **hate** that **Adobe**, **Adobe** RGB.



[View conversation](#)

# Conclusion

---

- Profile linking may be necessary for many organizations / needs
- Better profile linking is possible with past history of user handles



# Activity

---

- Take 2 of your accounts or any accounts that you know are same in 2 different social networks
- Find out various ways in which you can link these 2 accounts
  - List the features
- List down things that you will change in the profile to make it look as 2 different networks

# References

---

- Paridhi Jain's Ph.D. thesis work

# Thank you

[pk@iiitd.ac.in](mailto:pk@iiitd.ac.in)

[precog.iiitd.edu.in](http://precog.iiitd.edu.in)

[fb/ponnurangam.kumaraguru](https://www.facebook.com/ponnurangam.kumaraguru)