Managing Information Leakage

Steven Whang, Hector Garcia-Molina

Stanford University

Anonymity | Personal Contact Info | Privacy | Quora Privacy | Edit

How much can you figure out about me, personally, and my identity, based on what you can observe about me from Quora?

Edit

I ask this question for the value of its general applicability, yes, but also using myself as a particular example of someone who has made a degree of effort to cover his tracks. Sleuth me out. Edit

1 Comment • Add Follow-Up Question • Flag Question

2 Answers • Create Answer Summary

Anon User

7 votes by June Lin, Leslie Fine, Joseph Jegvan-Andrej Cir, (more)

- Your full name has 21 characters. It also includes your dad's first name.
- You are 24 and will be turning 25 very soon.
- Your mother's maiden name starts with the letter P.
- Your mother is named after her dad's mother.
- You likely live with your aunt (your father's sister).
- You currently work for C.P. (2 months already).
- It takes you about 36 mins to 1 hour and 10 mins to get to work (depending on traffic condition).
- You probably have visited a pub near where you work.
- You live in a wealthy neighborhood and nearby a T intersection.
- Your house was built in 1968. Your family likely has lived there for 12 years.
- Your phone number ends with the following three digits: 994.
- Your Facebook profile ID ends with the number 7.
- Your Facebook profile picture is related to an event that happened in LA back in '09.
- You have a good GPA, graduated with High Honors and made it to the Dean's List every year.
- You are fluent in French and studied Arabic for 2 years (it's also your minor).

Follow Question | Options

Related Questions Edit

How do you get in touch with someone who goes by a pseudonym?

What services can help me figure out: "Something fun to do today near my...
(continue)

How does Quora figure out what questions are most interesting to me?

See more related questions

Share Question or **Ask to Answer**

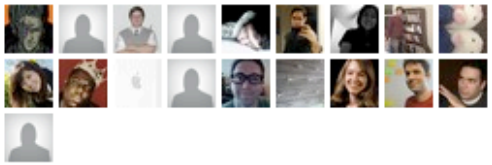
Facebook | Inbox | Email

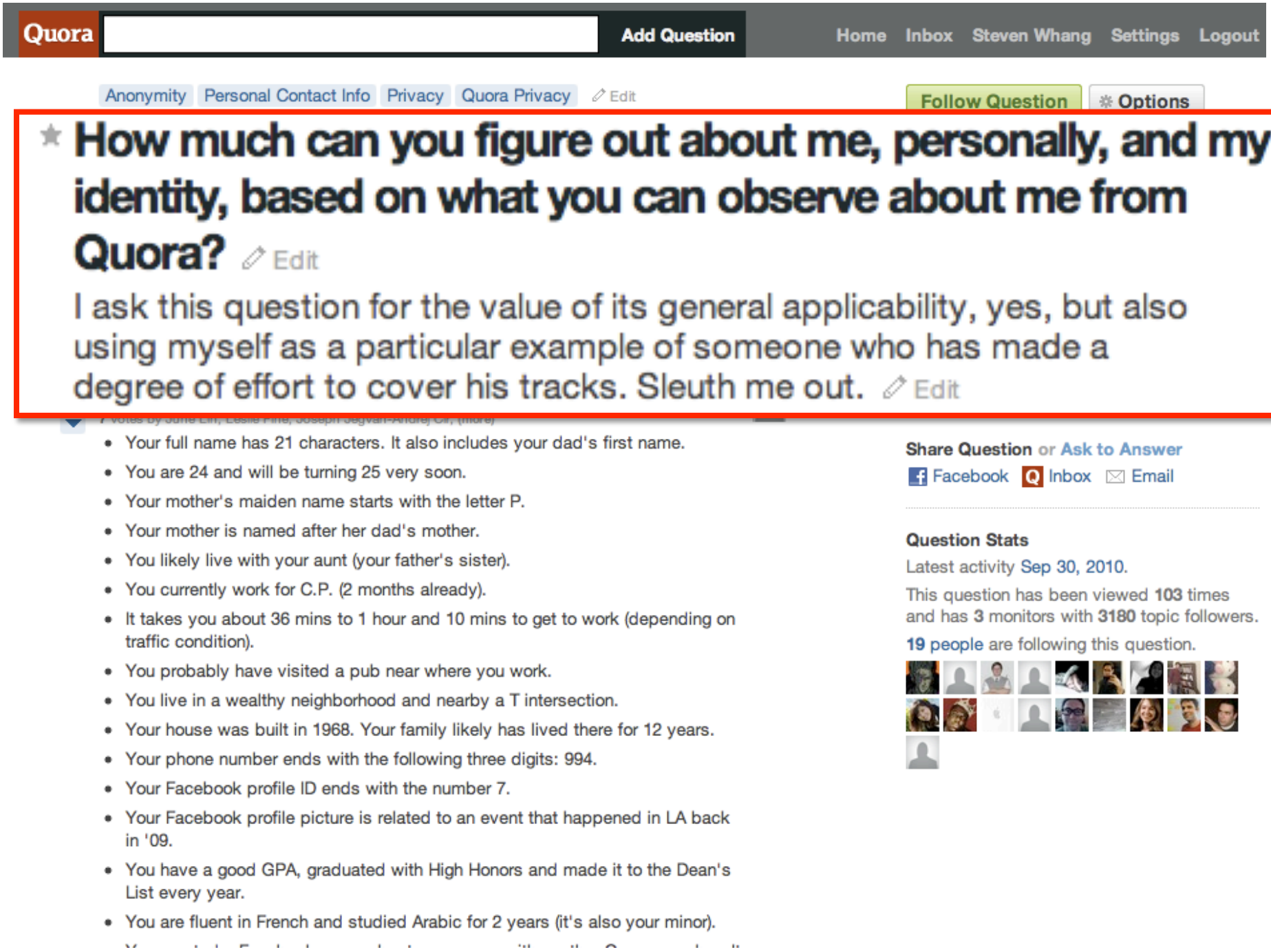
Question Stats

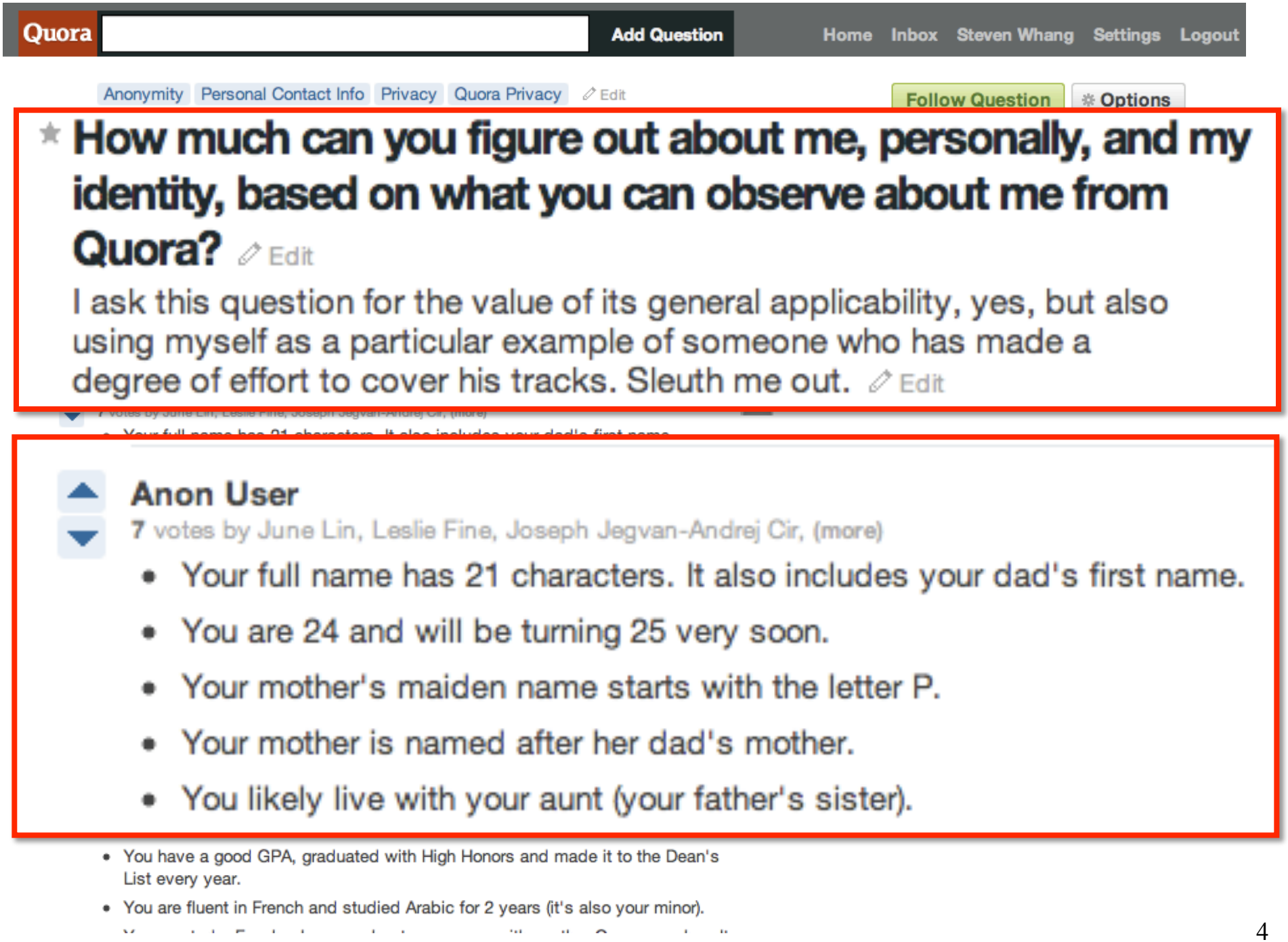
Latest activity Sep 30, 2010.

This question has been viewed **103** times and has **3** monitors with **3180** topic followers.

19 people are following this question.




Quora
Add Question
Home
Inbox
Steven Whang
Settings
Logout
Anonymity
Personal Contact Info
Privacy
Quora Privacy
Edit
Follow Question
Options
How much can you figure out about me, personally, and my identity, based on what you can observe about me from Quora?
Edit
I ask this question for the value of its general applicability, yes, but also using myself as a particular example of someone who has made a degree of effort to cover his tracks. Sleuth me out.
Edit
Your full name has 21 characters. It also includes your dad's first name.
You are 24 and will be turning 25 very soon.
Your mother's maiden name starts with the letter P.
Your mother is named after her dad's mother.
You likely live with your aunt (your father's sister).
You currently work for C.P. (2 months already).
It takes you about 36 mins to 1 hour and 10 mins to get to work (depending on traffic condition).
You probably have visited a pub near where you work.
You live in a wealthy neighborhood and nearby a T intersection.
Your house was built in 1968. Your family likely has lived there for 12 years.
Your phone number ends with the following three digits: 994.
Your Facebook profile ID ends with the number 7.
Your Facebook profile picture is related to an event that happened in LA back in '09.
You have a good GPA, graduated with High Honors and made it to the Dean's List every year.
You are fluent in French and studied Arabic for 2 years (it's also your minor).
Share Question or Ask to Answer
Facebook
Inbox
Email
Question Stats
Latest activity Sep 30, 2010.
This question has been viewed 103 times and has 3 monitors with 3180 topic followers.
19 people are following this question.


Quora
Add Question
Home
Inbox
Steven Whang
Settings
Logout
Anonymity
Personal Contact Info
Privacy
Quora Privacy
Edit
Follow Question
Options
How much can you figure out about me, personally, and my identity, based on what you can observe about me from Quora?
Edit
I ask this question for the value of its general applicability, yes, but also using myself as a particular example of someone who has made a degree of effort to cover his tracks. Sleuth me out.
Edit
Anon User
7 votes by June Lin, Leslie Fine, Joseph Jegvan-Andrej Cir, (more)
Your full name has 21 characters. It also includes your dad's first name.
You are 24 and will be turning 25 very soon.
Your mother's maiden name starts with the letter P.
Your mother is named after her dad's mother.
You likely live with your aunt (your father's sister).
You have a good GPA, graduated with High Honors and made it to the Dean's List every year.
You are fluent in French and studied Arabic for 2 years (it's also your minor).

Joseph Jegvan-Andrej Cir, mistranslated villain.

3 votes by Anon User, Ani Ravi and Rohit Khare

I am a fantastic questioner/answerer, but besides that...

Post mortem on this experiment:

Most of the correct answers from Anon are things I put up deliberately in the course of publicizing my work, but some of it is stuff that is floating around that I would rather wasn't - like my genealogy - kudos for getting that. Most of the stuff that is wrong is based on the few pieces of deliberate misdirection I have put out there, OR confusion with my similarly-named uncle. Therefore, I would say that the lesson here is that **obfuscation may be more useful than discretion** when it comes to protecting your privacy.

Correct:
scary

- You are 24 and will be turning 25 very soon.
- Your mother's maiden name starts with the letter P.
- Your mother is named after her dad's mother.
- You have accounts at the following sites: LinkedIn, Facebook, LiveJournal, CGSociety, and Photobucket.
- Your Facebook profile ID ends with the number 7.
- You created a Facebook game about a year ago with another Quora member. It has 0 daily active users, 0 weekly active users, and 2 monthly active users. It has one review (written by you) and 3 fans (which are your friends). [hey it was just for fun, come on!]
- Also, you don't know me.

deliberately public - discovered once he linked my professional identity to one or more others

- Your phone number ends with the following three digits: 994.
- You know how to program in C++, JavaScript, PHP, and ActionScript.

Joseph Jegvan-Andrej Cir, mistranslated villain.

3 votes by Anon User, Ani Ravi and Rohit Khare

I am a fantastic questioner/answerer, but besides that...

Post mortem on this experiment:

Most of the correct answers from Anon are things I put up deliberately in the course of publicizing my work, but some of it is stuff that is floating around that I would rather wasn't - like my genealogy - kudos for getting that. Most of the stuff that is wrong is based on the few pieces of deliberate misdirection I have put out there. OR confusion with my similarly-named uncle. Therefore, I would say that the

Correct:

scary

- You are 24 and will be turning 25 very soon.
- Your mother's maiden name starts with the letter P.
- Your mother is named after her dad's mother.

- You have accounts at the following sites: LinkedIn, Facebook, LiveJournal,

deliberately public - discovered once he linked my professional identity to one or more others

- Your phone number ends with the following three digits: 994.
- You know how to program in C++, JavaScript, PHP, and ActionScript.

more others

- Your phone number ends with the following three digits: 994.
- You know how to program in C++, JavaScript, PHP, and ActionScript.

Wrong:
Stuff about my semi-namesake uncle

- You have accounts at the following sites: Newsvine
- You are a Democrat. [No party can handle my idiosyncrasies!]
- You read lots of news online and frequently comment about them. [Sorta true, but that's Uncle Joe again. When I comment I rarely use my full name.]
- You like Joe Biden but do not like Hillary Clinton. [meh!]

other

- You likely live with your aunt (your father's sister). [Based on incorrect genealogical info, but gave a shock because coincidentally, this was once true]
- Your Facebook profile picture is related to an event that happened in LA back in '09. [Deliberate obfuscation]

Sorry if this has all seemed self-indulgent; I was hoping it was going to go nowhere but obviously I was wrong! Very eye-opening concerning means to protect oneself - let me reiterate that **it seems that far more effective to allow incorrect information to propagate than to try and stem the tide**. Another benefit of this is that it can be used to identify the sources that the information came from. Knowing my age and family seemed more scary before I found out how he found them, but **seeing the incorrect stuff at least tells me what he was looking at**.

I am actually personally acquainted with a well-known fine artist who has used this type of misdirection in order to make her biographical information harder to pin down, and thus to make it difficult for critics to contextualize her work based on her identity - one of her goals. Now I am beginning to see that anyone can benefit from this if they so choose.

Add Comment • Thank • Not Helpful • Jun 14, 2010

Wrong:

Wrong:
Stuff about my semi-namesake uncle

- You have accounts at the following sites: Newsvine
- You are a Democrat. [No party can handle my idiosyncrasies!]
- You read lots of news online and frequently comment about them. [Sorta true, but that's Uncle Joe again. When I comment I rarely use my full name.]
- You like Joe Biden but do not like Hillary Clinton. [meh!]

other

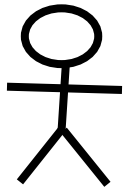
- You likely live with your aunt (your father's sister). [Based on incorrect genealogical info, but gave a shock because coincidentally, this was once true]
- Your Facebook profile picture is related to an event that happened in LA back in '09. [Deliberate obfuscation]

- let me reiterate that **it seems that far more effective to allow incorrect information to propagate than to try and stem the tide**. Another benefit of this

from this if they so choose.

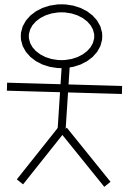
Add Comment • Thank • Not Helpful • Jun 14, 2010

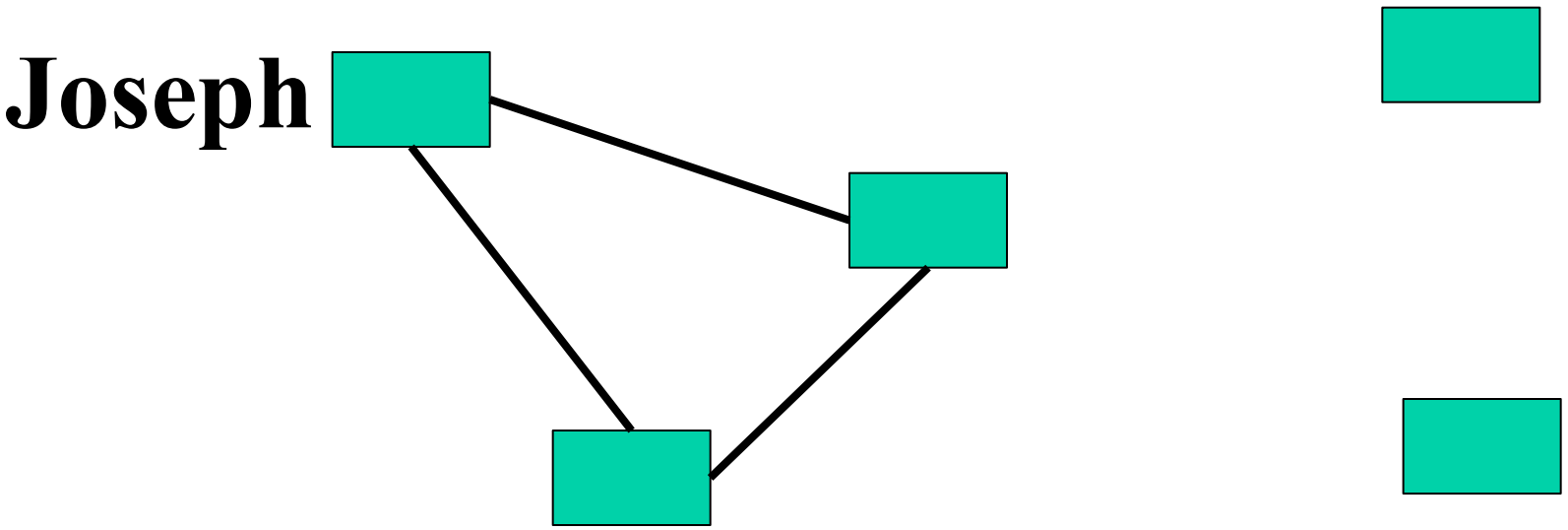
Information Leakage



Joseph

Information Leakage





Model

- p = {<N,n_1>,<C,c_1>,<C,c_2>,<P,p_1>,<A,a_1>}
- D ={s={<N,n_1>,<C,c_1>,<P,p_1>},
  t= {<N,n_1>,<C,c_2>}}
- M(x,y) = *true* if same N,C or N,P
- μ(x,y) = x ∪ y

Record Leakage

- $p = \{<N,n_1>,<C,c_1>,<C,c_2>,<P,p_1>,<A,a_1>\}$
- $r = \{<N,n_1>,<C,c_1>,<P,x>\}$
- $L_r(p, r) = |p \cap r| - |r - p| = 2 - 1 = 1$
  - In general, L_r can be any function

Query Leakage

$p = \{\langle N,n_1\rangle,\langle C,c_1\rangle,\langle C,c_2\rangle,\langle P,p_1\rangle,\langle A,a_1\rangle\}$

$D = \{s=\{\langle N,n_1\rangle,\langle C,c_1\rangle,\langle P,p_1\rangle\},$
$t= \{\langle N,n_1\rangle,\langle C,c_2\rangle\}\}$

$M(x,y)$ = *true* if same N,C or N,P

$\mu(x,y) = x \cup y$

- $q = \{\langle N,n_1\rangle,\langle C,c_1\rangle\}$

Query Leakage

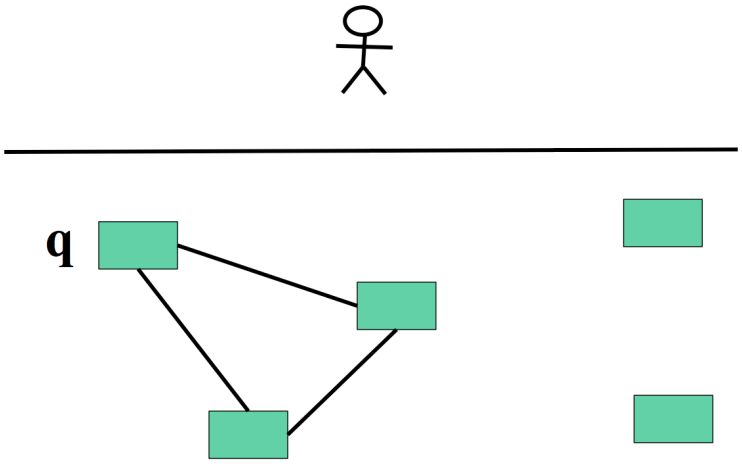
$p = \{<N,n_1>,<C,c_1>,<C,c_2>,<P,p_1>,<A,a_1>\}$

$D = \{s=\{<N,n_1>,<C,c_1>,<P,p_1>\},$
$t= \{<N,n_1>,<C,c_2>\}\}$

$M(x,y)$ = *true* if same N,C or N,P

$\mu(x,y) = x \cup y$

- $q = \{<N,n_1>,<C,c_1>\}$



Query Leakage

$p = \{\langle N,n_1\rangle,\langle C,c_1\rangle,\langle C,c_2\rangle,\langle P,p_1\rangle,\langle A,a_1\rangle\}$

$D = \{s=\{\langle N,n_1\rangle,\langle C,c_1\rangle,\langle P,p_1\rangle\},$
$t= \{\langle N,n_1\rangle,\langle C,c_2\rangle\}\}$

$M(x,y)$ = *true* if same N,C or N,P

$\mu(x,y) = x \cup y$

- $q = \{\langle N,n_1\rangle,\langle C,c_1\rangle\}$

q

s

t

Query Leakage

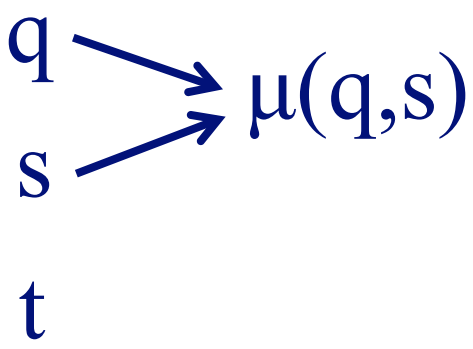
$p = \{\langle N,n_1\rangle,\langle C,c_1\rangle,\langle C,c_2\rangle,\langle P,p_1\rangle,\langle A,a_1\rangle\}$

$D = \{s=\{\langle N,n_1\rangle,\langle C,c_1\rangle,\langle P,p_1\rangle\},$
$t= \{\langle N,n_1\rangle,\langle C,c_2\rangle\}\}$

$M(x,y)$ = *true* if same N,C or N,P

$\mu(x,y) = x \cup y$

- $q = \{\langle N,n_1\rangle,\langle C,c_1\rangle\}$



Query Leakage

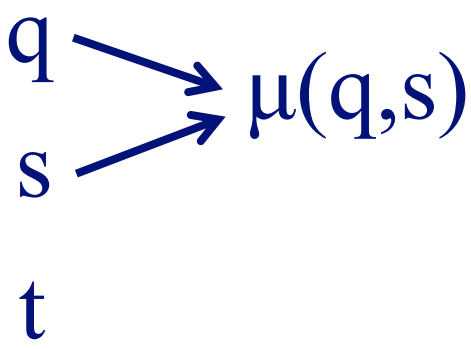
$p = \{\langle N,n_1\rangle,\langle C,c_1\rangle,\langle C,c_2\rangle,\langle P,p_1\rangle,\langle A,a_1\rangle\}$

$D = \{s=\{\langle N,n_1\rangle,\langle C,c_1\rangle,\langle P,p_1\rangle\},$
$t= \{\langle N,n_1\rangle,\langle C,c_2\rangle\}\}$

$M(x,y)$ = *true* if same N,C or N,P

$\mu(x,y) = x \cup y$

- $q = \{\langle N,n_1\rangle,\langle C,c_1\rangle\}$
- $L_q(p, q, D)$
  $= \max\{L_r(p, \mu(q,s))\}$
  $= \max\{3\}$
  $= 3$



Database Leakage

$p = \{<N,n_1>,<C,c_1>,<C,c_2>,<P,p_1>,<A,a_1>\}$

$D = \{s=\{<N,n_1>,<C,c_1>,<P,p_1>\},$
$t= \{<N,n_1>,<C,c_2>\}\}$

$M(x,y)$ = *true* if same N,C or N,P

$\mu(x,y) = x \cup y$

- $L_d(p, D)$
  $= \max\{L_q(p, s, D-\{s\}), L_q(p, t, D-\{t\})\}$
  $= \max\{L_r(p, s), L_r(p, t)\}$
  $= \max\{3, 2\}$
  $= 3$

Key Features

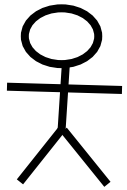
- Incorporated Entity Resolution
- Privacy: NOT all or nothing

- Uncertainty
- Incorrect Information

Interesting Problems

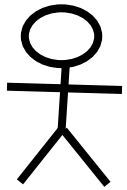
- Releasing Critical Information
- Comparing Entity Resolution Algorithms
- Releasing Disinformation

Releasing Disinformation



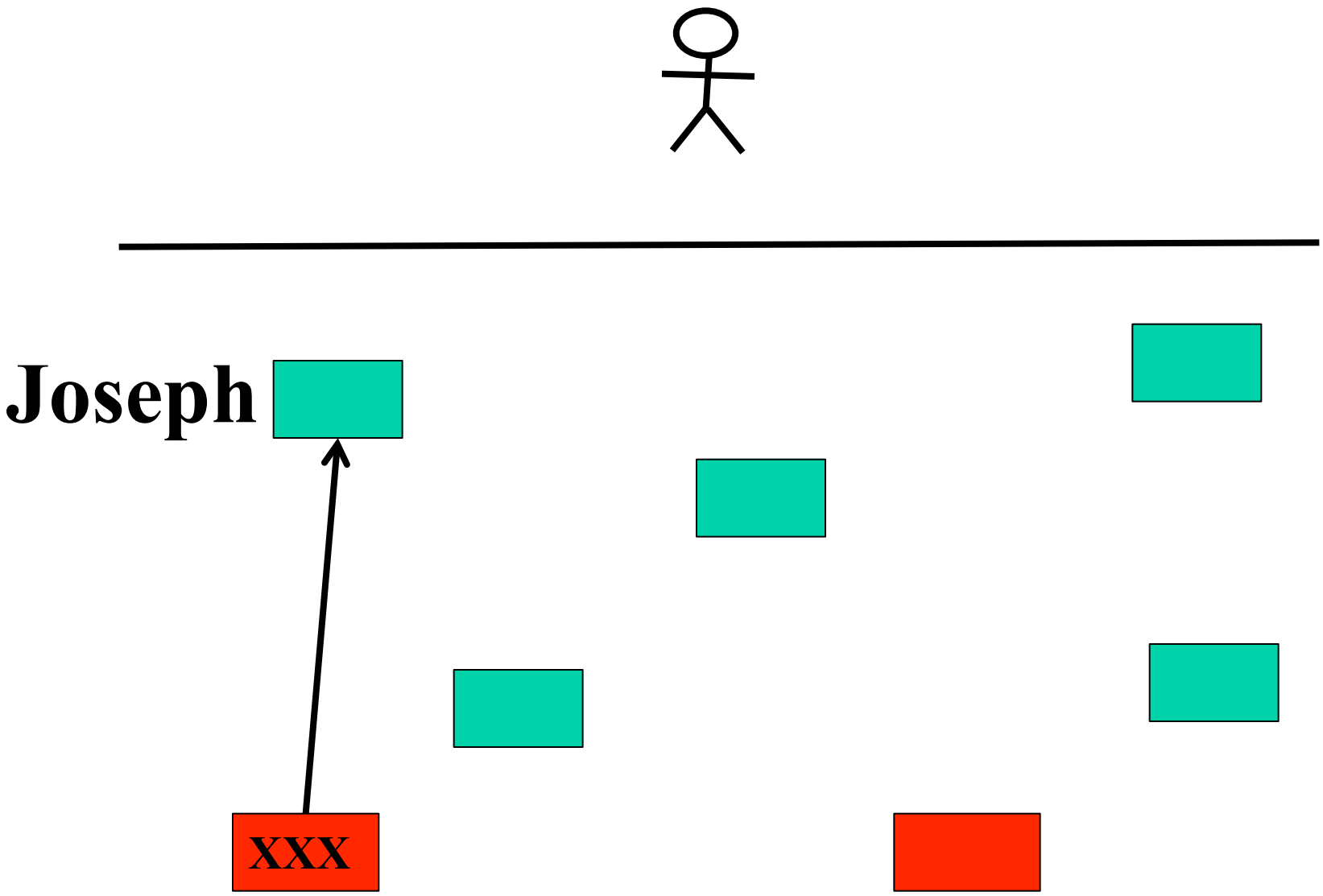
Joseph

Releasing Disinformation

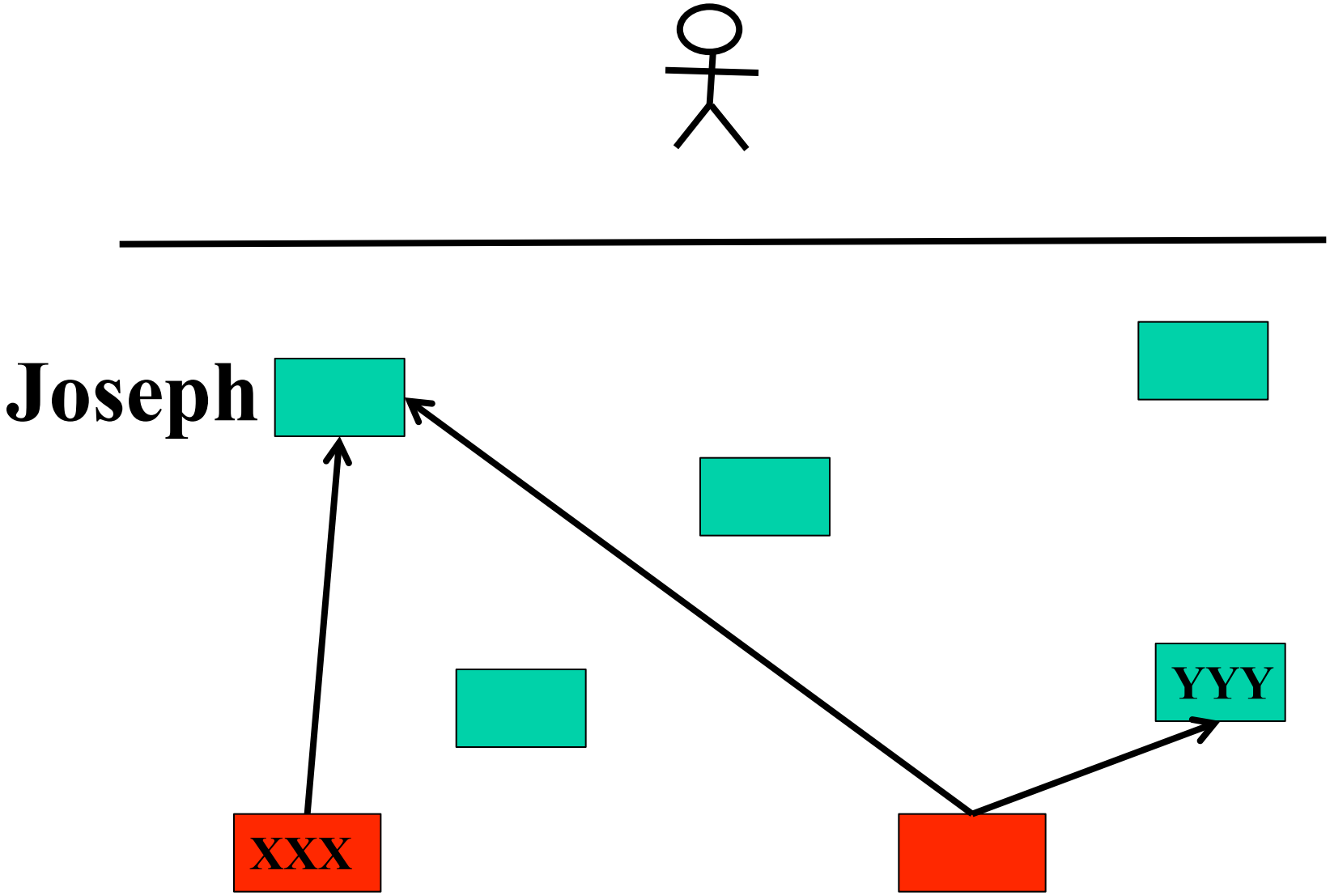


Joseph

Releasing Disinformation



Releasing Disinformation



Releasing Disinformation

- Minimize $L_d(p, DUS)$
  s.t. $\sum_{r \in S} Cost(r) \leq T$

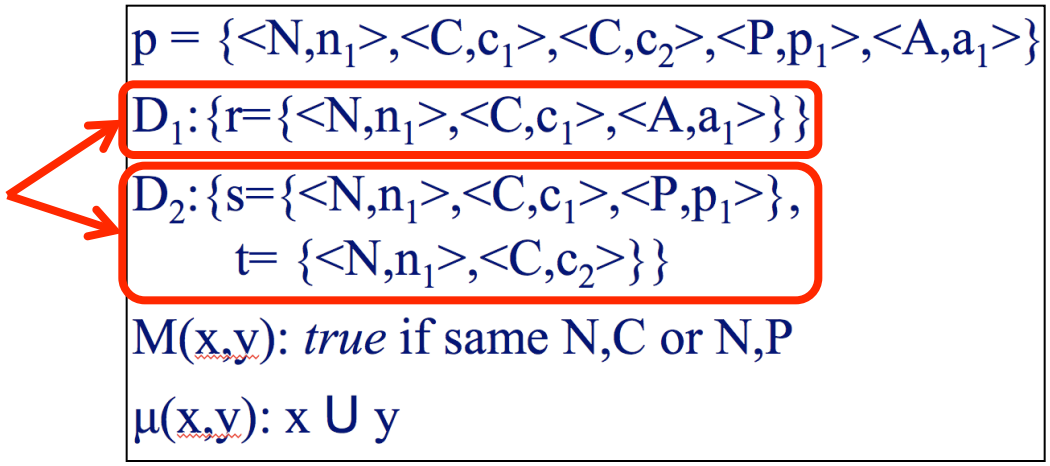
Conclusion

- We have formalized information leakage
  - Incorporated Entity Resolution
  - Privacy: NOT all or nothing
  - Uncertainty
  - Incorrect Information
- We have listed several challenges for managing information leakage

Thanks!

Releasing Critical Information

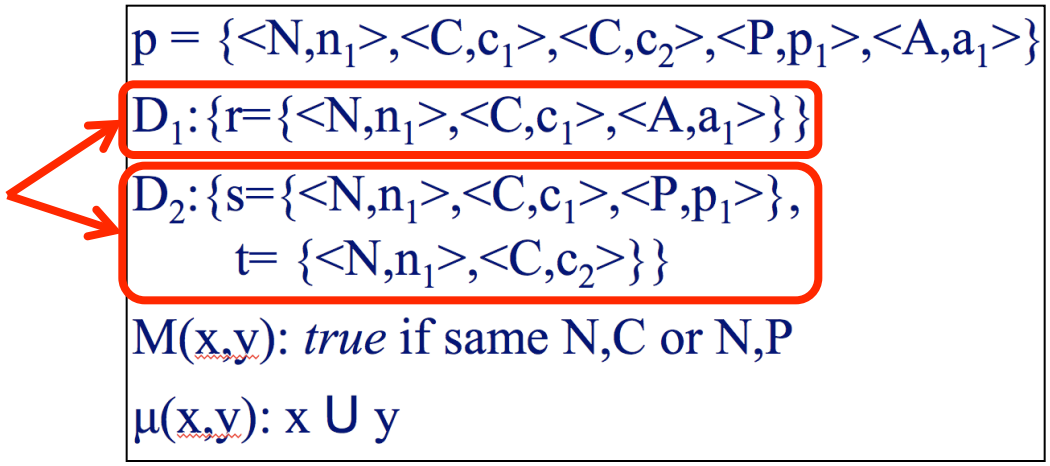
Online shopping websites



- $u=\{<N,n_1>,<C,c_2>,<P,p_1>\}$
- $L_d(p, D_1) = 3$
- $L_d(p, D_2) = 3$

Releasing Critical Information

Online shopping websites



- $u=\{<N,n_1>,<C,c_2>,<P,p_1>\}$
- $L_d(p, D_1) = 3 \rightarrow L_d(p, D_1 \cup \{u\}) = 3$
- $L_d(p, D_2) = 3 \rightarrow L_d(p, D_2 \cup \{u\}) = 4$

Related Work

- ReputationDefender
  - Promotes positive information
- Track-Me-Not
  - Obfuscates search queries

Current Work

- Model
  - Distinguish attributes, better leakage measures, update/delete, utility, privacy measure, …
- Implementation
  - Bogus creation, scalability, …
- More problems
  - Negative effect of disinformation, promoting good information, enhance record, check hypothesis, …