

# Web User Profiling using Data Redundancy

<http://aminer.org/profiling>

Xiaotao Gu, Hong Yang, Jie Tang, Jing Zhang

Tsinghua University



# Web User Profiling using Data Redundancy

- Introduction
- Traditional Way
- Basic Idea
- MagicFG
- Experiments
- Conclusion



Upload

Jiawei Han (韩家炜) Follow | 90

Department of Computer Science, University of Illinois at Urbana-Champaign

Professor

(217) 333-6903

hanj@cs.uiuc.edu

<http://www.cs.uiuc.edu/~hanj/>

Update

External Links



Overview

Papers 824

Lectures 13

Patents 1

Merge

Information

Name: Jiawei Han

名字(Chinese Name): 韩家炜

Position(Splitted by ";"): Professor

Phone: (217) 333-6903

Fax: (217) 265-6494

Email: hanj@cs.uiuc.edu

Affiliation: Department of Computer Science, University of Illinois at Urbana-Champaign

Address: 201 N. Goodwin Avenue, Urbana, IL 61801, USA.

Homepage: <http://www.cs.uiuc.edu/~hanj/>

Skill

- 10 D
- 5 in
- 4 kr
- 4 de
- 4 m
- 4 de
- 3 se
- 3 fr
- 3 sp

- Position
- Phone & Fax
- Email
- Homepage
- Affiliation
- Address

- Expert Finding
- Recommendation
- Getting in Touch
- ...

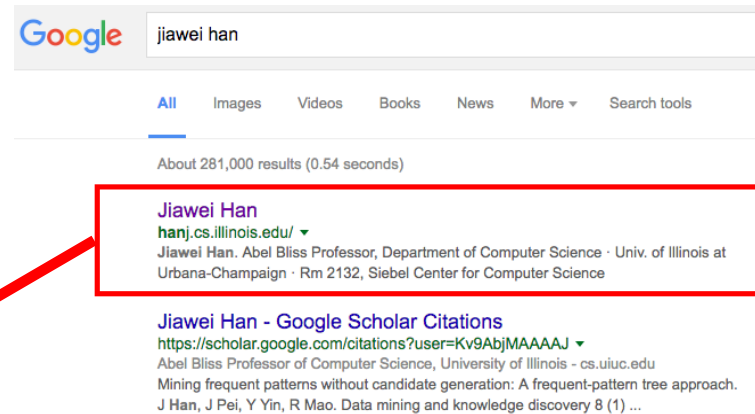


# Web User Profiling using Data Redundancy

- Introduction
- **Traditional Way**
- Basic Idea
- MagicFG
- Experiments
- Conclusion

# Traditional Way: Two-Step

- Source Finding
- Extraction



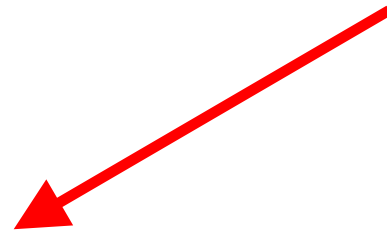
Google jiawei han

All Images Videos Books News More Search tools

About 281,000 results (0.54 seconds)

**Jiawei Han**  
[hanj.cs.illinois.edu/](http://hanj.cs.illinois.edu/)  
Jiawei Han, Abel Bliss Professor, Department of Computer Science · Univ. of Illinois at Urbana-Champaign · Rm 2132, Siebel Center for Computer Science

**Jiawei Han - Google Scholar Citations**  
<https://scholar.google.com/citations?user=Kv9AbjMAAAAJ>  
Abel Bliss Professor of Computer Science, University of Illinois - cs.uiuc.edu  
Mining frequent patterns without candidate generation: A frequent-pattern tree approach.  
J Han, J Pei, Y Yin, R Mao. Data mining and knowledge discovery 8 (1) ...



## Jiawei Han

Abel Bliss Professor, [Department of Computer Science](#)  
[Univ. of Illinois at Urbana-Champaign](#)  
[Rm 2132, Siebel Center for Computer Science](#)  
201 N. Goodwin Avenue  
Urbana, IL 61801, USA  
E-mail: [hanj\[at\]cs.uiuc.edu](mailto:hanj[at]cs.uiuc.edu)

Ph.D. (1985), Computer Science, Univ. Wisconsin-Madison

## Knowledge Discovery and Data Mining

[Data Mining Research Group](#)  
[Data and Information Systems Research Laboratory](#)  
UIUC Calendar: [\(15-16\)](#) ([Cites: Exchange](#)) ([CS](#))  
Office: (217) 333-6903  
Fax: (217) 265-6494  
Web: [hanj.cs.illinois.edu](http://hanj.cs.illinois.edu)  
[Schedule: Meetings and Appointments](#)

- Current Research ([Selected Publications](#))

# Traditional Way: Two-Step

- Source Finding
- Extraction

SVM

CRF

LR



**Jiawei Han**

Abel Bliss Professor, [Department of Computer Science](#)  
[Univ. of Illinois at Urbana-Champaign](#)  
[Rm 2132, Siebel Center for Computer Science](#)  
201 N. Goodwin Avenue  
Urbana, IL 61801, USA  
E-mail: [hanj\[at\]cs.uiuc.edu](mailto:hanj[at]cs.uiuc.edu)

Ph.D. (1985), Computer Science, Univ. Wisconsin-Madison

## Knowledge Discovery and Data Mining

[Data Mining Research Group](#)  
[Data and Information Systems Research Laboratory](#)  
[UIUC Calendar: \(15-16\) \(Cites: Exchange\) \(CS\)](#)

Office: (217) 333-6903

Fax: (217) 265-6494

Web: [hanj.cs.illinois.edu](http://hanj.cs.illinois.edu)

[Schedule: Meetings and Appointments](#)

• Current Research ([Selected Publications](#))



# Traditional Way: Two-Step

- Low Recall – single data source
- Low Precision – error propagation

# Traditional Way: Two-Step

- Low Recall – single data source
- **Low Precision – error propagation**







# Web User Profiling using Data Redundancy

- Introduction
- Traditional Way
- **Basic Idea**
- MagicFG
- Experiments
- Conclusion



# Basic Idea

- **A Uniform Framework**

- ✓ All in one step, avoiding error propagation

- ✓ Incorporate information from different data sources:

Homepage, Google Scholar, Twitter, Linkedin, Facebook, etc.



# Basic Idea

- A Uniform Framework
- Search Engine as the data source



# Basic Idea

Google philip s. yu email

About 951,000 results (1.05 seconds)

**Philip S. Yu - UIC - Computer Science - University of Illinois ...**  
<https://www.cs.uic.edu/PSYu/> University of Illinois at Chicago  
 May 20, 2009 - Philip S. Yu's main research interests include [data mining](#) (especially ... Dr. Yu is a Fellow of the ACM and the IEEE. ... e-mail [psyu@cs.uic.edu](mailto:psyu@cs.uic.edu)  
 Lab - Research - Teaching - Awards

**Philip S. Yu - Google Scholar Citations**  
[scholar.google.com/citations?user=D0IL1r0AAAAAJ](https://scholar.google.com/citations?user=D0IL1r0AAAAAJ) Google Scholar  
 Professor of Computer Science, University of Illinois at Chicago - cs.uic.edu  
 Philip S. Yu. Professor of Computer ... Verified [email at cs.uic.edu](mailto:psyu@cs.uic.edu) | Homepage | Scholar ... CC Aggarwal, JL Wolf, PS Yu, C Procopiuc, JS Park. ACM SIGMOD ...

**Philip S. Yu - Wikipedia, the free encyclopedia**  
[https://en.wikipedia.org/wiki/Philip\\_S.\\_Yu](https://en.wikipedia.org/wiki/Philip_S._Yu) Wikipedia  
 Philip S. Yu (born ca 1952) is an American computer scientist and Professor in Information Technology at the University of Illinois at Chicago, known for his work ...  
 Missing: email

**dblp: Philip S. Yu**  
[dblp.uni-trier.de](http://dblp.uni-trier.de) Home Persons University of Trier  
 List of computer science publications by Philip S. Yu.  
 Missing: email

**Jiawei Han**  
[hanj.cs.illinois.edu/](http://hanj.cs.illinois.edu/)  
 E-mail [hanj\[at\]cs.uic.edu](mailto:hanj[at]cs.uic.edu) Ph.D. ... Philip S. Yu, Jiawei Han, and Christos Faloutsos (eds), Link Mining: Models, Algorithms, and Applications, Springer, 2010. You've visited this page 2 times. Last visit: 8/14/15

**Philip Yu | Computer Science and Computer Engineering ...**  
[computer-science-and-computer-engineering.uark.edu/you.php](http://computer-science-and-computer-engineering.uark.edu/you.php)  
 Philip Yu distinguished speaker. ... Dr. Philip S. Yu, University of Illinois, Chicago Friday, November 6, 2015 3:05pm - 3:55pm. JBHT 144. Abstract: Philip Yu.

**Philip S. Yu - Journals, Conferences, Proceedings, Open ...**  
[www.scrip.org/journal/DetaildInforOfEditorialBoard.aspx?personID...](http://www.scrip.org/journal/DetaildInforOfEditorialBoard.aspx?personID...)  
 Philip S. Yu Computer Science Department, University of Illinois at Chicago, USA. Email [psyu@uic.edu](mailto:psyu@uic.edu) Qualifications. 1978 Ph.D., New York University, USA, ...

Google philip s. yu affiliation

About 516,000 results (0.72 seconds)

**dblp: Philip S. Yu**  
[dblp.uni-trier.de](http://dblp.uni-trier.de) Home Persons University of Trier  
 List of computer science publications by Philip S. Yu. ... Person information. affiliation: University of Illinois at Chicago. [+][-] ..... Charu C. Aggarwal, Philip S. Yu:

**Philip S. Yu - Google Scholar Citations**  
[scholar.google.com/citations?user=D0IL1r0AAAAAJ](https://scholar.google.com/citations?user=D0IL1r0AAAAAJ) Google Scholar  
 Professor of Computer Science, University of Illinois at Chicago - cs.uic.edu  
 Mining concept-drifting data streams using ensemble classifiers. H Wang, W Fan, PS Yu, J Han. Proceedings of the ninth ACM SIGKDD international conference ...

**Philip S. Yu - Wikipedia, the free encyclopedia**  
[https://en.wikipedia.org/wiki/Philip\\_S.\\_Yu](https://en.wikipedia.org/wiki/Philip_S._Yu) Wikipedia  
 Philip S. Yu (born ca 1952) is an American computer scientist and Professor in Information Technology at the University of Illinois at Chicago, known for his work ...

**Link Mining: Models, Algorithms, and Applications**  
<https://books.google.com/books?isbn=1441965157>  
 Philip S. Yu, Jiawei Han, Christos Faloutsos - 2010 - Science  
 Philip S. Yu, Jiawei Han, Christos Faloutsos ... Affiliation is a transitive relationship; therefore, all individuals sharing an affiliation form a clique.2 An affiliation ...

**Discovering High-Order Periodic Patterns - Springer**  
[link.springer.com/10.1007%2Fs101...](http://link.springer.com/10.1007%2Fs101...) Springer Science+Business Media  
 by J Yang - 2004 - Cited by 20 - Related articles  
 Authors. Jiong Yang - [jioyang@cs.uic.edu](mailto:jioyang@cs.uic.edu) (1); Wei Wang (2); Philip S. Yu (3).  
 Author Affiliations. 1. Computer Science Department, UIUC, Urbana, IL, USA; 2.

**On clustering massive text and categorical data streams ...**  
[link.springer.com/10.1007%2Fs101...](http://link.springer.com/10.1007%2Fs101...) Springer Science+Business Media  
 by CC Aggarwal - 2010 - Cited by 47 - Related articles  
 Aug 6, 2009 - Authors. Charu C. Aggarwal · [charu@us.ibm.com](mailto:charu@us.ibm.com) (1); Philip S. Yu (2).  
 Author Affiliations. 1. IBM T. J. Watson Research Center, 19 Skyline Drive, ...

**An Introduction to Privacy-Preserving Data Mining - Springer**  
[link.springer.com/10.1007%2F978-...](http://link.springer.com/10.1007%2F978-...) Springer Science+Business Media  
 by CC Aggarwal - 2008 - Cited by 7 - Related articles  
 (3); Philip S. Yu · [psyu@cs.uic.edu](mailto:psyu@cs.uic.edu) (4), Editor Affiliations. 1. IBM Thomas J. Watson Research Center; 2. Department of Computer Science, University of Illinois at ...

# Basic Idea

## - Search Engine as Data Source

### *Why snippets?*

#### ✓ Efficient

- Different from traditional methods that crawled each of the relevant pages, it is much faster and more stable, as different servers that host the relevant pages may have very different network speed.

#### ✓ Effective

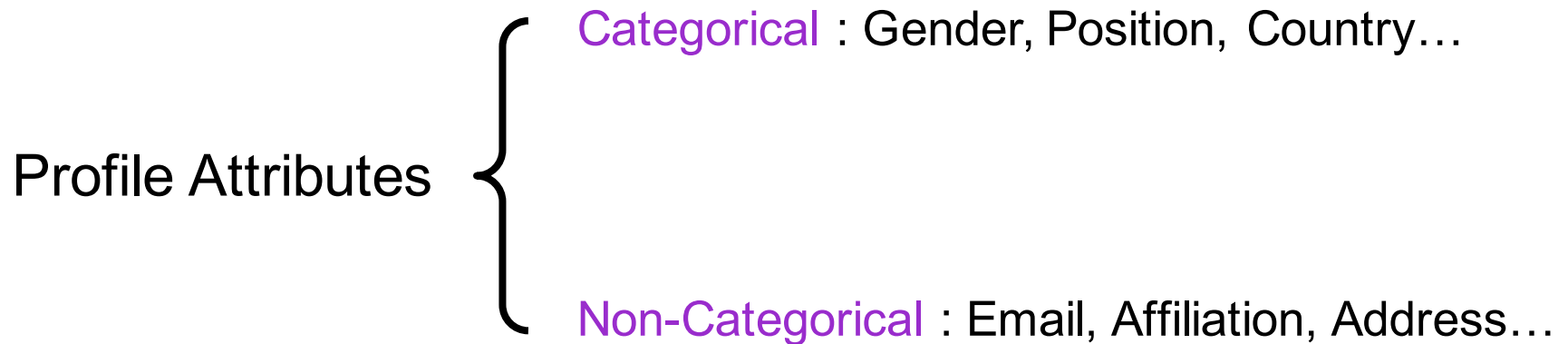
- we found with the constructed “smart” queries, more than 90% of the profile attributes are already contained in the snippets returned by the search engine.

#### ✓ Economical

- One additional advantage is that we do not need to maintain a large database to record all the relevant pages for all the query persons. This is very important, as, for example, in AMiner, we have more than 130,000,000 researchers—maintaining such a big database for all researchers itself is a challenging task.

# Basic Idea

- A Uniform Framework
- Search Engine as the Data Source
- **Smart Query Construction**



# Query Construction

Non-Categorical



Person\_Name + Attribute\_Name



Query = "Phillip S. Yu email"

Google philip s. yu email

All News Shopping Images Maps More Search tools

About 951,000 results (1.05 seconds)

**Philip S. Yu - UIC - Computer Science - University of Illinois ...**  
<https://www.cs.uic.edu/PSYu/> University of Illinois at Chicago  
May 20, 2009 - **Philip S. Yu's** main research interests include **data mining** (especially ... Dr. Yu is a Fellow of the ACM and the IEEE. ... e-mail **psyu@cs.uic.edu**.  
Lab - Research - Teaching - Awards

**Philip S. Yu - Google Scholar Citations**  
[scholar.google.com/citations?user=D0IL1r0AAAAJ](https://scholar.google.com/citations?user=D0IL1r0AAAAJ) Google Scholar  
Professor of Computer Science, University of Illinois at Chicago - cs.uic.edu  
**Philip S. Yu**. Professor of Computer ... Verified **email at cs.uic.edu** · Homepage · Scholar ... CC Aggarwal, JL Wolf, PS Yu, C Procopiuc, JS Park. ACM SIGMoD ...

**Philip S. Yu - Wikipedia, the free encyclopedia**  
[https://en.wikipedia.org/wiki/Philip\\_S\\_Yu](https://en.wikipedia.org/wiki/Philip_S_Yu) Wikipedia  
**Philip S. Yu** (born ca 1952) is an American computer scientist and Professor in Information Technology at the University of Illinois at Chicago, known for his work ...  
Missing: email

**dblp: Philip S. Yu**  
[dblp.uni-trier.de](https://dblp.uni-trier.de) Home Persons University of Trier  
List of computer science publications by **Philip S. Yu**.  
Missing: email

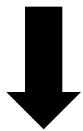
**Jiawei Han**  
[hanj.cs.illinois.edu/](https://hanj.cs.illinois.edu/)  
E-mail: [hanj\[at\]cs.uiuc.edu](mailto:hanj[at]cs.uiuc.edu). Ph.D. ... **Philip S. Yu**, Jiawei Han, and Christos Faloutsos (eds), **Link Mining: Models, Algorithms, and Applications**, Springer, 2010. You've visited this page 2 times. Last visit: 8/14/15

**Philip Yu | Computer Science and Computer Engineering ...**  
[computer-science-and-computer-engineering.uark.edu/you.php](https://computer-science-and-computer-engineering.uark.edu/you.php)  
Philip Yu distinguished speaker. ... Dr. **Philip S. Yu**, University of Illinois, Chicago Friday, November 6, 2015 3:05pm - 3:55pm. JBHT 144. Abstract: Philip Yu.

**Philip S. Yu - Journals, Conferences, Proceedings, Open ...**  
[www.scirp.org/journal/DetailedInforOfEditorialBoard.aspx?personID...](https://www.scirp.org/journal/DetailedInforOfEditorialBoard.aspx?personID...)  
**Philip S. Yu**. Computer Science Department. University of Illinois at Chicago, USA. Email [psyu@uic.edu](mailto:psyu@uic.edu). Qualifications. 1978 Ph.D., New York University, USA, ...

# Query Construction

Categorical



Person\_Name  
+ Representative Words



Query = "Phillip S. Yu his OR her"

Google philip s yu his OR her

All Images News Maps Videos More Search tools

About 21,400,000 results (0.41 seconds)

**Philip S. Yu - Wikipedia, the free encyclopedia**  
[https://en.wikipedia.org/wiki/Philip\\_S.\\_Yu](https://en.wikipedia.org/wiki/Philip_S._Yu)  
Philip S. Yu (born ca 1952) is an American computer scientist and Professor in Information Technology at the University of Illinois at Chicago, known for his work ...

**[PDF] About Me Philip S. Yu is a Professor in Computer Science ...**  
[scimaps.org/exhibit/images/130325/PR2-Semantics-Yu.pdf](https://scimaps.org/exhibit/images/130325/PR2-Semantics-Yu.pdf)  
Philip S. Yu is a Professor in Computer Science at the University of Illinois at Chicago and also ... Dr. Yu spent most of his career at IBM, where he was manager.

**Philip S. Yu - UIC - Computer Science - University of Illinois at ...**  
<https://www.cs.uic.edu/PSYu/>  
May 20, 2009 - Philip S. Yu's main research interests include data mining (especially ... He spent most of his career at IBM Thomas J. Watson Research Center, ... You've visited this page 3 times. Last visit: 5/8/16

**Philip S. Yu • IEEE Computer Society**  
<https://www.computer.org/web/awards/technical-philip-yu>  
Philip S. Yu has been awarded Recipient of "Year" Technical Achievement Award for ... Dr. Yu spent most of his career at IBM, where he was manager of the ...

**Philip S. Yu - Google Scholar Citations**  
[scholar.google.com/citations?user=D0IL1r0AAAAJ](https://scholar.google.com/citations?user=D0IL1r0AAAAJ)  
Professor of Computer Science, University of Illinois at Chicago - cs.uic.edu  
Mining concept-drifting data streams using ensemble classifiers. H Wang, W Fan, PS Yu, J Han. Proceedings of the ninth ACM SIGKDD international conference ...

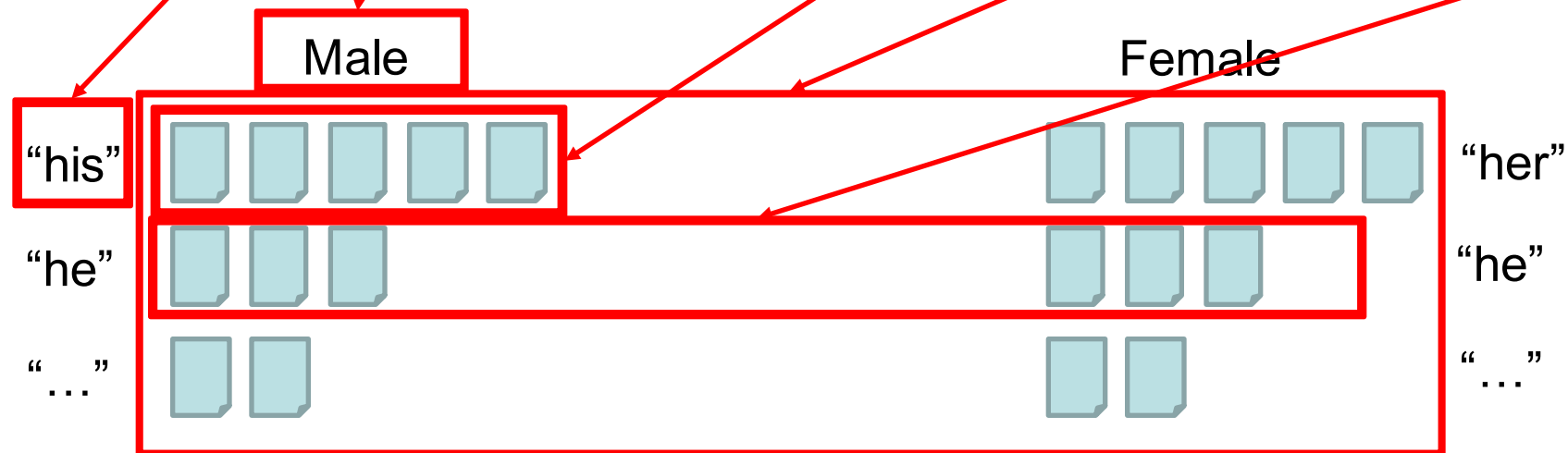
**On Mining Big Data - Philip S. Yu - Computer Science**  
[www.cs.iit.edu/~mbilgic/seminar/philip\\_yu.html](http://www.cs.iit.edu/~mbilgic/seminar/philip_yu.html)  
Philip S. Yu is currently a Professor in the Department of Computer Science at the ... His research interests include big data, data mining, privacy preserving data ...

**Philip Yu | Computer Science and Computer Engineering | Uni...**  
[computer-science-and-computer-engineering.uark.edu/you.php](http://computer-science-and-computer-engineering.uark.edu/you.php)  
Nov 6, 2015 - Philip Yu distinguished speaker. ... Dr. Philip S. Yu, University of Illinois, Chicago Friday ... Dr. Yu received his Ph.D. from Stanford University.



# Representative Words

$$\text{TF-IDF}(w, c) = (1 + \log n(S_c, w)) \log\left(1 + \frac{|S|}{n(S, w)}\right)$$



Query = "Phillip S. Yu his OR her"



# Basic Idea

- A Uniform Framework
- Search Engine as the Data Source
- Smart Query Construction
- **Basic Classification**



# Feature Definition

## Email

- First name in prefix
- Last name in prefix
- Initials in prefix
- ...

## Gender

- How many “his”
- How many “her”
- ...

# Basic Classification

## Email

Method	Precision	Recall	F1-score
CTRF	90.20	83.83	86.90
Rule	87.81	89.64	88.72
SVM	88.26	89.25	88.75
RF	90.76	90.58	90.56
LR	89.07	91.14	90.11

## Gender

Method	Precision	Recall	F1-score
FGNL	<b>94.66</b>	80.88	87.23
Rule	92.12	88.32	90.18
SVM	91.98	90.60	91.29
RF	90.17	89.99	90.08
LR	91.48	91.54	91.51

Uniformly outperform the baselines (CTRF, FGNL)

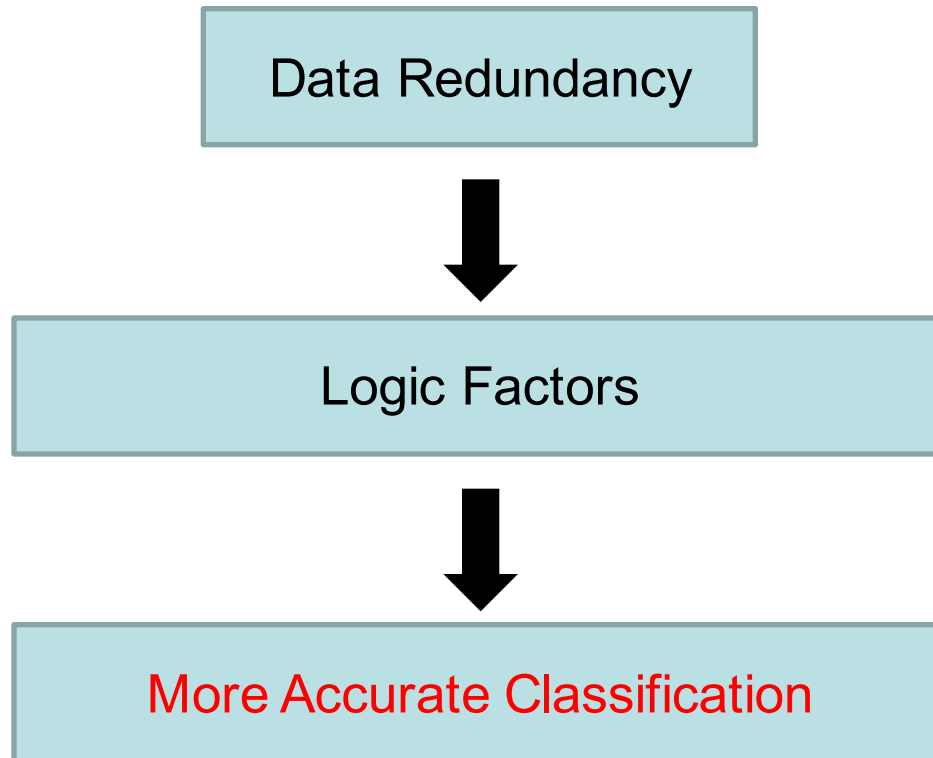


# Web User Profiling using Data Redundancy

- Introduction
- Traditional Way
- Basic Idea
- **MagicFG**
- Experiments
- Conclusion

# MagicFG

## - Markov Logic Factor Graph



Philip S. Yu - UIC - Computer Science - University of Illinois ...  
<https://www.cs.uic.edu/PSYu/> University of Illinois at Chicago  
May 20, 2009 - Philip S. Yu's main research interests include **data mining** (especially ... Dr. Yu is a Fellow of the ACM and the IEEE. ... e-mail [psyu@cs.uic.edu](mailto:psyu@cs.uic.edu).  
Lab - Research - Teaching - Awards

Philip S. Yu - Google Scholar Citations  
[scholar.google.com/citations?user=D0IL1r0AAAAJ](https://scholar.google.com/citations?user=D0IL1r0AAAAJ) Google Scholar  
Professor of Computer Science, University of Illinois at Chicago - [cs.uic.edu](https://www.cs.uic.edu)  
Philip S. Yu. Professor of Computer ... Verified e-mail at [cs.uic.edu](mailto:psyu@cs.uic.edu) · Homepage · Scholar ... CC Aggarwal, JL Wolf, PS Yu, C Procopiuc, JS Park. ACM SIGMoD ...

Philip S. Yu - Wikipedia, the free encyclopedia  
[https://en.wikipedia.org/wiki/Philip\\_S.\\_Yu](https://en.wikipedia.org/wiki/Philip_S._Yu) Wikipedia  
Philip S. Yu (born ca 1952) is an American computer scientist and Professor in Information Technology at the University of Illinois at Chicago, known for his work ...  
Missing: e-mail

dblp: Philip S. Yu  
[dblp.uni-trier.de](https://dblp.uni-trier.de) Home Persons University of Trier  
List of computer science publications by Philip S. Yu.  
Missing: e-mail

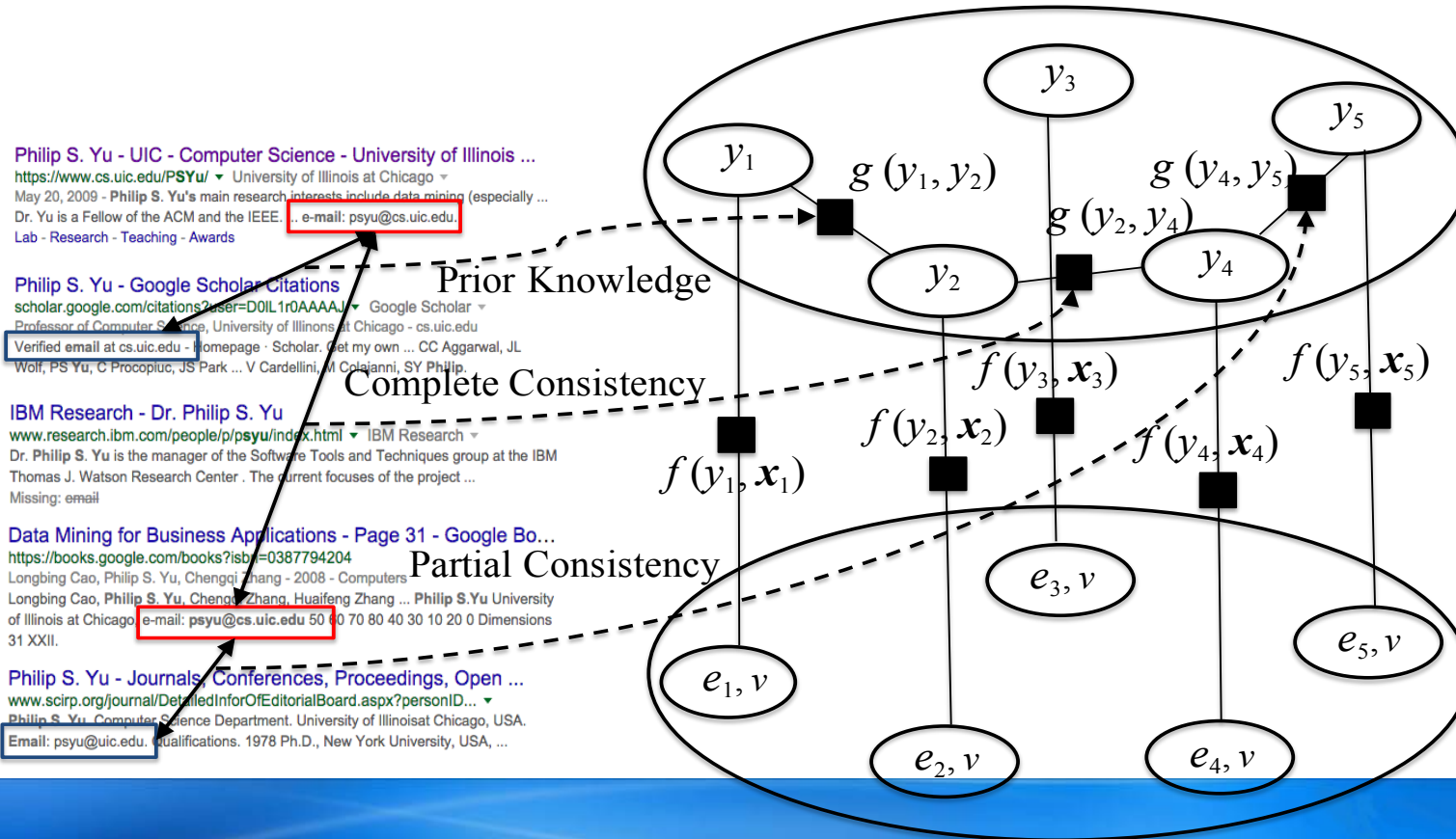
Jiawei Han  
[hanj.cs.illinois.edu/](http://hanj.cs.illinois.edu/)  
E-mail: [hanj\[at\]cs.uic.edu](mailto:hanj[at]cs.uic.edu), Ph.D. ... Philip S. Yu, Jiawei Han, and Christos Faloutsos (eds), **Link Mining: Models, Algorithms, and Applications**, Springer, 2010.  
You've visited this page 2 times. Last visit: 8/14/15

Philip Yu | Computer Science and Computer Engineering ...  
[computer-science-and-computer-engineering.uark.edu/yu.php](http://computer-science-and-computer-engineering.uark.edu/yu.php)  
Philip Yu distinguished speaker. ... Dr. Phillip S. Yu, University of Illinois, Chicago  
Friday, November 6, 2015 3:05pm - 3:55pm. JBHT 144. Abstract: Philip Yu.

Philip S. Yu - Journals, Conferences, Proceedings, Open ...  
[www.scirp.org/journal/DetailledInforOfEditorialBoard.aspx?personID...](http://www.scirp.org/journal/DetailledInforOfEditorialBoard.aspx?personID...)  
Philip S. Yu, Computer Science Department, University of Illinois at Chicago, USA.  
Email: [psyu@uic.edu](mailto:psyu@uic.edu), Qualifications. 1978 Ph.D., New York University, USA, ...

# Why logic factors?

- ✓ Depict and utilize correlations between possible candidates from redundant data.
- ✓ Incorporate human knowledge to guide and amend the classification model.

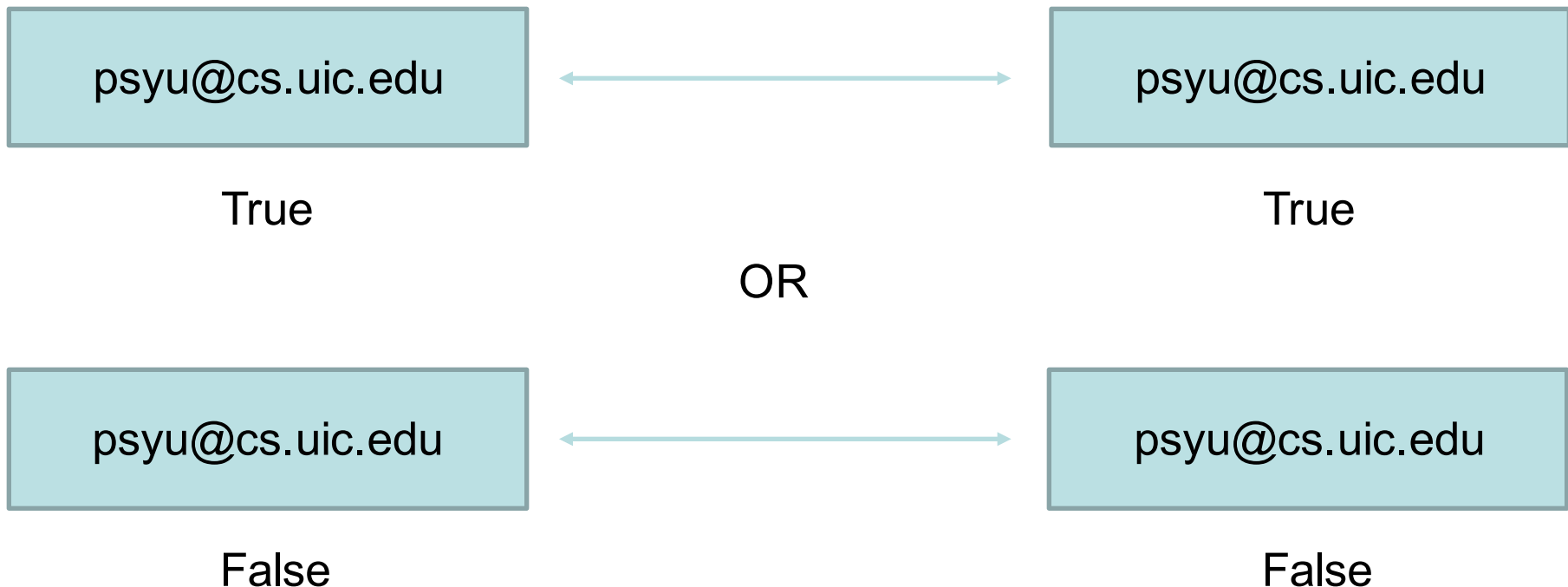


# Logic Factors

$$\text{Equals}(e_i, e_j) \Rightarrow \text{Equals}(y_i, y_j)$$

- Complete Consistency

Two same vertices must share the same label.





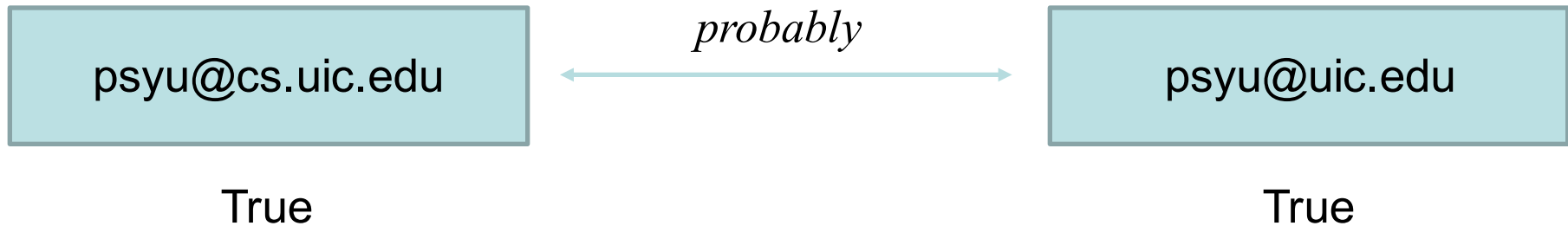
# Logic Factors

$$\text{SamePrefix}(e_i, e_j) \Rightarrow \text{True}(y_i) \wedge \text{True}(y_j)$$

- **Partial Consistency**

Two similar vertices probably share the same (preferred) label.

e.g. Two Emails sharing the same prefix are probably both credible for the target user.



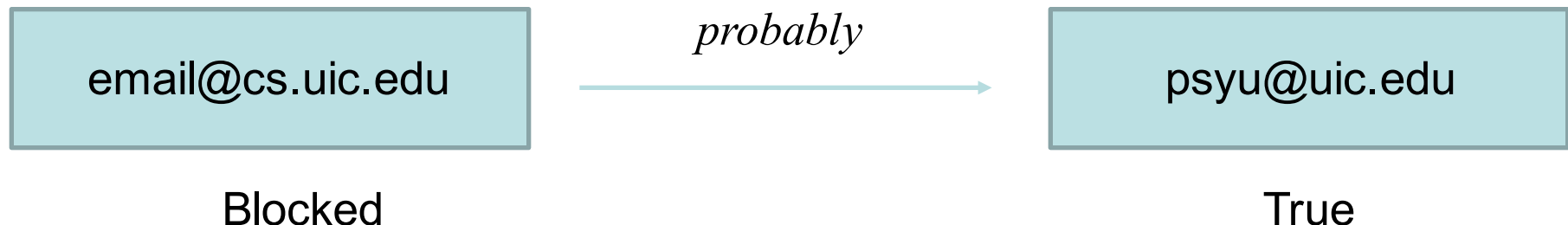
# Logic Factors

$$\text{IsBlocked}(e_j) \wedge \text{SameDomain}(e_i, e_j) \\ \Rightarrow \text{True}(y_i) \wedge \text{False}(y_j).$$

- **Prior Knowledge**

Some prior knowledge can be converted to logic factors.

e.g. Some Email addresses are modified (blocked) for some reason, whose domains are still visible and credible. Emails with the same domain with a blocked one are probably valid.



# Markov Logic Factor Graph

- **Attribute factor function**  $f(v, e_i, y_i) = \frac{1}{Z_a} \exp\left\{\sum_k \alpha_k \phi_k(y_i, \mathbf{x}_i)\right\},$
- **Logic factor function**  $g(y_i, y_j) = \frac{1}{Z_b} \exp\left\{\sum_m \beta_m \psi_m(y_i, y_j)\right\},$
- **Log-likelihood function** 
$$\begin{aligned} \log P(Y|X, \theta) &= \sum_{y_i \in Y} \sum_k \alpha_k \phi_k(y_i, \mathbf{x}_i) \\ &+ \sum_{e_i \sim e_j} \sum_m \beta_m \psi_m(y_i, y_j) - \log Z, \end{aligned}$$
- **Target parameter**  $\theta^* = \arg \max_{\theta} \log P(Y|X, \theta).$

# Markov Logic Factor Graph

- Training: Gradient Ascent

- Gradient: 
$$\frac{\partial L(\theta)}{\partial \alpha_k} = \mathbb{E}[\phi(y_i, \mathbf{x}_i)] - \mathbb{E}_P(y_i, \mathbf{x}_i)[\phi(y_i, \mathbf{x}_i)].$$

- Learning: 
$$\theta_{new} = \theta_{old} + \eta \cdot \frac{\mathcal{O}(\theta)}{\theta},$$

- Classification:

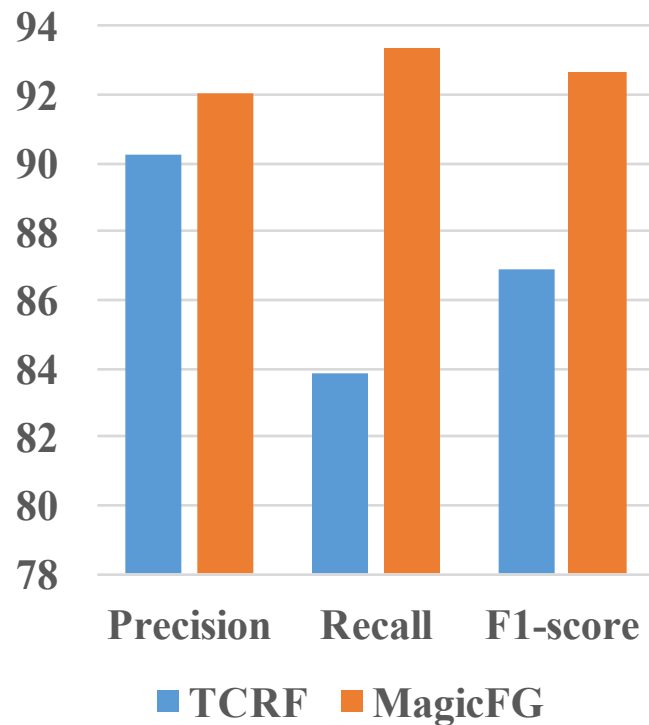
$$Y_v = \arg \max_{Y_v} P(Y_v | X_v, \theta)$$

# Web User Profiling using Data Redundancy

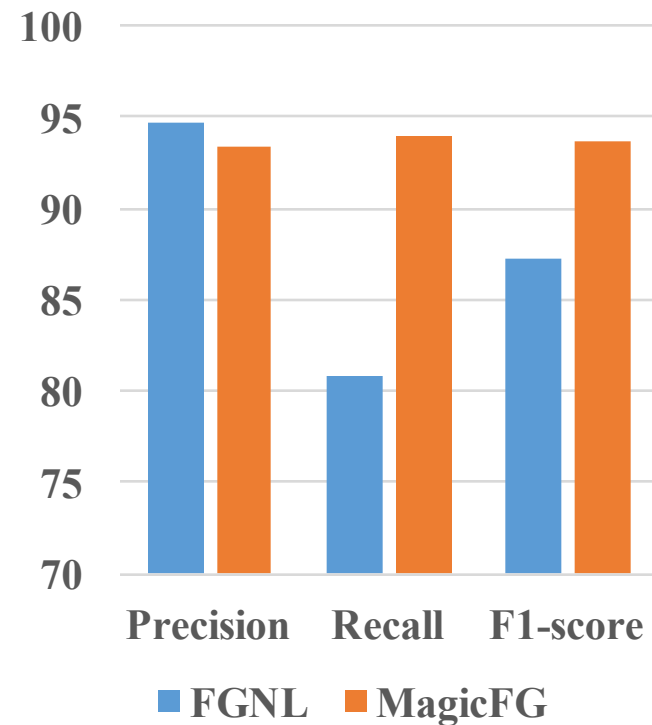
- Introduction
- Traditional Way
- Basic Idea
- MagicFG
- **Experiments**
- Conclusion

# Accuracy Performance

- Comparison between MagicFG and state-of-the-art methods for Email and Gender extraction



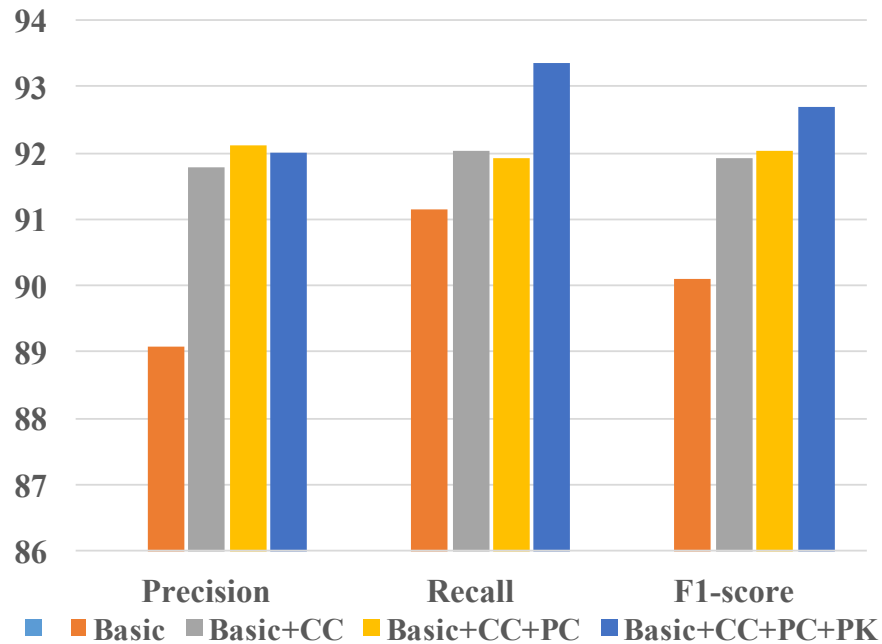
Email



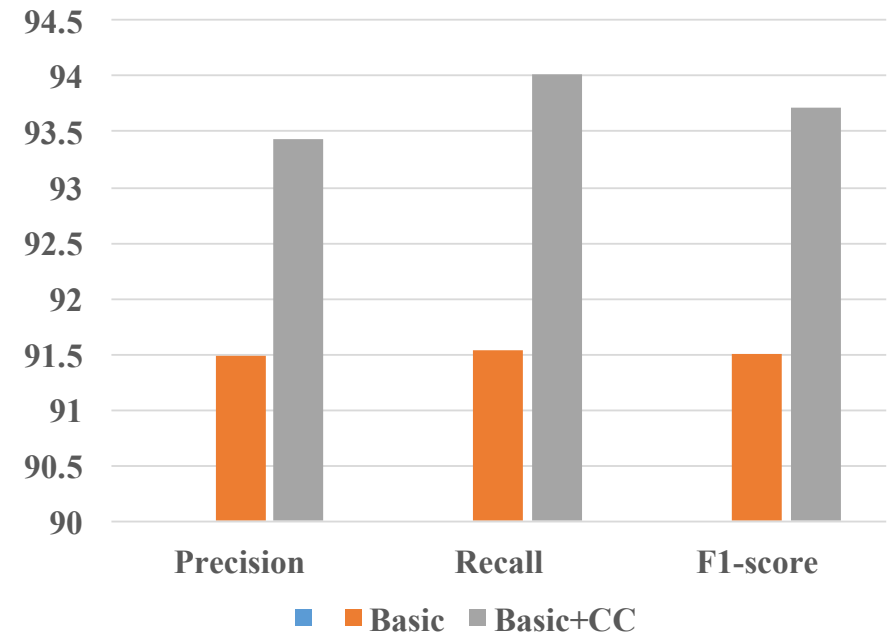
Gender

# Accuracy Performance

Logic factors do help!



Email



Gender



# Web User Profiling using Data Redundancy

- Introduction
- Traditional Way
- Basic Idea
- MagicFG
- Experiments
- **Conclusion**



# Conclusion

- **Motivation**
  - To solve the problem of **low recall** and **error propagation** in traditional **two-step** methods.
- **Basic Idea**
  - **Search engine** as the data source.
- **MagicFG**
  - Utilize correlations in **redundant data**.
  - Incorporate **human knowledge**

# Thank you!

Code & Data

<http://aminer.org/profiling>