

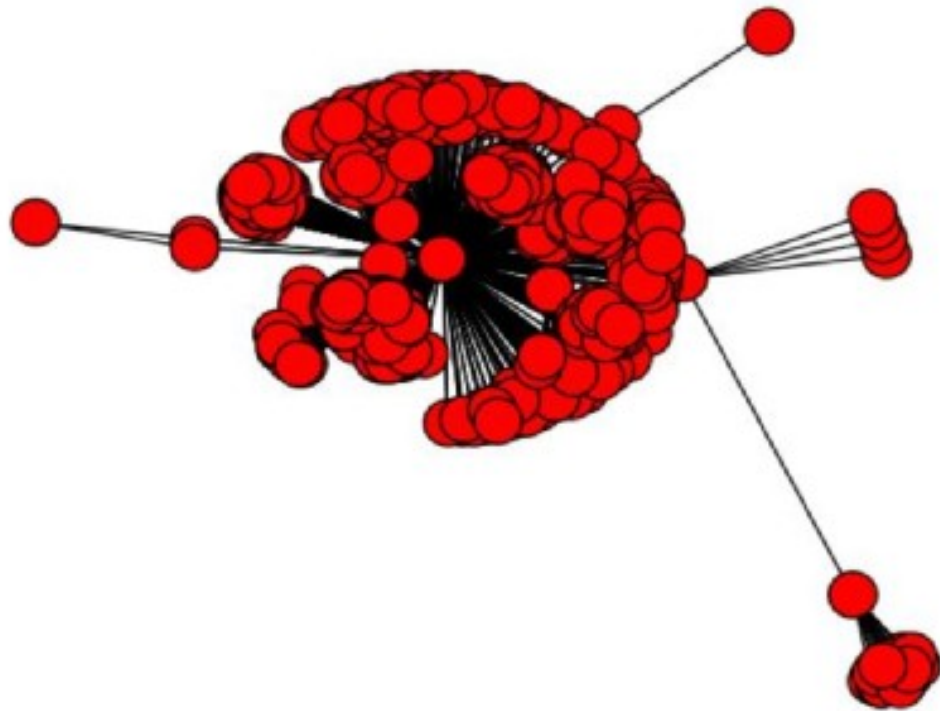


Why Finding Email Correspondents in Online Social Networks is Important

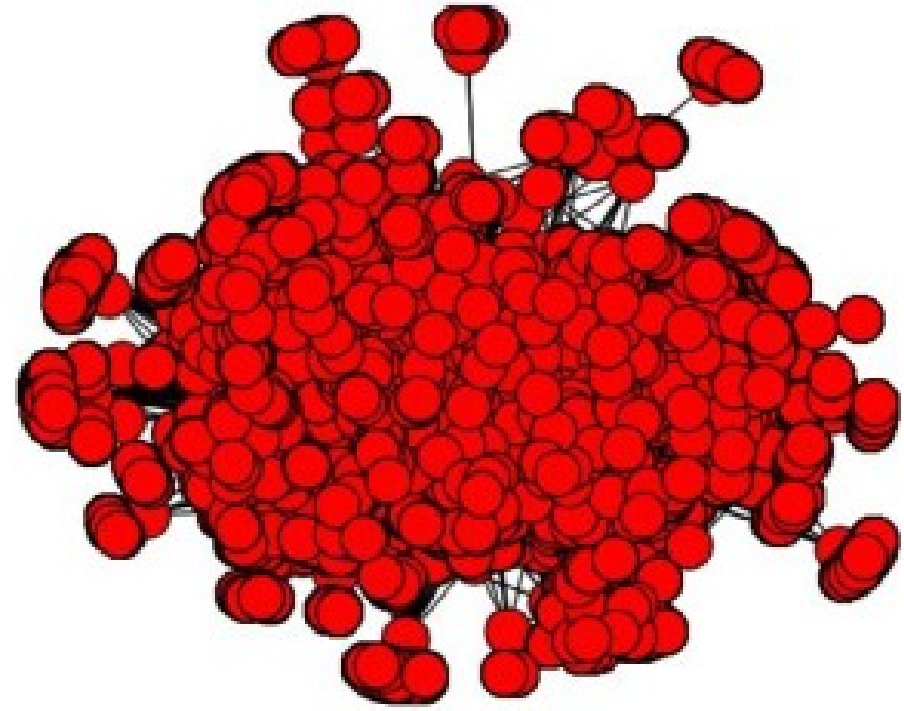
Yu Chen

Two Networks

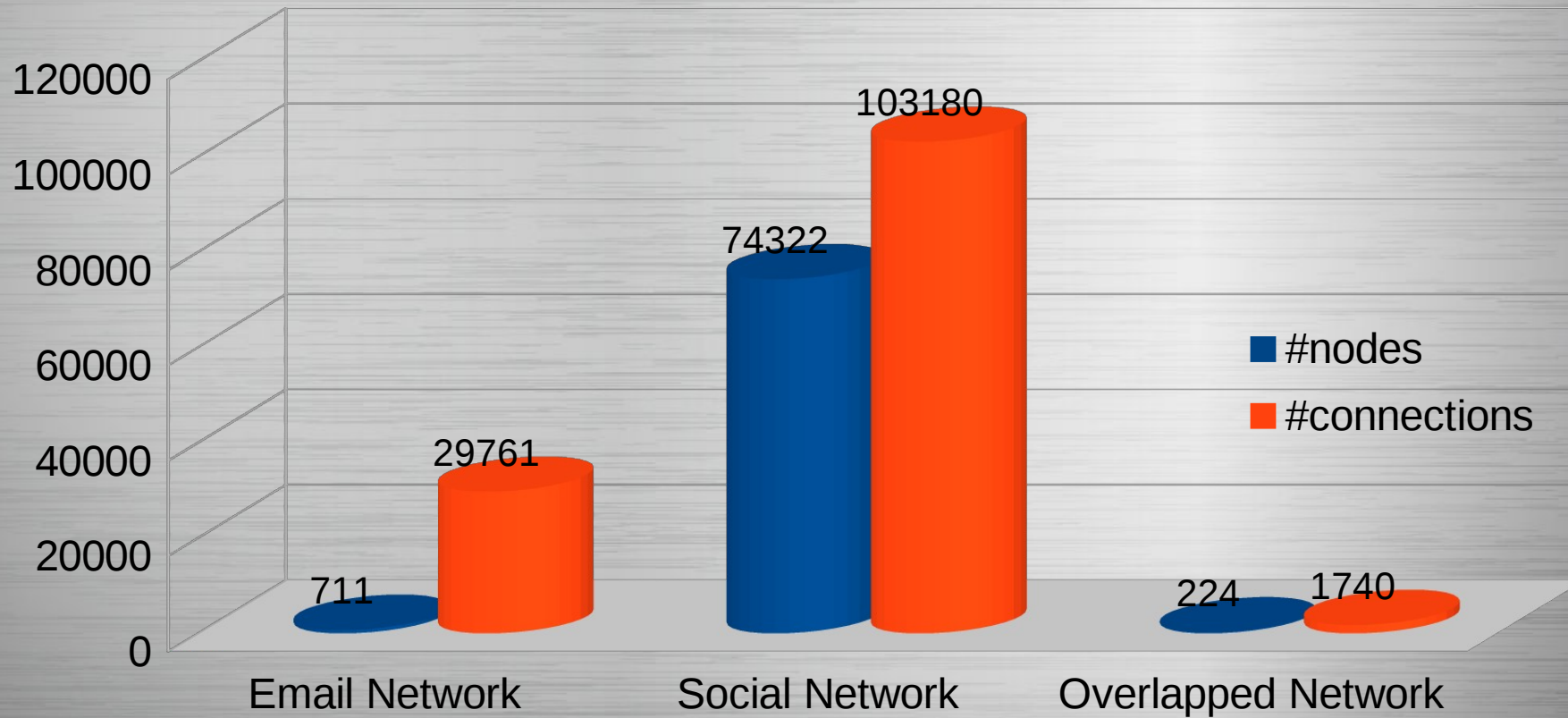
email network



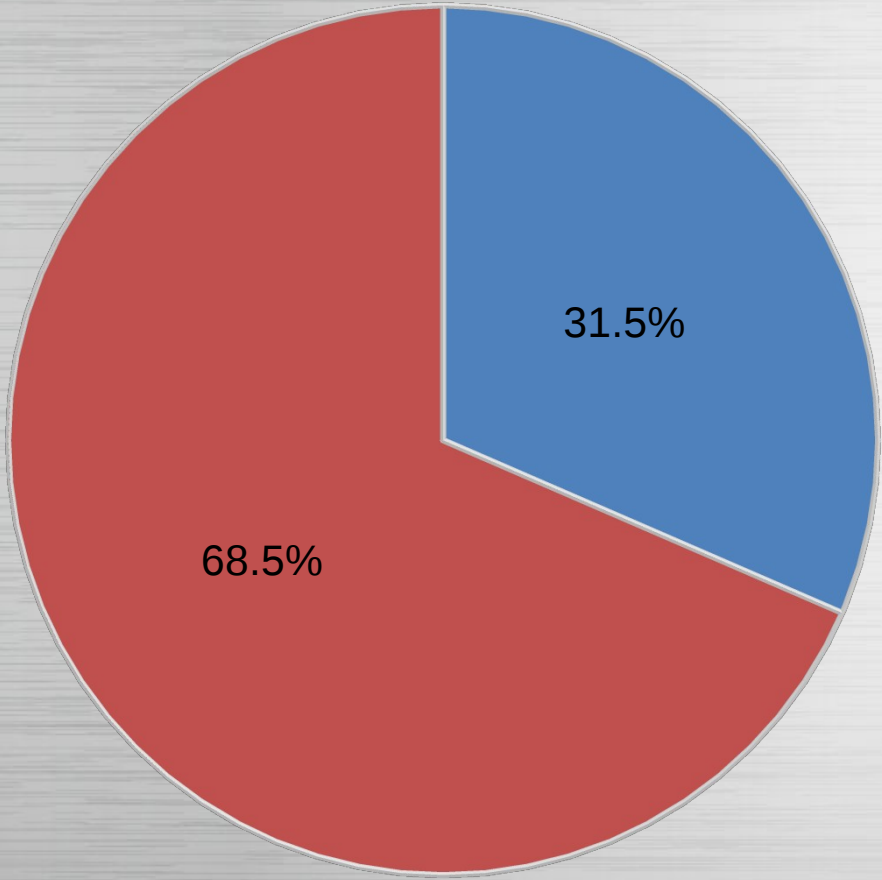
social network



Results of Statistics



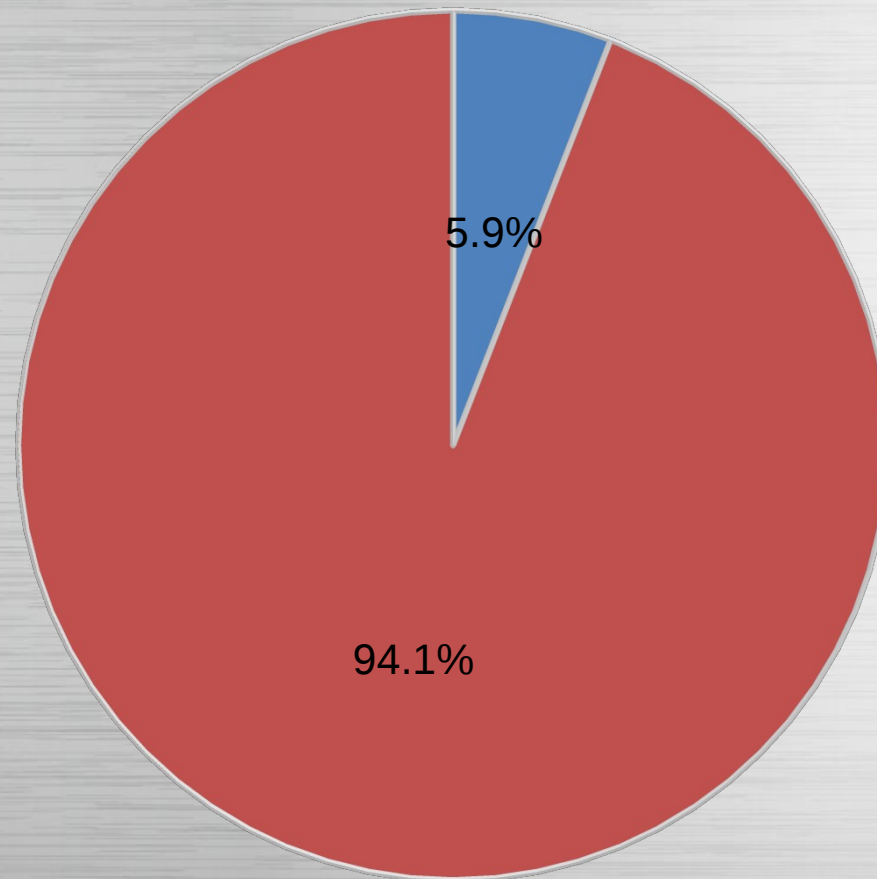
The proportion of overlapped users
in the **email network**



■ overlapped ■ non-overlapped

$$\frac{|V_M \cap V_N|}{|V_M|}$$

The proportion of overlapped connections
in the **social network**



■ overlapped ■ non-overlapped

$$\frac{|E_M \cap E_N|}{|E_M|}$$

The maximum overlap
proportion is

$31.5\% * 31.5\% \approx 9.9\%$

It is **meaningful**

to find email correspondents in online social networks...

Social benefits

More **connected** network world...

More **personalized** services...

The image shows the LinkedIn interface for connecting via email. At the top, there is a search bar with the text "Search for people, jobs, companies, and more..." and a magnifying glass icon. To the right of the search bar are the words "Advanced" and an envelope icon. Below the search bar is a navigation bar with links for "Home", "Profile", "Connections", "Education", "Jobs", "Interests", and "Business Services".

The main heading is "See Who You Already Know on LinkedIn" with a link "Manage imported contacts" to the right. Below this is a row of six email provider icons: Gmail, Hotmail, Outlook, Yahoo! Mail, AOL, and Any Email. The Gmail icon is highlighted with a blue border.

Below the icons is the text "Get started by adding your email address." followed by a text input field labeled "Your email". Below the input field is a blue "Continue" button.

At the bottom, there is a lock icon followed by the text "We'll import your address book to suggest connections and help you manage your contacts." and a link "Learn More".

The picture comes from LinkedIn.

Academic benefits

Cross-platform recommendation...

User identification across multiple social networks...

Finding Email Correspondents in Online Social Networks*

Yi Cui[†] Jian Pei[†] Guanting Tang[†] Wo-Shun Luk[†] Daxin Jiang[‡] Ming Hua[¶]

[†] Simon Fraser University, Burnaby, BC, Canada

[‡] Microsoft Research Asia, Beijing, China

[¶] Facebook Inc., Palo Alto, CA, USA

{cuiyic, jpei, gta9, woshun}@cs.sfu.ca, † djiang@microsoft.com, ¶ arceehua@fb.com

Abstract

Email correspondents play an important role in many people's social networks. email correspondents in social networks accurately, though may seem to be straightforward for a first glance, is challenging. Most of the existing online social networking sites recommend possible matches by comparing the information of email accounts and social network such as display names and email addresses. However, as shown empirically in this paper, methods may not be effective in practice. To the best of our knowledge, this problem has not been carefully and thoroughly addressed in research.

In this paper, we systematically investigate the problem and develop a practical data analysis approach. We find that using only the profiles or the graph structures is far from effective. Our method utilizes the similarity between email accounts and social network user profiles, the same time explores the similarity between the email communication network and the social network under investigation. We demonstrate the effectiveness of our method using two data sets on emails and Facebook.

Identifying Users in Social Networks with Limited Information

NORASES VESDAPUNT, HECTOR GARCIA-MOLINA

Computer Science Department, Stanford University
{norases, hector}@cs.stanford.edu

October 2014

1 Abstract

We study the problem of Entity Resolution (ER) with limited information. ER is the problem of identifying and merging records that represent the same real-world entity. In this paper, we focus on the resolution of a single node g from one social graph (Google+ in our case) against a second social graph (Twitter in our case). We want to find the best match for g in Twitter, by dynamically probing the Twitter graph (using a public API), limited by the number of API calls that social systems allow. We propose two strategies that are designed for limited information and can be adapted to different limits. We evaluate our strategies against a naive one on a real dataset and show that our strategies can provide improved accuracy with significantly fewer API calls.

What problems are really **important** ?

- > come from **life** ...
- > make a big **difference** ...
- > generalized & modeled **mathematically** ...
- > **relationships** with other problems ...

Thank you!