



## **LinkedIn Endorsements: Data Amplifies Desire**

Sam Shah

# LinkedIn: the professional profile of record

LinkedIn Account Type: Pro 7

Sydney Shoup Add Connections

Home Profile Contacts Groups Jobs Inbox Companies News More People Search... Advanced

High Yield Savings - Great Rate. No Monthly Fees. Open an Online Account at CIT Bank Today.

**Lauren Bowen**

Design is my lifelong craft.

San Francisco, California | Internet

Current LinkedIn, Napkin Sketches  
Previous LinkedIn, Advent Software, Tamale Software  
Education Georgia Institute of Technology

Improve your profile

Edit

500+ connections

www.linkedin.com/in/mosesting/

Contact Info

PEOPLE YOU MAY KNOW

Juliana Williams 2°

Director of Design

Connect

Ads by LinkedIn Members

Learn to design w/Agile

Full-day UX design workshop  
w/Anders Ramsay. Toronto: Nov 30.

Learn More »

Dreamers, Risers & You

Come join a growing group who are  
playing bigger in the Bay Area

Learn More »

Lauren Bowen via LinkedIn Today

1 How to Influence Your Company Culture (For the Better) [Read more](#)

2 Member Profiles [Read more](#)

250M Members

250M Member Profiles

LinkedIn BACKGROUND Share your profile!

SEE MORE

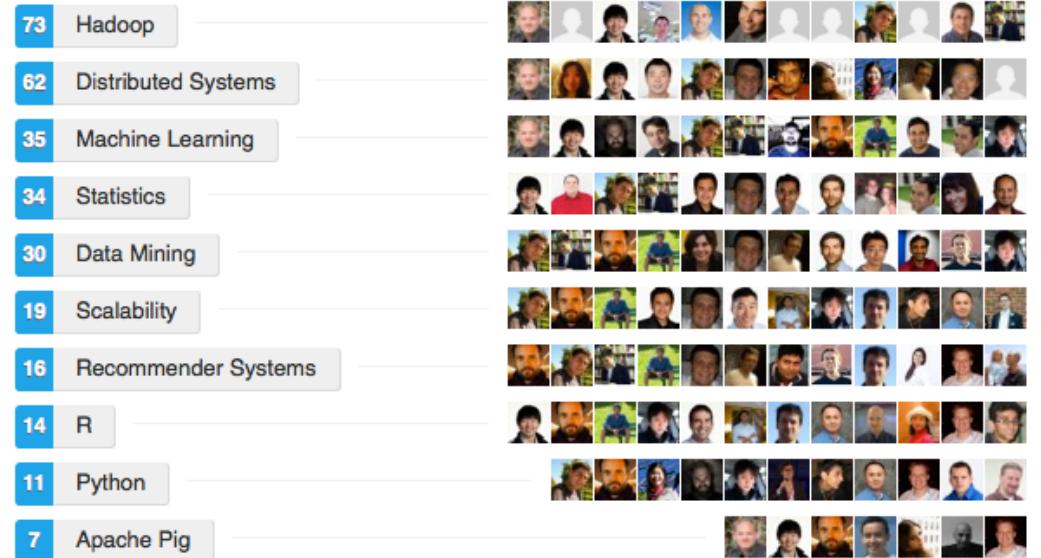
2

# LinkedIn's latest data product: Endorsements



## SKILLS & EXPERTISE

Most endorsed for...



Sam also knows about...

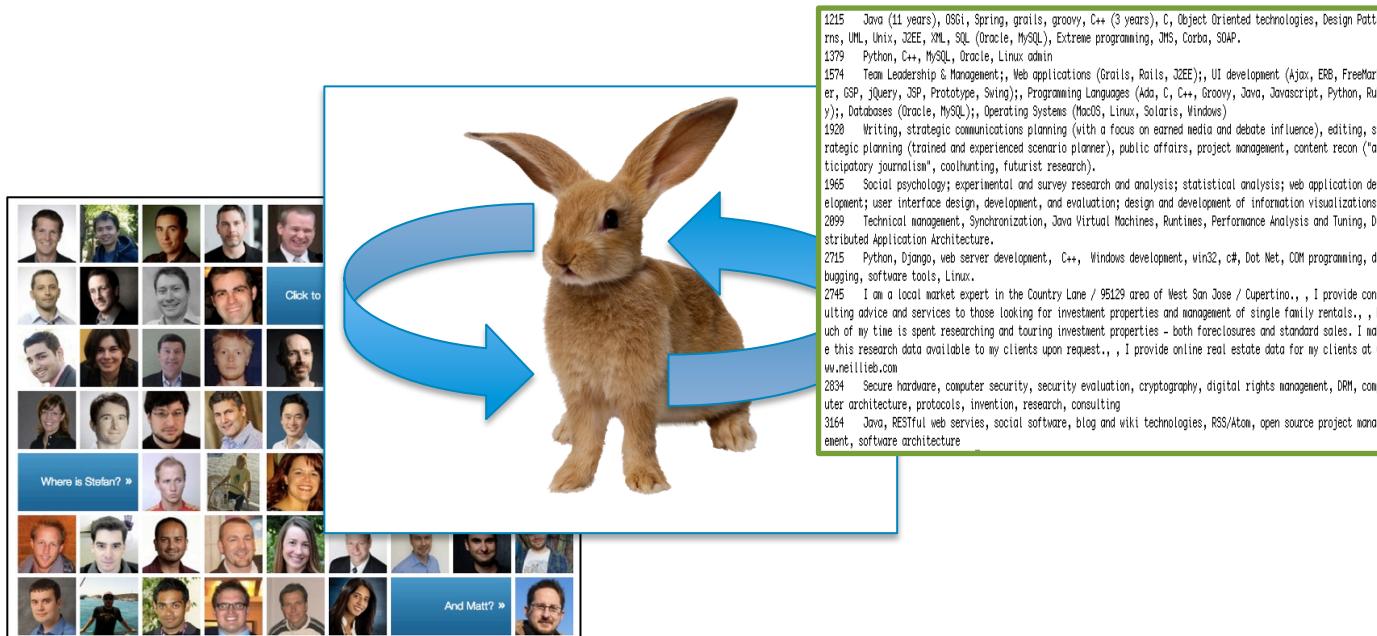


# Viral growth: 1B endorsements in 5 months



# Data amplifies desire

1. Desire + Social Proof
2. Viral Loops + Network Effects
3. Data Foundation + Recommendation Algorithms



# 1) Desire & social proof

Most endorsed for...

62	Machine Learning	+
40	Python	+
30	Data Mining	+
22	Data Science	+
16	Big Data	+
11	Django	+
8	Blogging	+
7	Natural Language...	+
5	Web Services	+
4	Artificial Intelligence	+

10 connections were endorsed for their skills and expertise.

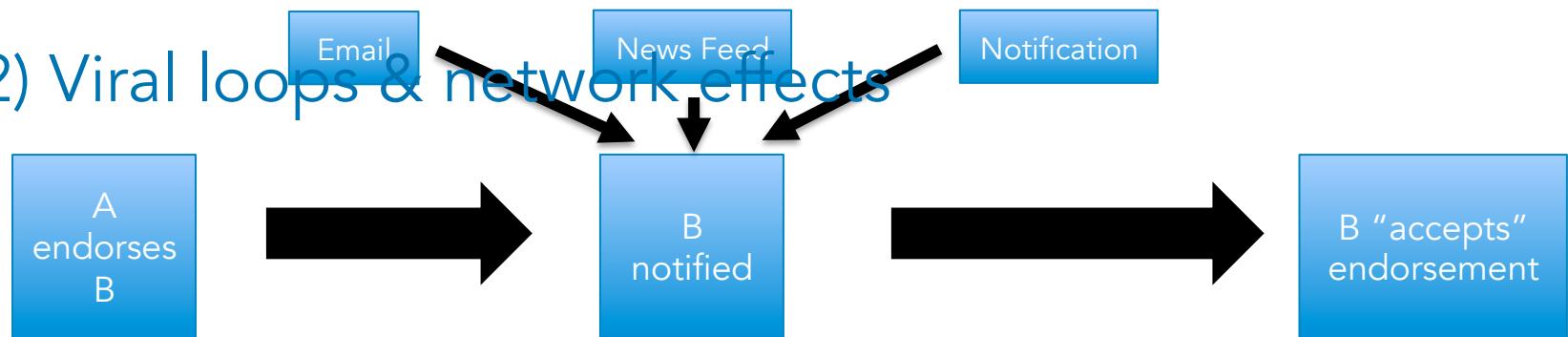
Monica Rogati was endorsed by Faisal Farooq for:  
Natural Language Processing  
Endorse your connections · 2 hours ago

Peter,

LinkedIn®

Congratulations! Your connection Julie Steele has endorsed you for the following skills and expertise:

## 2) Viral loops & network effects



Subject: Ann Howell has endorsed you for Mashups, Hadoop & Ruby!  
From: LinkedIn [???@linkedin.com]

LinkedIn Account Type: Pro

Home Profile Contacts Groups Jobs Companies News More

Hotel Art - Free Design - Free Design Service - Wholesale Framed Artwork and Mirrors - Install

Does Morgan have these skills or expertise?

- Music Appreciation
- Painting
- Creative Writing

Type another area of expertise...

Endorse

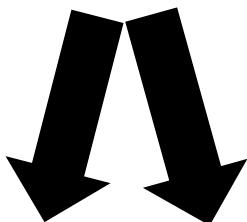
What is this?

Sug  
Send M  
Recom  
Endorse  
Save M

You are receiving this email because you are connected to Morgan Hale. You can change your weekly digest settings at any time.

This email was intended for Alexandre Lee (Product Manager at ZYX Corp). Learn how this information helps us protect your security. © 2011, LinkedIn Corporation. 2029 Stierlin Ct, Mountain View, CA 94043, USA

Endorsement recommendations



B endorses C      B endorses D

### 3) Data foundation: skills & suggested skills

What skills or expertise do your other connections have?

Does  
**DJ Patil**  
know about  
**Data Mining?**

Endorse

Does  
**Hirji Delhonte**  
know about  
**Entrepreneurship?**

Endorse

Does  
**Deep Nishar**  
know about  
**Business Strategy?**

Endorse

Does  
**Jay Kreps**  
know about  
**Distributed Systems?**

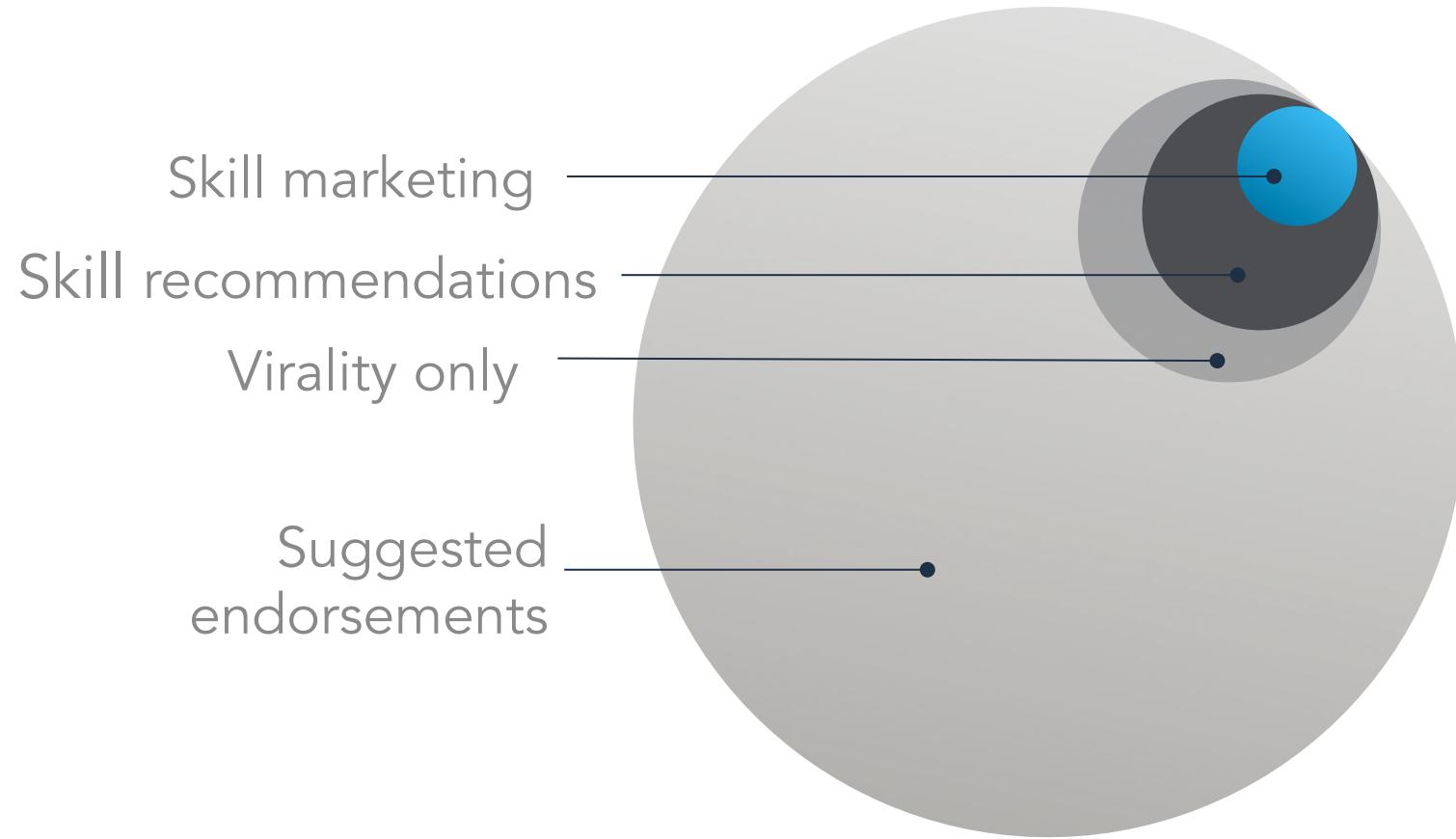
Endorse

Endorse all 4

Close

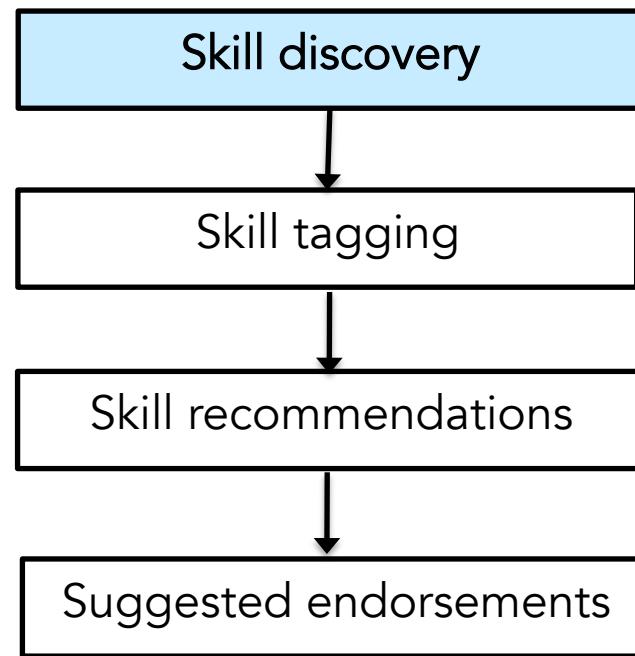
See more

## Social tagging accelerates adoption



Bootstrapping a data product  
Building LinkedIn endorsements

# Outline



# Data foundation: LinkedIn skills

Skills & Expertise beta > Hadoop

**Search Skills & Expertise**

**Related Skills**

- Hive
- Apache Pig
- HBase
- Oozie
- Sqoop
- MapReduce
- Mahout
- Flume
- Katta
- Cassandra
- Nutch
- Voldemort
- Avro
- Greenplum
- Vertica
- Data Science
- Recommender Systems
- Collaborative Filtering
- BigTable
- NoSQL

**Hadoop** ▲ 38% y/y

Primary Industry: Internet

✓ Listed on your profile See Suggested Skills

**Hadoop Professionals**

**Doug Cutting** 1st



I am a founder of the Lucene, Nutch, Hadoop and Avro open source projects.

**Arun C Murthy** 2nd



Founder and Architect at Hortonworks Inc., VP Apache Hadoop at ASF

I am a Founder and Architect at Hortonworks Inc. Hortonworks was formed by the key architects and core Hadoop...

**Tom White** 1st



Engineer at Cloudera

Tom White has been an Apache Hadoop committer since February 2007, and is a member of the Apache Software...

**Alan Gates** 1st



Co-founder at Hortonworks

**Zheng Shao** 2nd



Engineering Manager at Facebook

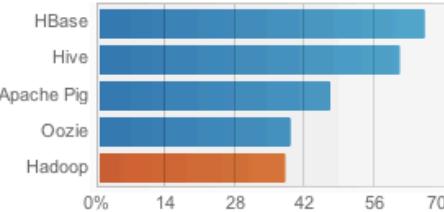
**Alejandro Abdelnur** 2nd



Software Engineer at Cloudera

Share Tweet

**Relative Growth** | [Size](#) | [Age](#) | [Gender](#)



Skill	Relative Growth (%)
HBase	68%
Hive	62%
Apache Pig	45%
Oozie	38%
Hadoop	35%

**Related Companies**

**Yahoo!** Internet, San Francisco Bay Area Follow

**cloudera** Computer Software, San Francisco Bay Area Stop following

**The Apache Software Foundation** Computer Software, United States Stop following

**LinkedIn** Internet, San Francisco Bay Area Stop following

**eBay Inc** Internet, San Francisco Bay Area Follow

# Unsupervised topic discovery from profiles

## Igor Perisic's Specialties:

Search, Statistics, Data Mining, Social Network Analysis, Search Algorithms, Semantic Engines, Unstructured Data Analysis, Knowledge Management, Information Retrieval, Hadoop



## Igor Perisic's Specialties:

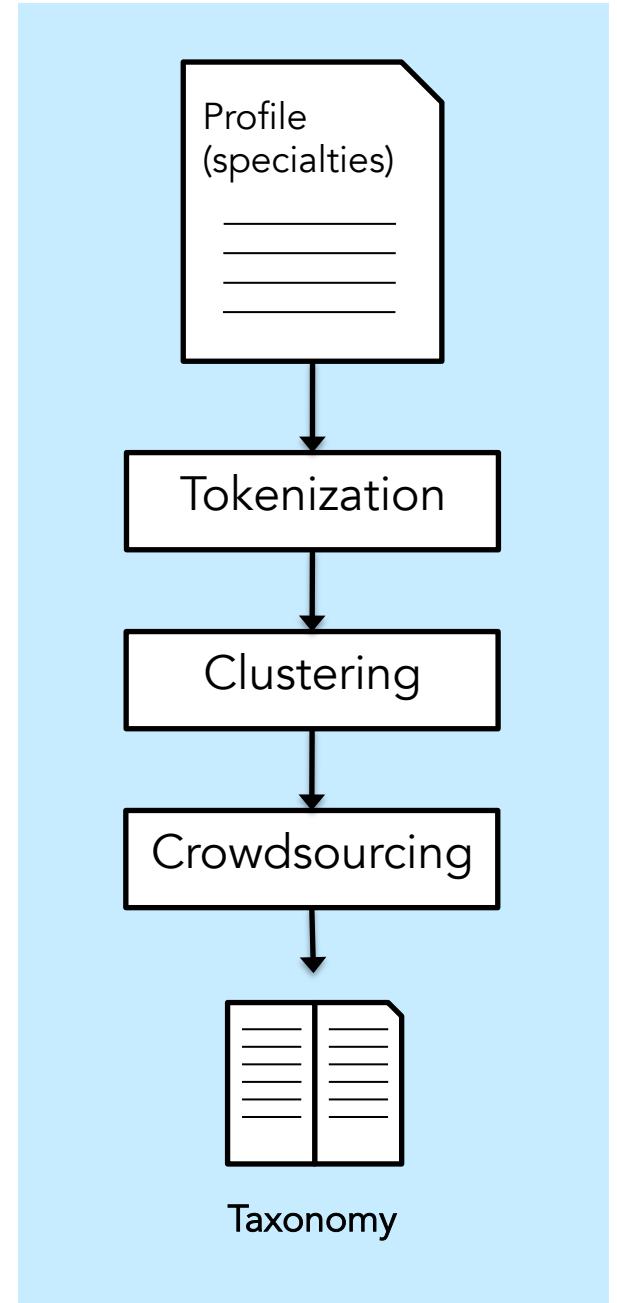
Search (Internet), Statistics, Data Mining, Social Network Analysis, Search Algorithms, Semantic Engines, Unstructured Data Analysis, Knowledge Management, Information Retrieval, Hadoop

# Building the skills dictionary

- What is the skills dictionary?
  - A growing taxonomy of skills



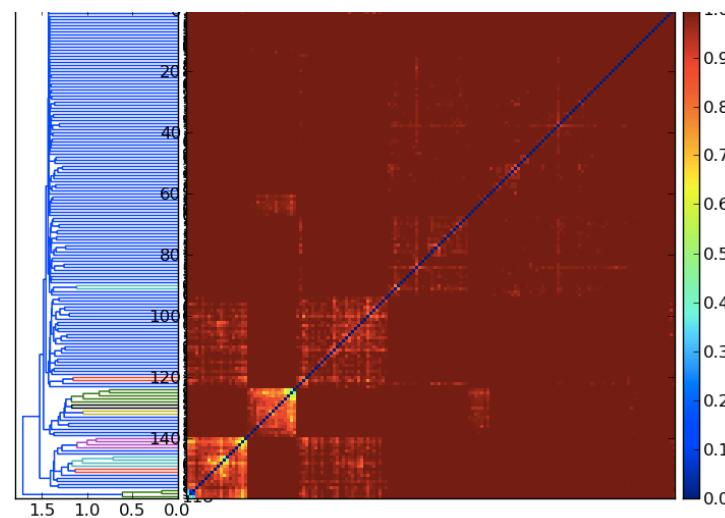
- Generated by mining profiles and maintained by the Skills team at LinkedIn
- Created using clustering and crowdsourcing.
- Multiple phrases, acronyms, and misspellings map to a single standardized skill
  - 250+ different phrases map to “Microsoft Office”



# Topic clustering & phrase sense disambiguation

**Angels (alternative medicine)** Clairsentient, Mediumship, Psychic Readings, Clairaudient, Soul Retrieval, Clairvoyant, Reiki II Practitioner, Psychic, Divination, Energy Healer, Spiritual Teacher, Metaphysical, Spiritual Healer

**Angels (venture capital)** Business Angels, Angel Investment, Angel Investing, Angel Investor, Deal Screening, Early Stage Investment, Business Angel, Angel Investments, Early Stage Funding, Angel Funding, Venture Capitalists, Early-stage Companies



## Skills dictionary: Microsoft Office

- ms office
- ms office suite
- computer skills including ms office
- office 97
- microsoft office user
- mac office
- microsoft office 2003 & 2007
- microsoft office suits
- microsoft ofice
- microsoft ofiice
- ms office certified
- office 98
- ...



Microsoft Office  
(Skill ID = 366)

# Deduplication signals from Mechanical Turk

amazon mechanical turk REQUESTER

Home Design Publish Manage Developer Help

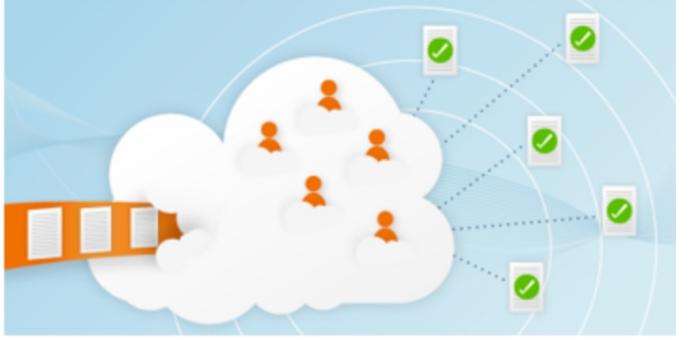
Overview Tour Case Studies Applications Solution Providers

## Work Distribution Made Easy

Mechanical Turk gives businesses and developers access to an on-demand, scalable workforce

- ✓ Flexibility: Scale your workforce up and down quickly
- ✓ Accuracy: Get high-quality, cost-effective results
- ✓ Price: Pay only when you are satisfied with the results
- ✓ Speed: Start receiving results in minutes

[Get Started ▶](#) [Take the Tour ▶](#)



### Your business can use Mechanical Turk to:

- Clean Your Data**
  - Data Verification & de-duplication
  - Data entry & collection
  - Algorithm training[Learn more »](#)
- Create & Moderate Content**
  - Moderate photos & content
  - Content creation & editing
  - Transcription[Learn more »](#)
- Categorize Items**
  - Classification
  - Sentiment analysis
  - Tagging[Learn more »](#)
- Get Feedback**
  - Test search relevancy
  - Product usability testing
  - Research[Learn more »](#)

### What Our Customers Say

**Joel Simpson**  
Principal, Cloud, Open Source and Emerging Technologies practice,  
*Statera*

**George Grey**  
CEO, *SnapMyLife*

[Explore Case Studies »](#)

# Sample task for Mechanical Turk workers

## Find the best Wikipedia article for this Topic: GPS

The Wikipedia page should be an article about this topic phrase. The topic phrase may be ambiguous, so use the following context to help choose the correct article:

**Primary Industry:** environmental services **Related Keywords:** Georeferencing, GPS data collection, Field Data Collection, Geodatabase, Geodatabases, ESRI ArcGIS 9.3, Map Production, Cartographic Design, ModelBuilder, Cartography

[CLICK HERE TO SHOW/HIDE EXAMPLES & ADDITIONAL INSTRUCTIONS](#)

1) Open the following search links for this phrase (links will open in new tab):

- "[Google site search for GPS](#)"
- "[Wikipedia site search for GPS](#)"

Examine the search results to find the best Wikipedia page about "GPS" based on the industry and related keyword context above.

Open the URL in your browser and verify the wiki page matches the context provided before submitting the URL. URLs that match the wrong sense of the phrase will be rejected. Wikipedia Disambiguation pages will be rejected. For example, if the context is the music industry and the phrase is "Organ", but you submit the Wikipedia page about "Organs" in medicine - your response will be rejected.

2) Were you able to find a Wikipedia Article about the topic "**GPS**"?

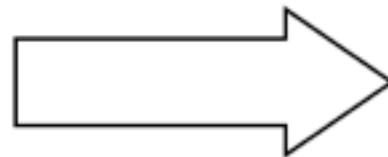
Yes  No

3) If "Yes", enter the URL of the Wikipedia Page here (it must begin with <http://en.wikipedia.org/wiki/>)

Please provide any comments you may have below, we appreciate your input!

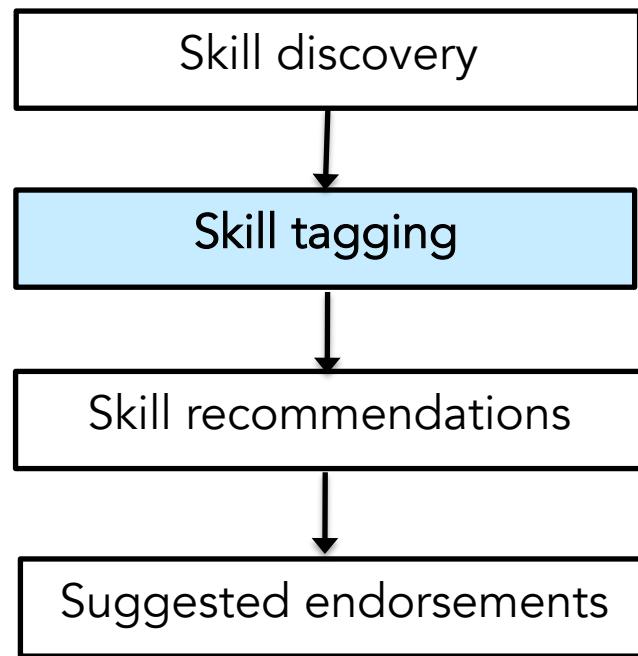
## Skill phrase deduplication

Ruby on Rails  
Rails  
RoR  
Ruby/Rails  
RubyOnRails  
Ruby on Rails development  
Ruby-on-Rails  
...  
Ruby on Rails (RoR)



Ruby on Rails

# Outline



# Skills classification

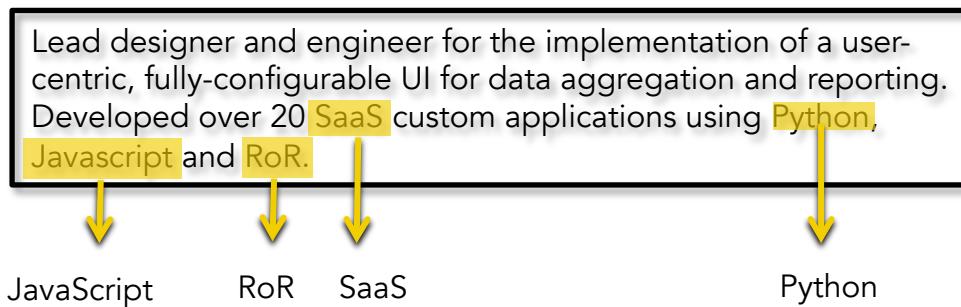
- Use skill dictionary metadata to **tag**, **standardize** and **infer** skills
- Run classifiers for each skill on member profiles

The screenshot shows a LinkedIn member profile for Lauren Bowen. At the top, there's a photo of Lauren, her name, and some basic information like her location (Bar Princeton, Ontario, Canada) and industry (Software Development). Below this is the 'ACTIVITY' section, which displays a post from Lauren about company culture. The 'SUMMARY' section follows, containing a bio and a 'SKILLS & EXPERTISE' section. The 'SKILLS & EXPERTISE' section lists 'Top 10' skills: User Experience, Interaction Design, UX, UI, Design, CSS, Design, Prototype, Adobe Creative Suite, and Processing. It also includes a 'More Skills & Expertise' section with items like General Development, Python, Java, Mobile Design, Problem Solving, Acting, Acting, Rapid Prototyping, and Agile Design. The 'EXPERIENCE' section shows two roles: 'Senior User Experience Designer' at LinkedIn from September 2010 to Present, and 'User Experience Designer' at Google from Aug 2009 to Nov 2011. Both roles mention 'Design for the web' and 'UX Designer'. The bottom of the profile shows a 'User Experience Designer' role at LinkedIn.

- Public Speaking
- Ruby on Rails
- Entrepreneurship
- Microsoft Office
- AP Style

# Tagging skill phrases

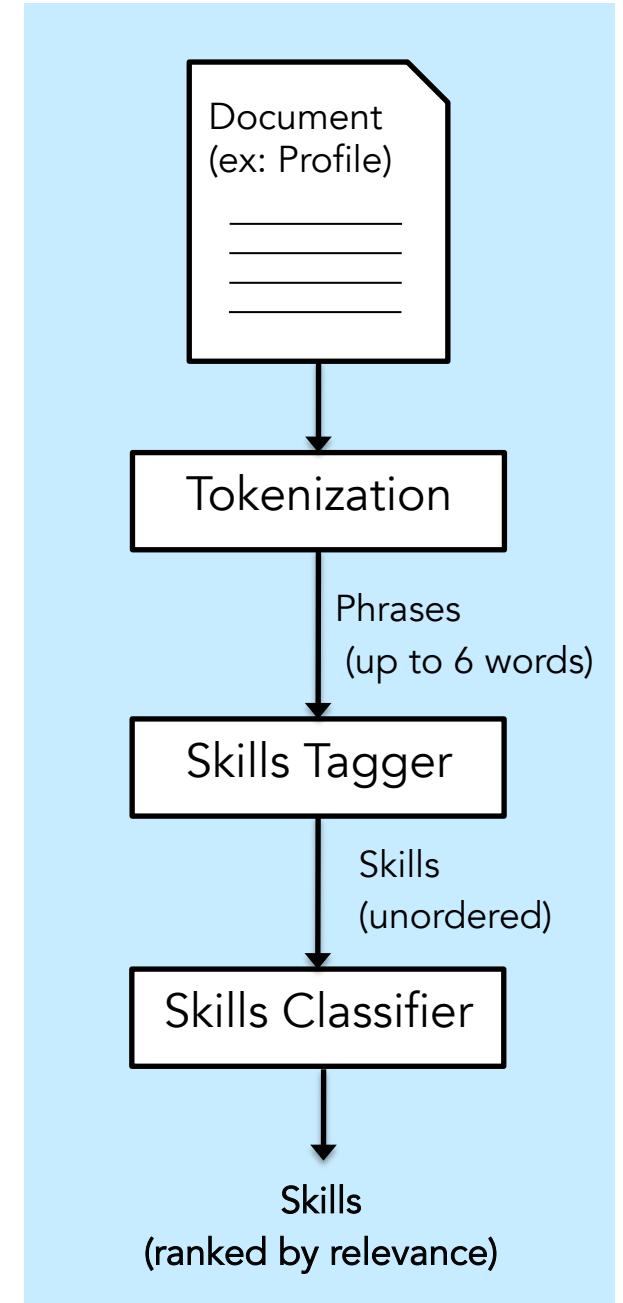
- Tagging: Extract potential skill phrases from text



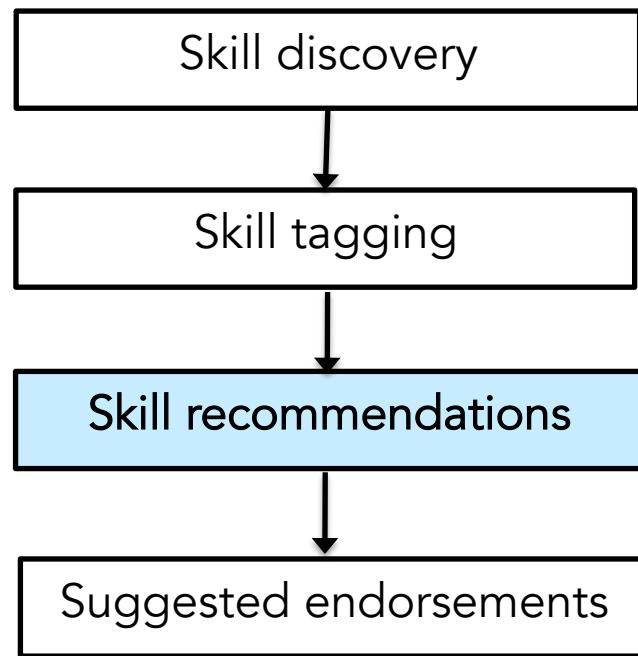
- Standardize unambiguous phrase variants

ror  
rubyonrails  
ruby on rails development  
ruby rails  
ruby on rail

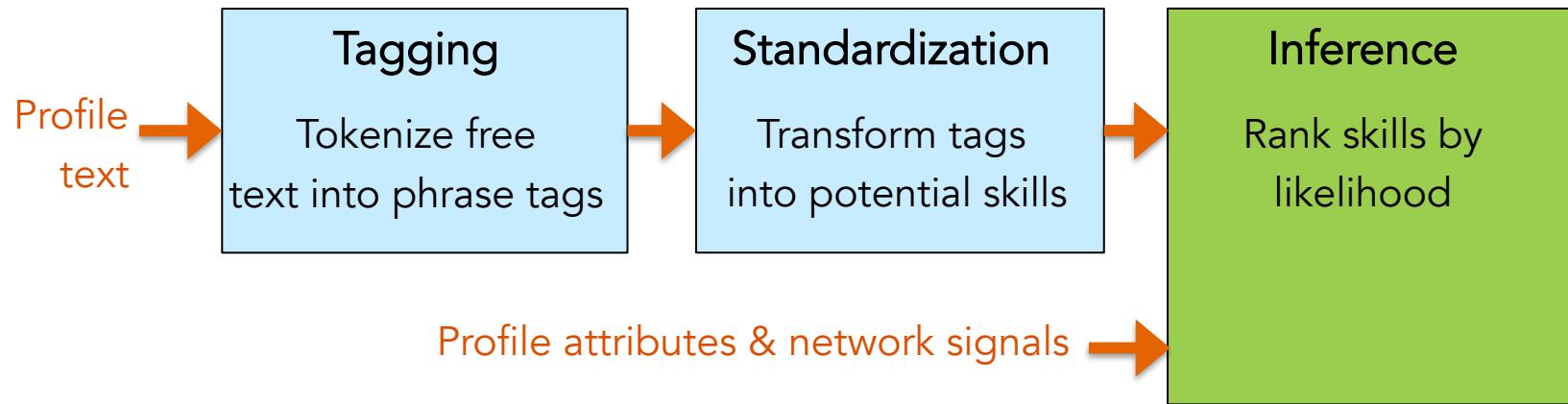
Ruby on Rails



# Outline



# Skills classification on member profiles



# Skill inference

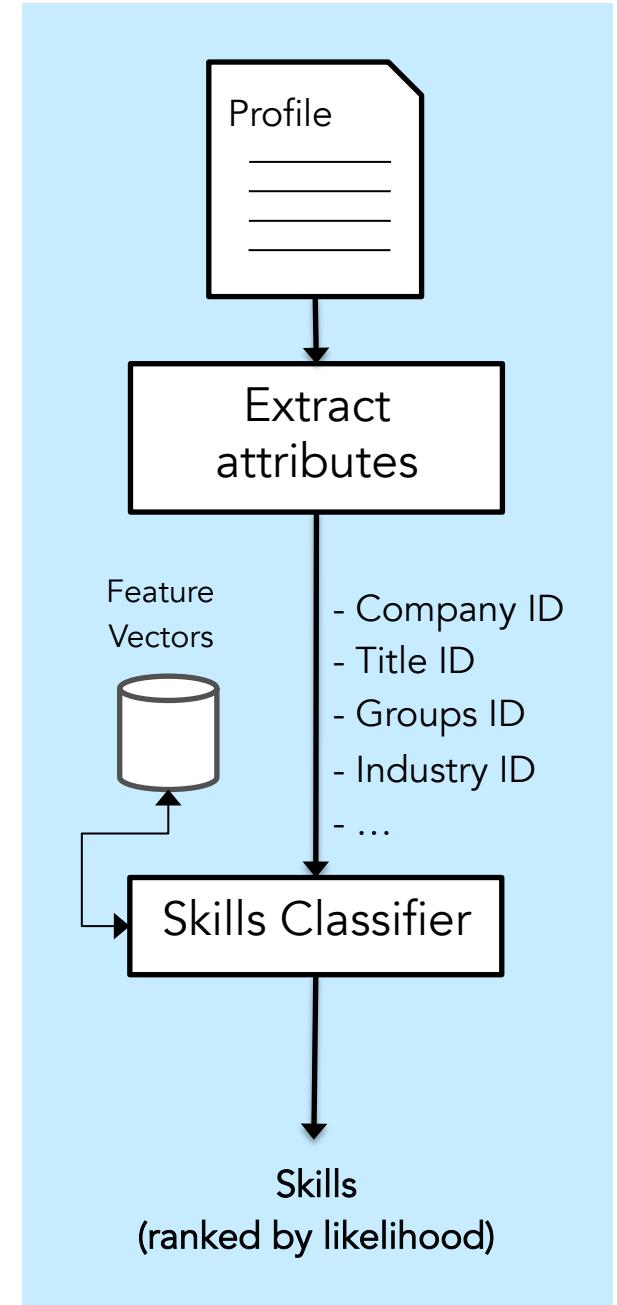
- How suggested/inferred skills work:
  - Profiles with skills help build a massive dataset of (attribute: skills).

$$P(\text{skill} \mid \text{attribute})$$

Example with a title:

Software Engineer	Java	100 000
Software Engineer	C++	88 000
...		
	↑	↑
	Title	Skill
		Occurrences

$$P(\text{Java} \mid \text{Software Engineer}) = \frac{100,000}{P(\text{Software Engineer})}$$



# Skill inference

- How suggested/inferred skills work:

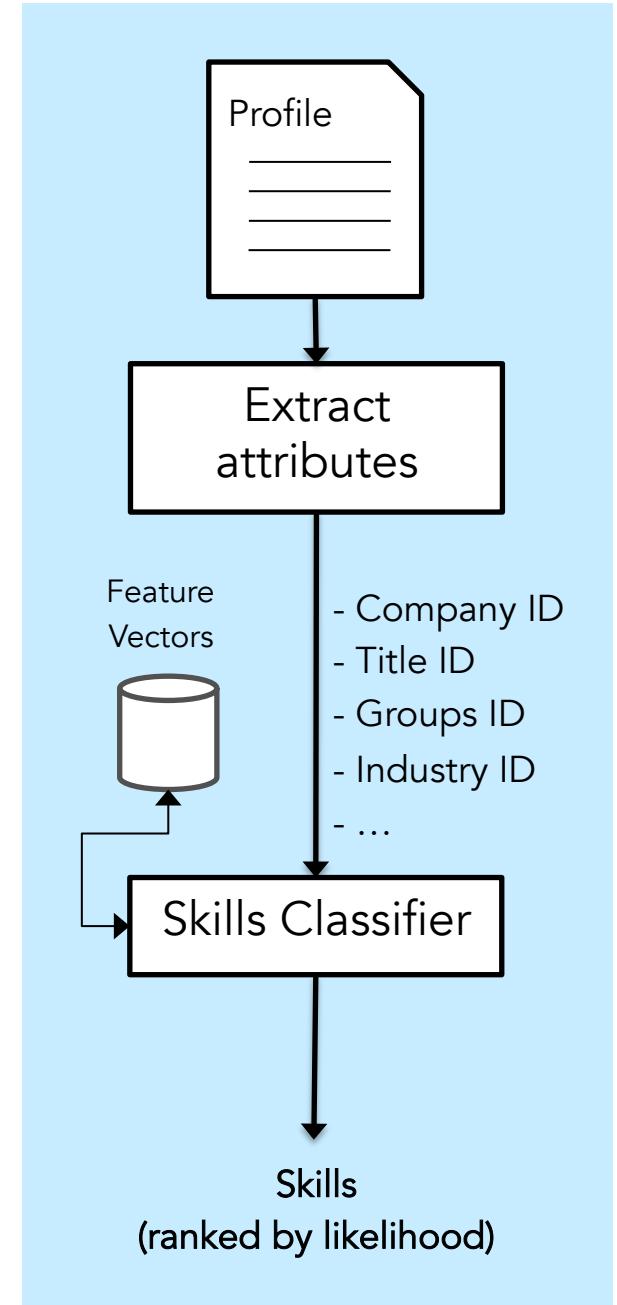
- The skill likelihood is a conditional model

$$P(\text{skill} \mid \text{profile}) = P(\text{skill} \mid \text{industry}, \text{company}, \text{title}, \dots)$$

- Probabilities are combined using a Naïve Bayes Classifier

$$p(C \mid F_1, \dots, F_N) = \frac{p(C)p(F_1, \dots, F_N \mid C)}{p(F_1, \dots, F_N)}$$

- If you are an engineer at Apple, you probably know about iPhone Development.





[Agile Process](#) - Trying to understand agile? Get your free agile process poster now!

## Vinod Khosla

Partner at Khosla Ventures

San Francisco Bay Area | Venture Capital & Private Equity

2nd



Current **Partner at Khosla Ventures**

Connections **74** connections

Public Profile <http://www.linkedin.com/pub/vinod-khosla/16/874/653>

[Share](#)[PDF](#)[Print](#)[Connect](#)[Ask Vinod For More Info](#)[Send InMail](#)[Get introduced through a connection](#)[Search for references](#)[Save Vinod's Profile](#)

## Experience

### Partner

**Khosla Ventures**

Currently holds this position

## Skills & Expertise

[Venture Capital](#) [Cleantech](#) [Private Equity](#) [Start-ups](#) [Entrepreneurship](#)[Investments](#) [Product Management](#) [Team Building](#) [Mergers](#)[Business Strategy](#) [Strategic Planning](#) [Business Planning](#)[Business Development](#) [Management Consulting](#) [New Business Development](#)[Renewable Energy](#) [Finance](#) [Strategic Partnerships](#) [Negotiation](#)[Product Development](#)[View All \(47\) Skills](#) ▾

Enter the  
**Small Business**  
**Mobile Makeover now**

T-Mobile® means Better Business

[Enter now](#)

Legal

## How you're connected to Vinod

You



Jeremy Howard



# The Midas List

## 2012's Top Tech Investors

[View complete list](#) | [Methodology](#) | In partnership with TrueBridge Capital Partners



## The 10 Savviest Financiers Of Silicon Valley

Atop this year's Midas List of technology's best investors is once again Accel Partner's Jim Breyer, with Greylock's Reid Hoffman (left) holding steady at number three. They head our definitive roster of the best-performing moneymen fueling a bull market for hot young companies. [Continue](#)

**F** Edited By Tomio Geron | 05.02.12



#1 Jim Breyer



#2 Michael Moritz



#3 Reid Hoffman



#4 Peter Fenton



#5 Scott Sandell



#6 Kevin Efrusy



#7 Peter Thiel



#8 Peter Barris



#9 David Sze



#10 Marc Andreessen



#11 Douglas Leone



#12 Fred Wilson



#13 Ron Conway



#14 John Doerr



#15 Aneel Bhusri



#16 Bryan Roberts



#17 Mike Maples



#18 Josh Kopelman

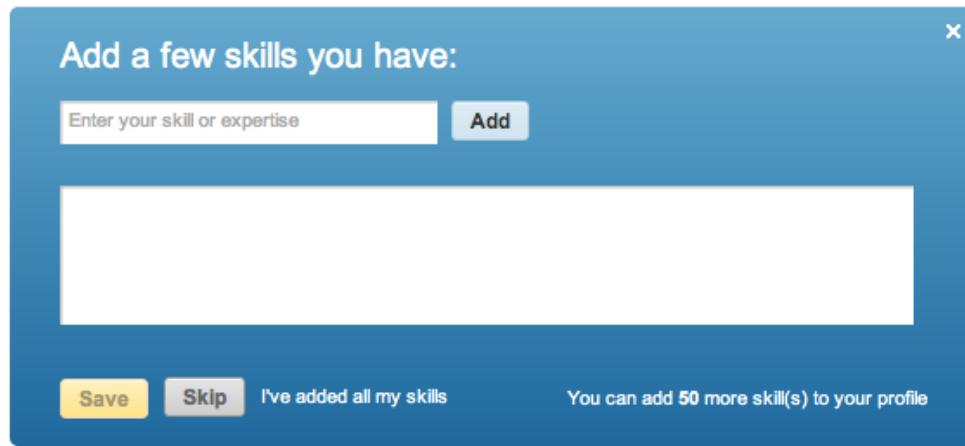


#19 Jeff Brody

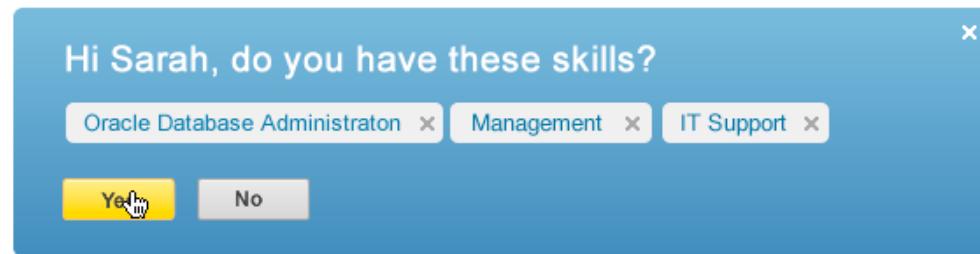


#20 Jeremy Levine

## Skill suggestions for your LinkedIn profile

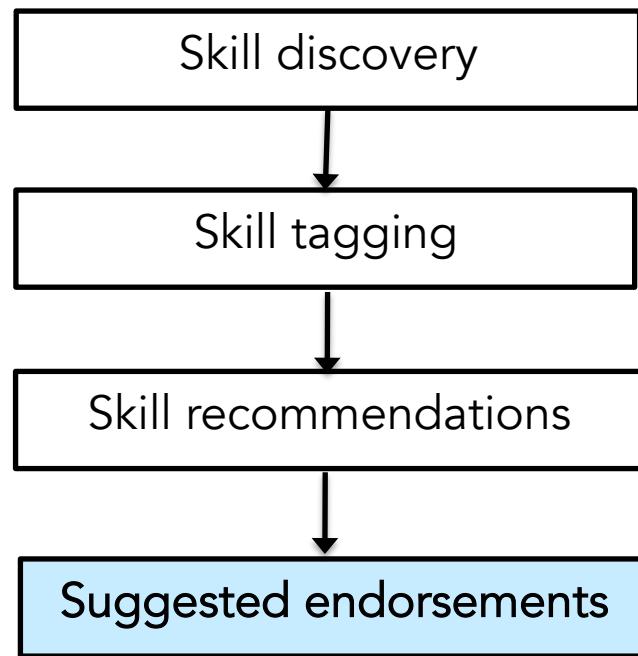


4% Conversion



49% Conversion

# Outline



# Social tagging via skill endorsements

What skills or expertise do your other connections have?

Does DJ Patil know about Data Mining?



Endorse

Does Hirji Delhonte know about Entrepreneurship?



Endorse

Does Deep Nishar know about Business Strategy?



Endorse

Does Jay Kreps know about Distributed Systems?



Endorse

Endorse all 4

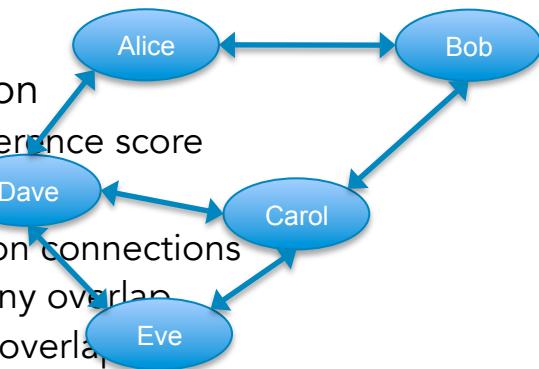
Close

See more

# Suggesting endorsements

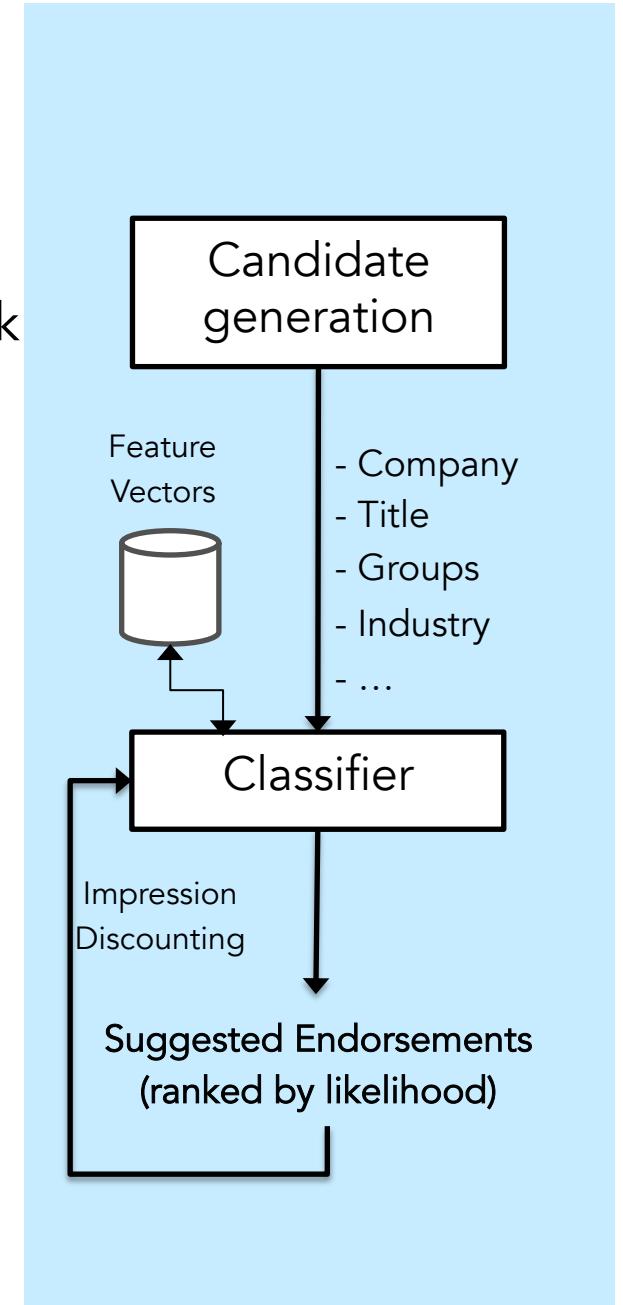
- People-skill combinations in a member's network
- Binary classification
- Features

- Classification
  - Skill inference score
- Edge features
  - Common connections
  - Company overlap
  - School overlap
  - Group overlap
  - Industry and functional area similarity
  - Title similarity



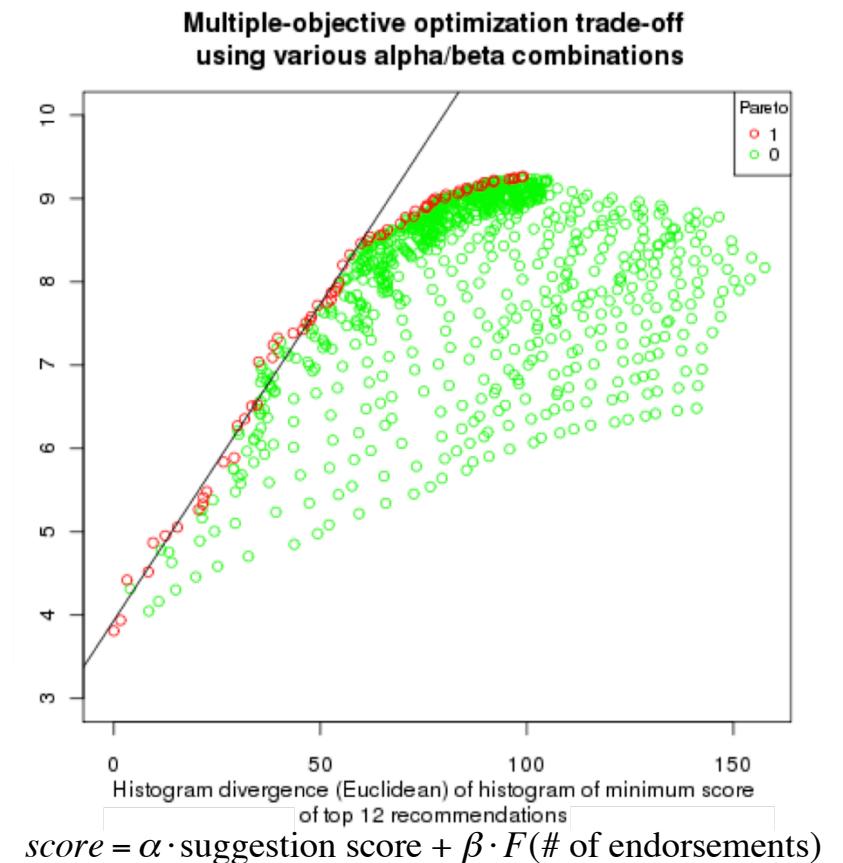
(Alice, Bob, Hadoop)  
(Alice, Bob, Machine Learning)  
(Alice, Dave, Natural Language Processing)  
(Alice, Dave, Text Mining)  
(Carol, Eve, Distributed Systems)  
(Carol, Eve, Scalability)  
(Carol, Eve, Hadoop)

...

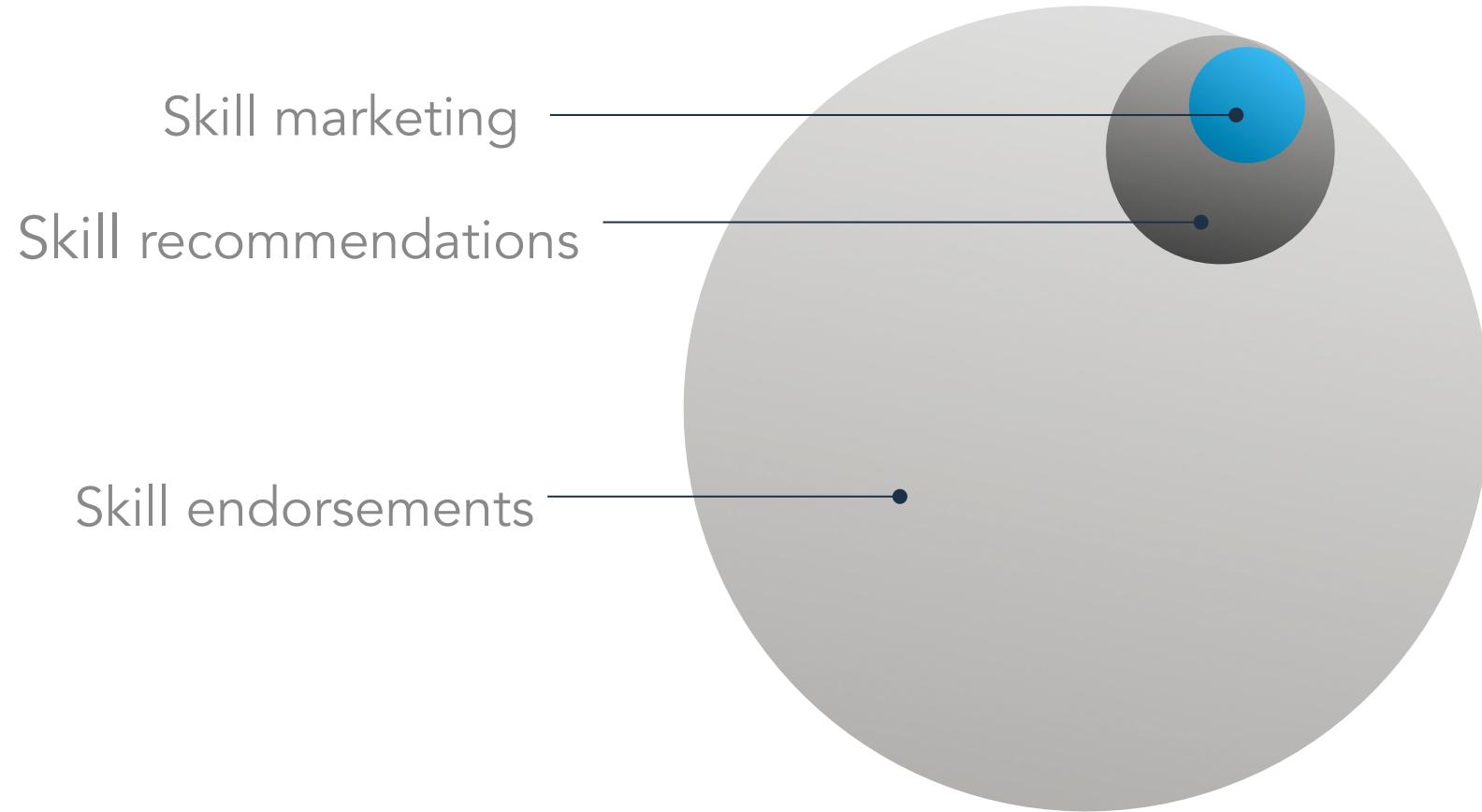


# Guns vs. butter: Helping the endorsement poor

- Want members to reach activation threshold
- Perturb suggestions to boost endorsement poor
- Multi-objective optimization
  - Likelihood of endorsement
  - # of endorsements of recipient
- Pick point on Pareto curve
- Validate via A/B testing



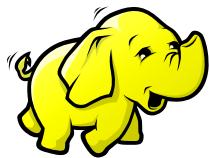
# Social tagging accelerates adoption



How did we build it?

## Endorsements Data Infrastructure

# Data Infrastructure



V



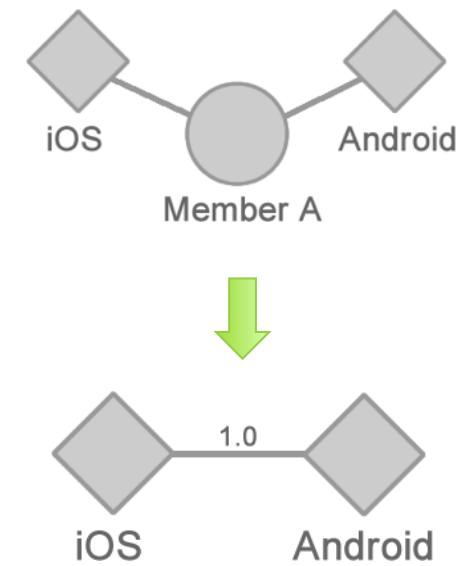
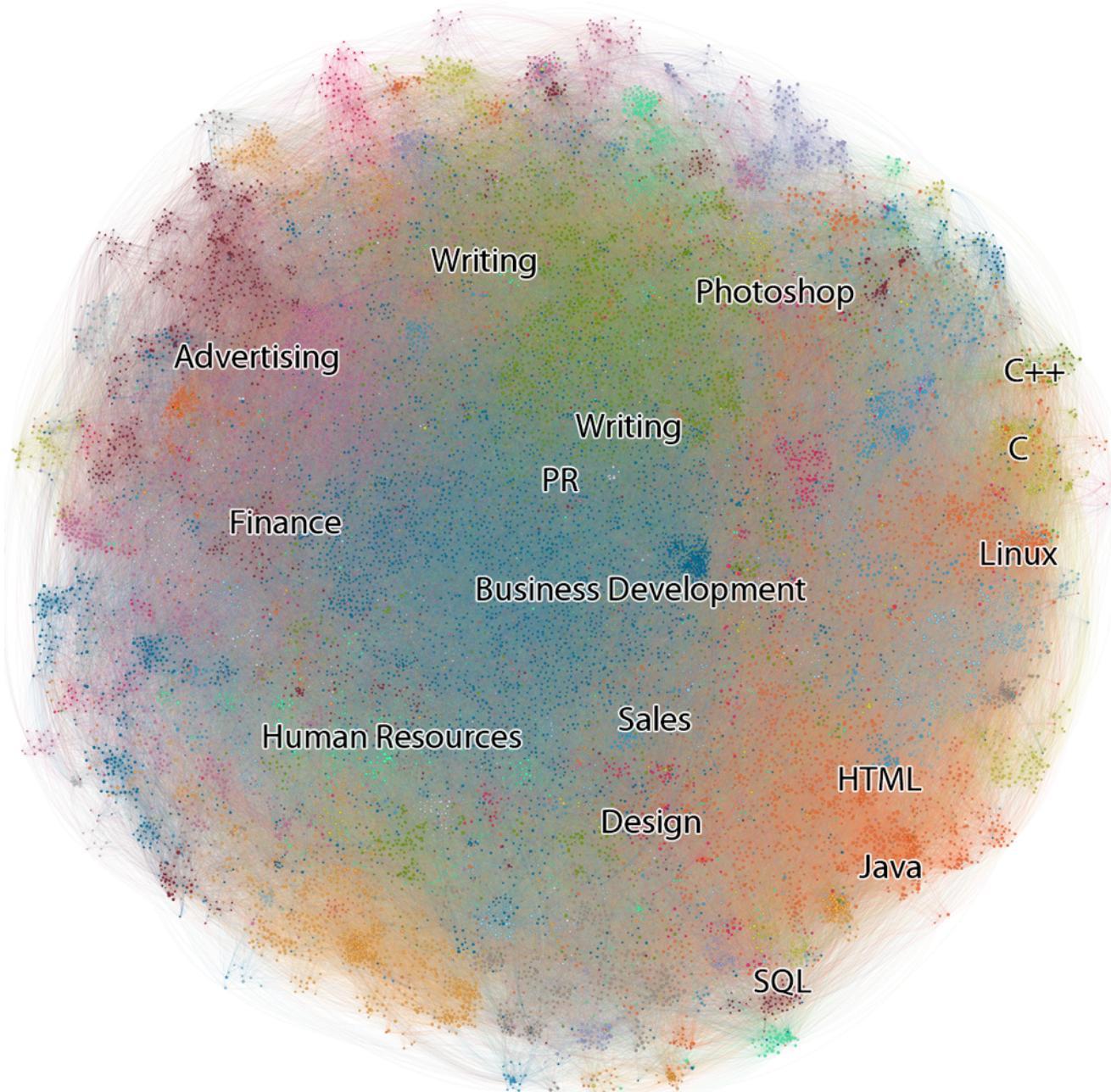
- **Apache Hadoop:** Parallel processing architecture
- **Apache Kafka:** Ingress pipes
- **Azkaban:** Hadoop scheduler
- **Voldemort:** Egress database
- **Apache Pig:** High-level MR language
- **DataFu:** Convenience routines

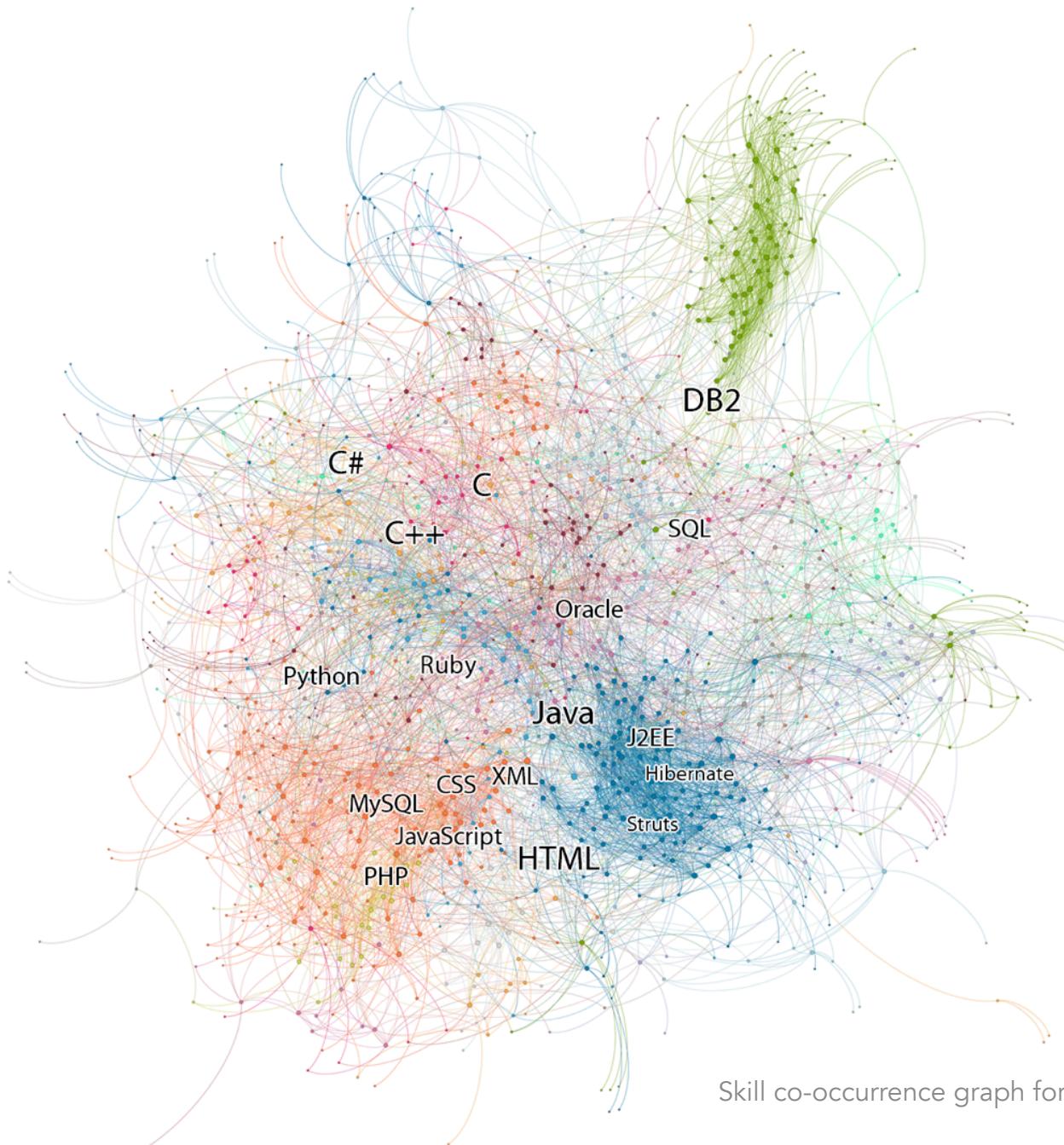
<http://data.linkedin.com>

R. Sumbaly, J. Kreps, and S. Shah. "The 'Big Data' ecosystem at LinkedIn". In SIGMOD 2013.

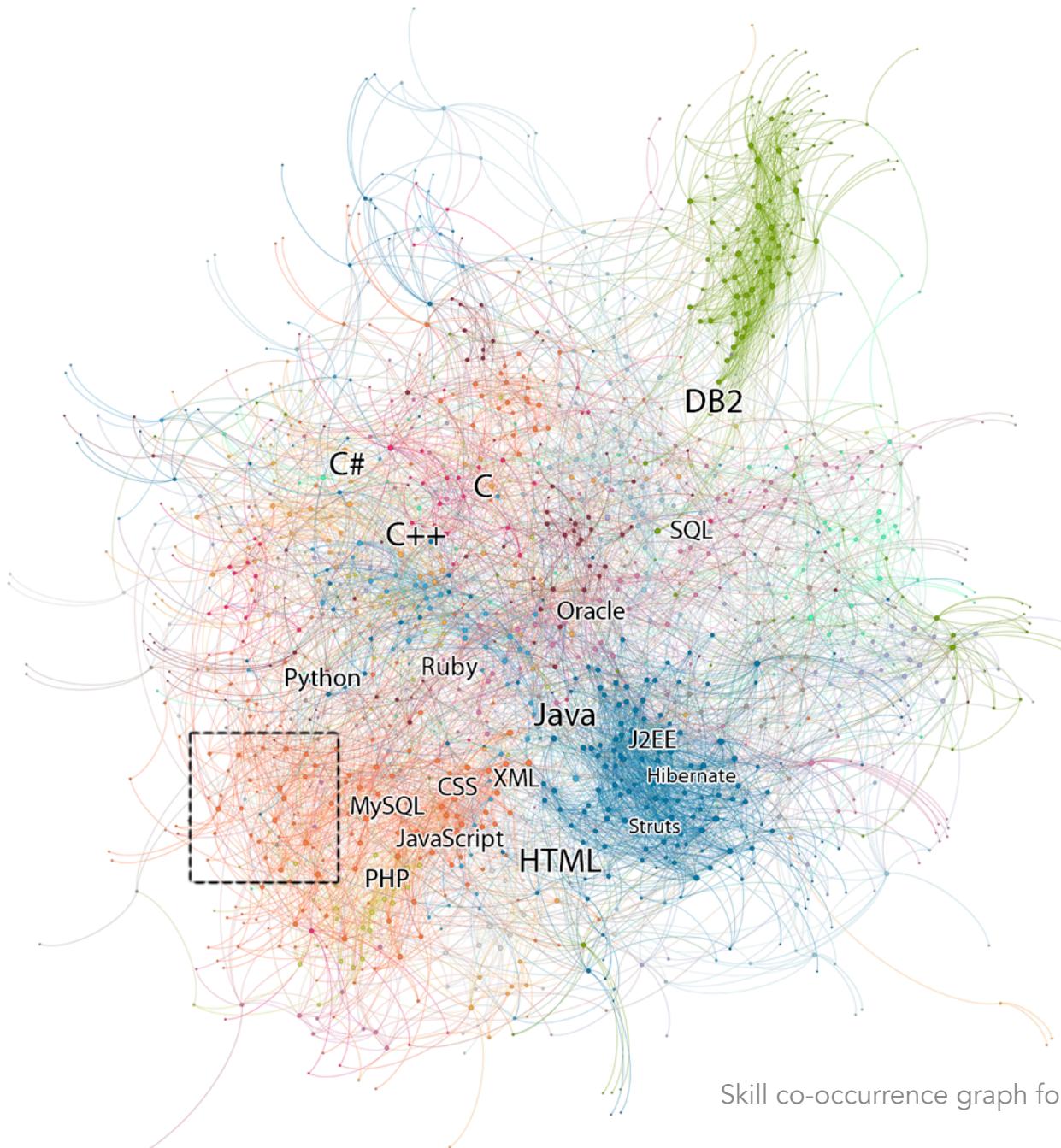
What does reputation look like?

## Applications of Endorsements data

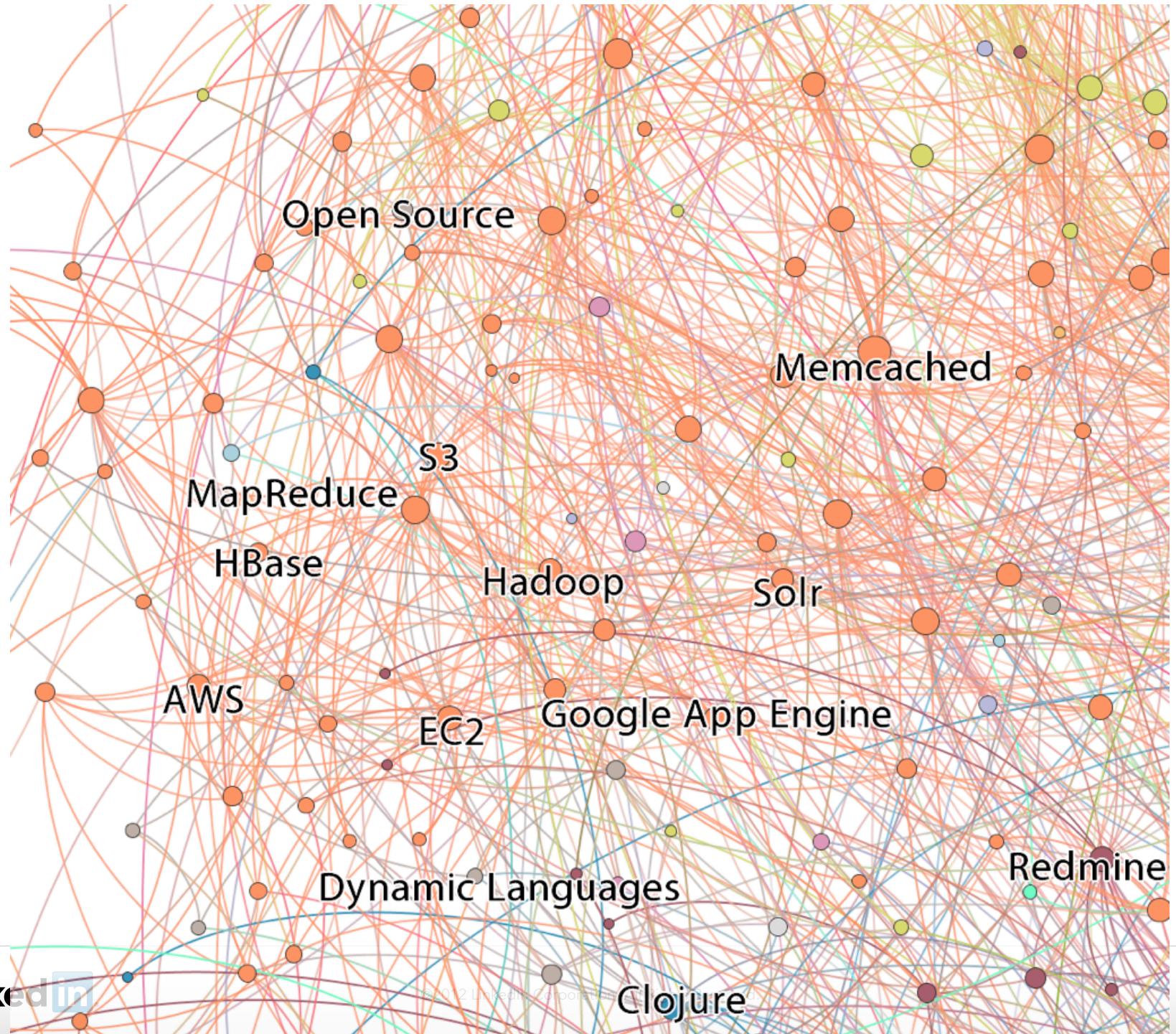


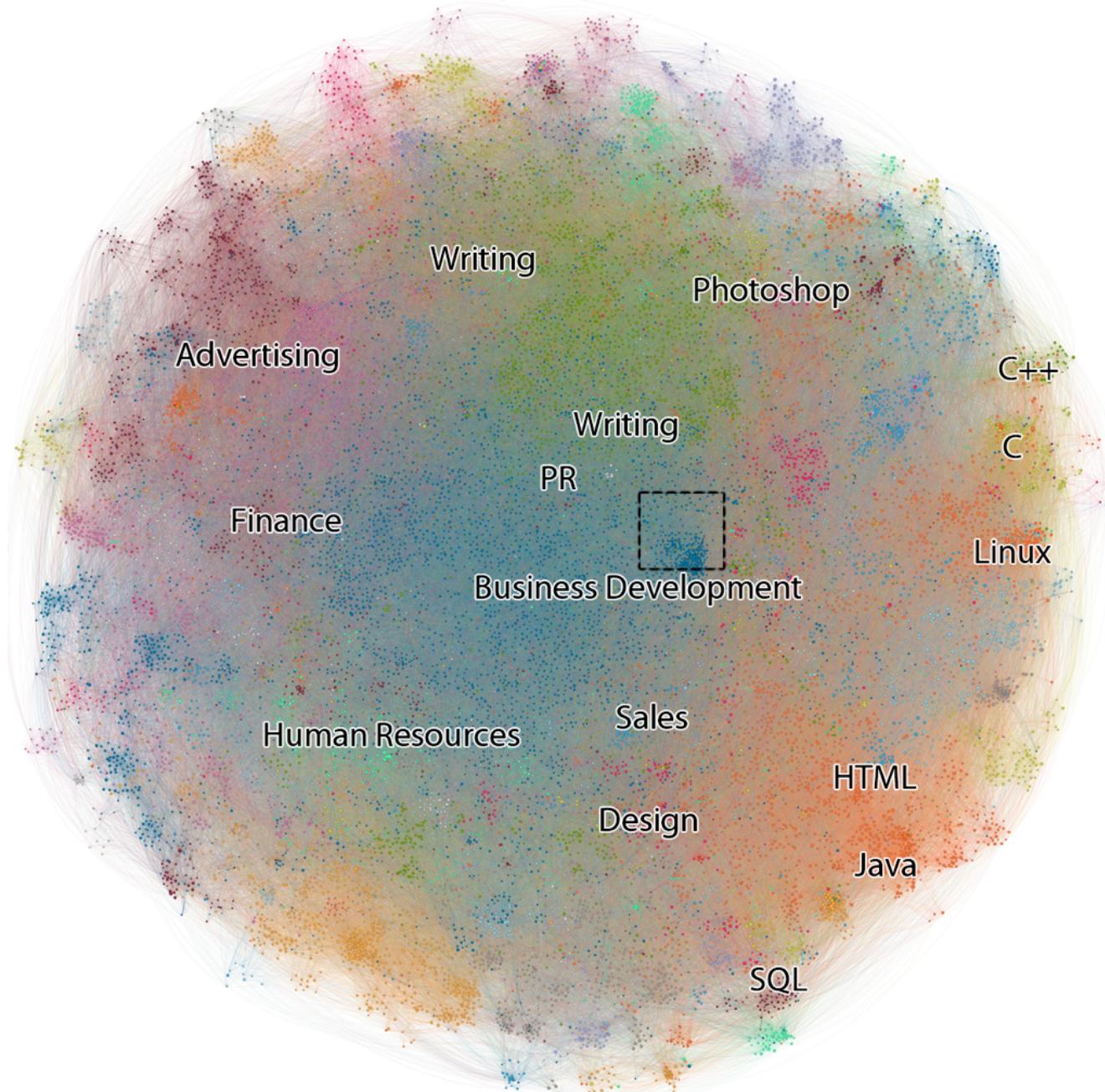


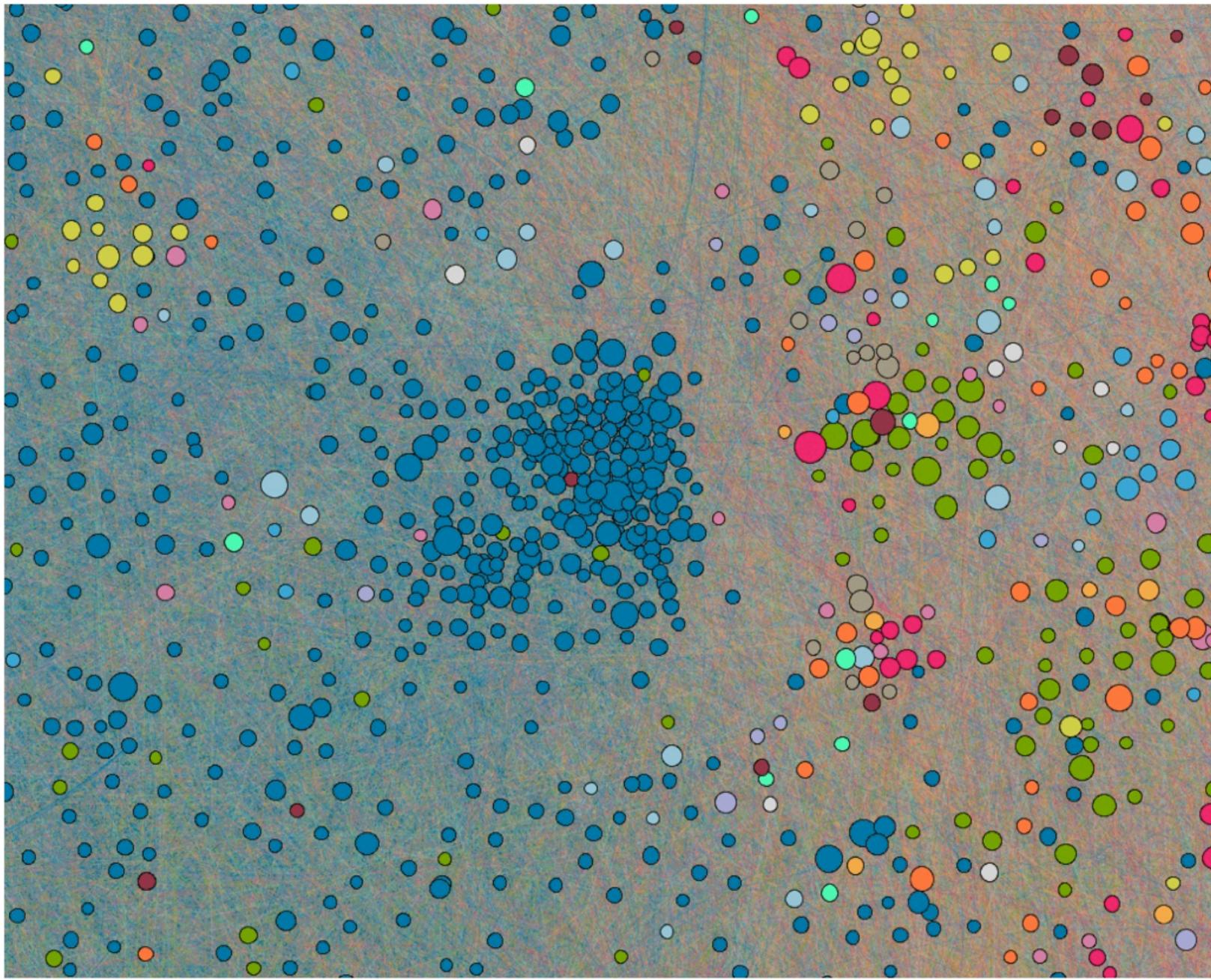
Skill co-occurrence graph for software developers

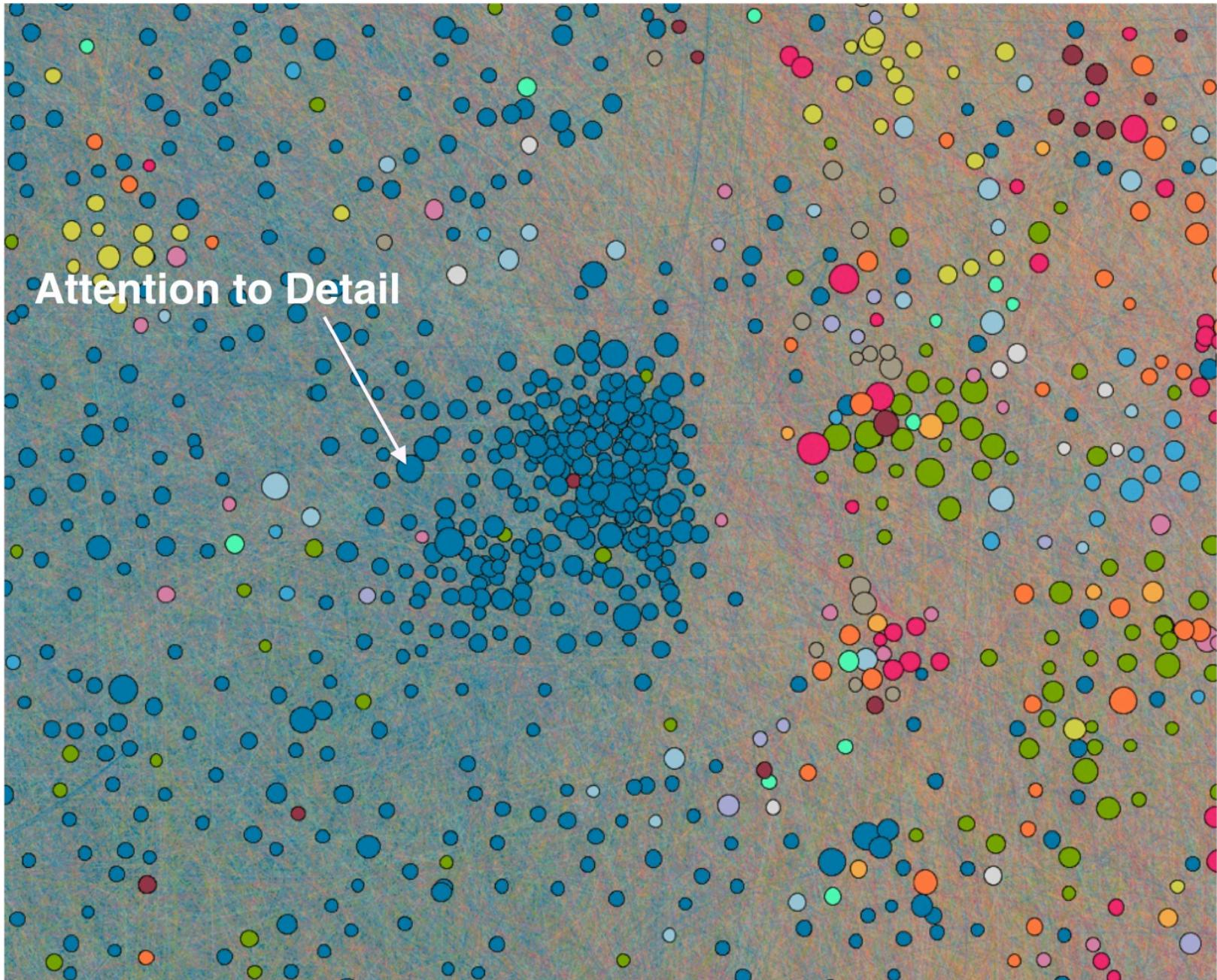


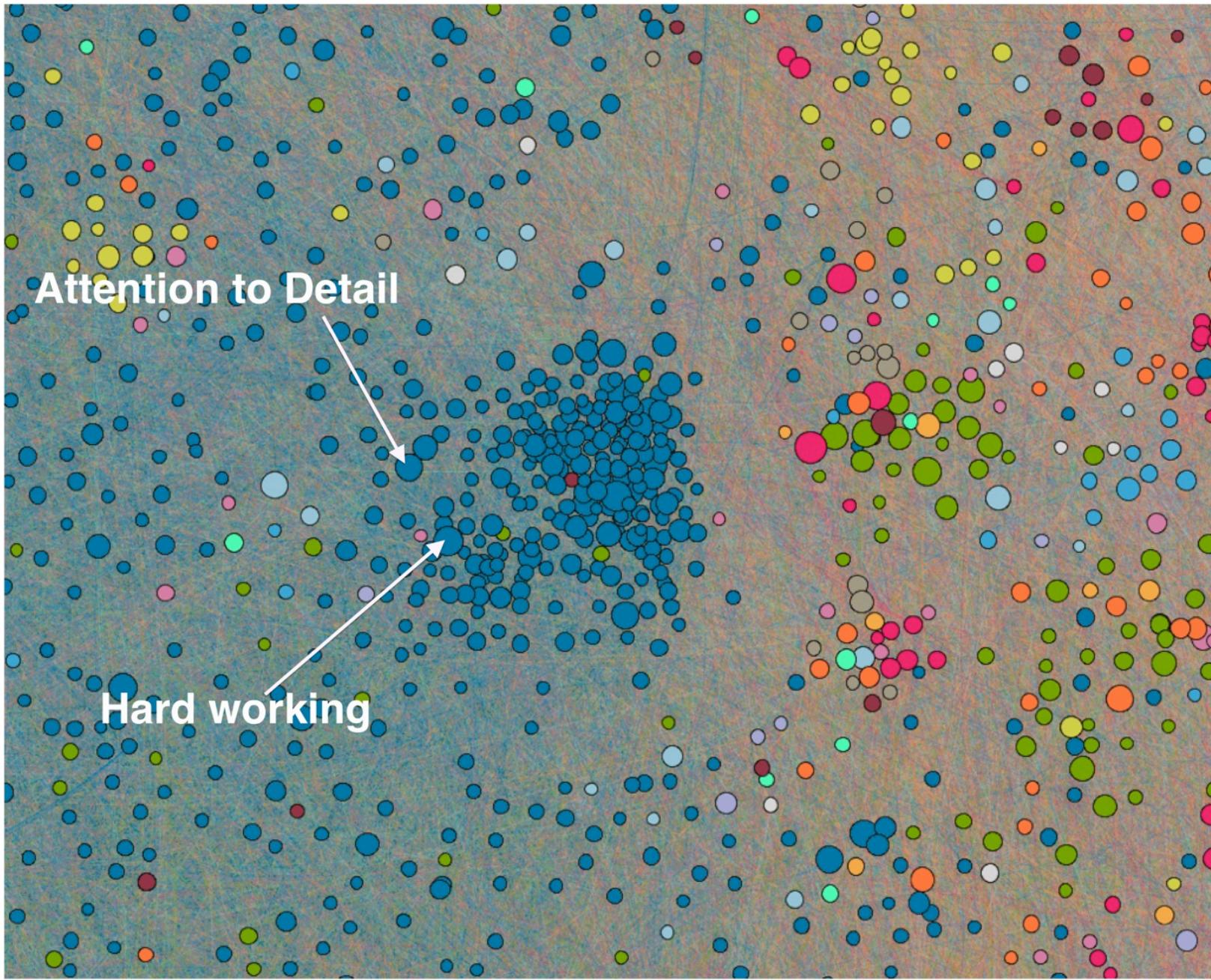
Skill co-occurrence graph for software developers

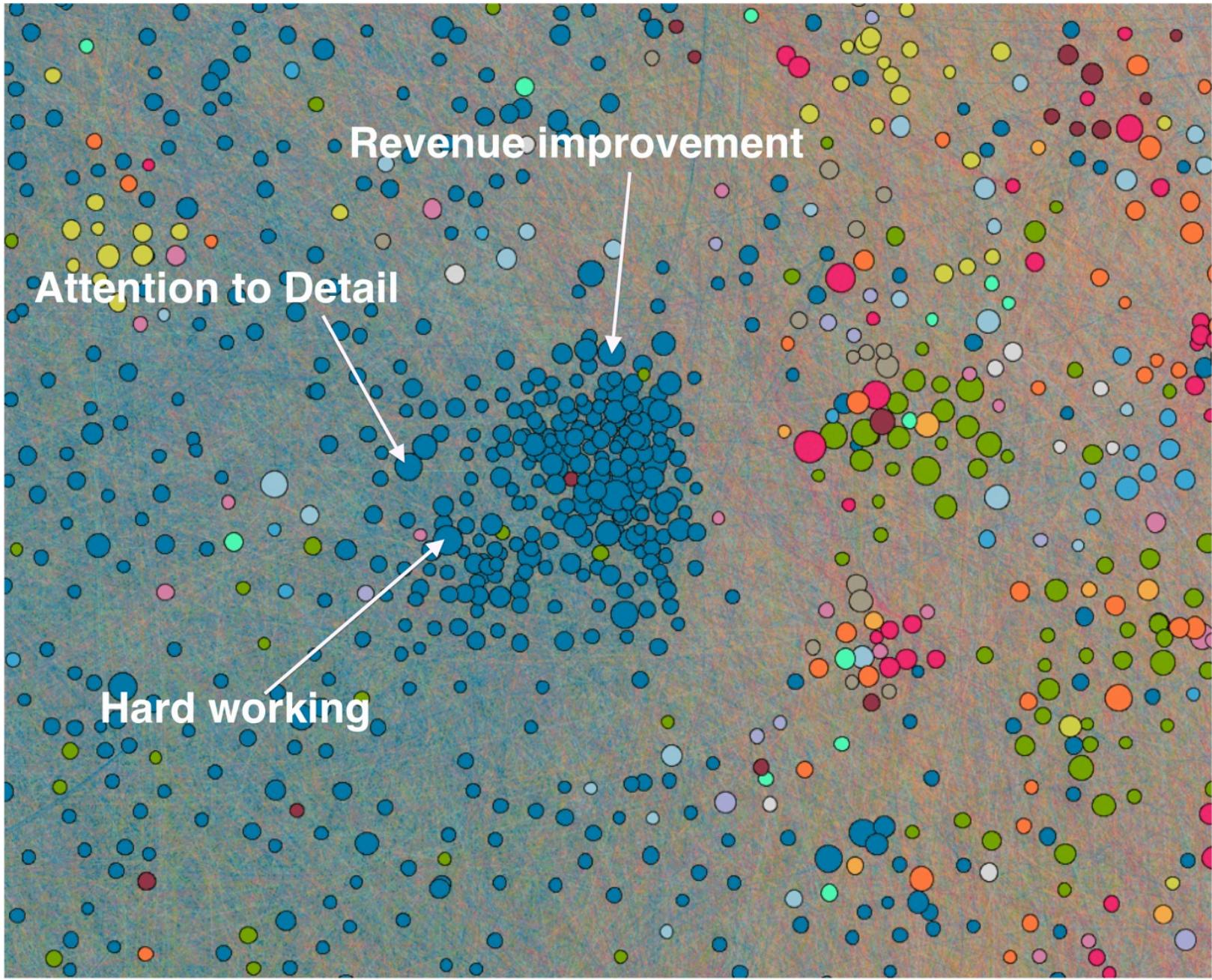


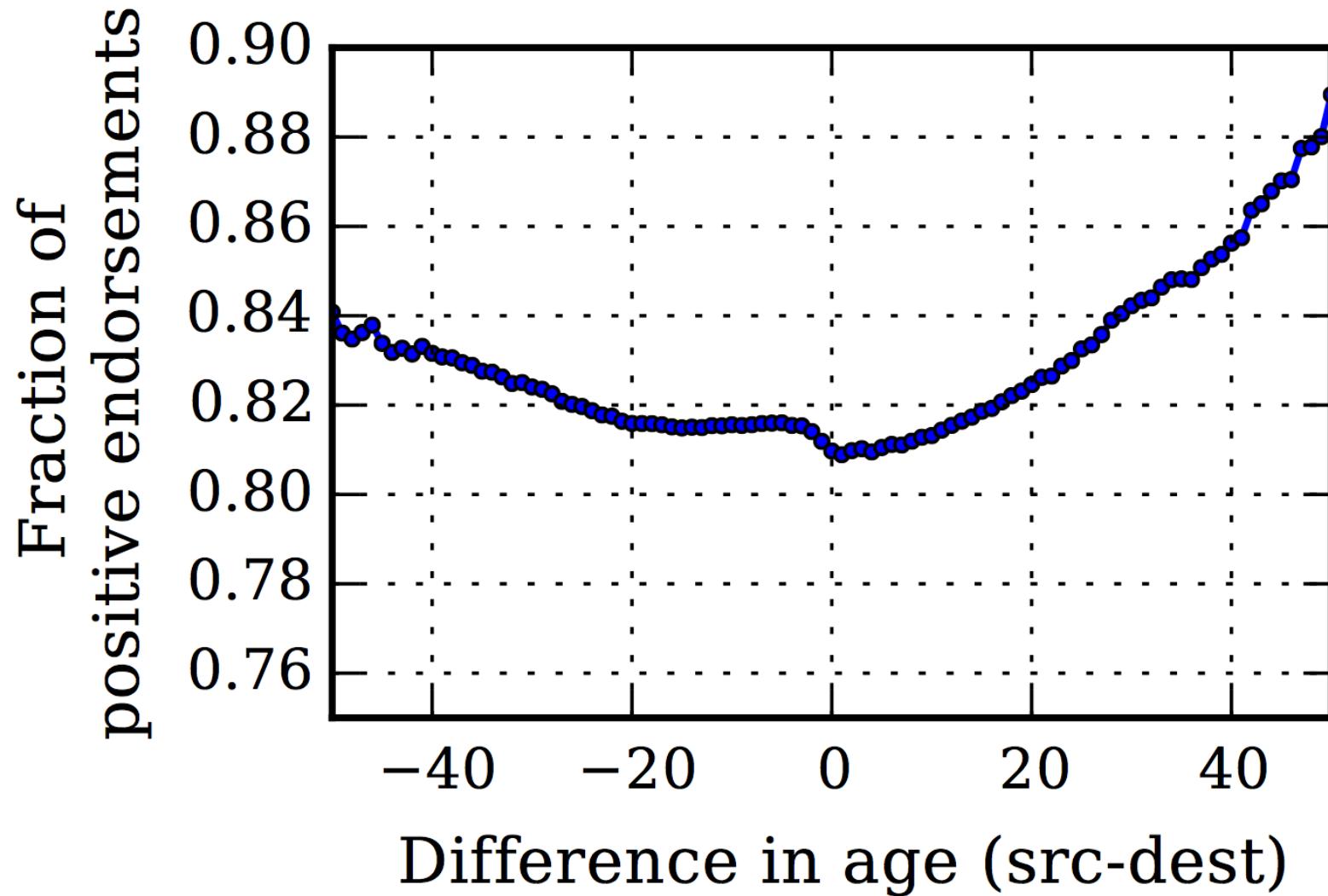




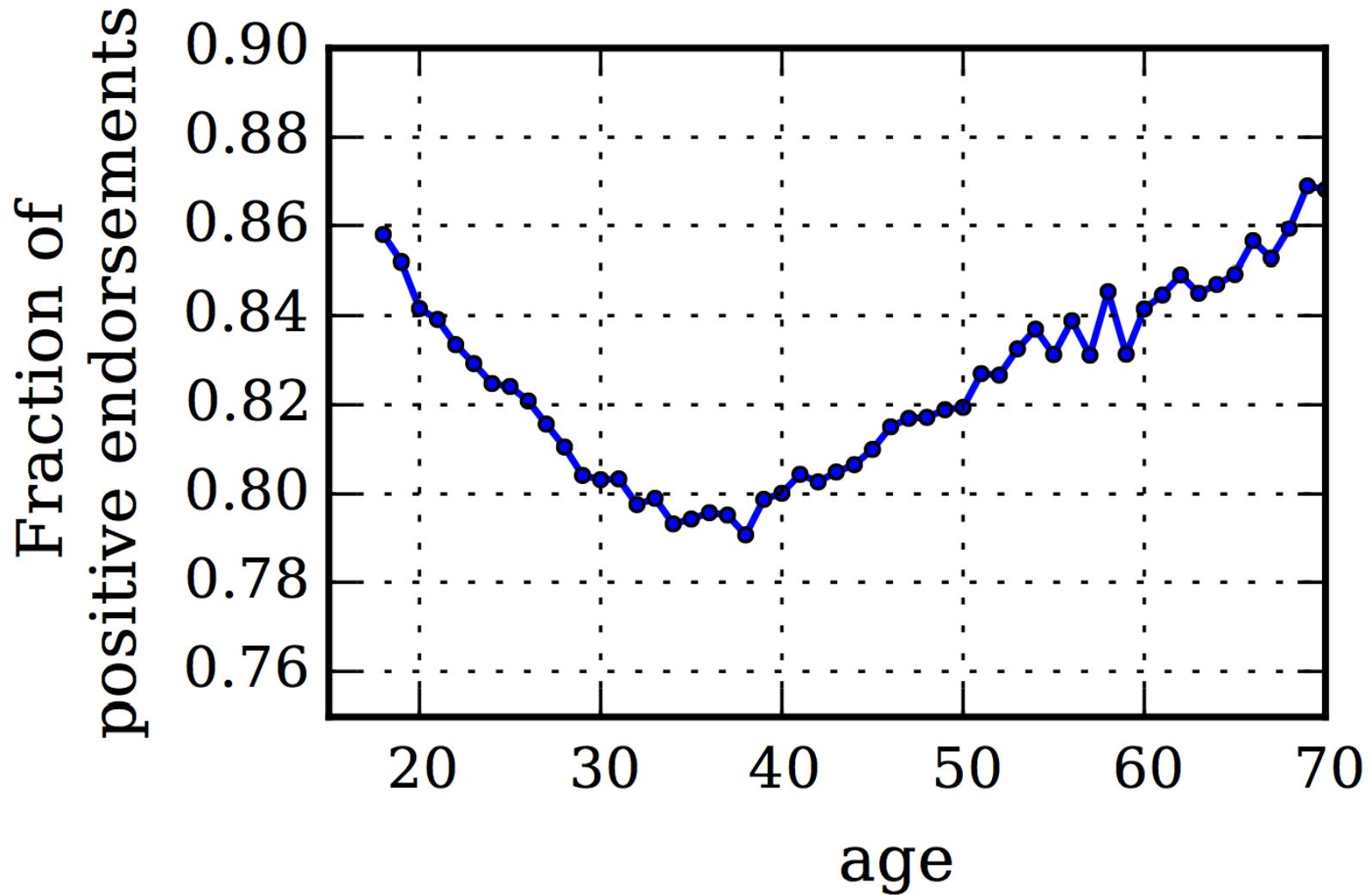




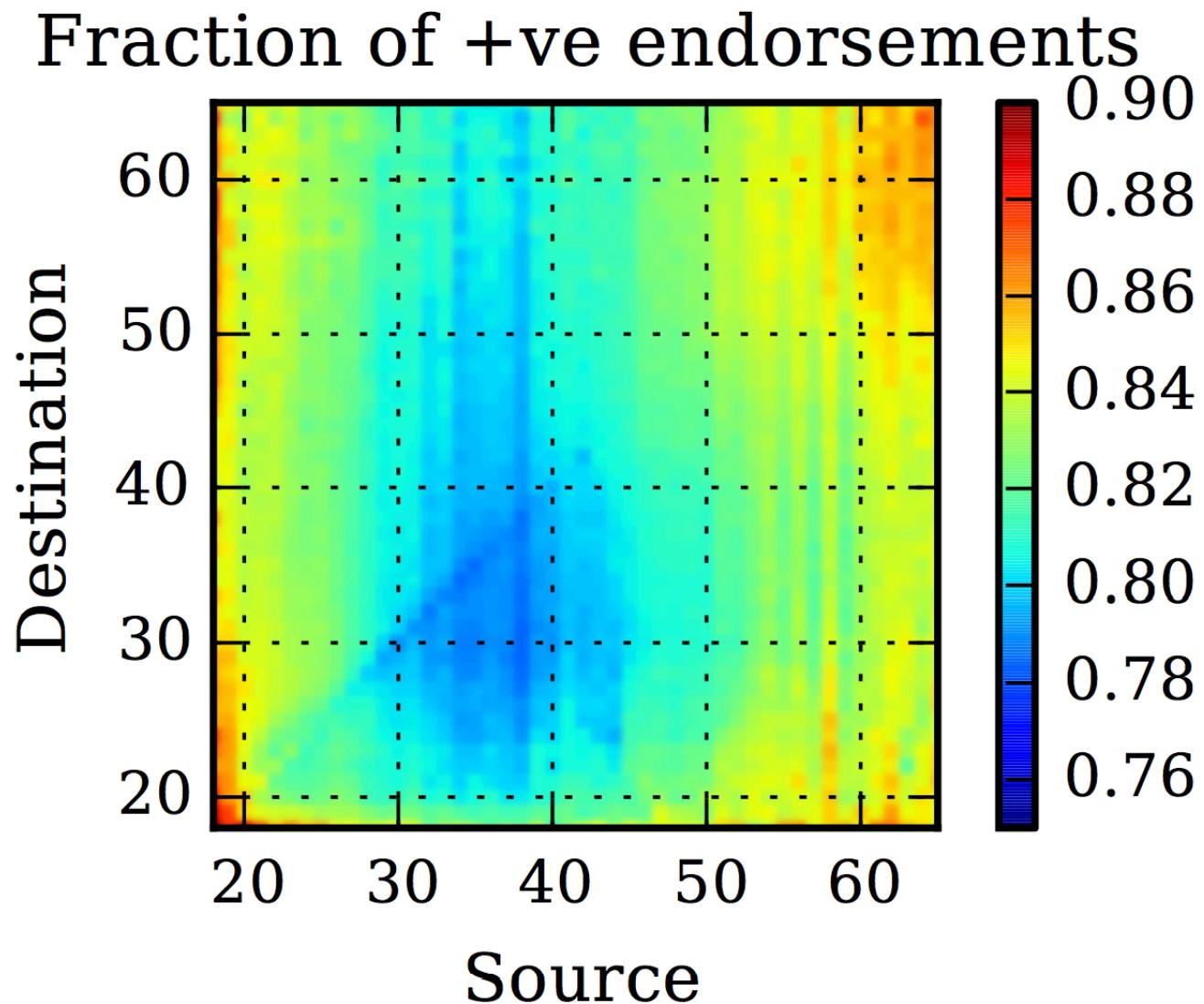




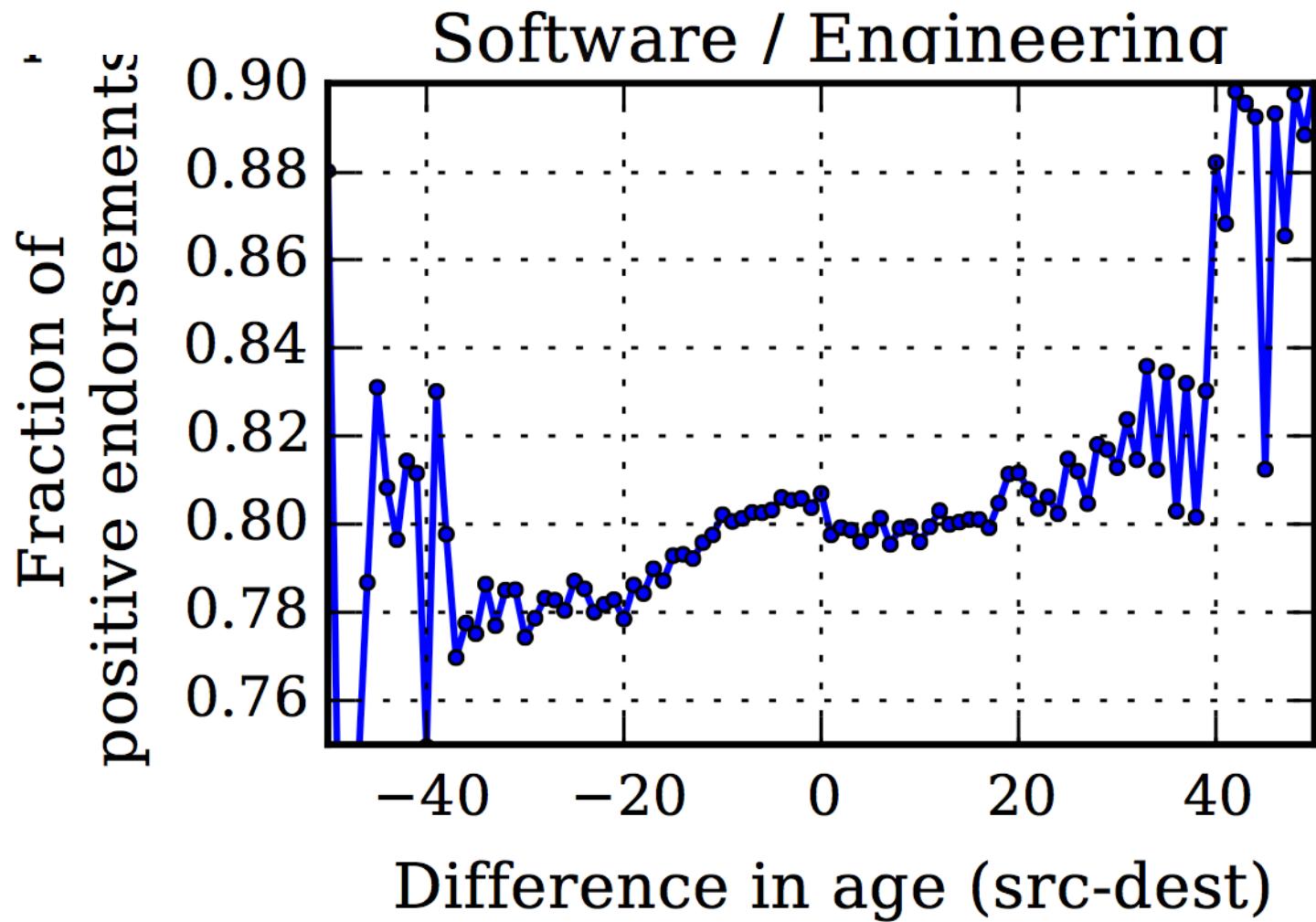
"Positive" endorsements by age difference



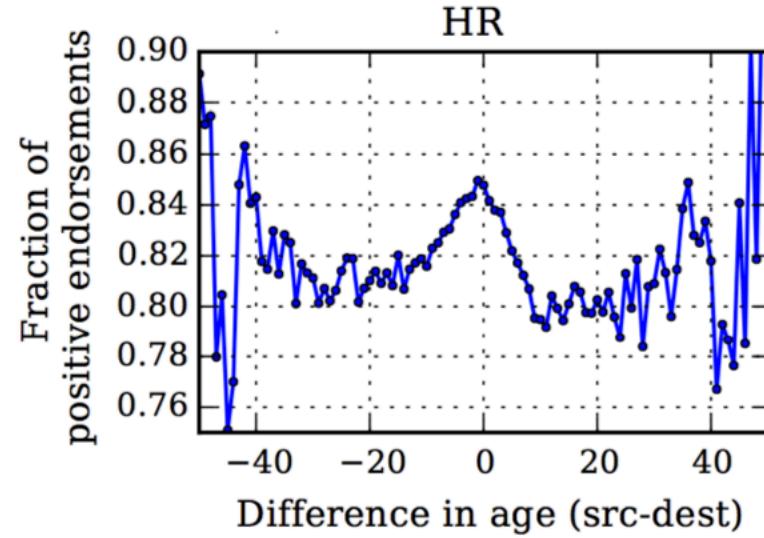
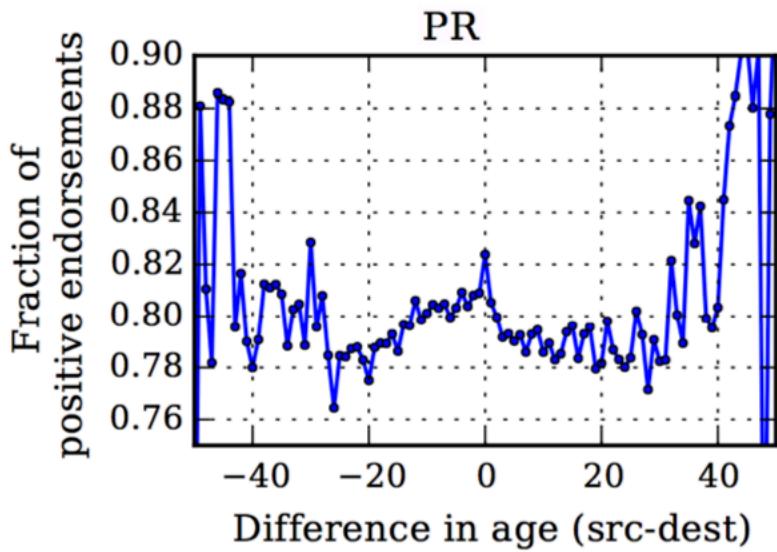
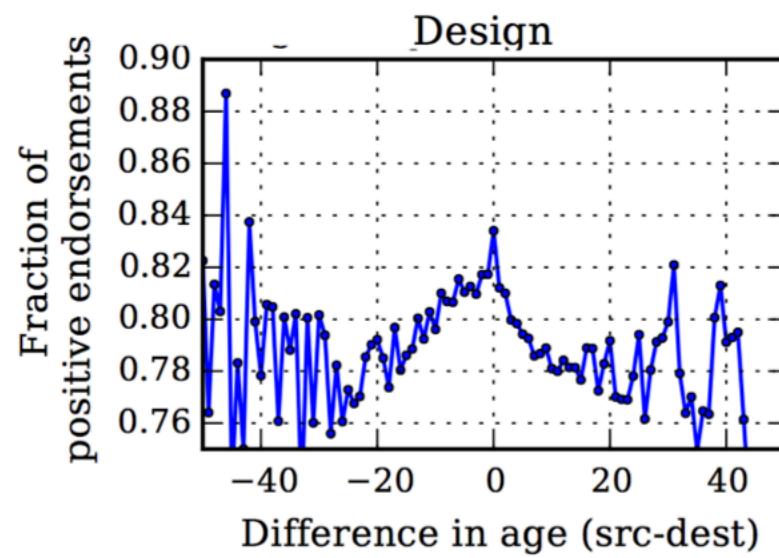
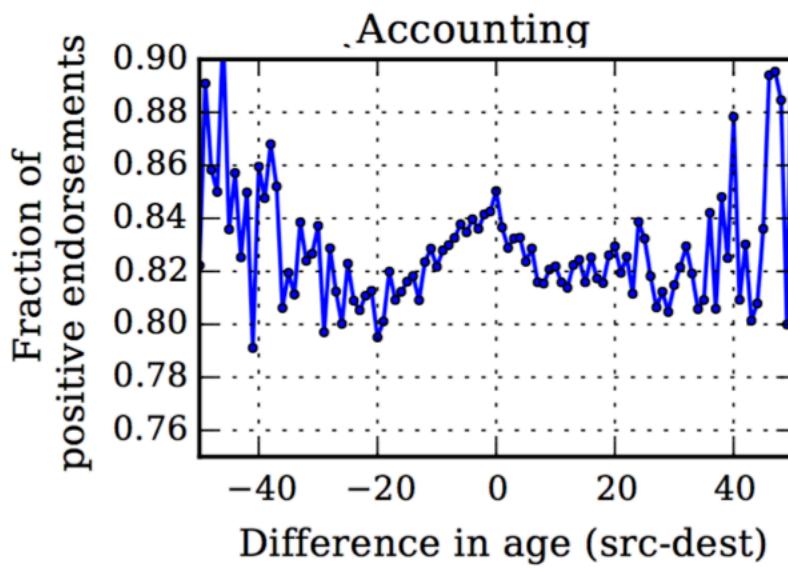
"Positive" endorsements by absolute age



"Positive" endorsements by endorser, endorsee age

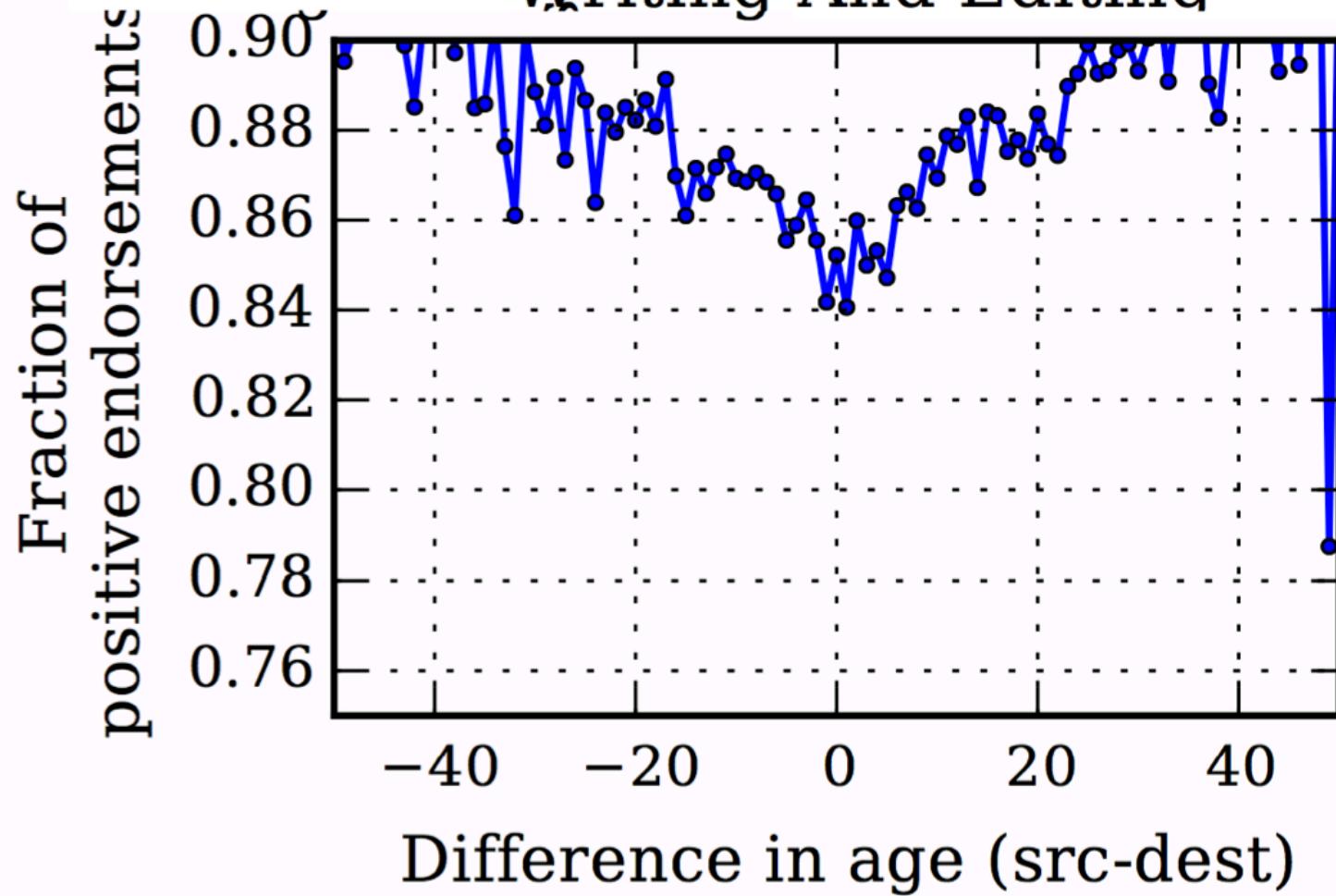


"Positive" endorsements by age difference for software professionals



"Positive" endorsements by age difference by industry

## Writing And Editing



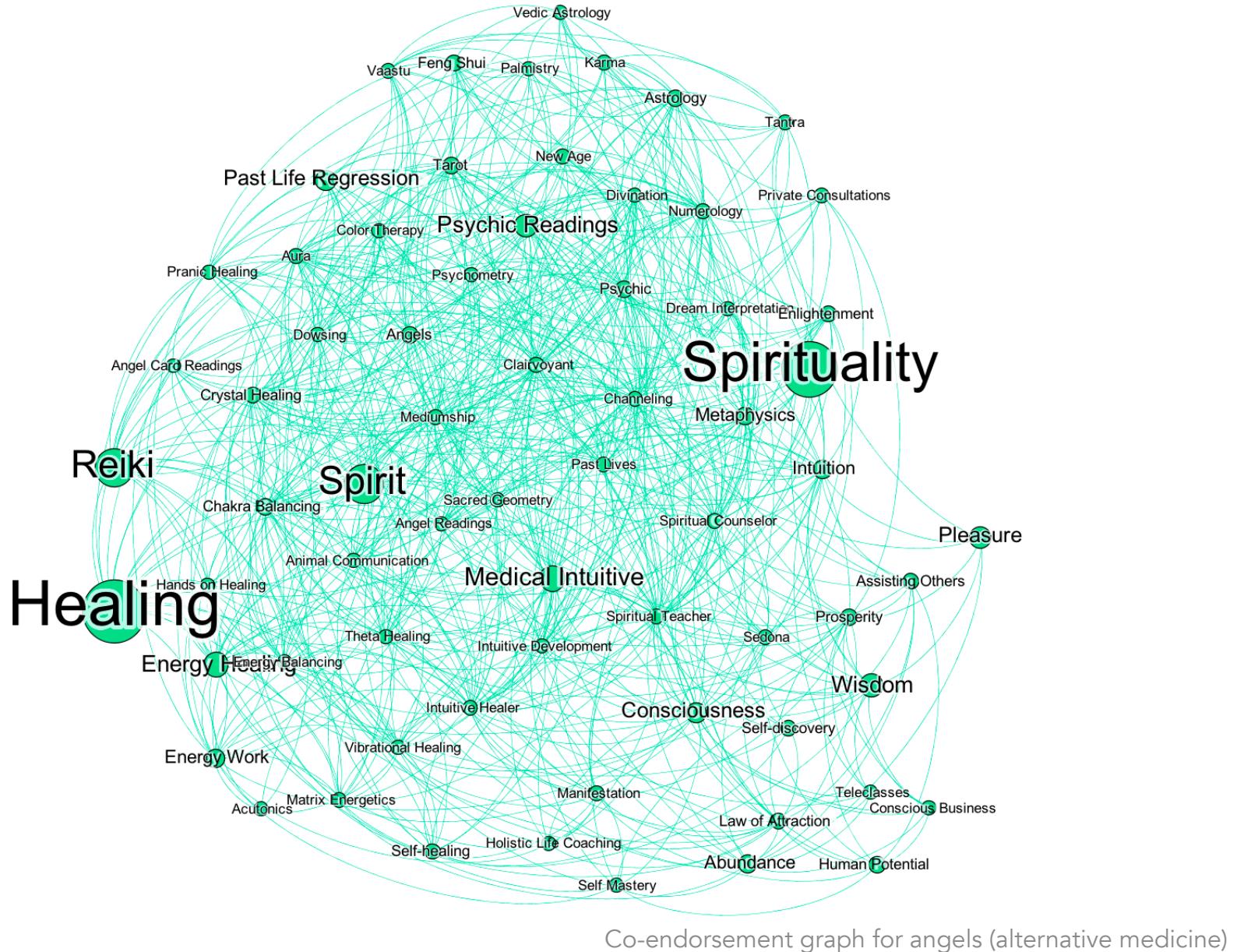
"Positive" endorsements by age for writers and editors

# Destination (endorsee)

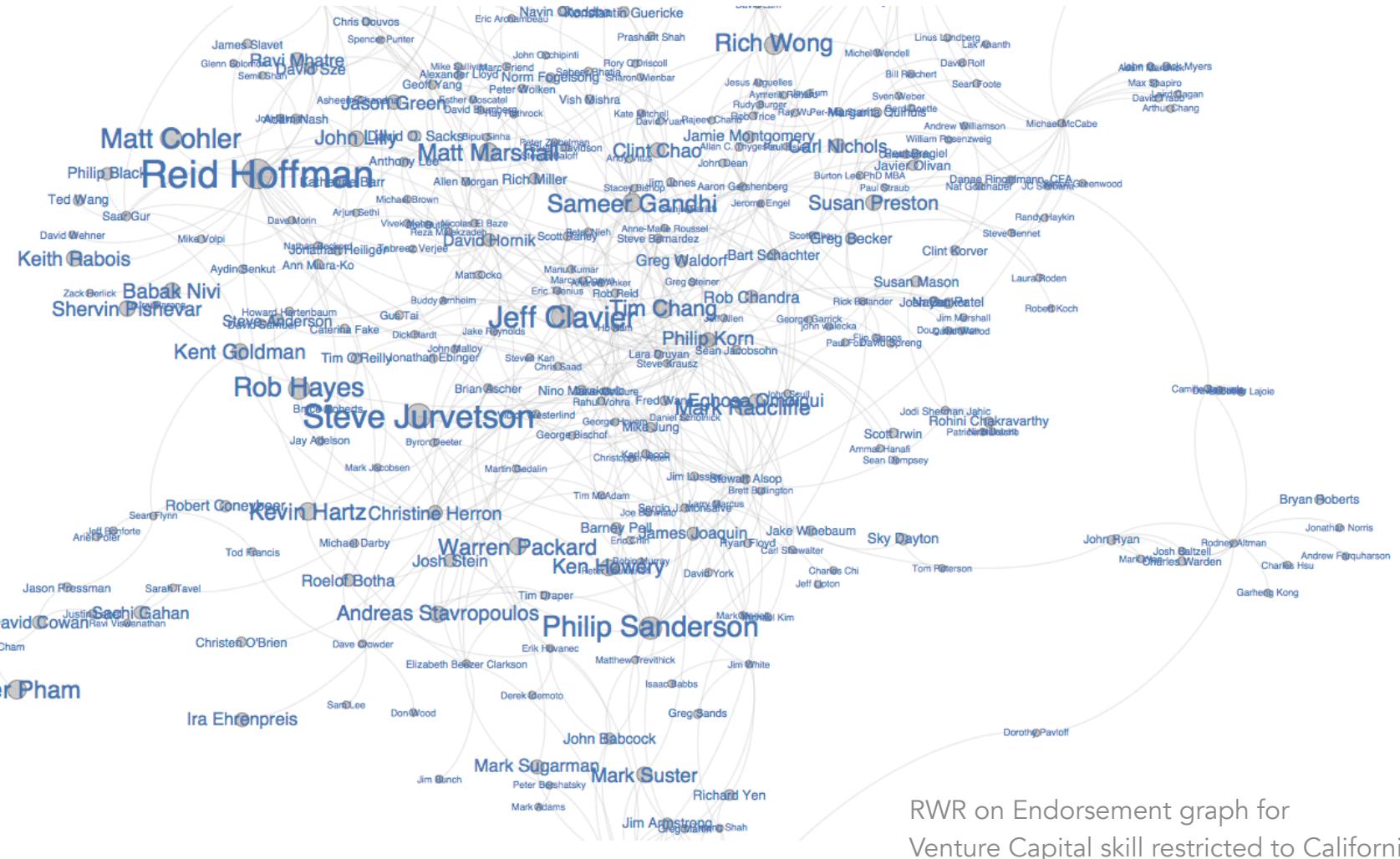
Source (endorser)

	M	F
M	83.4%	83.0%
F	81.0%	81.1%

"Positive" endorsements by gender



# Can we find influencers in venture capital?

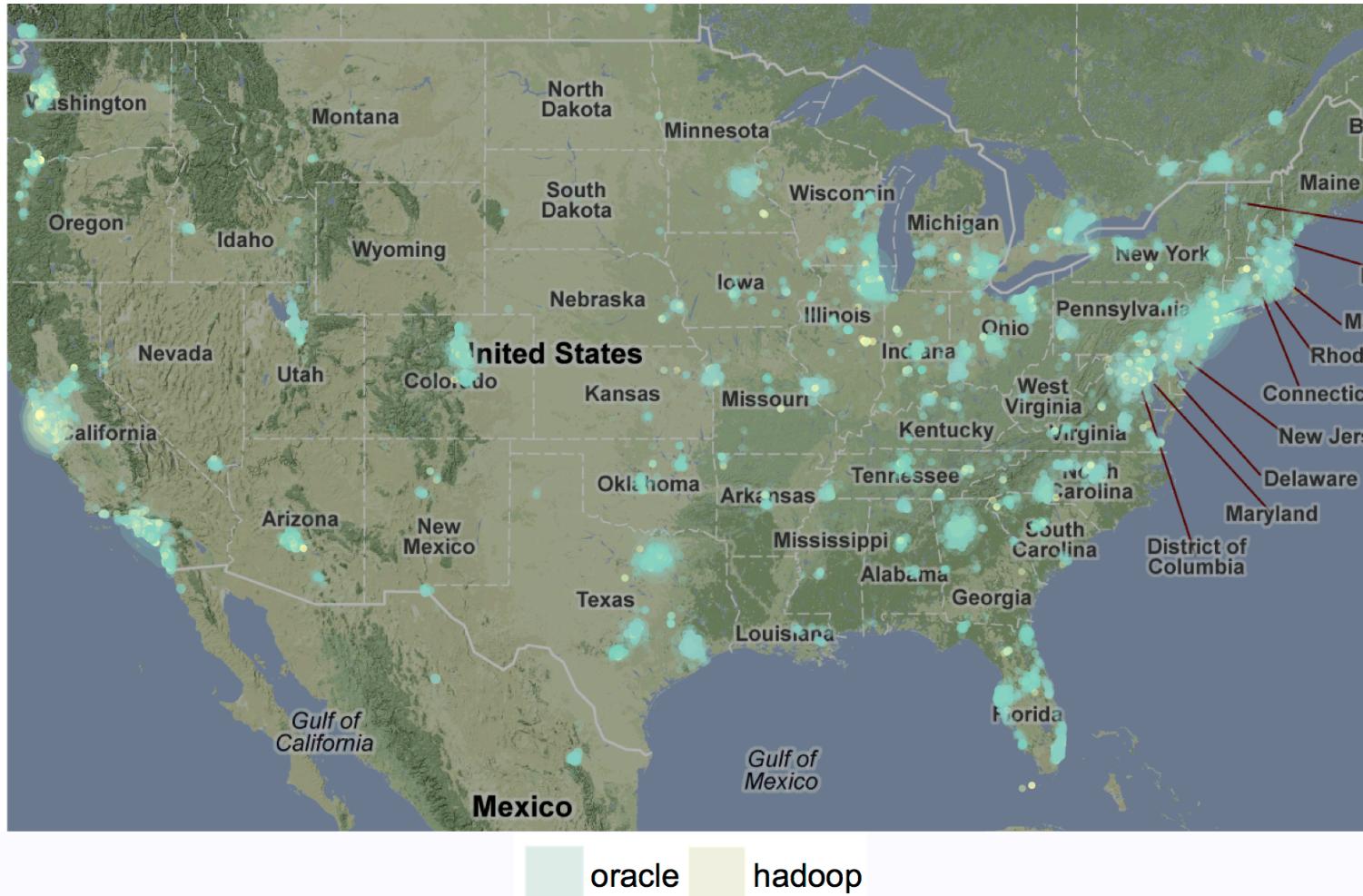


# What are the regional differences in software engineers?



Skill comparison between software technologists  
in California versus New York

# What is the adoption of Hadoop in the Enterprise?



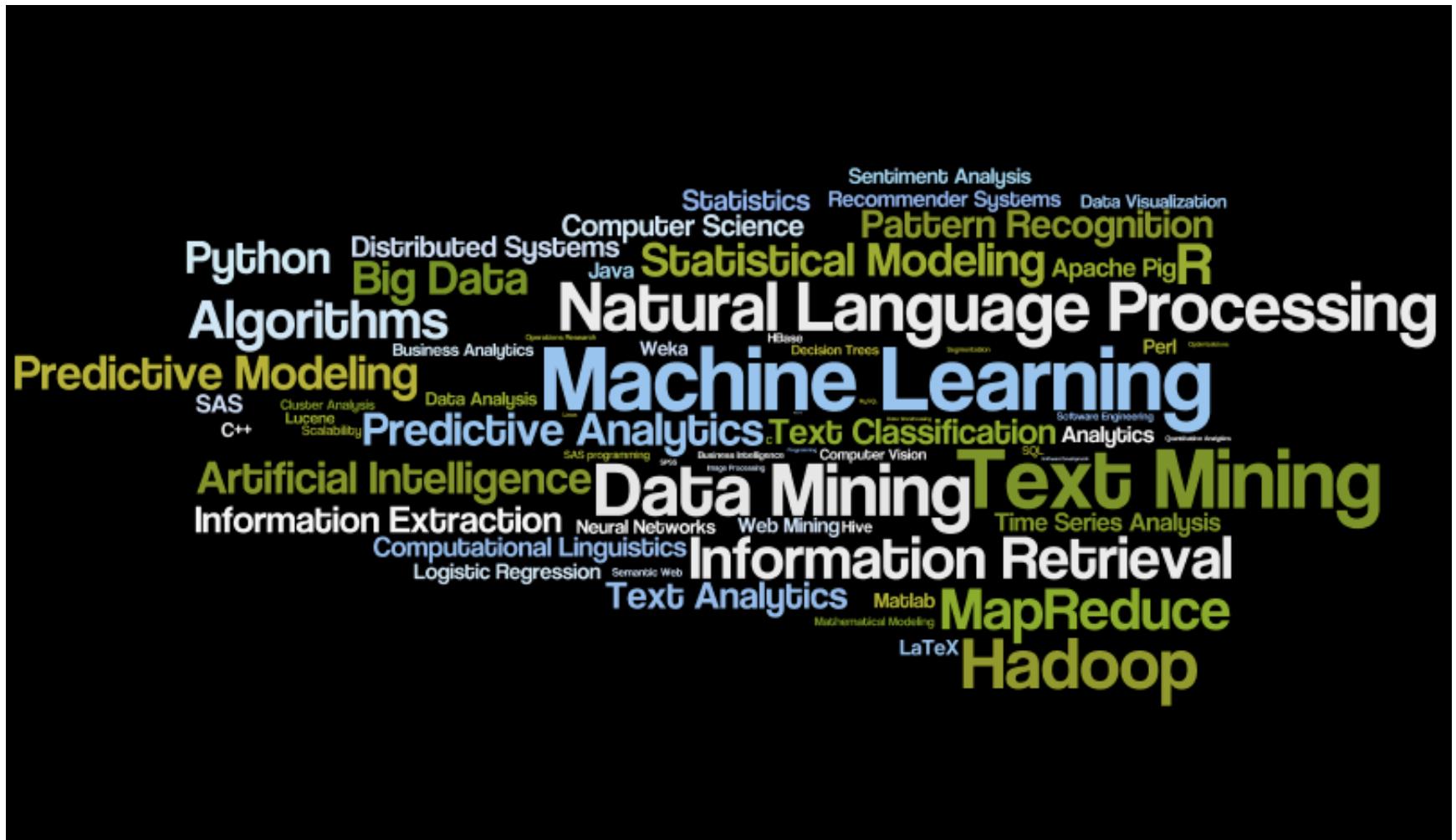
# Which technologies are software professionals adopting?



## New skills (by velocity) for software professionals



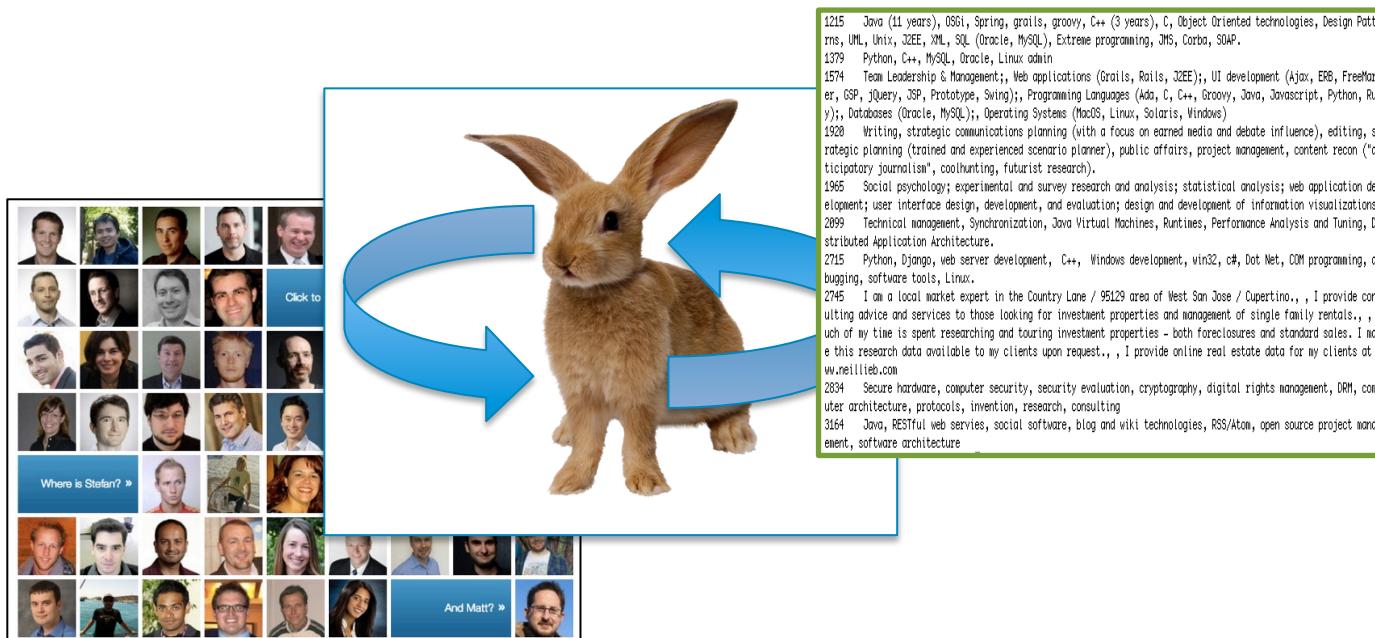
# Which skills are important for a data scientist?



Skills of members with the title "Data Scientist"

# Data amplifies desire

1. Desire + Social Proof
2. Viral Loops + Network Effects
3. Data Catalyst + Recommendation Algorithms



Learning More  
[data.linkedin.com](https://data.linkedin.com)

